NMR analysis of synthetic analogues of Ubiquitin

Helen McSparron

A thesis submitted for the degree of Doctor of Philosophy

University of Edinburgh

May 1998
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 has not previously been submitted, in whole or in part, for any degree at this or any other university.

University of Edinburgh
May 1998
To Cameron Toll
Acknowledgements

Firstly, I would like to thank my supervisor Paul Barlow for the provision of research facilities as well for his constant help and support. I would also like to acknowledge Professor Ramage for his encouragement throughout my Ph.D.

Thanks are due to Alan Cooper of Glasgow University and David Norman of Dundee University not only for the use of their laboratory equipment, but also for valuable discussion and advice. I am indebted to everyone in the Barlow group and Lab 29, in particular I am most grateful to Dusan Uhrin, Emma Beatty, for their NMR expertise, Alan Wiles for passing on his knowledge of XEASY and Alice Soteriou for her help and advice. I would also like to thank Jill Wilken for provision of the UbFleu(50,67) material and Craig Jamieson for his interest and contribution to the project.

I would like to thank all my fellow Ph.D students and colleagues for great nights out, good chat and for making my time in Edinburgh so unforgettable. A special thank you goes to Claire for being such a terrific friend and to my mam and dad and Paul for absolutely everything they have helped me through.
Abstract

NMR analysis of synthetic analogues of ubiquitin

Ubiquitin is a small globular protein (76 amino acids, MW 8565 Da) which occurs in both eukaryotes and prokaryotes. It is involved in many biological processes, its function to 'tag' other proteins, to signal degradation. Substrates include cyclin which is involved in cell regulation, the tumour repressor p53 and transcription factors, myc and fos. Ubiquitin is an ideal model for structural studies since it has a well defined secondary and tertiary structure and is stable to changes in temperature, pH and to a wide range of solvents. Synthetic techniques have been used to incorporate unnatural amino acids into the hydrophobic core of ubiquitin, the aim being to observe the structural defects incurred by these changes. Two analogues of ubiquitin have been under investigation.

The first analogue has the unnatural amino acid, 2S, 4S, 5-Fluoroleucine substituted into two leucine positions which, from studies of the native protein were directed towards each other, across the hydrophobic core. The structure of this fluorinated analogue has been determined using both $^1$H and $^{19}$F, 1D and 2D NMR methods. The NMR solution structure reveals that the fluoro-methyl groups adopt preferred conformations within the protein core.

The second analogue incorporates unnatural aminobutyric acid and norvaline into positions on the $\alpha$-helix, at positions i, i +4. These unnatural amino acids replace valine and isoleucine, therefore eliminating $\beta$-branching at these specific positions and also giving a net loss of two methyl volumes from the hydrophobic core. The structure has been determined using $^1$H 2D NMR and detailed comparisons made to the X-ray crystallographic structure of the analogue. The solution structure reveals that the substituted residues adjust their positions to compensate for the lost methyl volume.
Table of Contents

Abbreviations i
The Naturally Occuring Amino acids ii
The Unnatural Amino Acids Used ii

Chapter One
Introduction
1.0 Protein Folding 1
1.1 Protein Folding Strategies 2
1.2 Stabilising Effects within the Hydrophobic Core 4
  1.2.1 Hydrophobic Interactions 4
  1.2.2. Hydrogen Bonds 5
  1.2.3. Van der Waals and Electrostatic Interactions 6
1.3. Introduction to Ubiquitin 8
  1.3.1 Ubiquitin-Mediated Protein Degradation 9
  1.3.2. Ubiquitin Structure 12
  1.3.3. Ubiquitin as a ‘Superfold’ 13
  1.3.4. Investigation into the Ubiquitin Folding Pathway 15
  1.3.5. Exploration of Ubiquitin Stability and Conformation 17
1.4. The Ubiquitin Analogues in this Study 19
  1.4.1. Rationale behind the Study of UbFleu(50,67) 20
  1.4.2. Rationale behind the Study of UbAbu(26)Nva(30) 24

Chapter Two
Solid Phase peptide Synthesis and Purification of Ubiquitin analogues
2.1 Introduction to Solid Phase synthesis 27
2.2 Fmoc Solid Phase Peptide Synthesis 27
2.2.1. The Solid Support 29
2.2.2. Amide Bond Coupling 30
2.2.3. Fmoc-SPPS Monitoring 33

2.3 Synthesis and Purification of UbAbu(26)Nva(30)

2.3.1. Synthesis of UbAbu(26)Nva(30) 35
2.3.2. Purification of UbAbu(26)Nva(30)
   2.3.2.1. Purification using Tbfmoc-Cl 37
2.3.2. Size Exclusion Gel Filtration 40
2.3.3. Cation Exchange Chromatography 41
and characterisation.

2.4. Synthesis and Purification of UbFleu (50,67) 44

2.5 Materials and Methods 47

2.5.1. Synthesis Equipment and Reagents 47
2.5.2. Resin and Amino acids 48
2.5.3. Purification Equipment and Reagents. 49
2.5.4. Synthesis and Purification.
   2.5.4.1. Coupling of the C-terminal Amino Acid 51
to the 4-Alkoxybenzylalcohol resin
   2.5.4.2. Automated SPPS 52
   2.5.4.3. Tbfmoc Purification Methodology 53
   2.5.4.4. Size Exclusion Gel Filtration 55
   2.5.4.5. Cation Exchange Chromatography 55
   2.5.4.6. Refolding 55
   2.5.4.7. Yield 56
2.5.5. Simple Protein Characterisation Methods 57
   2.5.5.1 One Dimensional NMR 57
   2.5.5.2. Electrospray Mass Spectrometry 58
   2.5.5.3. Differential Scanning Calorimetry 59
Chapter Three

Nuclear Magnetic Resonance (NMR) Studies of Ubiquitin Analogues

3.1 Introduction to NMR 60

3.2. Theory of the Basic 1D NMR Experiment 62
   3.2.1. Nuclei in Magnetic Fields 62
   3.2.2. Application of a Radiofrequency Pulse 63
   3.2.3. Relaxation of Magnetisation, $M_{xy}$ 64
   3.2.4. Coupling 65
   3.2.5. Detection 67
   3.2.6. Solvent Suppression 68

3.3. Two Dimensional NMR 70

3.4. Homonuclear NMR 72
   3.4.1. The COSY Experiment 72
   3.4.2. The TOCSY Experiment 73
   3.4.3. The NOESY Experiment 75

3.5. Heteronuclear NMR 76
   3.5.1. 2D HMQC 76
   3.5.2. 1D HMQC Selective Excitation Techniques 77

3.6. Experimental 80
   3.6.1. Sample Preparation 80
   3.6.2 Determination of Hydrogen Bonds 82
   3.6.3. The NMR Spectrometer 82
   3.6.4. Spectral Considerations 83
      3.6.4.1. Pulse Width 83
      3.6.4.2. Frequency Lock 84
   3.6.5. Spectral Processing 84
      3.6.5.1. Fourier Transformation 85
      3.6.5.2. Deconvolution and Zero-filling 85
      3.6.5.3. Window Functions 86
## Chapter Four

**NMR Spectral Assignment of Ubiquitin Analogues, UbFLeu(50,67) and UBAbu(26)Nva(30)**

4.1 Introduction to Spectral Assignment 88

4.2 Assignment of the 2D $^1$H NMR of Ubiquitin Analogues 90

4.2.1 Assignment of Aromatic Residues 90

4.2.2 Assignment of Backbone Protons 91

4.2.3 Assignment of the Proline Residues 103

4.2.4 Assignment of Amino Acid Side Chains 108

4.2.5 Identification of Secondary Structure 116

4.2.6 Determination of Hydrogen Bonds 120

4.2.7 Assignment of Tertiary nOes 125

4.3 Assignment of 1D HMQC NMR spectra of UbFleu(50,67) 125

## Chapter Five

**Structure Determination**

5.1 Introduction 128

5.2 Tertiary Structure Determination 129

5.2.1 Compilation of nOe Table 129

5.2.2 Compilation of Hydrogen Bond Table 131

5.2.3 Measurement of the $\phi$ Angles from 2D $^1$H COSY Experiment 131

5.3 Molecular Dynamics Simulated Annealing 132

5.3.1 Structure Calculation Protocol 133

5.3.2 Structure Refinement 135

5.4 Structure Calculation Results 136

5.4.1 UbFleu(50,67) 136

5.4.2 UbAbu(26)Nva(30) 143
Chapter Six

Discussion

6.1. Discussion of Structural Studies of the Analogue

UbFleu(50,67)

6.1.1. Expectations of Project

6.1.2. Results and Observations

6.1.3. Purification and folding problems

6.1.4. Summary

6.2. Discussion of Structural Studies of the Analogue

UbAbu(26)Nva(30)

6.2.1. Expectations of Project

6.2.2. Results and Observations

6.2.3. Summary

References

Appendix A: UbFleu(50,67) Hydrogen bonds; UbAbu(26)Nva(30) Hydrogen bonds

Appendix B: UbFleu(50,67) nOe table

Appendix C: UbAbu(26)Nva(30) nOe table

Appendix D: UbFleu(50,67) assignment table

Appendix E: UbAbu(26)Nva(30) assignment table
**Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABI</td>
<td>Applied Biosystems</td>
</tr>
<tr>
<td>Ac</td>
<td>acetyl</td>
</tr>
<tr>
<td>Boc</td>
<td>tertiary-butyloxycarbonyl</td>
</tr>
<tr>
<td>Bu'</td>
<td>tertiary butyl</td>
</tr>
<tr>
<td>D</td>
<td>dimension</td>
</tr>
<tr>
<td>DCM</td>
<td>dichloromethane</td>
</tr>
<tr>
<td>DIC</td>
<td>N, N'- diisopropylcarbodiimide</td>
</tr>
<tr>
<td>DIEA</td>
<td>N, N- disopropylethylamine</td>
</tr>
<tr>
<td>DMF</td>
<td>dimethylformamide</td>
</tr>
<tr>
<td>DSC</td>
<td>Differential Scanning Calorimetry</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>EDT</td>
<td>1,2-ethanediol</td>
</tr>
<tr>
<td>ESI-MS</td>
<td>electro-spray ionisation mass spectrum</td>
</tr>
<tr>
<td>Fmoc</td>
<td>9-fluorenylmethoxycarbonyl</td>
</tr>
<tr>
<td>GuHCl</td>
<td>guanadinium hydrochloride</td>
</tr>
<tr>
<td>HOBt</td>
<td>1-hydroxybenzotriazole</td>
</tr>
<tr>
<td>HOCl</td>
<td>ethyl-1-hydroxy-1H-1,2,3,-triazole-4-carboxylate</td>
</tr>
<tr>
<td>HPLC</td>
<td>high performance liquid chromatography</td>
</tr>
<tr>
<td>MALDI MS</td>
<td>matrix assisted laser desorption ionisation mass spectrometry</td>
</tr>
<tr>
<td>Mbh</td>
<td>4, 4'- dimethoxybenzhydryl</td>
</tr>
<tr>
<td>MHz</td>
<td>mega hertz</td>
</tr>
<tr>
<td>MW</td>
<td>molecular weight</td>
</tr>
<tr>
<td>MWCO</td>
<td>molecular weight cut off</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>nOe</td>
<td>nuclear Overhauser effect</td>
</tr>
<tr>
<td>NOESY</td>
<td>nuclear Overhauser effect spectroscopy</td>
</tr>
<tr>
<td>PG</td>
<td>protecting group</td>
</tr>
<tr>
<td>PGC</td>
<td>porous graphitised carbon</td>
</tr>
<tr>
<td>Pmc</td>
<td>2, 2, 5, 7, 8 - pentamethylchroman-6-sulponyl</td>
</tr>
<tr>
<td>RP</td>
<td>reverse phase</td>
</tr>
<tr>
<td>Rt</td>
<td>retention time</td>
</tr>
<tr>
<td>SPPS</td>
<td>solid phase peptide synthesis</td>
</tr>
<tr>
<td>Tbfmoc</td>
<td>17-tetrabenzo[a, c, g, i]fluorenylmethoxycarbonyl</td>
</tr>
<tr>
<td>TFA</td>
<td>trifluoroacetic acid</td>
</tr>
<tr>
<td>Ub</td>
<td>ubiquitin</td>
</tr>
<tr>
<td>UV</td>
<td>ultra violet</td>
</tr>
<tr>
<td>VIS</td>
<td>visable</td>
</tr>
</tbody>
</table>
### The Naturally Occurring Amino Acids

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Three letter code</th>
<th>One letter code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>Ala</td>
<td>A</td>
</tr>
<tr>
<td>Arginine</td>
<td>Arg</td>
<td>R</td>
</tr>
<tr>
<td>Asparagine</td>
<td>Asn</td>
<td>N</td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>Asp</td>
<td>D</td>
</tr>
<tr>
<td>Asparagine</td>
<td>Asn</td>
<td>N</td>
</tr>
<tr>
<td>Cysteine</td>
<td>Cys</td>
<td>C</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>Glu</td>
<td>E</td>
</tr>
<tr>
<td>Glutamine</td>
<td>Gln</td>
<td>Q</td>
</tr>
<tr>
<td>Glycine</td>
<td>Gly</td>
<td>G</td>
</tr>
<tr>
<td>Histidine</td>
<td>His</td>
<td>H</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>Ile</td>
<td>I</td>
</tr>
<tr>
<td>Leucine</td>
<td>Leu</td>
<td>L</td>
</tr>
<tr>
<td>Lysine</td>
<td>Lys</td>
<td>K</td>
</tr>
<tr>
<td>Methionine</td>
<td>Met</td>
<td>M</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>Phe</td>
<td>F</td>
</tr>
<tr>
<td>Proline</td>
<td>Pro</td>
<td>P</td>
</tr>
<tr>
<td>Serine</td>
<td>Ser</td>
<td>S</td>
</tr>
<tr>
<td>Threonine</td>
<td>Thr</td>
<td>T</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>Trp</td>
<td>W</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>Tyr</td>
<td>Y</td>
</tr>
<tr>
<td>Valine</td>
<td>Val</td>
<td>V</td>
</tr>
</tbody>
</table>

### The Unnatural Amino Acids Used

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Three letter code</th>
<th>One letter code</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-Aminobutyric acid</td>
<td>Abu</td>
<td>-</td>
</tr>
<tr>
<td>Norvaline</td>
<td>Nva</td>
<td>-</td>
</tr>
<tr>
<td>2S,4S,5-Fluoroleucine</td>
<td>Fleu</td>
<td>-</td>
</tr>
</tbody>
</table>
Chapter one

Introduction

1.0. Protein Folding

Protein folding is the molecular mechanism by which a transformation from a random coil to a relatively fixed three-dimensional conformation is effected\(^1\). The folded conformation of a protein is critical for its specific biological function. A full understanding of the folding process would enable deduction of protein structure from the amino acid sequence\(^2\). However, this is currently not possible in most cases. Hence methods have been developed to experimentally determine structure through X-ray crystallography and NMR techniques. These have generated a body of information from which a set of basic principles of protein structure has emerged. These principles make it easier for us to understand how protein structure is generated and to identify common structural themes.

This project is concerned with aspects of conformation and stability of a chemically synthesised version of the protein, ubiquitin. The folding problem has been approached through modification of the protein's core residues and observation of the consequences, primarily using high-field NMR. This
investigation addresses an important folding issue: the contribution of the ordering of individual amino acid side chains to the overall structure and stability of the core.

1.1. Protein Folding Strategies

Most small proteins and some larger ones can be refolded from a denatured state which indicates that their primary amino acid sequences contain all the information necessary to determine the final folded structure.

Proteins fold under optimal conditions within $10^{-1}$ - $10^3$ seconds so it is impossible that refolding occurs by a random search of all possible conformations. Instead the folding process must be directed in some way. Two protein folding models have been described. These were initially proposed in the early 1970’s. The nucleation model suggests that protein folding begins with segments of about 15 residues of an unfolded peptide chain which ‘flicker’ in and out of native secondary structural forms. These transient structures come together by diffusion and stabilise each other by forming a complex. Each of these complexes acts as a nuclei to stabilise other transient elements of secondary structure. In the hydrophobic collapse model, the protein first undergoes aggregation of its non-polar groups to form a structure with a loose hydrophobic core. The secondary structural elements develop around this core and are stabilised by packing against each other. This is
hypothesised to be similar to the ‘molten globule’ which finally folds in a slow step to form a tightly packed native structure. \textit{In vivo}, specialised proteins and physiological conditions play a very important role in accelerating folding and in ‘steering’ the polypeptide chain toward the native state and away from kinetically stable, abortive conformations\textsuperscript{7,8,9}.

Peptide-bond formation on ribosomes is stereospecific, nearly always yielding \textit{trans} isomers. Spontaneous isomerisation of peptide bonds during the folding process of the polypeptide chain is unlikely. The \textit{trans} peptide bond isomer is favoured 1000-fold over the \textit{cis} form since \textit{cis} peptide bonds cause steric hindrance between adjacent side chains. The energy difference between the \textit{cis} and \textit{trans} isomers is about 2 kcal/mol in favour of \textit{trans}. This difference is smaller when the next residue is proline since a transition from \textit{trans} to \textit{cis} replaces the contact $C_i^\alpha$ and $C_{i+1}^\alpha$ only by a similar contact $C_i^\alpha$ and $C_{i+1}^\beta$, (where $i+1$ is Pro) and therefore in folded proteins Pro residues can be preceded by either isomer\textsuperscript{10} The kinetics of the folding process are usually complicated by \textit{cis-trans} isomerisation of peptide bonds preceding proline residues, which is very slow, but maybe enzyme-catalysed \textit{in vivo} by the peptidyl prolyl isomerases\textsuperscript{11}. These are ubiquitous proteins that are found in numerous subcellular compartments. This may be an important factor to consider when working with peptides \textit{in vitro}. 
1.2. Stabilising Effect within the Hydrophobic core

Proteins fold to attain stable conformations with low free energies. It is therefore important to consider the factors which contribute to the stability of the folded protein.

The 3D structure of a folded protein provides unique environments for, and orientations of, amino acid side chains giving rise to the many special properties of proteins. The physicochemical properties of proteins are also influenced by interactions with their environment, e.g. with water, salt, membranes, other proteins and nucleic acids. These inter- and intramolecular interactions arise mainly from a set of fundamental non-covalent attractions. Amongst these are hydrophobic interactions, hydrogen bonds, van der Waals forces and electrostatic interactions. These factors work together in combination and the net balance determines the overall stability.

1.2.1. Hydrophobic interactions

Thermodynamic studies of protein unfolding demonstrate that the burial of non-polar groups within the protein interior provides an important contribution to stability. This step is proposed to occur during the last stage of protein folding when the close packing of side chains is the predominant interaction. Such tight packing is related to the protein’s unique and highly complex three
dimensional shape. The packing process is likely to be energetically complex and difficult as side chains prefer to be disordered\textsuperscript{13,14}. Packing calculations were first done by Richards more than two decades ago\textsuperscript{15}. Subsequent calculations\textsuperscript{16} have revealed that the packing efficiency in the core is high and greater than that of crystals of organic molecules. The packing in the core seems to be a tight jigsaw puzzle of side chains, sometimes accommodating water molecules. Improving the packing of the protein interior has become a method of increasing conformational stability\textsuperscript{17,18}. For example, increasing the packing efficiency for ribonuclease H1\textsuperscript{19} and T4 lysozyme\textsuperscript{20,21} has led to an increase in protein stability. Variants of ribonuclease H1 (V74L and V74I) reveal a 2.1 °C - 3.7 °C increase in $T_m$ values. A corresponding decrease in stability is evident in the V74A mutant which shows a 7.4 °C fall in $T_m$ value. The T4 lysozyme mutant, S117F, was found to be more thermostable than wild-type by 1.1-1.4 kcal/mol.

1.2.2. Hydrogen bonds

The contribution from interactions between polar groups, especially the formation of hydrogen bonds, to protein stability was thought to be limited to only ‘locking’ the 3D conformation. It was assumed that the hydrogen bond energy within the folded protein was the same as that between water and polar groups within the unfolded protein. This, however, ignores the intramolecular
nature of the interactions within the folded state and assumes that they operate independently of each other.

The presence of numerous weak interactions gives a much greater contribution to net stability than is possible with individual interactions. Consideration of the thermodynamics of the model systems indicate that hydrogen bonds may play a greater role than non-polar interactions\textsuperscript{22,23,24}. The lengths of hydrogen bonds depend on acceptor and donor atoms, they range from 2.70 Å to 3.10 Å for oxygen and nitrogen respectively\textsuperscript{25}. The bond energies range from 3 - 7 kcal/mol. The strength of a hydrogen bond will depend upon the alignment of the three atoms involved. The optimal configuration exists when the donor, hydrogen and acceptor atoms are co-linear. The hydrogen bond becomes progressively weaker with increasing angle between the acceptor atom and the donor and hydrogen atom. Eleven of the twenty natural amino acids have functional groups which enable hydrogen bonding through their side chains.

1.2.3. \textbf{Van der Waals and Electrostatic Interactions.}

Van der Waals forces are weak non-specific interactions between any two atoms. The basis of a van der Waal interactions is the distribution of electronic charge around an atom as a function of time. This interaction comes into effect when the two atoms are within 3 - 4 Å of one another. The attraction energy (to the first order) varies as the inverse sixth power of the distance between the
Van der Waals interactions are very important factors within the hydrophobic core since the detailed packing in the core determines the strengths of the van der Waals interactions. The optimal distance for interaction of two atoms is usually 0.3 - 0.5 Å greater than the sum of their van der Waal's radii. These are described by the Lennard-Jones potential, equation 1:

\[
U(r) = 4\varepsilon_{LJ}\left[ \frac{\sigma_{LJ}}{r} \right]^{12} - \left( \frac{\sigma_{LJ}}{r} \right)^6
\]

where \(\varepsilon_{LJ}\) and \(\sigma_{LJ}\) are adjustable parameters which depend on the identity of the atoms and \(r\) is the internuclear distance. The first (12th-power) term increases as the intermolecular distance \(r\) decreases. This is the repulsive contribution to the potential energy. The second (6th-power) term produces a decrease in the potential energy as \(r\) decreases and corresponds to the attractive contribution. A typical non-bonded van der Waal interaction between two aliphatic protons measures -0.04 kcal/mol when 2.92 Å apart compared to -0.11 kcal/mol for two nitrogen atoms with an inter-nuclei distance of 3.51 Å. This is much less than electrostatic and hydrogen bonds which involve 3-7 kcal/mol at similar inter-nuclear separations.

Electrostatic interactions are the most fundamental non-covalent interactions. Covalent bonds between different types of atom lead to asymmetric bond electron distribution and so most atoms will carry a partial charge. Charged
groups within the same protein molecule can interact and this results in a stabilising effect. As atoms approach each other their electronic orbitals overlap and the repulsive energy increases enormously. The interaction also depends on the dielectric constant, \( \varepsilon \), of the surrounding medium. In the structure calculations involved in this project, we have used the value \( \varepsilon = 4 \) which is the macroscopic dielectric constant of amide polymers. The charges interact according to Coulomb’s Law, which is defined by equation 2:

\[
\Delta E = \frac{332}{\varepsilon} \cdot \frac{Z_A Z_B}{D_{rAB}} \quad (2)
\]

where \( Z_A \) and \( Z_B \) is the charge of the two groups concerned, \( \varepsilon \) is the dielectric constant and \( D_{rAB} \) is the distance between the two atoms.

1.3. Introduction to Ubiquitin

Ubiquitin is a small globular protein (76 amino acids, 8565 Da). It is abundant in all eukaryotes where it has an essential role in the degradation of unwanted proteins by a multi-subunit protein complex called the proteosome.

Ubiquitin is involved in numerous cell processes including DNA repair and in the stress response.
1.3.1. Ubiquitin Mediated Protein Degradation

Protein degradation requires the assembly of a multi-ubiquitin chain on the target protein. This facilitates the recognition of the substrate for degradation by the 26S proteosome\textsuperscript{36,37}.

The process to link ubiquitin to another protein occurs in three basic steps. This is presented schematically in figure 1.1. The first stage is the activation of the ubiquitin C-terminal carboxyl group and its linkage to a thiol group of the activating enzyme E1. Then the activated ubiquitin is transferred from E1 to a second protein E2. E2 can transfer ubiquitin to various proteins some of which, including histones, are reversibly linked to ubiquitin. In such cases, rather than degradation of the protein, ubiquitin has the effect of modifying some of the functions of the protein before it is released. However, when a protein is to be degraded, ubiquitin is finally transferred from E2 to a lysine of the target protein\textsuperscript{38}, either directly or with the assistance of a ubiquitin protein ligase (E3). Protein ligase is generally required for the formation of multi-ubiquitin chains on the substrate. These ubiquitin molecules are linked to each other in chains produced by an iso-amide bond between the α-carboxyl group of one ubiquitin molecule and the side chain of Lys 48 of another. The protein is degraded by the 26S proteosome, which is a very large ATP-dependant protease complex, specific for ubiquitin-protein conjugates. Ubiquitin is released from the degraded protein and can be reused.
Figure 1.1. Schematic pathway for the Ubiquitin mediated protein degradation pathway.
Figure 1.2. (a) Ribbon backbone representation of human ubiquitin with secondary structural elements as indicated. (b) Atom space filled representation of human ubiquitin.
1.3.2. Ubiquitin Structure

The primary structure of ubiquitin is highly conserved throughout all species. The amino acid sequence is identical in human, chicken, *Xenopus* and *Drosophila* ubiquitin. Two exceptions to the regular amino acid sequence have been found. These occur in oat and in yeast in each of which there are three amino acid changes. The changes observed in yeast are P19S, E24D, and A28S, while in oat ubiquitin they are P19S, E24D and S57A\(^{39}\). In each case the tertiary structure remains identical.

Ubiquitin is an ideal model for structural studies since its native form is stable to changes in temperature (23 °C to 80 °C) and to pH fluctuations (pH 1.18 to 8.43)\(^{40,41}\). Ubiquitin lacks di-sulphides, cis prolyl-peptide bonds and metal-binding sites which may complicate folding mechanics. Figure 1.2 displays the 3D conformation of Ubiquitin. The tertiary structure consists of a single turn of 3\(_{10}\) helix (residues 56-60) and five strands of mixed β sheet which pack against one side of three-and-half turns of an α-helix (residues 23-35). The β-strands positioned in the centre of the sheet (residues 1-8 and 65-71) are parallel and the other three strands (residues 11-18, 41-46 and 48-50) are anti-parallel to one another. These structural elements are linked by seven reverse turns. The structure of human ubiquitin has been characterised by both X-ray crystallography\(^{42}\) and by solution-state NMR\(^{43,44}\). The crystal structure of
chemically synthesised ubiquitin has been shown to be identical to the native folded form\textsuperscript{45,46}.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{ubiquitin_structure}
\caption{Primary structure of human ubiquitin with secondary structural elements highlighted.}
\end{figure}

1.3.3. Ubiquitin has a ‘Superfold’

Protein superfamilies have been studied through sequence and structure databases\textsuperscript{47}. So far there are only nine protein folds known to recur in proteins which are unrelated through function or their primary sequence. Ubiquitin is one of these, and adopts a superfold called the 'Ub $\alpha\beta$ roll'. The other proteins known to adopt this superfold are ferrodoxin\textsuperscript{48,49} and the B\textsubscript{1} domain of
Figure 1.4. Backbone Ribbon Representations of proteins sharing the 'Ub αβ roll' superfold: (a) Immunoglobulin Binding Domain of Streptococcal Protein G (B1 Domain); (b) Immunoglobulin Binding Domain of Streptococcal Protein G, type 7 (B2 Domain); (c) 2Fe-2S Ferredoxin and (d) Ras Binding Domain of C-Raf-1.
**streptococcal** protein G$^{50,51}$. Figure 1.4. compares these tertiary structures. Note how they all share a similar α-helix/β-sheet interface.

A superfold may have been reached through evolutionary convergence to an extra-stable conformation. Alternatively members of a superfold family may have diverged from a common ancestor and, despite extensive sequence changes, retained its stable topology.

1.3.4. Investigation into the Ubiquitin Folding Pathway.

NMR observation of amide proton-exchange rates has become one of the principal approaches for the structural characterisation of protein folding intermediates$^{52}$ and protein- ligand complexes$^{53}$.

The early hydrogen-bonding events in the folding mechanism of ubiquitin have been studied by Briggs *et al*$^{54}$. This technique relies on the ability to measure the protection against solvent exchange for individual amide protons involved in stable intra- or inter-molecular hydrogen bonds. Pulsed H-D exchange experiments performed on a quenched- flow apparatus make it possible to measure the time course of amide protection during protein folding. These studies show that the folding of ubiquitin occurs in three stages: a highly concerted early folding event followed by two minor folding stages.
The early folding event is indicated by those amide protons involved in hydrogen bonds which are 80% protected during the initial 8 ms. This shows that the α-helix and the β-sheet and the interface between them has formed. The secondary structural elements have formed at the same rate, demonstrating cooperativity in their formation. Other amide protons including those of residues Tyr 59, Ile 61 and Leu 69 in the C-terminal strand of the β-sheet are involved in hydrogen bond formation with residues in the parallel N-terminal strand. However these are not protected and fold approximately three times slower.

In comparison with other proteins studied, the early folding stage of ubiquitin (8 ms) occurs more rapidly than the early folding stage of cytochrome c\textsuperscript{55}, ribonuclease A\textsuperscript{56} and barnase\textsuperscript{57}, which all take 20 ms. The early event of folding in the mechanism of cytochrome c shows the two α-helical structures to be formed i.e. the secondary elements are formed in a common initial stage as found in ubiquitin. In the cases of ribonuclease A and barnase it was found that portions of the β-sheet structure were already formed in the same early folding period.

Following the early stage of ubiquitin folding, the two parallel β-strands remain misaligned. This leaves the amides on the C-terminal strand without hydrogen-bond acceptors. The subsequent rearrangement step forms the native hydrogen-bond pattern. Hydrogen-exchange experiments on native ubiquitin,
also show that protected amide protons on the C-terminal strand exchange more rapidly than the other hydrogen-bonded amide groups in the β-sheet. This suggests that the equilibrium fluctuations of the folded state lead to similar structural defects\(^{58}\). Approximately 9% of molecules remain unfolded, these are expected to contain one or more cis prolyl peptide bonds\(^{59}\).

1.3.5. Exploration of Ubiquitin Stability and Conformation

A common approach to the study of protein stability and conformation is through modification of the native protein primary sequence. The effects of the mutation on stability and conformation of any protein analogue depends on the role the original amino acid played in the folded structure. The strict evolutionary conservation of ubiquitin suggests that each residue must have an important role within the protein.

The core of a protein is very sensitive to change\(^{60}\) and a single mutation can decrease the stability by up to 9 kcal/mol. This is sufficient to unfold the protein. Substitution of an amino acid on the surface of the protein leads to little change in stability and conformation unless involved in a specific role. Site-specific mutagenesis of the surface residues of ubiquitin was investigated to assess changes in stability and also to assess their importance in the interaction of ubiquitin-activating enzyme (E1). The mono-substitutions made to ubiquitin in this study involved the replacement of leucines with arginines.
UbR42L and UbR72L were observed to bind ubiquitin activating enzyme (E1) with $10^3$ times lower affinity. This is a destabilisation of 7 kcal/mol, indicating that these arginine residues are important to the interaction between Ub and E1$^{61}$.

The effect of amino acid changes at the helix-sheet interface of ubiquitin have been studied by Korasanizaden and co-workers$^{62}$ employing NMR techniques. A variant of ubiquitin was synthesised incorporating a tryptophan (Trp) residue, to serve as a sensitive fluorescence probe in folding studies. The Trp residue replaced phenylalanine at position 45 (F45W). The structure and activity of this mutant was found to be very similar to the native protein structure$^{63}$. Val 26 is in the centre of this helix-sheet hydrophobic interface. To measure the importance of this residue the F45W variant was substituted at position 26 with other hydrophobic residues, Gly, Ala, Leu and Ile. All analogues except V26G gave good quality 2D NMR spectra. The V26G analogue was found to be unstable, aggregating under NMR concentrations. Equilibrium unfolding using a guanadinium hydrochloride-containing denaturing medium, and fluorescence as a probe, reveals that the less bulky substitutions V26A and V26G are highly destabilised, probably due to the creation of a cavity, whereas the folding transitions for V26, V26I and V26L appear to be similar, demonstrating the plasticity within the hydrophobic core and its ability to maintain stability when more bulky residues are introduced.
1.4. The Ubiquitin Analogues in this Study

In the past, protein variants have been made primarily through mutation using site-directed mutagenesis. The ubiquitin analogues studied in this project were chemically synthesised. This approach has the advantage of being able to incorporate unnatural amino acids at specific sites in the primary sequence, which otherwise would be impossible. It therefore represents a more sophisticated approach to probing factors that stabilise the tertiary structure. Incorporation of unnatural amino acids allows us to study the effects of a specific change rather than just a substitution with one of the limited set of coded amino acids. With the development of structural techniques such as high-field NMR it is possible to study subtle alterations in protein conformation at the atomic level.

This project is concerned with two analogues of ubiquitin, in each the hydrophobic core has been modified through replacement of two core residues. Packing within the hydrophobic core of ubiquitin involves three residues from the $\alpha$-helix (Ile 23, Val 26, and Ile 30) and eleven residues from the $\beta$-sheet (Ile 3, Val 5, Ile 13, Leu 15, Val 17, Ile 36, Leu 43, Leu 50, Leu 56, Ile 61 and Leu 67). In the first ubiquitin analogue to be considered here, Leu 50 and Leu 67 have each been replaced with 2S, 4S, 5-fluoroleucine (Fleu); this analogue will be referred to as UbFleu(50,67). In the second analogue Val 26 is replaced by
aminobutyric acid (Abu) and Ile 30 is replaced by norvaline (Nva); this analogue will be referred to as UbAbu(26)Nva(30).

1.4.1. Rationale behind the study of UbFleu(50,67)

The first analogue, UbFleu(50,67) has involved the incorporation of two heteroatoms into ubiquitin. The NMR-active isotopes, $^{13}$C and $^{15}$N, are regularly introduced into proteins to assist in resolution and the assignment of NMR spectra. However atoms that are not found naturally in a protein can also be used as NMR reporter groups so long as the nuclear spin gives a resolved and assignable signal. Fluorine nuclei are ideal for use in NMR since they offer sensitivities to detection that are comparable with that of protons. $^{19}$F is 100 % naturally abundant and its sensitivity to detection by NMR is 83% that of the sensitivity of $^1$H. The van der Waals radius of fluorine is 0.14 nm compared to 0.12 nm, the radius of hydrogen$^{64}$. $^{19}$F NMR of proteins cannot provide the highly detailed structural information obtainable from $^1$H NMR studies. The significant advantage of 'fluorine labelling', however, is the large range of chemical shifts obtained because of the high sensitivity of fluorine shielding to subtle changes in the local electronic environment of the protein structure. These are of significant benefit when detecting subtle conformational changes. Electron shielding in protons is affected primarily by the ring currents of nearby aromatic residues$^{65}$ causing local shifts of 0.5 ppm. Fluorine is a 'heavy' element and is surrounded by p-electrons and its shielding is
dominated by the paramagnetic term. The fluorine shift parameter is much larger than that due to aromatic ring currents. There have been $^{19}$F NMR studies on fluorine labelled proteins for more than 20 years. In these studies the local environment has been investigated mainly through incorporation of fluorinated aromatic amino acids. This is most popular because many of these are available commercially, including 2-, 3-, and 4- fluorophenylalanine.

Previous $^{19}$F protein NMR studies include 2-, 3- and 4- fluorophenylalanine and 4-fluorotyptophan-labelled egg white lysozyme, 4-fluorophenylalanine-containing carbonic anhydrase, and 6-fluorotyptophan-containing dihydrofolate reductase (DHFR). These and other fluorinated aromatic amino acid studies have been used to probe local conformational changes accompanying ligand binding. They have also yielded information about $^{19}$F shielding constants and $^{19}$F chemical shifts. A typical chemical shift range for fluorine labelled aromatic residue in proteins maybe very large, approximately 17 ppm, depending on the nature of the fluorinated amino acid and the protein. Studies have also shown that surface charge fields have little affect on $^{19}$F chemical shifts, whereas minor side chain environment changes cause 1-1.5 ppm shift differences in the majority of analogues studied.

Parallel studies using fluorinated aliphatic amino acids are required to provide a comparison to the fluorinated aromatic residue data. This work however has not been attempted until recently, because only fluorinated aromatic amino
acid residues are commercially available and the synthesis of fluorinated
aliphatic amino acids is non-trivial.

The stereospecifically fluorine labelled aliphatic amino acid used in this project
was synthesised by the group of Professor Douglas Young at the University of
Sussex\textsuperscript{73}, using an eleven step protocol. A single fluorine atom has been
stereospecifically substituted for a proton on a $\delta$ methyl group of leucine to
give 2S, 4S, 5-Fluoroleucine. (Fleu). The substitution of just one fluorine atom
into the amino acid should result in minimal steric perturbations to the protein.
This is an ideal mono-substituted isomer since the fluorine atom is in a
stereochemically well defined position in the isomer.

Previous studies have been carried out where the same 2S, 4S, 5-fluoroleucine
has been biosynthetically incorporated into dihydrofolate reductase and
analysed using $^1$H and $^{19}$F NMR\textsuperscript{74}. The $^{19}$F NMR displayed 12 peaks (one
containing two overlapped peaks) attributed to the 13 Fleu residues in the
DHFR analogue which are expected if all Leu sites are occupied by Fleu to
some extent. Yields of incorporation of Fleu at each leucine site ranged from
20\% - 25\%. This data provided an indication of the $^{19}$F chemical shift range
likely for fluorinated aliphatic amino acid; for DHFR it was found to be large
at 15.2 ppm compared to 17 ppm in a typical fluorinated aromatic amino acid.
In the current project, the novel amino acid 2S, 4S, 5-Fluoroleucine has been introduced into the hydrophobic core at two sequence specific positions and with 100% occupancy. This creates a highly unusual analogue since evolutionary selection has eliminated electronegative groups from the cores of most proteins.

![2S, 4S, 5-Fluoroleucine (Fleu)](image)

The Fleu residues were substituted into positions Leu 50 and Leu 67, to give the analogue UbFleu(50,67). Leu 50 and Leu 67 are situated on the fourth and fifth β-strands respectively and their side chains are orientated across the hydrophobic core in the native structure. The distance between the pro-substituted δC atoms of Leu 50 and Leu 67 is 4.2 Å.

Usually modification to proteins involve creation or filling of a cavity through replacement with another amino acid. However here the substitution is isosteric. Therefore any effects observed will not be due to size but electronic effects. This project has involved detailed structure determination and analysis using \(^1\)H and \(^19\)F NMR in both one- and two- dimensions.
1.4.2. Rationale behind the study of UbAbu(26)Nva(30).

The reasoning behind the synthesis of the analogue UbAbu(26)Nva(30) was based on previous work utilising natural amino acids to investigate the effects of subtle changes in a particular peptide or protein. The conclusions of these studies, some of which are described here will be tested using this analogue.

An investigation into helix forming tendencies of various non-polar amino acids has identified β-branched amino acids such as valine and isoleucine, as well as phenylalanine, to be helix destabilising. The side chain rotomer conformations permitted in an α-helix are severely restricted by β-branching and the bulky aromatic ring of Phe also restricts the side chain conformations that occur in α helix. Continuing these studies, side-chain interaction effects on α-helices were considered. An alanine based peptide (17 amino acids) was shown to form stable monomeric helices in water and was used as a model for further experiments. Aminobutyric acid and norvaline are unnatural straight chain and non-polar amino acids. These were substituted into positions i and i+4 of the alanine based α-helix and were shown not to interact such as to stabilize helix formation. Other replacements, using branched natural amino acids, in the same system were studied to provide a comparison. This study measured the stabilising effects of i, i + 4 and i, i+3 Leu-Leu, Val-Leu, Ile-Leu and Phe-Leu pairs. Leu spaced four residues from another Leu is measurably helix-stabilising relative to the corresponding i, i+3 pair but this is less true for
i, i+4 Val-Leu, Ile-Leu or Phe-Leu (cf. i,i+3 pairs). The results indicated that limited side-chain flexibility in an α-helix strongly favours the interaction between two non-polar residues to stabilise an isolated α-helix.

A separate stability study replaced a leucine residue in native lysozyme with norvaline using unnatural amino acid mutagenesis\textsuperscript{80}. This effectively eliminates a γCH\textsubscript{3} group from the leucine residue. Free energy calculations were conducted to determine the relative stability of the lysozyme mutant. The results suggest packing interactions have been lost in the mutant form. This is the major source of stabilisation. The subtle responses of the backbone also affected the magnitude of the loss of stability.

The second analogue to be considered in the current project has aminobutyric acid and norvaline substituted for valine 26 and isoleucine 30 respectively. These positions are separated by one turn of the α-helix in ubiquitin. In the native protein structure these residue side chains protrude from the same side of the α-helix into the hydrophobic core. These unnatural amino acid substitutions effectively eliminate the β-branching at positions Val(26) and Ile (30).

The net loss to the α-helix is two methyl groups; although this is a subtle modification, we are interested in fine structural and packing detail within the core as well as effects on stability.
Figure 1.5. *Unnatural amino acids substituted into the amino acid sequence of human ubiquitin in order to eliminate β-branching.*

The crystal structure of this analogue was published recently. In the current study the analogue has been re-synthesised and purified and the solution structure determined through high field $^1$H NMR. The crystal and solution structures have been compared in chapter six.
Chapter Two

Solid Phase Peptide Synthesis and Purification of Ubiquitin Analogues.

This chapter begins with a general introduction to solid phase synthesis (2.1) and the chemical theory and mechanisms involved (2.2). In section 2.3 and 2.4 the synthesis and purification of both analogues is described, together with results from each stage. Finally, the materials and methods (2.5), gives specific details of reagents and the synthesis, purification and preliminary characterisation methods used.

2.1. Introduction to Solid Phase Synthesis

Chemical synthesis of proteins has at least one clear advantage over recombinantly produced proteins. Protein synthesis can allow the incorporation of unnatural amino acids, D-amino acids and isotopic labels at specific sites in the primary sequence.

Solid phase peptide synthesis (SPPS) was first described in 1963 by Bruce Merrifield. SPPS uses a polymer to immobilise the growing peptide chain. The solid cross-linked polymer is functionalised with a linker to which the carboxy-terminal of the first amino acid is attached. The peptide chain is built up by a series
of coupling stages in which the carboxy-group is activated and reacted with the free amino group of the growing peptide. These are alternated with deprotection steps to remove the Nα protecting group. Once the required peptide chain is complete, it is cleaved from the resin with simultaneous removal of side chain protecting groups. SPPS is normally fully automated. It has advantages over solution synthesis since isolation and purification of intermediates are not required and unreacted reagents and side products are removed by filtration and washing.

Normally SPPS employs one of two Nα protected amino acids. These are Nα tertiary-Butyloxy carbonyl (Boc) (1) and 9-Fluorenylmethoxycarbonyl (Fmoc) (2) amino acids.

The Nα Boc group is cleaved using trifluoroacetic acid, whereas the Fmoc protecting group is cleaved using mild base conditions. Fmoc chemistry is more widely employed since in Boc chemistry special apparatus are necessary to facilitate the use
of strong acid when cleaving side-chain protecting groups and when cleaving from the resin.

In this study, peptides were synthesised using an orthogonal protection scheme whereby the base-labile Fmoc Nα-protection is complemented with the use of acid-labile side chain protection and an acid-labile peptide-resin linker. The side chain protection was as follows: t-butyl (Bu') ethers for serine, threonine and tyrosine; t-Bu esters for aspartic and glutamic acid; t-butoxycarbonyl (Boc) for lysine; τ-triphenylmethyl (trityl) for histidine; 2,2,5,7,8-pentamethyl chroman-6-sulphonyl (Pmc) for arginine and 4,4'-dimethoxybenzhydryl (Mbh) for asparagine and glutamine.

2.2. Fmoc Solid Phase Peptide synthesis

2.2.1. The Solid Support

A polystyrene-based resin with an acid-labile linkage is used for Fmoc SPPS. The peptides in this study were synthesised using 4-alkoxybenzylalcohol resin also called the ‘Wang’ resin; figure 2.1.
The first amino acid is loaded onto the linker manually, and subsequent amino acid couplings are automated. Cleavage of the peptide chain from this linker gives the C terminal carboxylic acid. For a successful synthesis, it is necessary that the resin swelling is effective in the solvents used since the majority of the functional sites are situated on the interior of the polymer matrix.

2.2.2. Amide bond coupling

The formation of an amide bond on the solid support involves prior deprotection of the terminal amino function that is held on the support. The attacking amino acid has a protected amino group and an activated carbonyl. Formation of the peptide bond involves nucleophilic attack of the free amine group followed by elimination of the electronegative activating substituent via a dihedral intermediate; figure 2.2.
Figure 2.2 *Direct electrophilic acylation mechanism.*

For efficient coupling, the protected carboxylic acid functionality is activated by substituting an electron withdrawing group. This enhances the electrophilicity of the carbonyl C and facilitates nucleophilic attack of the amine group. The formation of a symmetrical anhydride achieves this effect well. However the procedure is expensive since half of the protected amino acid species remains unused, (figure 2.3).
An alternative method for achieving the activation of carboxyl group is through the active ester formation. A commonly used coupling agents is 1-hydroxybenzotriazole (HOBt) (3). This methodology was developed by König and Geiger. Derivatives of this agent have been developed, including ethyl-1-hydroxy-1H-1,2,3-triazole-4-carboxylate (HOCt), (4). Currently, the latter is used in Professor Ramages synthetic peptide group (University of Edinburgh). Preparation of HOCt activated esters are made immediately before use in the presence of N, N'-diisopropylcarbodiimide (DIC).
2.2.3. Fmoc-SPPS monitoring

The extent of coupling of amino acid is assessed after each coupling cycle. Coupling efficiencies often drop as the synthesis progresses due to effects such as steric hindrance. The Fmoc group forms a dibenzofulvene adduct which is subsequently quenched with piperidine. UV monitoring at 302 nm of the fulvene-piperidine adduct gives an estimation of the coupling efficiency.
Figure 2.4. Schematic diagram of Fmoc SPPS: (i) attachment of first amino acid to the resin; (ii) capping with acetic anhydride; (iii) deprotection of Nα Fmoc with piperidine; (iv) coupling of the next amino acid and (v) simultaneous cleavage from the resin and from protecting groups.
2.3. Synthesis and Purification of UbAbu(26)Nva(30)

2.3.1 Synthesis of UbAbu(26)Nva(30)

Automated peptide synthesis of UbAbu(26)Nva(30) was carried out according to the strategy described previously. The synthesis was carried out on a 0.1 mmol scale and Fmoc deprotection was monitored at 302 nm. All amino acids were single-coupled using the HOCt ester (1 mmol) except for the His residue and also the final ten amino acids of the sequence. The progress of the synthesis was monitored by detection of the Fmoc-fulvene adduct at 302 nm, figure 2.5. Some coupling yields appeared to be greater than 100 %, this is an artefact caused by swelling of the resin.

![Figure 2.5. Progress of the UbAbu(26)Nva(30) synthesis, monitored by UV at 302 nm.](image-url)
His residues are prone to racemisation when activated due to intramolecular base catalysis by the lone pair of π electrons on the nitrogen of the imidazole ring. This effect is reduced by trityl protection and by using the HOBut active ester (1 mmol) which suppresses αCH abstraction.

Previous syntheses of ubiquitin indicated that the coupling efficiency had a tendency to fall during the coupling of the last ten residues. For this reason these residues were ‘double coupled’. The N-terminal Fmoc group was not removed at the completion of assembly, and the resin bound product was capped by acetic anhydride. This was then washed and dried to give 1.685 g of resin-bound product.

2.3.2. Purification of UbAbu(26)Nva(30)

The analogue UbAbu(26)Nva(30) was purified by following a three stage protocol. This involved an initial purification using the hydrophobic molecule, 17-tetrabenz(a,c,g,i) fluoromethoxycarbonyl (Tbfmoc) (5), in combination with porous graphitised carbon. The protein was then further purified using gel-filtration and finally ion-exchange chromatography yielded folded material.

Prior to eventually obtaining folded material by means of the cation exchange chromatography step, many denaturing and refolding routines were attempted. Efforts to renature the peptide from denaturing conditions using dialysis led to aggregation and precipitation. A refolding protocol was followed whereby the urea-
denatured protein was bound to a cation-exchange column. A slow gradient from 0 % urea to 100 % urea was passed through the column. The protein was subsequently eluted from the column via a salt gradient. UV monitoring at 280 nm revealed that only ca 20% of the total protein loaded was eluted from the column at the expected salt concentration. It is thought that the remainder of the protein either did not bind to the column in the first place or was eluted in a prolonged high-salt wash at the end of the experiment. (Details of these protocols are given in section 2.5.4.6.)

2.3.2.1. Purification using Tbfmoc-Cl.

Incomplete coupling stages in SPPS often leave acetylated deletion sequences bound to the resin. An efficient methodology has been developed to remove these\textsuperscript{88,89}. It uses the hydrophobic, base-labile molecule 17-tetrahydrobenzo(a,c,g,i)fluoromethoxycarbonyl (Tbfmoc) (5) group to derivatise the N-terminus of the protein. The high affinity that this polycyclic aromatic label has for porous graphitisied carbon (PGC) allows the Tbfmoc-peptide to be separated from the truncated sequences which are washed away. The target sequence is then cleaved from the solid support. A schematic representation of this protocol is shown in figure 2.6.
Figure 2.6. Schematic representation of the Tbfmoc purification method.
The use of Tbfmoc resulted in a significant degree of purification. Figure 2.7a and b compares a RP-HPLC traces of crude Tbfmoc-UbAbuNva with that after purification on PGC. The percentage of total protein recovered from this stage was around 40%. A major contributor to the low yield was that the PGC was difficult to remove from the suspension. This meant that some resin-bound product was discarded along with the deletions in the supernatant.
2.3.2. Size exclusion Gel Filtration

The protein was further purified by gel-filtration using a G30 Superdex column in a series of six separate runs. The RP-HPLC trace, shown as an inset in figure 2.8, demonstrates that the material was further purified after gel filtration. The mass obtained from MALDI MS data (8533 Da) was consistent with the target material (8537 Da). However a 1D $^1$H NMR spectrum of this material indicated that it was not properly folded, (figure 2.8). The resonances in this spectrum are broad and appear in clusters. There is a lack of signals >8.6 ppm and < 0.7 ppm which are where well-dispersed peaks would be expected if the protein were folded.
2.3.3. Cation Exchange Chromatography.

A cation exchange step using a SP-Sepharose column and running a 0 % to 100 % NaCl gradient yielded only folded peptide albeit in small yields of \( ca \) 20 %. The expected mass of the purified material from the cation exchange stage was confirmed through ESI-MS, this is displayed along with the RP-HPLC in figure 2.9a and b. The major peak in the mass spectrum (8536 Da), is within experimental error of the
expected mass (8537 Da). The other large peak (8542 Da) is a deuterated species caused by the ESI-MS sample being taken from a NMR sample containing 10% D₂O. Preliminary 1D ¹H NMR shown in figure 2.9c indicated that the material was folded with a good dispersion of resonances throughout the spectrum and distinct resonances in the amide region.
Figure 2.9. (a) ES - MS of; (b) RP - HPLC trace monitored at 240 nm and (c) 1D $^{1}H$ NMR of the purified UbAbu(26)Nva(30) after cation-exchange.
The thermodynamics of unfolding for UbAbu(26)Nva(30) was followed through Differential Scanning Calorimetry (DSC)\textsuperscript{90-93} and compared to that of bovine ubiquitin. The results (figure 2.10) were not as expected, since the trace revealed a very low (undetectable) thermal transformation (in conflict with the NMR data), suggesting that there is a complete absence of tertiary structure in the protein sample.

![DSC trace of bovine ubiquitin and UbAbu(26)Nva(30)](image)

**Figure 2.10.** *DSC trace of bovine ubiquitin and UbAbu(26)Nva(30)*

### 2.4. Synthesis and Purification of ubiquitin analogue UbFleu(50,67)

Synthesis and purification of the analogue UbFleu(50,67) was carried out by Jill Wilken\textsuperscript{94}. The analytical RP-HPLC of this purified analogue was as highly resolved as that of bovine ubiquitin. However the retention time from the RP-HPLC column
was less, \( R_t = 17.4 \) mins compared to \( R_t = 17.6 \) mins for bovine ubiquitin (figure 2.11).

![Graph](image)

**Figure 2.11.** *RP - HPLC monitored at 240 nm of UbFleu(50,67) after purification.*

Far-UV studies of UbFleu(50,67) were carried out to probe the secondary structure content\(^4\). The results were compared to those found for bovine ubiquitin. The \( \alpha \)-helical content was found to be *ca* 20\% for both the synthetic analogue and the native protein. However, the apparent \( \beta \)-sheet content reduced from 56\% in bovine ubiquitin to 30\% in the fluorinated analogue.

The ESI-MS indicated of the purified UbFleu(50,67) displayed a mass (8602 Da) which is close to the expected mass (8601 Da), figure 2.12a. The resulting spectrum of the analogue UbFleu(50,67), figure 2.12b, displayed good dispersion and sharp well resolved peaks. In
particular, the low field resonances at 8.5 - 9.5 ppm and the methyl resonances below 0.0 ppm indicate that the analogue has a folded conformation with well defined secondary and tertiary structural elements. A large resonance at 1.97 ppm is due to an acetate contamination. This impurity later proved to be a source of problems in the resonance assignments in the 2D spectra.

Figure 2.12. (a) ESI-MS and (b) 1D $^1$H NMR UbFleu(50,67) after purification.

Differential scanning calorimetry was performed on UbFleu(50,67) and bovine ubiquitin. The thermal transition for UbFleu(50,67) and bovine ubiquitin is identical, both having a mid-point temperature ($T_m$) of 70 °C, figure 2.13. However the height of the normalised peak, and the area under the peak are much lower. This is difficult to explain since the NMR spectra are
consistent with a predominantly folded conformation of the protein. The DSC results are discussed further in chapter six.

Figure 2.13. *DSC trace of bovine ubiquitin and UbFleu(50,67)*

2.5. Materials and Methods

2.5.1. Synthesis Equipment and Reagents

Peptide synthesis was carried out on an Applied Biosystems 430A automated peptide synthesiser.

A Decon FS300b sonic bath was used for sonication of samples.
Peptide synthesis-grade dimethylformamide (DMF), 1-4-dioxan and piperidine were supplied by Rathburn Chemicals.

N-N- Diisopropylethylamine (DIEA) and trifluoroacetic acid (TFA) were both peptide synthesis-grade and were purchased from Applied Biosystems.

1-Hydroxybenzotriazole (HOBt) was bought from Aldrich. Ethyl-1-hydroxy-1H-1,2,3-triazole-4-carboxylate (HOCt) was synthesised within the Chemistry Department at the University of Edinburgh.

25.2. Resin and amino acids.

4-Alkoxybenzylalcohol-functionalised polystyrene resin was supplied by Bachem.

9-Fluorenlymethoxycarbonyl (Fmoc)-protected amino acids were purchased from Bachem or Novabiochem.

2S, 4S, 5-Fluoreleucine (Fleu) was synthesised by B.A. Starkmann, University of Sussex, UK.

All amino acids were of L-configuration.
2.5.3. Purification equipment and reagents.

High performance liquid chromatography (HPLC) was carried out using either an ABI system having two 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 manomeric module and a 119 UV/VIS detector.

A Sanyo Mistral 2000R apparatus was used for centrifugation.

Gel-filtration and cation exchange chromatography was carried out using a Pharmacia GradiFrac apparatus.

A G30 HiLoad Superdex column (60 cm x 16 mm) was used for all gel filtration experiments. Cation exchange experiments used a SP-Sepharose column (16 cm x 10 mm).

Ultraviolet (UV) absorption measurements at 280 nm were recorded on a CECIL 1000 series spectrophotometer in the solvents indicated.

pH measurements were done using a Jenway 3420 electrochemistry analyser using a Walden precision pH probe.
Concentration was achieved using Centricons, Centripreps, Microcons or a pressure cell, all having a molecular weight cut off (MWCO) membrane of 3 KDa. These were all supplied by Amicon.

BioRad Econo-Pac 10 DG (10 ml) disposable columns were used for desalting protein samples.

The DSC experiments were performed on a Microcal MC-2D instrument.

17 - Tetrabenzo (a,c,g,i)fluorenylmethylchloroformate (Tbfmoc) was synthesised within the Chemistry Department at the University of Edinburgh.

Porous graphitised carbon (PGC) was supplied by Shandon Scientific.

HPLC-grade acetonitrile (CH$_3$CN) was supplied by Rathburn Chemicals and HPLC grade TFA was obtained from Fisons.

Refolding protocols used dialysis tubing purchased from CelluSep with MWCO 2000 Da.

All NMR spectra, including subsequent structural experiments described in Chapter three were recorded on either a Varian VXR 5000 or a Varian INOVA spectrometer, both operating at 600 MHz proton frequency.
The ESI-MS were carried out on a Micromass Quatro triple quadropole mass spectrometer and the DSC experiments were performed on a Micocal MC-2D instrument.

2.5.4 Synthesis and purification

2.5.4.1. Coupling of the C-terminal amino acid to the 4-alkoxybenzylalcohol resin.

The first amino acid was attached to the 4-alkoxybenzylalcohol functionalised polystyrene resin manually. Fmoc-Gly-OH amino acid (1.92 mM) was activated as the symmetric anhydride by treatment with N,N’-diisopropylcarbodiimide (DIC) (0.9 mM in 10 ml DMF and sonication for 15 minutes at room temperature. The 4-alkoxybenzylalcohol ‘Wang’ resin (1.0 g, 0.96 mM) was added together with a catalytic amount of 4-(N,N’-dimethylamino)-pyridine and left at room temperature for 60 mins. The loading of the amino acid was then determined by treating a small quantity (7.6 mg) with 20 % piperidine in DMF in a 10 ml volumetric flask for 15 minutes in a sonic bath. The UV absorbance was measured at 302 nm and the loading calculated using the Beer-Lambert law ($\epsilon_{302} = 15,400 \text{ dm}^3 \text{ mol}^{-1} \text{ cm}^{-1}$ for the fulvene-piperidine adduct.)
The synthetic procedures were pre-programmed into the synthesiser prior to the commencement of synthesis. The pre-programmed synthetic cycle is as follows:

1. Capping - The unreacted functional sites are capped. The resin is vortexed in a solution of 0.5 M acetic anhydride, 0.125 M DIEA, and 0.2 % w/w HOBt in DMF (10 ml) for 10 minutes. The capping solution is then drained from the vessel and the resin is washed with six portions of DMF/1,4-dioxan (1:1).

2. Deprotection - The Nα Fmoc protecting group is cleaved using 20 % v/v piperidine in DMF (10 ml) for 3 minutes before being drained. An aliquot of the deprotection solution is then sent to a UV detector in order to quantify the amount of fulvene-piperidine adduct present and hence an indication of the incorporation of the residue. Deprotection is repeated a second time for a period of one minute to establish if the Nα protecting group has been completely removed. Finally the resin is washed with six portions of DMF/1,4-dioxan (1:1).

3. Coupling - The amino acids were coupled as active ester of HOOb and Fmoc amino acid (1 equivalent) in DMF/1,4-dioxan (1:1 v/v) (8 ml). His was coupled as the HOOb ester. After vortexing for 30 minutes the reaction vessel is drained and washed with six portions of DMF/1,4-dioxan (1:1 v/v).
The synthesis of UbAbuNva was carried out on a 0.1 mmol scale using the functionalised resin Fmoc-Gly-(OCH₂C₆H₄OR) (0.57 g 1.92 mmol). All amino acid were single coupled using the HOOct ester (1 mmol) except for the His residue and also the final 10 amino acids of the sequence. The side chains were protected as previously described. The N-terminal Fmoc group remained attached at the completion of the assembly and the resin bound product was capped by sonication for 60 minutes in 0.5 M acetic anhydride, 0.125 M DIEA, and 0.2% w/v HOBt in DCM (20 ml). This was then washed sequentially with DMF, 1,4-dioxan and dichloromethane and dried to give 1.685 g of resin-bound product. This was stored in 1,4-dioxan at 4°C until required.

2.5.4.3. Tbfmoc purification methodology

The N-terminal Fmoc protecting group was first removed from 264 mg of resin-bound peptide using 20 % v/v piperidine/DMF (20 ml). The resin was then thoroughly washed with DMF (75 ml) and DCM (75 ml) and dried under suction.

The dry resin-bound protein was then added to a suspension of TbfmocCl (20.8 mg, 0.045 mmol) in dichloromethane (6 ml) and N, N’-DIEA (7.7 μl, 0.045 mmol). The reaction flask was sealed and covered in aluminium foil before being sonicated for three hours. The Tbfmoc-peptide-resin was filtered and washed with dichloromethane (75 ml) and then diethyl ether (75 ml). The resin was dried and the functionality found to be 0.042 mmol/g by UV at 364 nm.
The Tbfmoc resin was then added to a cleavage mixture of 0.757 g phenol, 0.5 ml thioanisole, 0.5 ml H$_2$O, 10 ml TFA and 2 ml EDT. This was stirred under nitrogen in the dark for 4 hours. The resin was separated from the Tbfmoc-peptide by filtration. The resin was washed with trifluoroacetic acid (5 ml) and dichloromethane (5 ml). The filtrate was concentrated in vacuo to give a yellow oil and the protein was precipitated with 2 % β-mercaptoethanol in diethyl ether (100 ml). The solid Tbfmoc-protein was then filtered and dissolved in 20 % v/v acetonitrile in H$_2$O and lyophilised.

The peptide was dissolved in 12.5 ml 6 M guanidine hydrochloride (GuHCl), to ensure solubility and to denature the protein, and diluted with 12.5 ml propan-2-ol. This was added to 600 mg of PGC. The absorption of the Tbfmoc label was monitored at 364 nm by HPLC. Once the product had been absorbed, the carbon was washed twice with 50 % v/v 6 M GuHCl/isopropanol (10 ml), to eliminate any deletion peptides.

The carbon-bound peptide was deprotected by addition of 10 % v/v piperidine in 6 M GuHCl/isopropanol (40 ml). This was mixed and centrifuged for 30 mins (4500 rpm, 3600 g, 25 °C). The peptide in the solution obtained was neutralised with 100 % acetic acid to pH 4.5. The isopropanol was removed in vacuo. The protein solution was then lyophilised.
2.5.4.4. Size Exclusion Gel Filtration

A 10 ml protein sample (concentration 2 mg/ml) was then applied to a G30 Superdex HiLoad size exclusion gel column (60 cm x 16 mm) using a 20 mM CH$_3$COONa buffer at pH 4.8. The flow rate of buffer was 0.5 ml/minute and 2 ml fractions were collected whilst monitoring at 280 nm. Fractions 57 to 60 were combined.

2.5.4.5. Cation exchange chromatography

A SP-Sepharose (16 cm x 10 mm) cation exchange column was equilibrated with 20 mM sodium acetate buffer at pH 4.8 and the peptide applied (5 ml of concentration approximately 2 mg/ml). A salt gradient of NaCl from 0 M to 0.5 M was run, the flow rate was 0.7 ml/min, and 1 ml fractions were collected over 2 hours. Fractions 31-35 were combined and lyophilised.

2.5.4.6. Refolding

Three separate refolding protocols were attempted.

1) The protein (ca. 20 mg) was dissolved in 20 ml N$_2$-saturated solution of 8 M urea; 50 mM NH$_4$OAc; 5 mM DTT at pH 4.5. The material was dialysed sequentially against a gradient of seven N$_2$-saturated solutions, (6 M, 5 M, 4 M, 3 M, 2 M, 1 M and 0 M urea in 50 mM NH$_4$OAc; 5 mM DTT (800 ml) ), each for 24 hours at 4 °C.
Finally the material was dialysed against nitrogen saturated 1 mM DTT; 0.1% v/v Na azide, pH 7.0 (3000 ml) for 24 hours.

II) Protein was dissolved in 20 ml N₂- saturated solution of 8 M urea; 50 mM NH₄OAc; 5 mM DTT at pH 4.5. This was dialysed against 50 mM NH₄OAc; 5 mM DTT (3000 ml) at 4 °C for 24 hours.

III) A SP-Sepharose cation exchange column was equilibrated with a degassed solution of 8M urea, 20 mM NaOAc, 10 mM DTT, pH 5.0 at 25 °C. The protein sample (ca. 10 mg) was dissolved in 1 ml of this solution and applied to the column. A urea gradient from 8 M to 0 M at a flow rate of 1 ml/min was run over two hours. A NaCl gradient from 0 M to 1.0 M was then used with a flow rate of 1 ml/min over 80 minutes.

2.5.4.7 Yield

From an initial loading of resin-bound Fmoc-Gly-OH of 1.92 mmol (0.57 g), the synthesis yielded 1.685 g of resin bound product.

A 264 mg portion of Tbfmoc resin bound product was cleaved and purified on PGC to give 120 mg of free peptide.
The size-exclusion column gave a yield of approximately 17% of target material for each of the six runs. These were pooled and 21 mg applied to the cation exchange column in two runs. A total of 4.7 mg of folded material was obtained. This equates to a yield of 4% as calculated below:

\[ \text{Mass of end product (4.7 mg)} + \text{Mass of free peptide (120 mg)} \times 100 \approx 4\%. \]

2.5.5. Simple Protein Characterisation Methods

2.5.5.1. One dimensional $^1$H NMR

Preliminary one dimensional $^1$H NMR experiments incorporating solvent (water) suppression indicate the quality of folding and aggregation. For example, a folded protein will display a well-dispersed array of peaks throughout the spectra since hydrophobic aromatic residues within the core of a protein can considerably shift the resonances of nearby protons. Particular indications of a folded peptide include the appearance of methyl resonances with chemical shifts less than 0.7 ppm and alpha protons of $\beta$-sheet with shifts between 5-6 ppm.

In an unfolded protein the resonances will coincide, grouping in distinct regions. Also, the line-width can indicate whether a protein is unfolded or not. For example a protein with no defined structure (i.e. random coil) will display sharp resonances since the polypeptide chain has free and rapid movement. In the case of a partly folded protein, the lines will be broad since several conformations are exchanging with one another.
The pulse sequences for these preliminary 1D \(^1\)H NMR experiments are shown in chapter three, (figures 3.4a and 3.4b) and the specific experimental details are given in figure 3.9.

2.5.5.2. Electrospray Ionisation Mass Spectrometry (ESI-MS)

Electrospray ionisation mass spectrometry (ESI-MS) is based on an idea originally proposed by Dole\(^9\) in 1968 but was not successfully implemented at the time. ESI-MS involves a process whereby molecules move directly from the liquid phase into the gas phase in an ionised state. The ESI-MS were carried out on a Micromass Quatro triple quadropole mass spectrometer. The protein concentration was 20 pM/\(\mu\)l in a \(\text{CH}_3\text{CN/H}_2\text{O} (1:1)\) buffer containing 0.1 % TFA.

A protein solution is passed through a needle that is kept at a high electrical potential. At the end of the needle, the solution disperses into a mist of small, highly charged droplets. The solvent evaporates quickly and protonated protein ions are released into the gas phase. These are then 'leaked' into the vacuum of a mass spectrometer, where they are separated and detected according to their mass-to-charge ratio\((m/z)\) on the basis of different trajectories in electric fields. The resulting mass spectrum is a series of peaks caused by multiple charging of the protein molecules. This series is 'deconvoluted' by a computer algorithm into a single peak which is the protein mass.
2.5.5.3. Differential Scanning Calorimetry (DSC)

Transformations of proteins between various conformational states maybe brought about by changes in temperature, pressure or pH. Temperature-induced protein unfolding (at equilibrium) arises from differences in enthalpy ($\Delta H$) between folded and unfolded states. The thermodynamics of unfolding can be followed using Differential Scanning Calorimetry (DSC)\(^96\). This was carried out on both protein analogues and the results compared to that of bovine ubiquitin.

In a DSC experiment, a solution of protein is heated at a constant rate in the calorimeter cell alongside an identical reference cell containing buffer. The differences in heat energy uptake between the sample and the reference cell required to maintain equal temperature correspond to the differences in heat capacity. It is these differences in heat capacity that give direct information about the energetics of thermally induced processes in the sample. The buffer used in each case was 50 mM glycine.HCl at pH 3.0. The DSC was performed on a Microcal MC-2E instrument at a rate of 60° per hour over 5 to 100 °C temperature range.

The mid-point temperature of the transition, ($T_m$), is the point at which 50% (on average) of the protein molecules are unfolded. The integrated area beneath the peak in the DSC thermogram divided by the total amount of protein in the calorimeter cell gives the calorimetric enthalpy (heat uptake, $\Delta H_{ca}$) for the unfolding transition.
Chapter Three

Nuclear Magnetic Resonance Studies of Ubiquitin analogues

3.1 Introduction to NMR

NMR spectroscopy is widely used in protein studies\textsuperscript{100}. The increased availability of superconducting magnets with high-fields, along with advances in computational capacity has made this technique very powerful\textsuperscript{101}. The development of NMR as a structural technique is important since it offers a solution structure for comparison with X-ray determined crystal structures. Moreover, an additional method for protein structure determination can only lead to an increase in the number of solved structures and hence a clearer understanding of the relationship between amino acid sequence and protein structure.

Nuclear magnetic resonance is based on the fact that nuclei of non-zero spin, placed in a strong magnetic field, will absorb applied radiation. Protons are usually the only atoms in proteins with spin quantum $I=1/2$, except for $^{13}\text{C}$ and $^{15}\text{N}$ which are in low natural abundance. All protons can be observed except labile protons of -NH, -NH$_2$, -OH and -SH groups which exchange rapidly with aqueous solvent.
Each proton within a structured protein molecule experiences a slightly different field strength since the local magnetic field is directly influenced by the electronic environment, e.g., involvement of the proton in hydrogen bonds, proximity to aromatic and carbonyl moieties. It is these small differences which are measured by NMR. Depending on the environment of the nucleus, a particular frequency of electromagnetic radiation will be absorbed, giving rise to a distinct spectral line. Interpretation of the spectra gives information on molecular structure and conformation. The structural determination of a protein usually involves the analysis of two (or higher), dimensional NMR spectra. However a preliminary one-dimensional (1D) NMR can reveal the extent to which the protein is structured, i.e. give an indication of the extent of folding and aggregation.

Proton $^1$H NMR is not a suitable method for the structural study of relatively large proteins (> 30 KDa). This is partly because larger proteins have more crowded spectra due to the presence of many resonances which will overlap. This results in the spectra being too complicated to interpret. Design of sophisticated experiments and introduction of heteronuclei into proteins means that larger proteins can be studied using three and four dimensional NMR techniques.

In addition to allowing solution structure determination, NMR can provide valuable information which cannot be derived from x-ray crystallography. This includes the determination of ionised states, measurement of dynamic motion within the protein such as the rate of ring flipping and the study of solvent exchange rates.
3.2 Theory of the 1D $^1$H NMR experiment

3.2.1 Nuclei in magnetic fields

Proton nuclei in a magnetic field adopt orientations where the nuclear magnetic moment is either parallel (the lower energy $\alpha$ state) or antiparallel (the higher energy $\beta$ state) to the applied magnetic field, $B_0$. At equilibrium, the population of states $\alpha$ and $\beta$ are determined according to the Boltzmann distribution. The bulk magnetisation is aligned with the applied field in the $\alpha$ state. The energy differences between these two spin states is proportional to the magnetic field.

Figure 3.1. Precession of nuclear spins around the magnetic field, $B_0$. The net magnetisation is aligned parallel to the field, and the net magnetisation $M_0$ indicated. Adapted from Derome (1991).
The system may be disturbed from this equilibrium causing a transition from the $\alpha$ to the $\beta$ state. This nuclear transition is made by applying a pulse of radiofrequency (rf) electromagnetic radiation, $B_1$, which is equal to the Larmor frequency, $\nu$. The Larmor frequency is related to the strength of the magnetic field, $B_0$, and expressed in equation (1) where $\gamma$ is the magnetogyric ratio.

$$\nu = \gamma B_0 / 2 \pi$$

(1)

3.2.2. Application of a radiofrequency pulse

A pulse of electromagnetic radiation, $B_1$ is applied along the $x$-axis perpendicular to $B_0$. The duration of the applied pulse is proportional to the magnetisation flip angle. Therefore a 90° pulse flips the net magnetisation, $M_z$, from the $z$-axis onto the $y$-axis. The spins do not remain aligned along the $y$-axis but under the influence of $B_0$ begin to precess and dephase in the $xy$-plane. The rotating vector is denoted $M_{xy}$. 
3.2.3. Relaxation of magnetisation, $M_{xy}$.

During a period after the pulse, the induced magnetisation $M_{xy}$ relaxes exponentially over a period of time, to its equilibrium state, $M_z$. There are two types of relaxation, spin-lattice (longitudinal) relaxation and spin-spin (transverse) relaxation. Spin-lattice relaxation involves the nuclei relaxing back to their original Boltzmann population, and occurs over time, $T_1$, the longitudinal relaxation time. Spin-spin relaxation refers to the disruption of phase coherence in the xy-plane and is denoted as $T_2$, the transverse relaxation time.

The line-width of a particular resonance can give an indication as to the value of $T_2$. These relaxation times are dependant on both the dynamic motion of the molecule and the rate at which the molecule tumbles in solution. For example since large proteins tumble more slowly,
their relaxation is more rapid and transfer of magnetisation less efficient. This leads to broader peaks and a loss in sensitivity.

3.2.4. Coupling

Scalar coupling, or J-coupling gives resonances that are split into multiplets of peaks in a way that is dependant upon their chemical nature. The local magnetic field of a particular nucleus, A, will affect the spin orientation of a near-by nucleus B. Since nucleus A has two possible orientations, nucleus B experiences two slightly different rf fields. Therefore in an NMR experiment the B resonance will appear as a doublet with a 1:1 ratio of intensities, representing the two Larmor frequencies. The separation between the peaks of the doublet is given by the coupling constant, J, measured in Hertz (Hz).

When there are more than two protons involved, the pattern of splitting becomes more complex. Usually, scalar coupling occurs between protons connected by up to three bonds. The value of the coupling constant depends on the type of nuclei, the bond length, and the dihedral angle, $\phi$. 
The relationship between coupling constant and dihedral angle is determined by the Karplus equation (2), where $P$, $Q$, and $R$ are constants. For coupling between the $\text{C}_\alpha\text{H}$ and $\text{C}_\beta\text{H}$ of amino acids, $P$, $Q$ and $R$ are 9.5, 1.6 and 1.8 respectively. For NH to $\text{C}_\alpha\text{H}$ couplings, $P$, $Q$ and $R$ are 6.4, 1.4 and 1.9.

$$J = P\cos^2\phi + Q\cos^2\phi + R \quad (2)$$

Dipolar interactions are involved in relaxation and are the basis of the so-called nuclear Overhauser effect (nOe). Nuclei which have undergone a transition to a higher energy state need to be stimulated in order to relax back to their original position (since spontaneous relaxation is very rare). Because there are a limited number of suitable stimuli present within the system, relaxation has to follow well-defined pathways. Local magnetic fields which are
created by neighbouring nuclei, cause a particular magnetic dipole of a nucleus to align itself with or against the applied field. This dipole remains this way as it tumbles in solution and thus any other neighbouring nuclei experience a fluctuating magnetic field which provides a relaxation pathway. This effect is important for nuclei within 5.5 Å of each other.

3.2.5. Detection.

The relaxation of magnetisation to equilibrium is recorded using a detector coil positioned close to the sample in the xy-plane. The current induced in the coil by the precessing magnetisation, varies sinusoidally and oscillates with the frequency of $M_{xy}$. As transverse relaxation occurs, the signal decays exponentially to zero and a 'free induction decay' (FID) is recorded as a function of time. The FID consists of a superimposition of all the separate signals, each with a different frequency arising from non-equivalent protons within the sample, together with noise.

An analogue to digital converter (ADC) operates to convert this signal into a digitised form which then can be used by the computer. The sample is pulsed many times and the FIDs detected and summed. This aids the cancellation of baseline noise and therefore enhances the signal. The component signals of the FID are extracted using a mathematical function called a 'fourier transformation' which produces the familiar frequency domain spectrum.
3.2.6. Solvent suppression

The solvent used in protein NMR is commonly aqueous. There are several reasons why the solvent signal has to be suppressed. The proton concentration in aqueous solution is high (110M) which means that the water resonance is so large that it swamps proton resonances from the protein. There are also problems concerning the solvent signal with regard to the instrumental capabilities. Receiver coils can only work over a limited dynamic range and will have difficulty detecting both very strong resonances and very weak resonances without introducing distortion. Also, since the detected signal must be converted into a binary number by the ADC, any protein signal that is smaller by a factor of, say, $2^{15}$ (depending on the ADC), than the largest solvent signal will be lost. The water resonance must therefore be eradicated in so far as possible. There are a wide range of methods available for water suppression. In this study, the water resonance has been removed from spectra either through selective saturation at the water resonance or by employing a ‘WATERGATE’ (WATER suppression by GrAdient-Tailored Excitation)$^{112}$ pulse sequence.

The simplest method involves saturation of the water resonance by prolonged irradiation at its Larmor frequency. This equalises the populations of the two possible energy states so that no signal is observed. This method suffers from the disadvantage that the saturation effect may be transferred to the amide protons causing their signals to disappear. Figure 3.4a shows a 1D presaturation pulse sequence using a prolonged low power pulse to irradiate the water signal.
The so called 'WATERGATE'\textsuperscript{112} sequence, (figure 3.4b), employs a more sophisticated pulsing method. This involves selective 90° pulses being applied (at the water frequency) either side of a 180° pulse. Essentially the water resonance experiences a net 360° magnetisation flip while the rest of the protons in the sample undergo a 180° rotation. The first pulsed field gradient (PFG) dephases the magnetisation in the xy-plane whilst the second rephases only the protons which have undergone a magnetisation flip of 180°. The water resonances are not refocused and not detected.

Figure 3.4a. 1D PRESATURATION experiment.

Figure 3.4b. 1D WATERGATE experiment
Two dimensional NMR was first introduced by Jenner in 1971\textsuperscript{113} and subsequently developed by Ernst and co-workers\textsuperscript{114}. Spectra of two or more dimensions are required for structure determination of large proteins.

All 2D experiments involve a preparation period, an evolution period, $t_1$, during which the spins are labelled according to their chemical shift, a mixing period during which the spins are correlated with each other, and an acquisition period, $t_2$. A basic 2D pulse sequence is shown in figure 3.5.

![Figure 3.5. A Basic 2D pulse sequence.](image)

During the preparation period, the sample reaches thermal equilibrium. A pulse of radiofrequency is applied to create phase coherence in the xy-plane. In the evolution period, $t_1$, the magnetisation is allowed to build up. A second radiofrequency pulse, the mixing pulse, allows the transfer of phase coherence to take place. The FID is recorded over the acquisition period, $t_2$. 
With each repetition of the pulse sequence the evolution period increases by a fixed increment. The first FID has a very small evolution period and subsequently a series of FIDs are collected each having a slightly longer evolution period than the previous one. As the evolution period increases, the time for which the magnetisation stays in the xy-plane and precesses about the z-axis also increases. The receiver coil detects the extent of movement in the xy-plane with increasing incremental delay and builds up a picture of a ‘pseudo-FID’ which reflects the evolutions of nuclei chemical shifts over t1. The importance of the second or ‘mixing’ pulse is to allow some type of magnetisation transfer between nuclei depending on what information is required from the experiment (i.e. through-bond or through-space information). The extent of magnetisation transfer depends on the length of the mixing pulse and also the strength of the nuclear interaction.

Homonuclear 2D spectra can be considered as correlation maps which link the protons of the molecule together. Homonuclear 2D spectra are represented as contour plots and are symmetrical about a diagonal line of peaks which corresponds to the 1D spectrum. There are two types of signal visible on a 2D NMR spectrum; these occur as diagonal peaks and cross-peaks. In order to explain how these peaks arise we must consider what is happening in the 2D experiment. During the evolution time the nuclei are precessing at their own Larmor frequency. When the mixing pulse is applied a particular nucleus, say Ha, may become ‘labelled’ with the magnetisation from a neighbouring nucleus, Hb, (whether coupled or close in space). The resulting frequency of Ha after the mixing pulse is therefore different than in t1. Sometimes magnetisation transfer does not occur and therefore Ha would have the same frequency after the mixing pulse. These two situations are manifested in a 2D spectrum (after
fourier transformation), as cross peaks where magnetisation transfer does take place and a diagonal peak where it does not.

The basic 2D pulse sequence is tailored to give different types of information about the molecule. It is the mixing period which distinguishes the various experiments. The COSY experiment gives cross-peaks derived from scalar couplings occurring between nuclei linked through less than three bonds. The TOCSY experiment is capable of displaying cross-peaks for an entire spin system. The most important structural information comes from the NOESY experiment where through-space information is obtained from the intensity of cross-peaks which correspond to nOes between the appropriate protons.

3.4. Homonuclear NMR

3.4.1. The COSY experiment

The COSY experiment\textsuperscript{115} shows the presence of through-bond coupled spins. The major application in biological NMR is the assignment of resonances in the spin systems of amino acid residues. Magnetisation is transferred to the proton on the adjoining carbon atom by the mixing pulse. Spin couplings are not transmitted through the peptide bond so the amide, CoH and side chain protons of each amino acid together form an isolated system of coupled spins called a spin system. Each spin system gives a characteristic pattern of crosspeaks due to the connectivity of the protons and their spread of chemical shifts.
In a COSY spectrum, crosspeaks are characterised by a pattern of anti-phase fine structure cross-peaks. The separation of a pair of anti-phase peaks is related to the coupling constant and line width. The double quantum filtered (DQF) COSY\textsuperscript{116} experiment has a major advantage over ordinary COSY experiments in that the spectra only display cross-peaks which are spins that are coupled to other spins. This eliminates most solvent peaks and the resolution is improved, especially near the diagonal. The pulse sequence for this experiment is shown in figure 3.6a. Application of a shaped pulse (SP) helps to overcome the bleaching of Hα signals in aqueous solutions of proteins\textsuperscript{117}.

3.4.2. The TOCSY experiment

Total Correlated Spectroscopy (TOCSY)\textsuperscript{118,119} is used to connect all nuclei within the same spin system. This occurs through an effect where transverse magnetisation is transferred from one spin to another, and is in turn transferred to the next spin and so on along the amino acid side chain, yielding crosspeaks corresponding to pairs of protons separated by up to five or six bonds.

Figure 3.6b displays the TOCSY pulse sequence. The key feature of the TOCSY experiment is that it uses a period of spin-locking in the mixing period to achieve coherence transfer. A single coherent radiofrequency is applied during the whole of the mixing time. This pulse acts as a ‘spin-lock’ and forces locked nuclei to precess at the same frequency and experience the same radiofrequency power. (This satisfies the Hartmann-Hahn\textsuperscript{120} condition). This means that they are in essence ‘strongly coupled’ and mutual coherence transfer will occur, thus
Figure 3.6a. *SCUBA DQF-COSY experiment with composite watergate*

Figure 3.6b. *Watergate TOCSY experiment with DIPSI spin lock*

Figure 3.6c. *Watergate NOESY experiment*
magnetisation is transferred along the amino acid chain. Theoretically, the longer the mixing
time, the further through the spin system the magnetisation will travel. Therefore by varying
the mixing time the information provided by the experiment can be tailored. A mixing time of
40-60 ms usually allows the entire side chain of a residue to be traced from the backbone
amide.

3.4.3. The NOESY experiment

The two dimensional nuclear Overhauser effect spectroscopy (NOESY)\textsuperscript{121} (figure 3.6c),
experiment relies on interactions which occur through space rather than through bonds. This
provides the most important information for structure calculations. The strength of the nOes is
proportional to $r^6$, where $r$ is the distance between the interacting nuclei. NOes observed in
proteins arise from protons no more than 5.5 Å apart. The nOe build up rate between any two
nuclei A and B can be approximated to equation 3:

$$k = \left[ \frac{34.2 \tau_c}{(1 + 4\omega^2 \tau_c^2)} - 5.7 \tau_c \right] \times 10 r_{ab}^{-6}$$

(3)

where:

- $r_{ab}$ - internuclear distance in Å
- $\omega$ - angular proton resonance frequency
- $\tau_c$ - rotational correlation time of the molecule

Magnetisation of spins precessing with one frequency during $t_1$ migrate by cross-relaxation to
nearby sites during the mixing time $T_m$. The magnetisation is then labelled with a different
frequency during $t_2$ and the 2D spectrum will contain a cross peak reflecting the chemical
shifts of two nuclei close in space. NOes build up during the mixing time $T_m$, and in general
longer mixing times will produce stronger nOes.

NOESY spectra are used to sequentially link the amino acid spin systems identified by COSY
and TOCSY experiments. The nOe data is also used in structure calculations since from them
it is possible to derive short and long distance restraints within a folded protein.

3.5. Heteronuclear NMR

3.5.1. 2D HMQC

Heteronuclear multiple quantum correlation (HMQC)$^{122,123}$ is concerned with nuclear
transitions which are forbidden by the usual NMR selection rule $\Delta m = \pm 1$, (where $m$ is the
magnetic quantum number), i.e. coherences corresponding to multiple transitions are not
directly detected. An HMQC experiment allows one to correlate the chemical shift of a proton
with its directly attached heteroatom, (e.g. $^{13}\text{C}$, $^{15}\text{N}$, $^{19}\text{F}$).

Although magnetic quantum coherences cannot be observed directly, indirect detection
methods allow them to be utilised. Here we use a 'proton-observe' method. This means the
protons are detected directly and the fluorine nuclei are detected indirectly, i.e. the proton
magnetisation which is detected during $t_2$ originated as proton magnetisation at the start of the
sequence. For example, the pulse sequence shown in figure 3.7, would give rise to a 2D
Figure 3.7. 2D HMQC experiment with DIPSI2 spin lock

3.5.1. HMQC experiment.

The HMQC experiment is used to determine the through-bond and through-space interactions of a particular fluorine atom within the hydrophobic core of the analogue UbFleu(50,67). The HMQC spectra is expected to be reasonably uncomplicated and therefore the distinguishing

spectrum where F\(_1\) is the chemical shift of the heteronucleus and F\(_2\) is the chemical shift of the directly attached proton. Proton-observe methods are usually more sensitive than other heteronuclear detected experiments since the relative sensitivity of protons is higher, (\(^1\text{H}, 1.00; \ ^13\text{C}, 1.59 \times 10^{-2}; \ ^15\text{N}, 1.04 \times 10^{-3}; \ ^19\text{F}, 0.83\)). If both excitation and detection are carried out on protons, full proton sensitivity is maintained even though the magnetisation is temporarily transferred to another type of nucleus. In the HMQC experiment the overall sensitivity depends on the product \(\frac{\gamma_e \gamma_d^{3/2}}{\omega}\) where \(\gamma_e\) is the magnetogyric ratio of the excited nucleus and \(\gamma_d\) is that for the detected nucleus.

3.5.2. 1D HMQC selective excitation techniques.

These techniques are useful in probing particular nuclei of importance within the sample. In this project the information we seek is the through-bond and through-space interactions of a particular fluorine atom within the hydrophobic core of the analogue UbFleu(50,67). The HMQC spectra is expected to be reasonably uncomplicated and therefore the distinguishing
resonances are easily identified using 1D spectra. 1D versions of 2D experiments also saves time and computer storage space. We have performed 1D analogues of the 2D HMQC NMR experiment incorporating a DANTE$^{124}$ pulse, which involves selectively exciting a single fluorine nucleus resonance. The selected spin and its interactions with other spins can be observed independently of all other interactions in the system. The 1D version is achieved by using a selective DANTE pulse on the nuclei of interest and replacing all variable delays with fixed delays. These principles are also applied to TOCSY and NOESY pulse sequences to create 1D HMQC-TOCSY and 1D HMQC-NOESY experiments. The former observes the interactions made by the individual F atom with protons in the side chain of the Fleu residue. The latter also identify the atoms and groups that are close to the F atom within the hydrophobic core. The pulse sequences for these experiments are displayed in figure 3.8a, b and c.
Figure 3.8a. 1D HMQC DANTE experiment

Figure 3.8b. 1D HMQC TOCSY experiment with DIPSI2 spin lock

Figure 3.8c. 1D HMQC DANTE NOESY experiment
3.6. Experimental

All spectra were collected on either a Varian VXR 5000 or Varian INOVA 600 MHz spectrometer. All spectral processing was carried out using the Felix95 (Biosym MSI) software package on either a Silicon Graphics or Sun workstation. Experimental and processing details are presented in figure 3.9. These experiments were all carried out at 25 °C and referenced to the water resonance which is 4.78 ppm at this temperature.

3.6.1. Sample preparation

Good quality NMR tubes (Wilmad 535-PP) were used to ensure a cylindrical sample. Between experiments the tubes were washed with distilled water and acetone and left to dry at room temperature since the cylindrical shape of thin walled NMR tubes will distort in direct heat.

Both synthetic peptides UbF(50,67) and UbAbu(26)Nva(30) were lyophilised after cation exchange column chromatography to yield 6.4 mg and 4.7 mg respectively. The protein samples were each dissolved in 550 µl of 25 mM C$_2$D$_3$O$_2$Na / 1 mM DTT / 0.1 % w/v Na azide / 10 % D$_2$O at pH 4.8.
<table>
<thead>
<tr>
<th>Experiment</th>
<th>Solvent</th>
<th>Solvent Suppression</th>
<th>Sweep Widths</th>
<th>Points Acquired</th>
<th>Processed Points</th>
<th>Window Functions</th>
<th>Miscellaneous</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preliminary</td>
<td>Na acetate</td>
<td>presaturation</td>
<td>12,000</td>
<td>12,000</td>
<td>32,000</td>
<td>line broadening</td>
<td></td>
</tr>
<tr>
<td>Presat 1D</td>
<td>buffer pH4.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preliminary</td>
<td>Na acetate</td>
<td>watergate</td>
<td>12,000</td>
<td>12,000</td>
<td>32,000</td>
<td>line broadening</td>
<td></td>
</tr>
<tr>
<td>Watergate 1D</td>
<td>buffer pH4.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DQF-COSY</td>
<td>Na acetate</td>
<td>watergate</td>
<td>10,000</td>
<td>2048</td>
<td>4096</td>
<td>D1 sinebell</td>
<td>SCUBA incorporation</td>
</tr>
<tr>
<td></td>
<td>buffer pH4.8</td>
<td></td>
<td>6,800</td>
<td>512</td>
<td>1024</td>
<td>D2 sinebell</td>
<td></td>
</tr>
<tr>
<td>NOESY</td>
<td>Na acetate</td>
<td>watergate</td>
<td>10,000</td>
<td>2048</td>
<td>4096</td>
<td>D1 Gaussian</td>
<td>mixing time: 150ms</td>
</tr>
<tr>
<td></td>
<td>buffer pH4.8</td>
<td></td>
<td>6,800</td>
<td>512</td>
<td>1024</td>
<td>D2 sinebell</td>
<td></td>
</tr>
<tr>
<td>TOCSY</td>
<td>Na acetate</td>
<td>watergate</td>
<td>10,000</td>
<td>2048</td>
<td>4096</td>
<td>D1 Gaussian</td>
<td>mixing time: 30ms &amp; 70ms</td>
</tr>
<tr>
<td></td>
<td>buffer pH4.8</td>
<td></td>
<td>6,800</td>
<td>512</td>
<td>1024</td>
<td>D2 sinebell</td>
<td></td>
</tr>
<tr>
<td>NOESY</td>
<td>D2O</td>
<td>presat</td>
<td>10,000</td>
<td>2048</td>
<td>4096</td>
<td>D1 Gaussian</td>
<td>mixing time: 150ms</td>
</tr>
<tr>
<td>TOCSY</td>
<td>D2O</td>
<td>presat</td>
<td>10,000</td>
<td>2048</td>
<td>4096</td>
<td>D1 Gaussian</td>
<td>mixing time: 30ms &amp; 70ms</td>
</tr>
<tr>
<td>2D HMQC</td>
<td>90% H2O 10% D2O</td>
<td>presat</td>
<td>9,600</td>
<td>1072</td>
<td>4096</td>
<td>D1 Gaussian</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>8,000</td>
<td>256</td>
<td>512</td>
<td>D2 Gaussian</td>
<td></td>
</tr>
<tr>
<td>1D HMQC</td>
<td>90% H2O 10% D2O</td>
<td>N/A</td>
<td>10,000</td>
<td>4,000</td>
<td>32,000</td>
<td>Exponential</td>
<td>DANTE incorporation</td>
</tr>
<tr>
<td>1D HMQC-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Multiplication</td>
<td></td>
</tr>
<tr>
<td>TOCSY</td>
<td>90% H2O 10% D2O</td>
<td>N/A</td>
<td>10,000</td>
<td>4,000</td>
<td>32,000</td>
<td>Exponential</td>
<td>DANTE incorporation</td>
</tr>
<tr>
<td>1D HMQC-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Multiplication</td>
<td></td>
</tr>
<tr>
<td>NOESY</td>
<td>90% H2O 10% D2O</td>
<td>N/A</td>
<td>10,000</td>
<td>4,000</td>
<td>32,000</td>
<td>Exponential</td>
<td>DANTE incorporation</td>
</tr>
</tbody>
</table>

**Figure 3.9. Summary of the parameters used for the acquisition and processing of all NMR experiments**
3.6.2. Determination of hydrogen bonds.

Amide protons which are involved in stable secondary structural elements are protected from exchange with solvent since they are participating in hydrogen bonds. These amide protons can be identified through an experiment in which a fully protonated protein sample is lyophilised and then resuspended in D₂O. Spectra are recorded immediately. Only amide protons involved in hydrogen bonds are observed since all other protons are rapidly exchanged with D atoms.

The NMR samples, (prepared in 3.7.1.), of UbFleu(50,67) and UbAbu(26)Nva(30) were freeze dried and resuspended in 550 μl D₂O (the pH of the samples was not adjusted). A series of 1D ¹H NMR experiments figure 3.9, were run as quickly as possible. This procedure was repeated using bovine ubiquitin (5.0 mg / 550 μl D₂O).

3.6.3. The NMR Spectrometer

High-field spectrometers have superconducting magnets which observe proton frequencies of greater than 200 MHz. The applied field must be very homogeneous. A niobium alloy-based superconducting magnet is cooled in a bath of liquid helium. An outer encasement of liquid nitrogen acts as a heat-loss shield; this means the liquid helium need only be ‘topped up’ every two to nine months. Within the bore of the magnet there is a set of field gradient coils. These are the ‘shims’ and adjustment of the passage of current through these can cancel out
errors in the static field and produce near perfect field homogeneity. ‘Shimming’ reduces the peak base-width and improves the measured signal-to-noise ratio.

The sample is lowered into the spectrometer on a cushion of air and comes to rest in the upper part of the probe. The probe converts the applied electromagnetic radiation into an oscillating magnetic field via a coil which surrounds the sample. The spectrometer functions are operated using a computer which controls all the hardware for generating the radio-frequency pulse sequences and to detect the signals.

3.6.4. Spectral considerations

3.6.4.1. Pulse width

The pulse-widths for 90° and 180° pulses must be set accurately in an NMR experiment. It is also essential that enough time is allowed between pulses for full relaxation to take place. To set the pulse-width a series of spectra are collected with increasing pulse-widths, then the intensity of the resulting spectra is plotted as a function of pulse-width. Since the coils for detection of magnetisation are located in the xy-plane, the pulse-width required to tip the magnetisation through 360° will appear as a null in the plotted spectra. The pulse width is set to one-quarter of this value for the experiment.
3.6.4.2. Frequency lock

The signal frequency must be constant during an NMR experiment, however the magnetic field of any magnet has a tendency to drift with time. This is due to a small electrical resistance of the magnet and also to fluctuations in room temperature and movement of metal objects near the magnet. A small amount (ca 5-10%) of D$_2$O added to the sample serves as a deuterium lock. The deuterium lock means long term stability of the magnetic field can be achieved. It works by comparing a constant frequency to an observed resonance. Adjustments are then made so that the magnetic field maintains that frequency. The intensity of the lock signal depends on the field homogeneity and so can be used to observe progress when adjusting the shims.

3.6.5. Spectral processing

Data-processing requires optimisation of the various functions applied to the data in order to yield the best spectral quality. This varies according to the experiment and some experience is required to obtain good results. Spectral processing converts the FID, which is a plot of intensity as a function of time, into an interpretable spectrum, which is a plot of intensity verses frequency. The processing stage, however is also an opportunity to suppress any artefacts and to enhance sensitivity or improve resolution. The manipulation of data in this way is carried out before and after the Fourier transformation.
3.6.5.1. Fourier Transformation

A Fourier transformation (FT) converts the FID in the time domain, F(v), into the frequency domain, I(v). The Cooley Tukey FT applied to each FID is expressed as:

\[ I(v) = \int e^{-2\pi i vt} F(v) \]  

(4)

The FT is a routine computational task. The Cooley - Tukey Transformation is generally used by the computer since it is reasonably fast and makes good use of computer memory which was an important consideration when NMR was being developed.

3.6.5.2. Deconvolution and Zero-filling

Prior to the FT any residual water in the FID is computationally removed through a mathematical process called deconvolution\textsuperscript{125}. The line-shape of the resulting spectrum can be slightly improved by ‘zero filling’. Zero filling involves adding points of zero intensity to the end of the FID, where it has almost decayed to zero. This works according to the idea that if the acquisition period was extended then it would be recording zeros anyway. Therefore we manufacture extra acquisition time by adding them artificially\textsuperscript{126}. 
Window functions are mathematical functions which are applied to the FID to improve the resolution and the signal-to-noise ratio. Noise levels are constant throughout the acquisition period whereas the signal decays exponentially. A window function works by adjusting the decay envelope of the FID. Application of a window function normally suppresses the latter part of the FID relative to the early part. An improvement in the signal-to-noise ratio is obtained but at the expense of the line-width. Commonly used window functions are exponential multiplication, the sine bell and Gaussian functions.

The most common of these is the exponential multiplication, (or line broadening function), where the FID is multiplied by an exponential decay. The larger the value of the line broadening function, the more rapidly the FID will decay. This will lead to broader lines i.e. a decrease in resolution. The signal-to-noise ratio however is improved. A negative line broadening factor has opposite effects - sharper lines but a poorer signal-to-noise ratio.

The sinebell window functions available include sinebell and sinebell-squared. The basic sinebell window function is a sine wave from zero to $\pi$, mapped onto the number of points in the FID. This window function can be shaped according to the number of data points and the position of the sine-function maximum. The sinebell-squared has an advantage over the sinebell window in that it approaches zero more smoothly. A sinebell squared function is shifted by $90^{\circ}$ and this is particularly good to use when insufficient acquisition time has been allowed and the resulting FID is truncated.
The best function for enhancement of resolution with minimal loss of signal-to noise ratio is a Gaussian multiplication. This function had two variable parameters. The first is the line broadening parameter and works in the same way as described for exponential decay. The second parameter determines the location of the maximum of the function along the FID.
Chapter Four

NMR Spectral Assignment of Ubiquitin analogues, UbFLeu(50,67) and UbAbu(26)Nva(30).

4.1. Introduction to Spectral Assignment.

There are three types of information available from NMR spectroscopy. These are through-bond interaction (via scalar couplings), through space interactions (via dipolar couplings) and the chemical environment (via the chemical shift).

The goal of spectral assignment is to correlate each nuclear magnetic resonance to a proton or protons in the protein. Scalar correlation experiments such as COSY, DQF-COSY and TOCSY, are used to identify resonance positions within each amino acid spin system. NOESY experiments are used to sequentially connect the amino acid spin systems.

The $^1$H resonances are categorised on the basis of chemical shift. Typically chemical shifts are as follows; backbone amide protons resonate between 10 and 7.0 ppm, the backbone alpha protons resonate between 6.0 and 3.5 ppm, aliphatic side chain
protons resonate between 3.5 and 1.0 ppm and methyl protons have chemical shifts less than 1.5 ppm.

A systematic procedure for the determination of spectral resonances was developed by Wüthrich\textsuperscript{127}. This is known as the 'sequential assignment strategy'. The analysis uses scalar couplings to identify NH– CαH and aliphatic side chain resonances belonging to the same residue spin system. These are assigned to an amino acid type.

The spin systems are connected using CαH(i) to NH(i+1) and Cβ(i) to NH(i+1) through-space dipolar couplings. Once the chain of spin systems is long enough, the residues are identified through correlation with the amino acid sequence. However this standard systematic procedure is not strictly followed in every spectral assignment.

Although the 2D NMR spectral analysis of human and synthetic ubiquitin have been published\textsuperscript{43,44} the spectral assignments of the synthetically prepared analogues used in the current study were undertaken almost independently of previous work. The backbone assignments for both UbFLeu(50,67) and UbAbu(26)Nva(30), however, were seen to correlate well with those found for human ubiquitin. Since the primary backbone assignments for both analogues of ubiquitin were almost identical, the examples shown here use only spectra of the analogue UbFleu(50,67). The spectral assignment package XEASY\textsuperscript{128} was used to display spectra throughout the assignment.
4.2 Assignment of the 2D $^1$H NMR of ubiquitin analogues.

Initial assignment of the aromatic residues in the amino acid sequence was used as a starting point for assignment of the sequential backbone $\text{CaH}(i)$ to HN $(i+1)$ connectivities. The proline residues, absent from the string of alpha to amide resonances, were then identified. All side chain residues were assigned including those of the unnatural amino acids. Secondary structural elements were then identified based on typical nOe patterns. Once this was complete, nOe assignments critical for determination of tertiary structure were made.

The labelling on the figures uses different colours for different residues or to indicate nOe resonances to other residues. Assignment of methyl or some methylene residues uses the label 'Q', which designates assignment of all protons on the specified carbon atom.

4.2.1. Assignment of aromatic residues.

The aromatic region of the 2D $^1$H spectra was a good starting point since this region is highly resolved. The assignments for Phe 4, Phe 45 and Tyr 59 in the aromatic region of the 2D $^1$H TOCSY spectra are shown in figure 4.1. (The identity of these aromatic residues was aided by reference to the human ubiquitin spectral analysis$^{43,44}$). The corresponding NOESY spectra mapping the full spin system of each of these residues is displayed in figure 4.2 a, b and c. This assignment was made
using combined information from DQF-COSY, TOCSY and NOESY spectra. The ring protons give unambiguous nOes to the Hα and Hβ of the same residue. In turn these resonate with NH of the same residue and make nOe connectivities to adjacent residues as indicated.

4.2.2. Assignment of backbone protons

The full assignment of the aromatic residues led to adjacent backbone assignments. Figure 4.3 (a-g) demonstrates the CαH to HN ‘backbone walk’. Examples of β(i) to HN(i+1) nOes which supported backbone assignment are shown. Breaks in the peptide chain occur at position Cα14 to NH15 and Cα32 to NH33 where resonances are overlapping, and also at Gly 53, Glu 24 and Cα26 to NH27 where the resonances are too weak to be seen at the contour level used for these plots. There are also breaks for the proline residues at positions 19, 37 and 38.
Figure 4.1. TOCSY spectrum displaying the aromatic ring proton resonances of Phe 4, Phe 45 and Tyr 59.
Figure 4.2a NOESY spectrum showing the full spin system of Phe 4 and nOe connectivities to adjacent residues
Figure 4.2b NOESY spectrum showing the full spin system of Phe 45 and nOe connectivities to adjacent residues.
Figure 4.2c  NOESY spectrum showing the full spin system of Tyr 59 and nOe connectivities to adjacent residues.
Figure 4.3a Region of the 2D $^1H$ NOESY spectrum showing CαH (i) to NH (i+1) connectivities from residues 1 to 12.
Figure 4.3b Region of the 2D $^1$H NOESY spectrum showing CaH (i) to NH (i+1) connectivities from residues 12 to 22.
Figure 4.3c Region of the 2D $^1$H NOESY spectrum showing CaH (i) to NH (i+1) connectivities from residues 22 to 34.
Figure 4.3d Region of the 2D $^1$H NOESY spectrum showing CaH (i) to NH (i+1) connectivities from residues 34 to 46.
Figure 4.3e Region of the 2D $^1H$ NOESY spectrum showing $\text{CaH}$(i) to $\text{NH}(i+1)$ connectivities from residues 46 to 56.
Figure 4.3f Region of the 2D $^1$H NOESY spectrum showing $\alpha$H (i) to NH (i+1) connectivities from residues 56 to 68.
Figure 4.3g Region of the 2D $^1$H NOESY spectrum showing CaH (i) to NH (i+1) connectivities from residues 68 to 76.
4.2.3. Assignment of the Proline residues.

Since proline residues have no amide protons (i.e. $\mathrm{C\alpha H}_{\mathrm{p-1}}$ to NH resonance), the proline residues are linked into the sequential assignment in an indirect way. Assignment was achieved by identifying cross-peaks between Pro C$\delta$H and C$\alpha$H of the preceding residue. These peaks were obvious since their likely chemical shift is near the diagonal, away from crowded regions. These resonances led to the full spin system assignment shown in figure 4.4a, b and c. The identification of a $\mathrm{C\alpha H}_\text{pro}$ to HN$_{\text{pro}+1}$ connectivity confirmed the position of the proline spin system within the sequential assignment.

A similar procedure was applied to the double prolines at Pro(37)Pro(38). The spin systems were first identified and connectivites C$\alpha$H (37) to C$\delta$H (38) and C$\beta$H (37) to C$\delta$H(38) linked the two spin systems together. NOes to HN (Ile 36) and HN (Asp 39) confirmed their position within the primary amino acid sequence.

The intensity of sequential nOes give an indication of the conformational type of X-proline peptide bond. A strong C$\alpha$H$_{\text{pro} \cdot 1}$ to C$\delta$H$_\text{pro}$ nOe is indicative of a trans peptide bond; whereas a strong C$\alpha$H$_{\text{pro} \cdot 1}$ to C$\alpha$H$_{\text{pro}}$ is typical of a cis peptide bond (see figure 4.5). The NOESY spectra (figure 4.4a,b and c) reveals that all three X-proline bonds in the analogues (as in human ubiquitin) are trans peptide bonds.
Figure 4.4a NOESY spectrum showing the spin system of Pro 19 with nOe connectivities to adjacent residues.
Figure 4.4b NOESY spectrum showing the spin system of Pro 37 with nOe connectivities to adjacent residues.
Figure 4.4c NOESY spectrum showing the spin system of Pro 38 with nOe connectivities to adjacent residues.
Figure 4.5. Diagram demonstrating (a) trans and (b) cis Proline peptide bonds.
4.2.4. Assignment of amino acid side chains.

Identification of side chain spin-system confirmed the nature of residues at specific positions and provided information that would be used in assembling a list of nOe constraints. Scalar correlation experiments (DQF-COSY and TOCSY experiments) in both H₂O and D₂O facilitated this process. Most protons were assigned to particular resonances. The assignments are more complete in the analogue UbFLeu(50,67) than in UbAbu(26)Nva(30).

Assignment of unnatural amino acid side chains.

a) 2S, 4S, 5-Fluoroleucine.

The expected resonance pattern observed in the 1D ¹⁹F NMR of the free amino acid, figure 4.6, is reflected in the 2D ¹H NMR. The fluorinated methyl group is prochiral. Each proton within this group resonates separately and is spin coupled to the fluorine nuclei to give a total of four peaks. In the 2D HMQC experiment for the ubiquitin analogue, (figure 4.7.), the fluorine atoms of Fleu 50 and Fleu 67 resonate at -218.7 ppm and -222.0 ppm respectively and each displays a $^{2}J_{HF}$ of 54 Hz. The same underlying splitting pattern is observed for both Fleu residues, however the pattern for Fleu 67 appears to be a triplet because the two central peaks of the pattern of four are overlapped. This implies that the two protons of the fluoromethyl group in Fleu 67 are less well resolved than the equivalent pair in Fleu 50. The full spin system assignment in the TOCSY for Fleu 50 and Fleu 67 are shown in figure 4.8 a and b.
Figure 4.6. $^{19}$F spectrum of Fmoc-2S, 4S, 5-Fluoroleucine (referenced to CFCI$_3$). The large coupling measures 49 Hz and the smaller $^2J_{HH}$ coupling measures 18 Hz.
Figure 4.7 2D $^{19}$F/$^1$H NMR showing the splitting pattern of the $\text{CH}_2F$ group
Figure 4.8a 2D TOCSY spectrum showing the spin system of Fleu 50
Figure 4.8b 2D TOCSY spectrum showing the spin system of Fleu 67
b) Aminobutyric acid (Abu) and c) Norvaline (Nva)

A combination of overlap, poor signal-to-noise and artefacts due to small molecule contamination made the assignment of these unnatural amino acid residues difficult. Exhaustive analysis of the DQF-COSY and TOCSY spectra in both H2O and D2O led to the assignments made as shown in figure 4.9 a and b.

The spin assignment of Abu 26 (figure 4.9 a) displays strong CαH → NH, CγH3 → NH and CβH2 → CγH3. However resonances CβH2 → NH and CβH2 → CαH are less obvious and are either very weak (CβH2 → NH) or lying in a confused area (CαH→CβH2). A position is marked by a box in figure 4.9 a, this represents the Cβ2H → NH resonance, where the height of the peak is too low to be observed.

The full spin system assignment of Nva 30 is shown in figure 4.9 b. The spectrum displays strong CαH → NH and CαH → CδH3, resonances, however the CβH2 → CγH2 resonance is very weak and the CγH2 → CδH3 resonance occurs in a crowded region and therefore appears to be weak.
Figure 4.9a 2D $^1$H TOCSY spectrum showing full spin system assignment of Abu 26
Figure 4.9b 2D $^1H$ TOCSY spectrum showing full spin system assignment of Nva 30
4.2.5. Identification of secondary structure

Secondary structure in proteins give rise to characteristic nOe patterns. An α-helix will always display strong nOes between sequential backbone amide protons, i.e. NN(i,i+1), weak αN(i,i+1) nOes and weak-medium nOes between αN(i,i+3) and αN(i,i+4). Both α and 3₁₀ helices show good nOe connectivities due to the short inter-residue dβN(i, i+1) and dNN(i,i+1) distances.

β strands display characteristic nOes with weak dNN(i,i+1) and strong dαN(i,i+1). Interstrand nOes enable the strands to be placed in order within a β-sheet. β-strands align with respect to each other in either a parallel or anti-parallel fashion. Typically strong αHi-NHj peaks are observed for the parallel strand pair and strong αHi-αHj peaks are observed in D₂O spectra (i.e. in the absence of large residual H₂O peaks) for the antiparallel strands.

The short-range nOes observed in the case of UbFleu(50,67) and UbAbu(26)Nva(30) are summarised in figure 4.10 a and b respectively. Characteristic nOes defined an α-helix stretching from residue 25 to 35 and a four-residue helical stretch from residues 56-60 in both analogues. The additional medium nOes, dαN(i,i+1) indicate that the smaller helical fragment is one turn of a 3₁₀ helix. The five stranded β-sheet is made up of three antiparallel strand pairs and one parallel pair. The β-sheet structure of UbFleu(50,67) is presented in figure 4.11.
Figure 4.10a Summary of the sequential and short range nOes for the protein analogue UbFleu(50,67). The height of the bar used denotes nOe intensity.
Figure 4.10b Summary of the sequential and short range nOes for the protein analogue UbAbu(26)Nva(30). The height of the bar used denotes nOe intensity.
Figure 4.11. The β-sheet secondary structure of UbFleu(50,67). Cross strand nOes are represented by dotted lines. The horizontally dashed lines represent hydrogen bonds.

The pattern of short-range and inter-strand nOes defining all the secondary structural components are very similar to those found for human ubiquitin$^{43,44}$. 
4.2.6. Determination of hydrogen bonds.

The identification of amide protons involved in hydrogen bonds in both analogues was undertaken and results compared to the hydrogen bonding found in bovine ubiquitin. Bovine ubiquitin displayed very slow exchange of protected amide protons with some amide protons in the region 8.2 - 9.4 ppm still unexchanged after eighteen hours, figure 4.12 a. However, the D₂O-exchange rate for both ubiquitin analogues was accelerated with all the amide protons, (in the region 8.2 - 9.4 ppm), being fully exchanged after twelve and eight minutes for UbFleu(50,67) and UbAbu(26)Nva(30) respectively. This high rate of exchange meant that the use of 2D TOCSY experiments to identify the core amide protons was impossible. Instead the 1D ¹H NMR at 8 minutes and 6 minutes for UbFleu(50,67) and UbAbu(26)Nva(30) respectively, (figure 4.12b and 4.12c), were assigned in so far as possible based on the previously assigned 2D ¹H TOCSY. However these 1D spectra are not well-resolved and the amide resonances were overlapped and difficult to assign unambiguously.

Despite the rapid rate of exchange with solvent of these amide protons relative to bovine ubiquitin, they are protected relative to other amides in the synthetic proteins. It was therefore decided to treat them as amide protons involved in hydrogen bonds. The carboxyl group to which the amide proton is hydrogen bonded to cannot be deduced directly from this experiment. However from the pattern of cross-strand
nOes and hence some knowledge of the secondary structure, an acceptor carbonyl oxygen atom can be inferred.
Figure 4.12a. 1D $^1$H NMR of bovine ubiquitin showing amides present 10 minutes, 40 minutes and 18 hours after resuspension in $D_2O$. 
Figure 4.12b. 1D $^1H$ NMR of UbFleu(50,67) showing amides present 8 minutes after resuspension in D$_2$O.

- Val 5 / Leu 67
- His 68
- Val 70 / Ile 44
- Val 17
- Phe 45
- Thr 7
- Glu 34
- Gln 31
- Phe 4
- Leu 15
- Lys 27
- Leu 69
- Ile 3
- Ile 30
- Val 26
- Ala 28
- Lys 29
- Lys 48
- Ser 65
- Lys 33
- Lys 11

8 minutes

ppm
Figure 4.12c. $1D^{1}H$ NMR of UbAbu(26)Nva(30) showing amides present 6 minutes after resuspension in $D_{2}O$. 
4.2.7. Assignment of nOes defining tertiary structure

Long range nOes are vital to the calculation since they determine the overall fold of the protein. These define pairs of protons that are close in space in the folded peptide but distant in the primary amino acid sequence.

Long-range nOe assignment began with connectivities made from the α−helix across the hydrophobic core to the β-strands. Once this was systematically completed as fully as possible, structure calculations were initiated. Viewing a family of calculated structures allows reasonable decisions to be made about the assignment of ambiguous nOes and subsequently more nOe input. The long range nOe assignments were therefore on-going throughout a series of the structure calculation.

4.3. Assignment of 1D HMQC NMR spectra of UbFleu(50,67)

Assignment of the 1D HMQC spectra shown in figure 4.13, was carried out with the aid of the 2D 1H TOCSY and NOESY data. The HMQC spectra provided confirmation of the assignments previously made, as well as clarifying which specific groups are in close proximity to each fluorine atom within the hydrophobic core.

The 1D HMQC spectrum shows the F atom of Fleu50 correlating with the δ−fluoromethyl protons and the γC proton of the same residue. The 1D TOCSY HMQC experiment enables us
to see further along the amino acid side chain to a C3H and also C8H3 of Fleu 50. These results provide confirmation of previously made assignments in the 2D 1H TOCSY (figure 4.7 a). The 1D NOESY HMQC reveals additional information as to which groups are within 5 Å through-space of the F50 atom, these being the β–methylene of Phe 45, the δ-methyls of Fleu 67, Ile 3 and Ile 61.

The 1D HMQC spectra for Fleu 67 can be interpreted in the same way. The simple HMQC experiment correlates the fluorine atom of Fleu 67 with the δC-protons of the same methyl group. This appears to be a 1:2:1 triplet, however the splitting pattern has been identified to be a 1:1:1:1 quartet (as observed in the case of Fleu 50) with the central peaks overlapped. A finer splitting can be observed on close inspection. This is the small 2JHH coupling and is more obvious in the 19F 1D NMR of the free Fmoc-2S, 4S, 5-fluoroleucine amino acid, figure 4.6. The HMQC-TOCSY spectrum correlates the fluorine atom of Fleu 67 with additional protons within the same spin system, i.e. C3H and C8H3 protons. The ‘extra’ resonances observed in the HMQC-NOESY spectrum enable the identification of groups which are close in space in the centre of the hydrophobic core. These are C8H3 of Leu 56, Ile 3 and Ile 61 as well as C3H3 of Ile 61. Both Fleu 50 and Fleu 67 show common correlations through-space to C8H3 of Ile 61 and Ile 3.
Figure 4.13  1D HMQC spectra for (a) Fleu 50 and (b) Fleu 67
5.1. Introduction

NMR structure determination is a global optimisation problem performed (in this case) in distance space. The goal of the optimisation is to obtain three-dimensional (3D) structures that are consistent with the NMR data.

NMR is an experimental tool which yields structural information indirectly. NMR data is used as input for a structure calculation which results in a 3D structure. There are many methods for structure calculation available, all of which aim to sample conformational space whilst satisfying a set of constraints. A popular structure calculation protocol is simulated annealing using molecular dynamics, a process introduced by Clore and Brünger et al (1986)\textsuperscript{129}. The data used as input for the calculations in this project were the nOe-derived distances and the experimentally identified hydrogen bonds, (Appendix A and B).
5.2. Tertiary structure determination

5.2.1. Compilation of nOe table

The input file for the structure calculation is a list of distance restraints derived from the NOESY spectral data. The intensity of an nOe depends on several factors, however the most significant is the rate of nOe build up, which has a $r^6$ dependency, where $r$ is the internuclear distance. Thus, the closer two nuclei are in space the stronger the nOe after a given build-up time. The intensity of assigned nOes in this study are categorised into strong, medium, weak and very weak. These are then allocated an upper distance bound; i.e. strong 2.5 Å, medium 3.5 Å, weak 5.0 Å and very weak 5.5 Å. Due to the low activation energy barriers to aromatic ring flipping and methyl group rotation, some protons do not have a defined location but can lie within a range of conformational space. To account for this a pseudoatom term is added to the distance restraint. This was 1.5 Å for methyl groups and 2.0 Å for either the δ or the ε protons of aromatic rings. Degenerate methylene protons have a pseudoatom correction of 1.0 Å. Where the two proton-resonances of a methylene were resolved but had different strengths and were not stereospecifically assigned, the distance restraint for both protons was set to a value corresponding to the weaker nOe. Each interproton nOe assignment and corresponding distance restraint was entered in an assignment table, divided into categories of intra-residue, sequential (i-i+1), short range ($2 \leq |i-j| \leq 4$) and long range ($|i-j| < 4$) assignments. Once a sufficient number of nOes were included, this list forms the basic input for the structure
Figure 5.1: The distribution of nOes in (a) UbFleu(50,67) and (b) UbAbu(26)Nva(30). The filled squares indicate that one or more distance constraints were observed between the corresponding residues on the vertical and horizontal axis. Above and to the left are shown restraints involving only main chain protons CaH to NH. Below and to the right of the diagonal are restraints involving at least one side chain proton.
calculation. A representation of the distribution of nOes is given in figure 5.1 and the nOe tables for both analogues are displayed in Appendix B.

5.2.2. Compilation of the hydrogen bond table

Each hydrogen-bond identified from amide exchange data is incorporated into the structure calculation as a pair of distance restraints (Appendix A). This ensures that the atoms involved are relatively fixed in a co-linear orientation. The distances used are the average N-H–O= and N=O– distances\(^{131}\) found in typical protein secondary structure. These are measured to be 0.42 nm and 1.32 nm respectively.

5.2.3. Measurement of $\phi$ angles from 2D $^1$H COSY experiment.

Dihedral angle-restraints based on measurements of $^3$JNH\(_\alpha\) couplings were unfortunately not included in the calculations, since the protein samples were too weak to afford the required level of signal to noise. This meant that the computer program designed to extract the dihedral angle information, despite several attempts, could not satisfactorily distinguish the COSY peaks from the noise\(^a\).

\(^a\) Further samples of both analogues are presently being produced and once good COSY spectra are recorded this work can be completed.
5.3. Molecular Dynamics Simulated Annealing

In this project the structure calculations have involved a molecular dynamics protocol incorporating simulated annealing within the computer program XPLOR. The purpose of a molecular dynamics calculation in an NMR structure-determination is to survey the conformational space available to the polypeptide chain. Simulated annealing mimics the processes of heating followed by cooling. This is performed using a simplified force-field which treats atoms as soft spheres without attractive or long range (i.e. electrostatic) non-bonded interactions, and the calculation excludes solvent considerations.

Molecular dynamics can create a starting structure from scratch, however it is acceptable to start from a random coil structure. The polypeptide chain thus incorporates chemically reasonable bond lengths, bond angles and chirality. This starting structure is subjected to a carefully devised force-field. The force-field is described by the total potential energy of the bonded and the non-bonded interactions which occur in the molecule.

\[ P_E^{(\text{total})} = P_E^{(\text{covalent})} + P_E^{(\text{non-bonded})} \]

\[ P_E^{(\text{covalent})} = P_E^{(\text{bond})} + P_E^{(\text{angle})} + P_E^{(\text{dihedral})} + P_E^{(\text{improper})} \]

\[ P_E^{(\text{non-bonded})} = P_E^{(\text{VdW})} + P_E^{(\text{electrostatic})} + P_E^{(\text{hydrogen bonds})} \]
The hydrogen-bond and electrostatic components are optional, while the van der Waals component may be treated as a simple repel function during much of the calculation. Distance constraints derived from NMR data also need to be satisfied. This is achieved by adding a $P_E(nOe)$ term to the non-bonded term.

$$P_E(\text{non-bonded}) = P_E(VdW) + P_E(\text{electrostatic}) + P_E(\text{hydrogen bonds}) + P_E(nOe)$$

5.3.1. Structure calculation protocol

A general overview of the simulated annealing protocol is as follows. The starting structure is heated to a temperature of 1000 K under the influence of the force field, $P_E(\text{total})$. This energy term is then minimised by a combination of temperature-gradient descent, simulated annealing and conformational search procedures. Firstly, the randomised starting structure undergoes 15 ps of high energy dynamics. During this time the 'repel function', which is a simplified van der Waals function, is switched off. The force constants for geometric and experimental constraints are also set to a low value. The potential energy-well describing nOe-derived distance restraints is shallow with gently sloping sides at this stage. The presence of kinetic energy in molecular dynamics simulations allows barriers of potential energy surfaces to be crossed. With the reduction of all repulsive forces, domains of the polypeptide chain can interact closely or even move through each other, thereby reducing the problem of misfolded conformations becoming trapped in local minima.
During the following 10 ps the temperature is lowered and the repel function is gradually increased, ensuring there are no close non-bonded contacts. The sides of the nOe functions are steeepened producing a ‘square-well’ potential. The system is gradually cooled in a series of small steps and the structures that emerge are subsequently energy-minimised and are selected on the basis that they satisfy restraints for good covalent and experimental data. This process can be repeated, each time starting from a different, initial, randomised structure.

A family of calculated structures is generated, each member arising from a different starting structure. The convergence of the structures that emerge depends upon the number of nOe assignments and their lack of ambiguity. A low root-mean-square deviation (r.m.s.d) between calculated structures is normally evidence of a good data set. To reach a good level of agreement between structures it is necessary to have, on average, about ten experimental constraints per residue.

The family of structures can subsequently be superposed, and regions where atom positions vary can be identified. Also some nOes may be consistently violated by significant amounts. This might mean they have either been mis-assigned in the spectra, or insufficient nOes have been included in the calculation. In these cases, the spectra are re-examined and corrections and/or more nOe-assignments made. A new family of structures is then calculated incorporating the new information.
During the simulated annealing protocol, a fraction of the molecules will follow an inappropriate trajectory. These will have a completely different fold to the majority of the molecules and are normally recognised since their nOe violation energies are very high. If these are only present as a very small fraction of the total number of structures they are ignored and discarded.

Simulated annealing by molecular dynamics requires substantially more computational time than any other method of structure calculation. However this is compensated for by the higher success rate - generally 40 - 100% of the structures generated are acceptable. This success may be attributed to the ability of the algorithm to ‘escape’ from local minima.

5.3.2. Structure Refinement.

Both UbAbu(26)Nva(30) and UbFleu(50,67) were refined using a ‘slow-cooling’ simulated annealing process. The protocol allows initial strain or bad contacts within the structural co-ordinates to be released by softening of the van der Waals repulsions. This allows the atoms to move gently though one another. The high starting temperature is gradually lowered, and the molecule undergoes subtle adjustments to arrive at its refined conformation.

Since the analogue UbF(50,67) potentially contains electronegative atoms (fluorine atoms) within its hydrophobic core, refinement requires incorporation of electrostatic
parameters. Polar solvents, usually H$_2$O, surrounding a protein molecule have a stabilising effect upon the polar and charged amino acid side chains which project into it. Since the computational structure calculations ignore the presence of solvent, the electrostatic potential for all hydrophilic residues were switched off. This avoids electrostatic interactions being over-emphasised by the lack of solvent thus giving a misleading impression of the tertiary structure. Thus the refinement protocol used in the case of UbFleu(50,67) was modified to incorporate electrostatic parameters for fluoroleucine residues and non-polar hydrophobic core residues only. This refinement uses the standard Lennard-Jones function, (equation 1, chapter one) and electrostatic interactions, (equation 2, chapter one), in addition to conformational energy terms.

5.4. Structure calculation results

5.4.1. UbFleu(50,67)

A calculation of 115 UbFleu(50,67) structures yielded an energy profile as presented in figure 5.2. The energy distribution of these initial calculated structures shows a spread of energies with approximately two-thirds having low energies (<100 kJmol$^{-1}$) and a distinct population of very high energy conformations. There are very few structures between these extremes.
Figure 5.2. Structure Calculation Energy profile. Plot demonstrates the energy distribution of the 115 initial calculated structures of UbFleu(50,67).

Of these 115 structures 48 were accepted on the basis of having no nOe violation distance greater than 0.4 Å. The accepted structures ranges in energy from 34.58 kJ/mol to 65.62 kJ/mol. An analysis of the experimental restraints and structural statistics of the final refined structures is given in figure 5.3. The atomic r.m.s. deviations were calculated using a structure averaging programme within XPLOR\textsuperscript{122}. The selected structures demonstrated reasonably good convergence with an average r.m.s.d. from the backbone atoms of 1.16 Å, and for all heavy atoms, 1.94 Å.

The overall structural definition of the analogue UbFleu(50,67) is satisfactory, see figure 5.4. The correlation between the number of nOes per residue and the backbone Ca r.m.s.d. clearly reveals that a greater number of restraints in a particular stretch of sequence reflects increased convergence in the final minimised structures. Figure 5.4 shows good definition throughout the structured part of the protein (residues 1 to 73)
<table>
<thead>
<tr>
<th>Category of Restraint</th>
<th>Number of restraints</th>
<th>$\text{rmsd} \pm \text{sd}$ for 48 structures $\text{Å}$ or $^\circ$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All nOe derived restraints</td>
<td>765</td>
<td>0.045±0.005</td>
</tr>
<tr>
<td>Intraresidue nOe's</td>
<td>51</td>
<td>0.032±0.016</td>
</tr>
<tr>
<td>Sequential nOe’s</td>
<td>254</td>
<td>0.040±0.005</td>
</tr>
<tr>
<td>Short range nOe’s</td>
<td>83</td>
<td>0.046±0.009</td>
</tr>
<tr>
<td>Long range nOe’s</td>
<td>337</td>
<td>0.044±0.007</td>
</tr>
<tr>
<td>H-bond derived</td>
<td>40</td>
<td>0.074±0.021</td>
</tr>
<tr>
<td>Chemical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths</td>
<td></td>
<td>0.013±0.037</td>
</tr>
<tr>
<td>Impropers</td>
<td></td>
<td>0.671±0.377</td>
</tr>
<tr>
<td>Convergence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All non H atoms</td>
<td>1.94</td>
<td></td>
</tr>
<tr>
<td>Backbone atoms</td>
<td>1.16</td>
<td></td>
</tr>
</tbody>
</table>

(a) Energy term Mean Value (kJ/mol)

<table>
<thead>
<tr>
<th>Energy term</th>
<th>Mean Value (kJ/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E(noe)</td>
<td>39.38</td>
</tr>
<tr>
<td>E(bond)</td>
<td>15.25</td>
</tr>
<tr>
<td>E(angle)</td>
<td>156.42</td>
</tr>
<tr>
<td>E(improper)</td>
<td>25.39</td>
</tr>
<tr>
<td>E(vdw)</td>
<td>41.45</td>
</tr>
<tr>
<td>E(total)</td>
<td>277.91</td>
</tr>
</tbody>
</table>

Figure 5.3. a) Summary of experimental restraints and structural statistics observed in the final 48 refined NMR derived structures of UbFleu(50,67)

b) Energy statistics for the final 48 refined NMR derived structures of UbFleu(50,67)
Figure 5.4. Correlation between structural definition and experimental restraints. The top panel shows the number of nOes per residue. The lower panel is a line graph displaying the average Cα r.m.s.d. of the 48 accepted and refined UbFleu(50,67) structures from a calculated average structure.
with low Cα rms deviations, especially in the α-helical region (residues 23-32). The least defined region in the main body of the protein is the first reverse turn involving residues 9-10. The C-terminal tail, residues 74-76, is unrestrained and highly mobile and hence displays a very high Cα r.m.s.d. The 3D conformation of all 48 refined and accepted structures are displayed in figure 5.5, where the backbone Cα (residues 2-70) atoms of each structure are superimposed onto a calculated average. The definition of these structures correlates well with the data described in figure 5.4.
Figure 5.5. Backbone representation of the 48 accepted and refined structures of UbFleu(50,67). The calculated structures were superimposed from residues 2 to 70 on a calculated average structure, using the Cα atoms; (a) View of the molecule with the α-helix in front of the β-sheet; (b) view shown in (a) rotated through 180° along the y-axis.
Figure 5.6. Ramachandran\textsuperscript{136} plot of the minimised average structure of Ubfleu(50,67) for residues 1-73 only.
Figure 5.6 is a plot of ψ against φ, referred to as a Ramachandran plot, for residue 1-73 of the calculated average structure of UbFleu(50,67). It can be seen that 91.5% of residues are clustered into the allowed areas of the plot with only one residue, Lys 63 in a disallowed region.

5.4.2 UbAbu(26)Nva(30)

An energy profile of the original 60 UbAbuNva calculated structures before refinement is presented in figure 5.7. As expected with a good set of nOe data, there is a high proportion of calculated structures with low nOe energy and a few with misfolded, high energy conformations.

Figure 5.7. Structure Calculation Energy profile. Plot demonstrates the energy distribution of the 60 initial calculated structures of UbAbu(26)Nva(30).
A converged family of 36 low energy structures were accepted on the basis that no structure had a distance violation greater than 0.4 Å. The structures all satisfy the experimental data and display good chemical geometry and low nOe violations. The structures range in energy from 38.54 kJ/mol to 58.63 kJ/mol. An analysis of the experimental restraints and structural statistics of the 36 final structures (i.e. after minimisation) is given in figure 5.8. The average r.m.s.d. for all backbone was found to be 1.27 Å and for all heavy atoms was 2.09 Å. The correlation of structural definition to experimental data is demonstrated in figures 5.9, where the number of nOes for each residue and the average rmsd of the Cα atoms from a calculated average structure are compared.

The secondary structural elements of UbAbu(26)Nva(30) are generally well defined. The areas with the lowest r.m.s.d. correlate with the residues involves in the α-helical and β-strand secondary structural elements , i.e. residues 11-17 (β2 strand), 41-45(β3 strand) and 64-72 (β5 strand) and 23-31(α-helix). More mobile regions (high r.m.s.d.) are less well defined by the data. They occur where the residues are part of reverse turns such as residues 9 to 11 and residues 36-41. The free C-terminal is expected to be a completely mobile region and the extremely high r.m.s.d is an obvious indication of this. The structural definition is directly correlated with the number of nOe restraints incorporated for each residue. The 3D backbone image of the 36 refined and accepted UbAbu(26)Nva(30) structures are displayed in figure 5.10, where the backbone Cα (residues 2-70) atoms are superimposed onto the
### Table

<table>
<thead>
<tr>
<th>Category of Restraint</th>
<th>Number of restraints</th>
<th>rmsd ± sd for 36 structures Å or °</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experimental</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All nOe derived restraints</td>
<td>695</td>
<td>0.041±0.003</td>
</tr>
<tr>
<td>Intraresidue nOe’s</td>
<td>675</td>
<td>0.006±0.009</td>
</tr>
<tr>
<td>Sequential nOe’s</td>
<td>251</td>
<td>0.037±0.004</td>
</tr>
<tr>
<td>Short range nOe’s</td>
<td>75</td>
<td>0.036±0.007</td>
</tr>
<tr>
<td>Long range nOe’s</td>
<td>320</td>
<td>0.045±0.004</td>
</tr>
<tr>
<td>H-bond derived</td>
<td>20</td>
<td>0.053±0.015</td>
</tr>
<tr>
<td><strong>Chemical</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths</td>
<td>0.004±0.002</td>
<td></td>
</tr>
<tr>
<td>Impropers</td>
<td>0.51±0.043</td>
<td></td>
</tr>
<tr>
<td><strong>Convergence</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All non H atoms</td>
<td>2.090</td>
<td></td>
</tr>
<tr>
<td>Backbone atoms</td>
<td>1.266</td>
<td></td>
</tr>
</tbody>
</table>

### Table

<table>
<thead>
<tr>
<th>Energy term</th>
<th>Mean Value (kJ/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E(noe)</td>
<td>49.08</td>
</tr>
<tr>
<td>E(bond)</td>
<td>14.04</td>
</tr>
<tr>
<td>E(angle)</td>
<td>145.98</td>
</tr>
<tr>
<td>E(improper)</td>
<td>23.87</td>
</tr>
<tr>
<td>E(total)</td>
<td>271.57</td>
</tr>
</tbody>
</table>

**Figure 5.8:**

(a) Summary of experimental restraints and structural statistics observed in the final 36 refined NMR derived structures of UbAbu(26)Nva(30)

(b) Energy statistics for the final 36 refined NMR derived structures of UbAbu(26)Nva(30)
Figure 5.9: Correlation between structural definition and experimental restraints. The top panel shows the number of nOes per residue. The lower panel is a line graph displaying the average Cα r.m.s.d of the 36 accepted and refined UbAbu(26)Nva(30) structures from a calculated average structure.
Figure 5.10: Backbone representation of the 36 accepted and refined structure of UbAbu(26)Nva(30). The calculated structures were superimposed from residues 2 to 70 on a calculated average structure, using the Cα atoms; (a) View of the molecule with the α-helix in front of the β-sheet; (b) view shown in (a) rotated through 180° along the y-axis.
calculated average backbone atoms. The definition of these structures correlates well with the data described in figure 5.9.

The plot of $\psi$ against $\phi$ for the 36 refined and accepted structures of UbAbu(26)Nva(30) is shown in figure 5.11. This has $86.3\%$ of its residues in allowed regions and no residue (apart from glycine) in a disallowed region.
Figure 5.11. Ramachandran$^{136}$ plot of the minimised average structure of UbAbu(26)Nva(30) for residues 1-73 only.

Plot statistics

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residues in most favoured regions [A,B,L]</td>
<td>24</td>
<td>40.7%</td>
</tr>
<tr>
<td>Residues in additional allowed regions [a,b,l,p]</td>
<td>27</td>
<td>45.8%</td>
</tr>
<tr>
<td>Residues in generously allowed regions [-a,-b,-l,-p]</td>
<td>8</td>
<td>13.6%</td>
</tr>
<tr>
<td>Residues in disallowed regions</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>Number of non-glycine and non-proline residues</td>
<td>59</td>
<td>100.0%</td>
</tr>
<tr>
<td>Number of end-residues (excl. Gly and Pro)</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Number of glycine residues (shown as triangles)</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Number of proline residues</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Total number of residues</td>
<td>71</td>
<td></td>
</tr>
</tbody>
</table>

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.
Chapter Six

Discussion

6.1. Discussion of structural studies of the analogue UbFleu(50,67)

6.1.1 Expectations of the project.

Two Leu residues in native ubiquitin were substituted by Fleu, to create the analogue UbFleu(50,67). Significant structural changes in side-chain packing order were anticipated within the hydrophobic core as a result of these substitutions. Although the volume occupied by a covalent carbon-fluorine bond is similar to that occupied by the corresponding carbon-hydrogen bond, the C-F bond is much more polar than a C-H bond. It has also to be considered that a fluorine atom is nineteen times the mass of a proton, and therefore the frequency and amplitude of molecular vibrations and intra-molecular motions involving the fluorine-substituted group will be changed relative to those of the native system. A possible consequence of this is that fluorine substitution for hydrogen could alter the time-averaged positions of atoms in the protein.

The formation of a hydrogen bond from the fluorine atom to a proton with a positive partial charge is a small possibility. It has been reported in an extensive study of
intermolecular interactions in fluorine-containing organic compounds that fluorine hardly ever participates in a hydrogen bond\textsuperscript{141,142}. This contrasts with the strong H—F bond found in HF and inorganic salts. The inability of covalently bonded fluorine to form hydrogen bonds was first observed by Pauling\textsuperscript{140a,b} in 1960. He reported; ‘It is interesting that in general fluorine atoms attached to carbon do not have significant power to act as proton acceptors in the formation of hydrogen bonds in the way that would be anticipated from the large difference in electronegativity of fluorine and carbon’. More recently statistical analysis of data from the Cambridge Structural Database shows that out of 5947 C-F bond occurrences only 0.6% were involved in possible C-F—H-X, (X= N,O) hydrogen bonds\textsuperscript{141,142}. The scarcity of F—H bonds can be explained because for hydrogen bond formation to happen the proton affinities of the donor and acceptor must be closely matched, however the binding affinity of 2p electrons of fluorine is 3 eV greater than that of O (2p) and 6 eV greater than that of N (2p)\textsuperscript{143}. An additional factor is that fluorine is unable to modify this low basicity and tightness of electron shell through electron delocalisation; fluorine can only form single bonds and therefore cannot attract electrons through the p system in the same way as carbonyl oxygen or imine nitrogen\textsuperscript{142}.

Carbon-fluorine bonds have been shown to be more hydrophobic than C-H bonds and it also appears that fluorine substitutions can strengthen interactions between hydrophobic side chains\textsuperscript{144}. This is observed, for example, in the lower solubility of fluorocarbon gases in water\textsuperscript{145,146} and in the superior ability of fluoro-surfactants in lowering the surface tension of water as compared to hydrocarbon surfactants\textsuperscript{147}. 
6.1.2. Results and Observations

Detailed $^1$H and $^{19}$F NMR studies in both one- and two-dimensions were performed on the purified analogue, UbFleu(50,67). Extensive analysis of NMR data provided a set of constraints which generated a well-converged family of three dimensional structures. Each structure satisfied the experimental data and displayed good chemical geometries. The backbone atom r.m.s.d. (residues 2-70) for 48 accepted structures is 1.16 Å. The secondary structural elements and the overall conformation of the UbFleu(50,67) analogue closely resemble native ubiquitin, with a backbone r.m.s.d. for residues 2-70 of 1.59 Å. With reference to the MOLSCRIPT representation (figure 6.1), the deviations from the native structure can be identified. The analogue displays five β-strands with three-and-a-half turns of well formed α-helix. The native-like $3_{10}$ helical turn (residues 56-60) is also present in the analogue. The major difference in secondary structure occurs where β-strands 1 and 2 are twisted inward toward the C-terminus of the α-helix.

Figure 6.2 displays an overlay of the Fleu side chains for all 48 accepted and refined structures. In (a), all structures have been superimposed on Cα(2-70), but only the average ribbon backbone and side chains of residues 50 and 67 are shown. In (b) all structures have been superimposed on Cα, Cβ, Cγ, and Cδ of Fleu 50 and Fleu 67; protons have been omitted for clarity. The Fleu residues are shown to converge well with the fluorine atoms fixed in a particular orientation, although the convergence for Fleu 50 is better than for Fleu 67. The positioning of Fleu side chains are clarified in
Figure 6.1. Comparison of the secondary and tertiary structure of (a) bovine ubiquitin and (b) the calculated average structure of UbFleu(50,67), the positions of residues 50 and 67 are indicated.
Figure 6.2. (a) View of the hydrophobic core, highlighting positions of the unnatural amino acid 2S, 4S, 5-fluoroleucine 50 and 67 of all 48 accepted structures. (a) All structures have been superimposed on the backbone atoms only of the calculated average structure. (b) Stick representation of the accepted structures after superimposition on all carbon atoms at these positions. (c) View of the average orientation of fluoroleucine residues and (d) average orientation of unnatural amino acid residues with Van der Waal radii.
the average structure figure 6.2(c) and the van der Waals radii of the fluoro-methyl
groups (average structure) are shown in figure 6.2(d). The high definition of
orientation of fluorine within the core demonstrates that the rotation of the native δ-
methyl group is hindered when mono-substituted with a fluorine atom. The higher
degree of resolution in side chain orientation in the case of Fleu 50 was predicted
since the 2D HMQC spectrum (fig 4.7) displayed a more resolved resonance pattern.

In response to this finding, the factors governing the particular alignment of the
fluorine atoms were investigated for each site. One possibility, although unlikely on
the basis of the literature, is that the fluorine atom of Fleu 50 is stabilised through a
hydrogen bond. A potential hydrogen bond donor was identified as the amide
nitrogen of Ile 44 situated centrally on β-strand 3. The average measured distance
from Ile 44 (NH) to Fleu 50 (F) is 2.93 Å. This is slightly greater than the average C-
F—H—N distance measured in organic compounds which is 2.5 to 2.6 Å\textsuperscript{141,142}, and
close to the sum of van der Waals radii of hydrogen and fluorine (2.67 Å)\textsuperscript{149}. The
evidence to support the existence of this hydrogen - fluorine bond is a slowly
exchanging amide proton resonance, attributed to the amide proton of Ile 44.
However this protection was originally believed to arise from a hydrogen-bond in the
β-sheet secondary structure. The same amide proton cannot participate in both
hydrogen bonds.

To investigate the correct position of the hydrogen bond a new family of structures
were calculated and refined in exactly the same way, except the pseudo-nOe
representing the β-sheet hydrogen bond involving the amide proton of Ile 44 was replaced with an amide proton to fluorine hydrogen bond. The position of Fleu 67 with respect to Fleu 50 and Ile 44, before and after the relocation of the hydrogen bond is shown in figure 6.3. (The fluorine atom of Fleu 50 is barely visible in (ai) and (aii) since its projection is into the plane of the page). The convergence of Fleu 67 has not significantly improved with the introduction of the F−H bond, figure bii.

The newly generated structures (structure set B) converged well with a r.m.s.d. on all backbone atoms (residues 2-70) of 1.06 cf. 1.16 Å for the previous set of structures (structure set A). The number of structures accepted under the same criteria for structure set A and B were 42 % and 43 % respectively. The most commonly occurring nOe violations were investigated and it was revealed that structure set B displays about half the number of nOe violations of structure set A, suggesting that the nOe data are more consistent with the modified position of the hydrogen bond. However since a hydrogen bond of the type, N-H−F-C has been proved to be very rare\textsuperscript{41,42}, this possibility continues to be doubtful. A further, more concentrated, sample of UbFleu(50,67) is essential to fully establish this outcome.

If the assignment of the H−F bond is not correct then the orientational preference of the Fleu residues could be ascribed to the influence of favourable electrostatics. The calculated contact distance N-H−F-C is 2.67 Å (sum of Van der Waal radii\textsuperscript{149}) and the measured distance in the average structure (structure set A) is 2.93 Å. Since van der Waals forces are significant within 1.0 Å greater then the contact distance this interpretation is viable. The fluorine atom of Fleu 67 is pointing away from Fleu 50
Figure 6.3. Summary of the orientation of the Fleu residues within the hydrophobic core (ai) Demonstration of a possible hydrogen bond between HN of Ile 44 and the F atom of Fleu 50 and (aii) the corresponding orientation of Fleu 67 in comparison to HN 44 to F50. (bi and ii) Detailed view of the protein molecule demonstrating the Fleu convergence once the H-F bond has been incorporated.
and into the centre of the hydrophobic core. Fleu 67 appears to have buried its side chain into the heart of the hydrophobic core, in the same fashion as Leu 67 in native ubiquitin. Detailed analysis failed to suggest any possible electron acceptor which could facilitate hydrogen-bond formation, and so explain the convergence in Fleu 67 conformation. However, the F atom of Fleu 67 being directed into the very centre of the core supports previous observations that fluorocarbons have an increased hydrophobic nature\textsuperscript{144,147} and therefore favour conformations which optimises the extent of hydrophobic burial.

The existence of a favourable interaction involving either a fluorine to hydrogen-bond or van der Waals contact promotes a favoured conformation for the fluoro-methyl group of Fleu 50. This could in turn induce a defined arrangement of the other fluoro-methyl within the core. It has been suggested that if an imaginary bond was placed between the fluoro-methyl carbon atoms of Fleu 50 and Fleu 67, the arrangement of the F atoms might reflect the preferred gauche conformation of the fluorine atoms in 1,2,-difluoroethane\textsuperscript{150,151}. The optimum FCCF dihedral angle in 1,2,-difluoroethane is measured to be $\pm 73^\circ$. 


The occurrence of the gauche conformation in 1,2-difluoroethane has been explained in electronic terms\textsuperscript{152}. The electronegativity of the fluorine atom withdraws electrons, leaving a low electron density at the attached carbon. The other carbon gives a hyperconjugative response, which corresponds to resonance of the following type:

\[
\begin{align*}
FCH_2CH_2F & \rightleftharpoons \quad -FCH_2 & \quad +CHFH & \rightleftharpoons \quad +HFHC & \quad CH_2F
\end{align*}
\]

This effect hinders the fluoro-methyl rotation and therefore conformational preferences are likely to occur. A co-operative resonance effect (in both directions) is only possible when the fluorines are 90° to one another so that both pairs of 2p \( \pi \) orbitals can be simultaneously involved. These resonance forms are however unusual and not expected to be of major importance in bonding since in all cases studied they
have a combined effect of $< 2 \text{ kcal mol}^{-1}$. This barrier, however, is large enough to have some significance in determining conformational behaviour.

The electronic explanation for a gauche conformation attributed to 1,2-difluoroethane is impossible in UbFleu(50,67) for two reasons. Firstly, the fluoro-methyl groups are not directed towards one another across the hydrophobic core. The measured angle $\text{C}_7 - \text{C}_8(50)$ to $\text{C}_8 - \text{C}_9(67)$ (figure 6.7(a) is 66° and 56° for structure set A and B respectively; this means the $2p\pi$ orbitals of the fluorines will not be 90° to one another which is required for this resonance effect. Secondly, the average $\text{C}_8 - \text{C}_9$ distance across the hydrophobic core ($3.52 \pm 0.26 \text{ Å}$) is too large to accommodate this gauche effect.

Electrostatic interactions are the most fundamental non-covalent effect. The partial charges carried by asymmetric covalent bonds allow attractive and repulsive effects between non-bonded groups. The charges interact according to Coulomb’s Law (equation 2, chapter one). The average binding energy for the $\text{F} \cdots \text{F}$ interaction for structure set A is calculated to be $+2.22 \text{ kcal mol}^{-1}$, (where $\varepsilon$, the dielectric constant is assumed to be 4 and $Z$, the partial charge of fluorine is -0.35). This demonstrates that there is an evident repulsive energy between the two fluorine atoms as expected. For comparison, the binding energy of hydrogen bonds in ice, $\text{H} \cdots \text{O}$-H is $-4.0 \text{ kcal mol}^{-1}$ and in a protein backbone $\text{N} \cdots \text{O}$ is $-3.0 \text{ kcal mol}^{-1}$. The repulsive energy between two fluorine atoms is sufficient to push them away from each other within their own fluoro-methyl volume. A larger effect involving the displacement...
Figure 6.7. (a) Visual description of the angle measured between 2S, 4S, 5-fluoroleucine side chains, (the protons have been omitted for clarity). (b) 2S, 4S 5-fluoroleucine indicating the fluoromethyl volume and C\textgamma-C\delta1 rotation axis.
of other atoms of the hydrophobic core is not expected due to the small amount of repulsive energy. Measurements reveal that the fluorine atoms are neither as close nor as far apart as is possible, if only the fluoro-methyl groups are rotated about the Cδ₁-Cγ bond, (figure 6.7); the closest and furthest possible distances between fluorines being 1.94 Å and 4.93 Å respectively. The average F–F distance in structure set A lies within this range, at 4.57 Å. This implies there is insufficient repulsive energy available to repel the fluorine atoms away from each other as far as possible.

A more rigorous investigation into the conformational preferences of the fluoro-methyl groups were analysed through a grid-search operation. This performs a systematic search on a set of torsion angles and calculates the energetics of each conformation. The torsion angle F-Cδ₁-Cγ-Cδ₂ was defined in each Fleu residue, as indicated in figure 6.7b. Fleu residue 50 was held static whilst Fleu 67 was rotated through 360° about the Cδ₁-Cγ axis, (figure 6.7b) in 72 steps of 5° increments. Fleu 50 was then rotated in 5° increments about its Cδ₁-C7 axis and Fleu 67 rotated by a full 360° for each increment. This operation was performed within XPLOR employing a ‘charm’ forcefield with electrostatics. The output from the grid-search is a 2D contour plot (figure 6.8) which represents the energetics (kJ mol⁻¹) of the molecule at the corresponding defined torsion values. The plot is divided into an area of low energy conformations (shown in orange) which are preferred to the high energy conformations (shown as red ‘hot spots’). The measured torsion angles of the average calculated UbFleu(50,67) structure are 233° and 351° for Fleu 67 and 50
Figure 6.8. 2D contour plot representing the energetics, $E_{\text{total}}$, of UbFleu(50,67) as the torsion angle $F-C\delta_1-C\gamma-C\delta_2$ at residues Fleu 50 and 67 is varied.
respectively. These angles were used as the 'starting conformation' of the search experiment. Therefore the energy of the calculated average conformation of UbFleu(50,67) is found at the (0,0) position in figure 6.8, which falls into an energetically favoured area (yellow shading). Nearby, there are regions of slightly lower energy, e.g. if the Fleu 67 methyl group were rotated by 180° the total energy of the molecule would be 20 kJ mol⁻¹ lower. However, given the semi-empirical nature of the force field used, it is difficult to attach significance to these kinds of energy differences. Overall, we can probably conclude that the conformation as determined by experimentally derived distance restraints is consistent with theoretical low energy rotomers of the fluoro-methyl groups.

The ¹⁹F chemical shift range is much reduced in UbFleu(50,67) compared to that reported in Fleu enriched DHFR³⁴ (15.2 ppm) and what is typically found in proteins containing fluorinated aromatic amino acids (17 ppm). In UbFleu(50,67) the ¹⁹F chemical shift range is 3.3 ppm. This small range is due to the fact that only two Fleu residues are present in the protein and these are buried within the hydrophobic core. This environment protects them from fluctuating electronic fields arising form aromatic and polar residues. The 13 Fleu enriched sites in DHFR are distributed throughout the protein, therefore these Fleu ¹⁹F will experience a wide range of electronic environments resulting in large shifts and a wide ¹⁹F chemical shift range.
6.1.3. Purification and folding problems

The UbFleu(50,67) material provided was purified as far as possible. The final amount of material obtained after purification was so low (6.4 mg) that any extra purification stage would have depleted the sample to a point where insufficient material would be remaining for NMR analysis.

An initial 1D $^1$H NMR of the UbFleu(50,67) appeared promising, displaying resonances typical of a folded peptide. However, the proportions of folded and unfolded material is difficult to estimate solely through NMR. The amide exchange NMR experiment gave an indication of the folded nature of the UbFleu(50,67). As well as a reduced number of hydrogen bonds involved in the secondary structure, there is an accelerated D$_2$O exchange rate. From this information the fold appears to be more open, lacking the rigidity of native ubiquitin.

A DSC experiment was used primarily, to determine the thermal transition of UbFleu(50,67), as well as to reveal the amount of folded material in the sample. The experiment shows that the thermal transition of UbFleu(50,67) occurs at the same temperature as bovine ubiquitin (70 °C). However, the integral of the transition peak is 20 % the size anticipated and this reflects the estimated amount of folded material within the sample.
As a consequence of the low folded peptide content, the time required to collect 2D $^1$H NMR spectra was very long (36 hours per experiment). Despite resonances from unfolded peptide molecules and a large artefact due to acetate, it was possible to pursue spectral assignment. Problems in the production of more material of high purity and in a folded state has slowed the development of this project. Another more concentrated sample is crucial to fully confirm and strengthen the validity of these results.

The yield of synthetic ubiquitin analogue purifications has been (in this instance and also in subsequent purifications) disappointingly very low. This can possibly be attributed to the difficulty in folding synthetic peptides in vitro. In biological systems, folding of the ubiquitin molecule is enzyme-catalysed using peptidyl prolyl isomerases. No such enzyme has been used in a folding protocol for synthetic ubiquitin analogues so far but this could be attempted in future synthetic ubiquitin studies.

Within a chemically pure sample, however, there will always be a proportion of molecules which are by chance, correctly folded. The final cation-exchange stage of purification gave a very low yield of folded peptide (as a percentage of the total peptide loaded onto the column). It is probable that the small proportion of folded UbFleu(50,67) molecules present were separated from the rest of the material in this final stage. This is possible since the folded molecules behave as expected in the cation-exchange column, i.e. they bind to the column and are eluted in the expected
position, whereas, the misfolded or aggregated peptide molecules seem to adhere to the column much tighter than anticipated, making recovery of these molecules very difficult.

6.1.4. Summary

It can be concluded from this study that synthetically manufactured UbFleu(50,67) resembles the overall conformation of native ubiquitin. The unnatural amino acid residues, 2S, 4S, 5- fluoroleucine behave differently to the original Leu residues. It is observed that the fluoro-methyl groups of Fleu adopt preferred conformations within the hydrophobic core. This convergence can be partly attributed to a small electrostatic repulsion between the two F atoms. There is also some evidence to support the possibility of a fluorine to amide proton hydrogen bond. Future work should include production of a good folded sample of UbFleu(50,67) to carry out essential $^{3}J_{NH_{2}}$ measurements. Once this information is included in the structure calculations, firmer conclusions as to the conformational behaviour of Fleu within the hydrophobic core of ubiquitin can be made.
6.2. Discussion of structural studies of the analogue UbAbu(26)Nva(30)

6.2.1. Expectations of the project

Two α-helical residues in native ubiquitin were replaced with the unnatural amino acids aminobutyric acid and norvaline, to give the analogue UbAbu(26)Nva(30). These changes resulted in a net loss of two methyl groups from the hydrophobic core. Figure 6.9 shows the position and van der Waals volume of the relevant methyl groups in the crystal structure of human ubiquitin. The effect of the loss of this steric bulk on the stability and tertiary structure of ubiquitin may lead us further to understand the importance of individual residues within the core, and their role in hydrophobic side chain packing.

The X-ray crystal structure of UbAbu(26)Nva(30), has been determined. This study reported that the loss of the two methyl groups has caused the whole hydrophobic core to contract slightly, however the authors did not discuss any specific α-helical distortions. The most apparent differences involve the amino acid residues of the β-strand opposite the α-helix (Val 17, Leu 15 and Ile 13) which move towards the α-helix. This has been attributed to the partial filling of the vacated methyl volume. However, this is only a small compensation since a net destabilisation of 14 kJ mol⁻¹ is also reported. The X-ray-determined UbAbu(26)Nva(30) structure will be compared closely with the solution structure obtained through NMR.
Figure 6.9. Ribbon representation of native ubiquitin with residues to be substituted highlighted. The van der Waal radii of the methyl groups to be deleted in the analogue are shown.
6.2.2. Results and Observations

We have successfully determined a 3D structure on the basis of high-field NMR spectroscopy. The family of 36 accepted low energy structures satisfy the experimental data and display good chemical geometry and low nOe violations. The Ramachandran plot shows the structure to be legitimate with 86.5% of residues in most favoured regions and no residues at all in disallowed regions. This implies that the nOe data has been correctly analysed.

The final averaged structure of UbAbu(26)Nva(30) is shown in figure 6.10 as a MOLSCRIPT representation. The main difference in secondary structure compared to the native structure is that the α-helix is short by one residue (residue 23) at the N-terminal. Although there are characteristic α-helical nOes involving Ile 23, including a typically weak $\alpha N_{(i+1)}$ nOe, most of them ($NN_{(i+1)}$, $NN_{(i+2)}$, $\alpha N_{(i+3)}$, $\alpha \beta_{(i, i+3)}$, and $\alpha N_{(i, i+3)}$, where $i = 23$) are weaker in intensity than expected from an α-helical residue. The final turn at the C-terminal is also not completely α-helical, appearing more ‘open’ in comparison to that of native ubiquitin. Another obvious difference is that β-strands 1 and 2 are twisted with respect to one another in the analogue. The backbone atom r.m.s.d. of the average solution structure compared with the solution-structure of native ubiquitin is 1.27Å. This deviation in backbone conformation most likely arises from structural compensations made within the core in order to accommodate the cavities created by the deletion of the two methyl groups in the
Figure 6.10. Comparison of the secondary and tertiary structure of (a) bovine ubiquitin and (b) the calculated average structure of UbAbu(26)Nva(30), the positions of residues 26 and 30.
core. They may however also result from too few nOe and the lack of hydrogen bond distance restraints within the structure calculation.

Differences in amino acid side chain packing are expected to be localised around the sites of substitution i.e. α-helical positions, 26 and 30. The calculated, NMR-derived structures are superimposed onto the backbone atoms of the average UbAbu(26)Nva(30) structure and presented in figure 6.11, where the average ribbon backbone is shown together with the side chains of all the calculated structures. This figure demonstrates that the overall convergence in orientation of the unnatural amino acid, Abu 26 and Nva 30 is modest. As shown in figure 6.11(a) the side chain of residue Abu 26 has only one preferred orientation, whereas the side chain of Nva 30 has two preferred orientations, (i and ii). Nva 30 conformation (i) is directed away from Abu 26, whereas Nva 30 conformation (ii) points towards Abu 26.

The relative orientation of Abu 26 and Nva 30 in comparison to the native Val 26 and Ile 30 residues is made in figures 6.11(b) and (c), where each calculated UbAbu(26)Nva(30) structure is shown in green, the average UbAbu(26)Nva(30) is shown in blue and native ubiquitin in red. Figure 6.11(b) displays the average Abu 26 side chain position to be equi-distance between the native positions of Val 26 CyH₃ groups. This shift has positioned the Abu 26 side chain where it can effectively balance areas of empty methyl volume.
Figure 6.11. Diagram to show orientations of Abu 26 and Nva 30 with respect to native ubiquitin. The side chain of Abu and Nva of each calculated structure is shown in green and the average calculated structure in blue; native ubiquitin is shown in red. (a) Orientations of Abu and Nva within the hydrophobic core. (b) Detailed view of Abu and (c) Nva side-chain conformation with respect to native ubiquitin.
The positions of the two calculated orientations for Nva 30, (figure 6.11 c) correlate well with each of the original β-branches of the Ile 30 in native ubiquitin.

The Nva side-chain adopts a conformation which is positioned almost identically to the Cγ1-Cδ branch of Ile 30 in the native protein, (orientation i), else it displays a full 180° rotation about the Cα-Cβ bond which effectively occupies the vacant CγH₃ volume of Ile 30, (orientation ii). The average picture of the Nva 30 orientation would produce an image with the side chain lying in between these two states i.e.

Figure 6.12. Orientation of Nva 30 in UbAbu(26)Nva(30) in comparison to Ile 30 in native Ubiquitin.
between the Cγ₁ and Cγ₂ atoms of the original Ile 30 position. Interestingly this is the observation made in the X-ray determined UbAbu(26)Nva(30) structure.

A comparison of orientation of residues occupying positions 26 and 30 in UbAbu(26)Nva(30), (both NMR and X-ray structures), as well as in native ubiquitin is made in figure 6.13. The figure displays the NMR-derived ribbon backbone of UbAbu(26)Nva(30) with all residues 26 and 30; native ubiquitin is shown in red and the NMR and X-ray structure of UbAbu(26)Nva(30) are shown in blue and green respectively. These three forms superimposed on the Cα atoms of the α–helix residues only (residues 22-35). The orientations of residues Abu and Nva in both the X-ray and NMR determined structures differ slightly. This is more pronounced in the case of position Abu 26 where the Cβ-Cγ of both Abu residues are orientated approximately 120° to one another each following a separate Cβ-CγH₃ branch of the native Val 26 position, (figure 6.13a).
Figure 6.13. Comparison of NMR and X-ray-derived UbAbu(26(Nva(30) structure and native ubiquitin structure. Native ubiquitin is shown in red, while the X-ray crystallographic structure of UbAbuNva is green and NMR determined structure of UbAbuNva is shown in blue. (a) All three structures are superimposed on the Ca atoms of the α-helix only; (b) All structures are superimposed on Ca atoms of residues 2 to 70.
Figure 6.14. Schematic representations to illustrate the relative orientations of Abu 26 of UbAbu(26)Nva(30) in both NMR-determined and X-ray-determined structures in comparison to Val 26 in native ubiquitin.

The orientation of Nva 30 in both NMR- and X-ray- determined UbAbu(26)Nva(30) structures display a similar orientations, although in the case of the X-ray determined structure Nva 30 appears to be ‘twisted’ further from the native Ile 30 position, figure 6.13(a).

A more realistic impression is displayed in figure 6.13(b), where the backbone Ca atoms (2-70) (rather than the α-helix atoms only), are superimposed and the side
chains at positions 26 and 30 shown; native ubiquitin is shown in red, 
UbAbu(26)Nva(30) of NMR and X-ray determined structure are represented in blue and green respectively. The backbone atom r.m.s.d of native ubiquitin and X-ray determined structure of UbAbu(26)Nva(30) is very small (0.13 Å). However, the α-helix of the NMR-derived UbAbu(26)Nva(30) diverges away from this conformation with a more pronounced inward shift.

A minor ‘contractive’ effect of the hydrophobic core was observed in the X-ray structure. This primarily involved inward movement of the residues involved in the α-helix to β-sheet interface, (figure 6.15 b). This contraction was investigated in the NMR - determined structure and compared to the X-ray structure and that of native ubiquitin. Distance measurements were made between residues 26 and 30 and residues on the facing β-strand, (figure 6.16) for both UbAbu(26)Nva(30) structures (X-ray and NMR derived) and native ubiquitin. The Cα(i) to Cα(j) distance across the core was measured in each case. The results are presented in a bar chart, (figure 6.17), which reveals that a larger overall contraction of the hydrophobic core is evident in the case of the NMR structure compared to the X-ray structure. However the contractions occur in different parts of the hydrophobic core by significant amounts. Although the secondary structural components of UbAbu(26)Nva(30) are largely intact in comparison to the native protein, the overall conformation displays large distortions. Figure 6.18(a) illustrates the structural differences between the backbones; the NMR structure is shown as a blue solid oval ribbon, and the X-ray derived ribbon backbone as green lines. The NMR derived backbone is able to be
Figure 6.15. (a) The native ubiquitin backbone is shown; residues 26 and 30 of both native (red) and UbAhu(26)Nva(30) (green) are highlighted. (b) Detailed view of the hydrophobic core with residues contributing to the α-helix to β-sheet interface highlighted.
Figure 6.16. Model of native ubiquitin demonstrating the distances measured across the hydrophobic core from the Cα atoms of postions 26 and 30 to Cα atoms of core residues on the facing β-sheet.
Figure 6.17. Distance measurements from the Cα of (a) Abu 26 and (b) Nva 30 to the Cα of each β-sheet hydrophobic residue.
Figure 6.18. Comparison of 3D structure of UbAbuNva derived through X-ray crystallography (green) and NMR (blue); (a) superimposition of Cα atoms of the secondary structural elements; (b) Van der Waal radii of all core residues.
traced inside the backbone of the X-ray crystal structure showing hydrophobic core volume of the NMR structure is significantly reduced. The van der Waal radii of all side chains of the core residues are shown in figure 6.18b; the green spheres are those of the X-ray determined UbAbu(26)Nva(30) structure and the blue spheres are those of the NMR derived structure. Careful examination indicates that the core hydrophobic side-chains of the X-ray determined structure pack into a smaller volume than the core residues of the NMR structure, even though the apparent core contraction is more pronounced in the NMR-derived structure. Less dense packing in the core of the NMR structure can be explained since the larger contraction of the core has produced other distortions including 'squeezing out' hydrophobic residues from the periphery of the core. This expulsion is inevitable since pronounced collapse of the hydrophobic core can not accommodate the original atom volume.

The lack of thermal transition in the DSC experiment indicates lack of rigid tertiary structure within the sample. In agreement with this data there is poor evidence of a structured hydrogen bonding network within the molecule and a collapsed tertiary structure.

A possible 3D solution structure has been reached through NMR. However, this structure significantly differs from the X-ray determined structure, the largest difference being the core volume. A test was carried out to check the validity of the nOe data used in the structure calculation. Each distance restraint used in the XPLOR calculation was compared to the distance in the X-ray derived structure of
UbAbu(26)Nva(30). The results revealed that only 8 out of 697 nOes differed by more than 3.0 Å. These are listed below in figure 6.19. The 2D $^1$H NOESY spectra was consulted to re-check if each of these nOes had been correctly assigned, figure 6.20 a-f.

<table>
<thead>
<tr>
<th>nOe violated</th>
<th>nOe intensity</th>
<th>X-ray-NMR structure distance difference</th>
<th>Validity of nOe</th>
</tr>
</thead>
<tbody>
<tr>
<td>I3 CγH₂ → F45 Cβ1H</td>
<td>weak</td>
<td>4.0 Å</td>
<td>correct, very weak</td>
</tr>
<tr>
<td>L8 NH → L15 CδH₃</td>
<td>very weak</td>
<td>6.7 Å</td>
<td>incorrectly assigned</td>
</tr>
<tr>
<td>V17 CγH₃ → F45 CεH</td>
<td>very weak</td>
<td>3.7 Å</td>
<td>correct</td>
</tr>
<tr>
<td>V17 CγH₃ → F45 CεH</td>
<td>very weak</td>
<td>3.6 Å</td>
<td>correct</td>
</tr>
<tr>
<td>V17 CαH → L56 CδH₃</td>
<td>strong</td>
<td>3.5 Å</td>
<td>correct</td>
</tr>
<tr>
<td>I23 NH → L69 Cδ1H₃</td>
<td>weak</td>
<td>8.1 Å</td>
<td>correct</td>
</tr>
<tr>
<td>I23 NH → L69 Cδ2H₃</td>
<td>weak</td>
<td>5.7 Å</td>
<td>correct very weak</td>
</tr>
<tr>
<td>Q41 NH → V70 Ha</td>
<td>medium</td>
<td>3.6 Å</td>
<td>medium</td>
</tr>
</tbody>
</table>

**Figure 6.19.** Heavily violated nOes produced through comparisons of NMR and X-ray derived UbAbu(26)Nva(30) structures.

All but one of these nOes were re-confirmed as being assigned correctly in the spectra and the therefore large changes in side chain packing order are evident. In addition, seven out of eight of these nOes involve at least one residue which is a main contributor to the packing within the hydrophobic core. This solution structure of UbAbu(26)Nva(30) may possibly represent a species which has been generated through the assessment of nOe intensity being over-estimated and therefore distance restraints being assigned values which are too small. The NMR-derived UbAbu(26)Nva(30) structure may however represent a ‘collapsed’ folded form or a
Figure 6.20. Assessment of heavily violated noes (a) 13 CyH2 → F45 Cβ1H and (b) L8 NH → L15 C82H₃
Figure 6.20. Assessment of heavily violated nOes (c) V17 Cγ2H → F45 CeH; V17 Cγ2H → F45 CδH and (d) V17 CaH → L56 Cδ1H,
Figure 6.20. Assessment of heavily violated nOes (e) I23 NH → L69 C81H3 and I23 NH → L69 C82H3 and (f) V70 Hα → Q41 NH
trapped folded intermediate. The reason for having obtained such a species may lie in the purification methodology followed. A common problem of synthetic peptide chemistry is the difficulty in folding. It has been observed that employing the purification method using Tbfmoc to derivitise the N-terminal of ubiquitin more commonly produces incorrectly folded peptides, compared to previous purifications of ubiquitin unmodified at the N-terminal. The purification protocol of UbAbu(26)Nva(30) in this instance involved the use of the Tbfmoc molecule as an affinity tag. Although the protocol yielded folded material, the effect of the large hydrophobic molecule may have produced a misfolded peptide.

6.2.3. Summary

The $^1$H NMR studies of the analogue UbAbu(26)Nva(30) reveal a defined 3D conformation. As predicted the $\alpha$-helix is not fully formed, being reduced by one residue (at the N-terminus) and having an incomplete turn occurring at the C-terminus. The unnatural amino acids within the core however, converge well, having well defined orientations in each case. The side chain of Abu 26 has not moved from its original position. The side chain of Nva 30 has two equally preferred orientations, one representing its original position and the second filling the vacated volume of the deleted $\beta$-branch of Ile 30. The main structural difference between the X-ray and the NMR determined structure of UbAbu(26)Nva(30) is that the latter displays a more pronounced collapsed hydrophobic core.
Future work required is to refine the NMR-derived structure of UbAbu(26)Nva(30) using backbone dihedral angle measurements ($^{3}J_{	ext{NHis}}$) and to confirm observations made in this thesis through repetition of the work.
References


3  Levinthal, C., J. Chem. Phys, (1968), 65, 44


6  Kim, PS., Baldwin RL., Annu. Rev. Biochem, (1990), 59:631-60


11 Fisher, G. and Schmid, F. X., Biochemistry, (1990), 29, 2205

12 E. Fisher, Einfluss der Configuration auf die Wirkung der Enzyme. Chem. Ber., (1894), 27, 2985-2993

13 Dill KA, Biochemistry, (1990), 29,7133-55


63 S. Korasanizaden *et al.*, *Biochemistry*, (1993), 32, 7054


75 Padmanabhan S., Marqusee. S., Ridgeway, T., and Laue, T. M. *Nature*, (1990), 344, 268-270


83 L.A. Carpino and G.Y.Han, *J. Org. Chem.*, (1972), 37, 3404


94 J.Wilken; PhD thesis (Edinburgh) (1995)


105 Kay, L.E., Clore, G.M., Bax, A., Gronenborn, A. M. *Science: 4-dimensional heteronuclear triple resonance NMR spectroscopy of interleukin-1-beta in solution*, (1990), 249, 411-414


130  Wüthrich, K., Billeter, M., and Braun, W., *J. Mol. Biol.* (1983), **169**, 949-961


(b) ibid 3rd ed., (1960), p464 A footnote acknowledges V. Schomaker as a source of this observation.


APPENDIX A

Hydrogen bonds for β-strands
assign (resid 1 and name O ) (resid 17 and name HN ) 1.88 0.3 0.42
assign (resid 1 and name O ) (resid 17 and name N ) 1.88 0.3 1.32
assign (resid 2 and name O ) (resid 65 and name HN ) 1.88 0.3 0.42
assign (resid 2 and name O ) (resid 65 and name N ) 1.88 0.3 1.32
assign (resid 3 and name O ) (resid 15 and name HN ) 1.88 0.3 0.42
assign (resid 3 and name O ) (resid 15 and name N ) 1.88 0.3 1.32
assign (resid 4 and name O ) (resid 67 and name HN ) 1.88 0.3 0.42
assign (resid 4 and name O ) (resid 67 and name N ) 1.88 0.3 1.32
assign (resid 6 and name O ) (resid 69 and name HN ) 1.88 0.3 0.42
assign (resid 6 and name O ) (resid 69 and name N ) 1.88 0.3 1.32
assign (resid 7 and name O ) (resid 11 and name HN ) 1.88 0.3 0.42
assign (resid 7 and name O ) (resid 11 and name N ) 1.88 0.3 1.32
assign (resid 11 and name O ) (resid 7 and name HN ) 1.88 0.3 0.42
assign (resid 11 and name O ) (resid 7 and name N ) 1.88 0.3 1.32
assign (resid 13 and name O ) (resid 5 and name HN ) 1.88 0.3 0.42
assign (resid 13 and name O ) (resid 5 and name N ) 1.88 0.3 1.32
assign (resid 15 and name O ) (resid 3 and name HN ) 1.88 0.3 0.42
assign (resid 15 and name O ) (resid 3 and name N ) 1.88 0.3 1.32
assign (resid 42 and name O ) (resid 70 and name HN ) 1.88 0.3 0.42
assign (resid 42 and name O ) (resid 70 and name N ) 1.88 0.3 1.32
assign (resid 44 and name O ) (resid 68 and name HN ) 1.88 0.3 0.42
assign (resid 44 and name O ) (resid 68 and name N ) 1.88 0.3 1.32
assign (resid 45 and name O ) (resid 48 and name HN ) 1.88 0.3 0.42
assign (resid 45 and name O ) (resid 48 and name N ) 1.88 0.3 1.32
assign (resid 48 and name O ) (resid 45 and name HN ) 1.88 0.3 0.42
assign (resid 48 and name O ) (resid 45 and name N ) 1.88 0.3 1.32
assign (resid 65 and name O ) (resid 4 and name HN ) 1.88 0.3 0.42
assign (resid 65 and name O ) (resid 4 and name N ) 1.88 0.3 1.32

Hydrogen bonds for α-helix
assign (resid 25 and name O ) (resid 28 and name HN ) 1.88 0.3 0.42
assign (resid 25 and name O ) (resid 28 and name N ) 1.88 0.3 1.32
assign (resid 28 and name O ) (resid 29 and name HN ) 1.88 0.3 0.42
assign (resid 28 and name O ) (resid 29 and name N ) 1.88 0.3 1.32
assign (resid 27 and name O ) (resid 30 and name HN ) 1.88 0.3 0.42
assign (resid 27 and name O ) (resid 30 and name N ) 1.88 0.3 1.32
assign (resid 26 and name O ) (resid 29 and name HN ) 1.88 0.3 0.42
assign (resid 26 and name O ) (resid 29 and name N ) 1.88 0.3 1.32
assign (resid 28 and name O ) (resid 31 and name HN ) 1.88 0.3 0.42
assign (resid 28 and name O ) (resid 31 and name N ) 1.88 0.3 1.32

Additional hydrogen bonds
Structure set A
assign (resid 68 and name O ) (resid 44 and name HN ) 1.88 0.3 0.42
assign (resid 68 and name O ) (resid 44 and name N ) 1.88 0.3 1.32
Structure set B
assign (resid 50 and name F ) (resid 44 and name HN ) 1.88 0.3 0.82
assign (resid 50 and name F ) (resid 44 and name N ) 1.88 0.3 1.72

UbAbu(26)Nva(30)

Hydrogen bonds for β-strands
assign (resid 1 and name O ) (resid 17 and name HN ) 1.88 0.3 0.42
assign (resid 1 and name O ) (resid 17 and name N ) 1.88 0.3 1.32
assign (resid 13 and name O ) (resid 5 and name HN ) 1.88 0.3 0.42
assign (resid 13 and name O ) (resid 5 and name N ) 1.88 0.3 1.32
assign (resid 15 and name O ) (resid 3 and name HN ) 1.88 0.3 0.42
assign (resid 15 and name O ) (resid 3 and name N ) 1.88 0.3 1.32
assign (resid 48 and name O ) (resid 45 and name HN ) 1.88 0.3 0.42
assign (resid 48 and name O ) (resid 45 and name N ) 1.88 0.3 1.32
assign (resid 65 and name O ) (resid 4 and name HN ) 1.88 0.3 0.42
assign (resid 65 and name O ) (resid 4 and name N ) 1.88 0.3 1.32

Hydrogen bonds for α-helix
assign (resid 25 and name O ) (resid 28 and name HN ) 1.88 0.3 0.42
assign (resid 25 and name O ) (resid 28 and name N ) 1.88 0.3 1.32
assign (resid 27 and name O ) (resid 30 and name HN ) 1.88 0.3 0.42
assign (resid 27 and name O ) (resid 30 and name N ) 1.88 0.3 1.32
assign (resid 28 and name O ) (resid 31 and name HN ) 1.88 0.3 0.42
assign (resid 28 and name O ) (resid 31 and name N ) 1.88 0.3 1.32
### APPENDIX B: UbFleu(50,67) nOe table

#### B1: Intraresidue nOes

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Distances</th>
<th>Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>resid 3 and name HN</td>
<td>resid 3 and name HG2*</td>
<td>0.0 1.5 6.5 lw</td>
</tr>
<tr>
<td>resid 4 and name HA</td>
<td>resid 4 and name HE*</td>
<td>0.0 2.0 5.5 lm</td>
</tr>
<tr>
<td>resid 4 and name HA</td>
<td>resid 4 and name HD*</td>
<td>0.0 2.0 4.5 Is</td>
</tr>
<tr>
<td>resid 4 and name HD*</td>
<td>resid 4 and name HG</td>
<td>0.0 2.0 4.5 Is</td>
</tr>
<tr>
<td>resid 4 and name HE*</td>
<td>resid 4 and name HG</td>
<td>0.0 2.0 7.0 tw</td>
</tr>
<tr>
<td>resid 5 and name HG1*</td>
<td>resid 5 and name HG2*</td>
<td>0.0 1.5 6.5 lw</td>
</tr>
<tr>
<td>resid 5 and name HE*</td>
<td>resid 5 and name HG2*</td>
<td>0.0 1.5 6.5 lw</td>
</tr>
<tr>
<td>resid 7 and name HG1*</td>
<td>resid 7 and name HG2*</td>
<td>0.0 1.5 6.5 lw</td>
</tr>
<tr>
<td>resid 9 and name HG1*</td>
<td>resid 9 and name HG2*</td>
<td>0.0 1.5 4.0 ls</td>
</tr>
<tr>
<td>resid 13 and name HN</td>
<td>resid 13 and name HG2*</td>
<td>0.0 1.5 4.0 ls</td>
</tr>
<tr>
<td>resid 13 and name HG2*</td>
<td>resid 13 and name HG2</td>
<td>0.0 1.5 4.0 ls</td>
</tr>
<tr>
<td>resid 15 and name HG</td>
<td>resid 15 and name HE*</td>
<td>0.0 0.0 2.5 ls</td>
</tr>
<tr>
<td>resid 15 and name HG</td>
<td>resid 15 and name HG1*</td>
<td>0.0 1.5 5.0 lm</td>
</tr>
<tr>
<td>resid 17 and name HG</td>
<td>resid 17 and name HG2*</td>
<td>0.0 1.5 4.0 ls</td>
</tr>
<tr>
<td>resid 17 and name HG1*</td>
<td>resid 17 and name HG2*</td>
<td>0.0 1.5 5.0 lm</td>
</tr>
<tr>
<td>resid 23 and name HN</td>
<td>resid 23 and name HG2*</td>
<td>0.0 1.5 4.0 ls</td>
</tr>
<tr>
<td>resid 26 and name HG1*</td>
<td>resid 26 and name HG2*</td>
<td>0.0 1.5 6.5 lw</td>
</tr>
<tr>
<td>resid 26 and name HG1*</td>
<td>resid 26 and name HG2</td>
<td>0.0 1.5 6.5 lw</td>
</tr>
<tr>
<td>resid 30 and name HN</td>
<td>resid 30 and name HG2*</td>
<td>0.0 1.5 5.0 lm</td>
</tr>
<tr>
<td>resid 30 and name HG2*</td>
<td>resid 30 and name HG2*</td>
<td>0.0 1.5 5.0 lm</td>
</tr>
<tr>
<td>resid 36 and name HG2*</td>
<td>resid 36 and name HG2*</td>
<td>0.0 1.5 5.0 lm</td>
</tr>
<tr>
<td>resid 36 and name HG2*</td>
<td>resid 36 and name HG2</td>
<td>0.0 1.5 5.0 lm</td>
</tr>
<tr>
<td>resid 40 and name HG</td>
<td>resid 40 and name HG2*</td>
<td>0.0 1.0 3.5 ls</td>
</tr>
<tr>
<td>resid 40 and name HG</td>
<td>resid 40 and name HG2</td>
<td>0.0 1.0 3.5 ls</td>
</tr>
<tr>
<td>resid 44 and name HG1*</td>
<td>resid 44 and name HG1*</td>
<td>0.0 1.5 4.0 ls</td>
</tr>
<tr>
<td>resid 45 and name HG</td>
<td>resid 45 and name HG1*</td>
<td>0.0 2.0 7.0 tw</td>
</tr>
<tr>
<td>resid 45 and name HG</td>
<td>resid 45 and name HG1*</td>
<td>0.0 2.0 7.0 tw</td>
</tr>
<tr>
<td>resid 45 and name HG</td>
<td>resid 45 and name HG1*</td>
<td>0.0 2.0 7.0 tw</td>
</tr>
<tr>
<td>resid 49 and name HG</td>
<td>resid 49 and name HG1*</td>
<td>0.0 1.0 4.5 lm</td>
</tr>
<tr>
<td>resid 49 and name HG</td>
<td>resid 49 and name HG1*</td>
<td>0.0 1.0 3.5 ls</td>
</tr>
<tr>
<td>resid 50 and name HG</td>
<td>resid 50 and name HG</td>
<td>0.0 0.0 5.0 lw</td>
</tr>
<tr>
<td>resid 50 and name HG</td>
<td>resid 50 and name HG</td>
<td>0.0 0.0 5.0 lw</td>
</tr>
<tr>
<td>resid 50 and name HG2*</td>
<td>resid 50 and name HG2*</td>
<td>0.0 1.5 6.5 lw</td>
</tr>
<tr>
<td>resid 50 and name HG2*</td>
<td>resid 50 and name HG2*</td>
<td>0.0 1.5 6.5 lw</td>
</tr>
<tr>
<td>resid 61 and name HG</td>
<td>resid 61 and name HG2*</td>
<td>0.0 1.5 5.0 lm</td>
</tr>
<tr>
<td>resid 61 and name HG</td>
<td>resid 61 and name HG2*</td>
<td>0.0 1.5 5.0 lm</td>
</tr>
<tr>
<td>resid 64 and name HG1</td>
<td>resid 64 and name HG1</td>
<td>0.0 0.0 3.5 lm</td>
</tr>
<tr>
<td>resid 64 and name HG2</td>
<td>resid 64 and name HG2</td>
<td>0.0 0.0 3.5 lm</td>
</tr>
<tr>
<td>resid 67 and name HG</td>
<td>resid 67 and name HG</td>
<td>0.0 0.0 3.5 lm</td>
</tr>
</tbody>
</table>

**Assignments:**
- resid: residue number
- name: atom name

#### B2: Sequential nOes

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Distances</th>
<th>Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>resid 1 and name HA</td>
<td>resid 2 and name HG</td>
<td>0.0 0.0 2.5 ls</td>
</tr>
<tr>
<td>resid 1 and name HB*</td>
<td>resid 2 and name HG</td>
<td>0.0 1.0 4.5 lm</td>
</tr>
<tr>
<td>resid 2 and name HA</td>
<td>resid 3 and name HG</td>
<td>0.0 0.0 3.5 ls</td>
</tr>
<tr>
<td>resid 2 and name HG</td>
<td>resid 3 and name HG</td>
<td>0.0 0.0 3.5 ls</td>
</tr>
<tr>
<td>resid 3 and name HA</td>
<td>resid 4 and name HG</td>
<td>0.0 0.0 3.5 lm</td>
</tr>
<tr>
<td>resid 3 and name HB</td>
<td>resid 4 and name HG</td>
<td>0.0 0.0 3.5 lm</td>
</tr>
<tr>
<td>resid 4 and name HA</td>
<td>resid 5 and name HG</td>
<td>0.0 0.0 3.5 ls</td>
</tr>
<tr>
<td>resid 4 and name HB</td>
<td>resid 5 and name HG</td>
<td>0.0 0.0 3.5 ls</td>
</tr>
<tr>
<td>resid 4 and name HB*</td>
<td>resid 5 and name HG</td>
<td>0.0 2.0 7.0 lw</td>
</tr>
<tr>
<td>resid 4 and name HB*</td>
<td>resid 5 and name HG</td>
<td>0.0 2.0 7.0 lw</td>
</tr>
<tr>
<td>resid 5 and name HA</td>
<td>resid 6 and name HG</td>
<td>0.0 0.0 3.5 lm</td>
</tr>
<tr>
<td>resid 5 and name HB</td>
<td>resid 6 and name HG</td>
<td>0.0 0.0 3.5 lm</td>
</tr>
<tr>
<td>resid 5 and name HB</td>
<td>resid 6 and name HG</td>
<td>0.0 1.5 5.0 lm</td>
</tr>
<tr>
<td>resid 5 and name HG</td>
<td>resid 6 and name HG</td>
<td>0.0 1.5 5.0 lm</td>
</tr>
<tr>
<td>resid 6 and name HA</td>
<td>resid 7 and name HG</td>
<td>0.0 0.0 2.5 ls</td>
</tr>
<tr>
<td>resid 6 and name HG</td>
<td>resid 7 and name HG</td>
<td>0.0 1.5 5.0 lm</td>
</tr>
<tr>
<td>resid 7 and name HA</td>
<td>resid 8 and name HG</td>
<td>0.0 0.0 2.5 ls</td>
</tr>
<tr>
<td>resid 7 and name HB</td>
<td>resid 8 and name HG</td>
<td>0.0 0.0 2.5 ls</td>
</tr>
<tr>
<td>resid 8 and name HA</td>
<td>resid 9 and name HG</td>
<td>0.0 0.0 3.5 lm</td>
</tr>
<tr>
<td>resid 8 and name HB</td>
<td>resid 9 and name HG</td>
<td>0.0 0.0 3.5 lm</td>
</tr>
<tr>
<td>resid 8 and name HB</td>
<td>resid 9 and name HG</td>
<td>0.0 1.5 6.5 lw</td>
</tr>
<tr>
<td>resid 8 and name HB</td>
<td>resid 9 and name HG</td>
<td>0.0 1.5 6.5 lw</td>
</tr>
<tr>
<td>resid 8 and name HB1</td>
<td>resid 9 and name HG</td>
<td>0.0 0.0 3.5 lm</td>
</tr>
<tr>
<td>resid 8 and name HB2</td>
<td>resid 9 and name HG</td>
<td>0.0 0.0 3.5 lm</td>
</tr>
<tr>
<td>resid 8 and name HB2</td>
<td>resid 9 and name HG</td>
<td>0.0 0.0 3.5 lm</td>
</tr>
<tr>
<td>resid 9 and name HA</td>
<td>resid 10 and name HG</td>
<td>0.0 0.0 3.5 lm</td>
</tr>
</tbody>
</table>

**Assignments:**
- resid: residue number
- name: atom name
assign (resid 3 and name HD1*) (resid 26 and name HB) 0.0 1.5 6.5 lw
assign (resid 3 and name HD1*) (resid 50 and name HD11) 0.0 3.0 8.0 lw
assign (resid 3 and name HD1*) (resid 50 and name HD12) 0.0 3.0 8.0 lw
assign (resid 3 and name HD1*) (resid 50 and name HD1*) 0.0 3.0 6.5 lm
assign (resid 3 and name HD1*) (resid 61 and name HG1*) 0.0 2.5 8.0 lw
assign (resid 3 and name HD1*) (resid 61 and name HD1*) 0.0 2.5 6.0
assign (resid 3 and name HD1*) (resid 61 and name HD2*) 0.0 3.0 5.5
assign (resid 3 and name HD1*) (resid 61 and name HD1*) 0.0 3.0 6.5
assign (resid 3 and name HD1*) (resid 61 and name HG1*) 0.0 2.5 8.0
assign (resid 3 and name HD1*) (resid 61 and name HD1*) 0.0 2.5 7.5
assign (resid 3 and name HD1*) (resid 61 and name HD1*) 0.0 2.5 6.0
assign (resid 3 and name HA) (resid 65 and name HDN) 0.0 0.0 5.0 lw
assign (resid 3 and name HB) (resid 65 and name HB2) 0.0 0.0 5.0 lw
assign (resid 3 and name HB) (resid 65 and name HB1) 0.0 0.0 5.5 lwv
assign (resid 3 and name HG2*) (resid 65 and name HB1) 0.0 0.0 5.0 lw
assign (resid 3 and name HG2*) (resid 65 and name HB2) 0.0 0.0 5.0 lwv
assign (resid 3 and name HD1*) (resid 67 and name HD12) 0.0 1.5 6.5 lw
assign (resid 3 and name HD1*) (resid 67 and name HD11) 0.0 1.5 7.0 lwv
assign (resid 4 and name HD*) (resid 15 and name HNG1) 0.0 0.0 5.0 lw
assign (resid 4 and name HD*) (resid 12 and name HGI1) 0.0 3.5 6.0 ls
assign (resid 4 and name HE*) (resid 12 and name HGI1) 0.0 3.5 6.0 ls
assign (resid 4 and name HA) (resid 12 and name HGI1) 0.0 1.5 7.0 lwv
assign (resid 4 and name HB1) (resid 12 and name HGI1) 0.0 1.5 5.0 lm
assign (resid 4 and name HB2) (resid 12 and name HGI1) 0.0 1.5 4.0 ls
assign (resid 4 and name HA) (resid 14 and name HGI1) 0.0 1.5 6.5 lwv
assign (resid 4 and name HA) (resid 14 and name HGI1) 0.0 1.5 6.5 lwv
assign (resid 4 and name HB1) (resid 14 and name HA) 0.0 0.0 5.5 lwv
assign (resid 4 and name HB2) (resid 14 and name HA) 0.0 0.0 5.5 lwv
assign (resid 4 and name He*) (resid 14 and name HA) 0.0 2.0 7.0 lw
assign (resid 4 and name HD*) (resid 14 and name HB) 0.0 2.0 7.0 lw
assign (resid 4 and name HD*) (resid 14 and name HGI1) 0.0 3.5 6.0 ls
assign (resid 4 and name HD*) (resid 14 and name HGI1) 0.0 3.5 6.0 ls
assign (resid 4 and name HD*) (resid 14 and name HGI1) 0.0 2.0 5.5 lm
assign (resid 4 and name HD*) (resid 15 and name HN) 0.0 2.0 7.5 lwv
assign (resid 4 and name HD*) (resid 15 and name HN) 0.0 2.0 7.5 lwv
assign (resid 4 and name HD*) (resid 16 and name HN) 0.0 2.0 7.5 lwv
assign (resid 4 and name HD*) (resid 16 and name HN) 0.0 2.0 7.5 lwv
assign (resid 4 and name HD*) (resid 64 and name HB*) 0.0 1.5 6.5 lwv
assign (resid 4 and name HD*) (resid 64 and name HB*) 0.0 3.0 8.0 lw
assign (resid 4 and name HD*) (resid 64 and name HB*) 0.0 2.0 7.0 lw
assign (resid 4 and name HD*) (resid 64 and name HA) 0.0 2.0 7.0 lw
assign (resid 4 and name HD*) (resid 66 and name HGI1) 0.0 1.5 6.5 lwv
assign (resid 4 and name HN) (resid 66 and name HNA) 0.0 0.0 3.5 lm
assign (resid 4 and name HB1) (resid 66 and name HNA) 0.0 0.0 5.0 lw
assign (resid 4 and name HB2) (resid 66 and name HNA) 0.0 1.5 6.5 lw
assign (resid 4 and name HB1) (resid 66 and name HGI2) 0.0 1.5 6.5 lw
assign (resid 4 and name HB2) (resid 66 and name HGI2) 0.0 1.5 6.5 lw
assign (resid 5 and name HGI1) (resid 12 and name HNA) 0.0 1.5 6.5 lw
assign (resid 5 and name HGI2) (resid 12 and name HNA) 0.0 1.5 6.5 lw
assign (resid 5 and name HGI1) (resid 13 and name HGI1) 0.0 1.5 6.5 lw
assign (resid 5 and name HGI2) (resid 13 and name HGI1) 0.0 1.5 6.5 lw
assign (resid 5 and name HGI1) (resid 13 and name HGI2) 0.0 1.5 6.5 lw
assign (resid 5 and name HGI2) (resid 13 and name HGI2) 0.0 1.5 6.5 lw
assign (resid 5 and name HGI1) (resid 13 and name HNA) 0.0 1.5 7.0 lwv
assign (resid 5 and name HGI2) (resid 13 and name HNA) 0.0 1.5 7.0 lwv
assign (resid 5 and name HGI1) (resid 14 and name HNA) 0.0 0.0 5.0 lw
assign (resid 5 and name HGI2) (resid 30 and name HNA) 0.0 1.5 7.0 lwv
assign (resid 5 and name HGI1) (resid 30 and name HGI1) 0.0 3.0 6.5 lm
assign (resid 5 and name HGI2) (resid 30 and name HGI1) 0.0 3.0 6.5 lm
assign (resid 5 and name HGI1) (resid 66 and name HGI2) 0.0 1.5 7.0 lwv
assign (resid 5 and name HGI2) (resid 66 and name HGI2) 0.0 1.5 7.0 lwv
assign (resid 5 and name HGI1) (resid 67 and name HNA) 0.0 0.0 5.0 lw
assign (resid 5 and name HGI2) (resid 67 and name HNA) 0.0 2.5 8.0 lwv
assign (resid 5 and name HGI1) (resid 68 and name HDP) 0.0 2.5 7.5 lw
assign (resid 5 and name HGI2) (resid 68 and name HDP) 0.0 1.5 7.0 lwv
assign (resid 5 and name HGI1) (resid 68 and name HNA) 0.0 1.5 7.0 lwv
assign (resid 5 and name HGI2) (resid 68 and name HNA) 0.0 1.5 7.0 lwv
assign (resid 5 and name HNA) (resid 68 and name HGI1) 0.0 0.0 5.5 lw
assign (resid 5 and name HNA) (resid 68 and name HGI2) 0.0 0.0 5.5 lw
assign (resid 5 and name HGI1) (resid 69 and name HNA) 0.0 1.5 6.5 lw
assign (resid 5 and name HGI2) (resid 69 and name HNA) 0.0 1.5 6.5 lw
assign (resid 6 and name HNA) (resid 12 and name HNA) 0.0 1.5 6.5 lw
assign (resid 6 and name HNA) (resid 12 and name HGI1) 0.0 1.5 6.5 lw
assign (resid 6 and name HNA) (resid 13 and name HNA) 0.0 0.0 5.0 lw
assign (resid 6 and name HNA) (resid 67 and name HNA) 0.0 0.0 5.0 lw
assign (resid 6 and name HNA) (resid 68 and name HNA) 0.0 0.0 3.5 lm
assign (resid 6 and name HNB) (resid 68 and name HNA) 0.0 1.0 6.0 lw
assign (resid 6 and name HGI1) (resid 68 and name HNA) 0.0 1.5 6.5 lw
assign (resid 6 and name HGI2) (resid 68 and name HNA) 0.0 1.5 6.5 lw
assign (resid 6 and name HGI1) (resid 68 and name HNA) 0.0 1.0 6.5 lwv
assign (resid 7 and name HG*) (resid 11 and name HN ) (resid 12 and name HA ) (resid 69 and name HD1*) 0.0 0.0 3.5 5m
assign (resid 7 and name HB) (resid 69 and name HD1*) 0.0 1.5 7.0 1vw
assign (resid 7 and name HB) (resid 69 and name HD2*) 0.0 0.0 3.5 5m
assign (resid 7 and name HB) (resid 69 and name HB*) 0.0 0.0 5.0 5w
assign (resid 7 and name HB) (resid 69 and name HB*) 0.0 1.0 6.5 1vw
assign (resid 7 and name HB) (resid 69 and name HB*) 0.0 1.0 6.5 1vw
assign (resid 7 and name HB) (resid 69 and name HD1*) 0.0 1.5 7.0 1vw
assign (resid 7 and name HB) (resid 71 and name HN ) 0.0 1.5 7.0 1vw
assign (resid 8 and name HD*) (resid 68 and name HB1) 0.0 3.0 8.0 1w
assign (resid 8 and name HD*) (resid 68 and name HB2) 0.0 3.0 8.0 1w
assign (resid 8 and name HD1*) (resid 69 and name HD2*) 0.0 1.5 5.0 1m
assign (resid 8 and name HD1*) (resid 70 and name HN ) 0.0 1.5 5.0 1m
assign (resid 8 and name HD1*) (resid 71 and name HN ) 0.0 1.5 6.5 1w
assign (resid 8 and name HD2*) (resid 70 and name HN ) 0.0 1.5 5.0 1m
assign (resid 8 and name HD1*) (resid 71 and name HN ) 0.0 1.5 5.0 1m
assign (resid 8 and name HD2*) (resid 71 and name HN ) 0.0 1.0 6.5 1vw
assign (resid 8 and name HD1*) (resid 73 and name HN ) 0.0 1.5 6.5 1vw
assign (resid 8 and name HG) (resid 73 and name HN ) 0.0 0.0 5.5 1m
assign (resid 11 and name HG*) (resid 34 and name HG2) 0.0 1.0 6.5 1vw
assign (resid 11 and name HG*) (resid 34 and name HG1) 0.0 1.0 6.5 1vw
assign (resid 11 and name HG1) (resid 34 and name HN ) 0.0 0.0 5.5 1m
assign (resid 11 and name HG2) (resid 34 and name HN ) 0.0 0.0 5.5 1m
assign (resid 11 and name HG1) (resid 41 and name HG*) 0.0 1.0 6.5 1vw
assign (resid 11 and name HG2) (resid 41 and name HG*) 0.0 0.0 5.5 1m
assign (resid 17 and name HG1*) (resid 23 and name HB) 0.0 1.5 6.5 1vw
assign (resid 17 and name HG2*) (resid 23 and name HB) 0.0 0.0 5.5 1m
assign (resid 17 and name HG2*) (resid 23 and name HN ) 0.0 0.0 5.5 1m
assign (resid 17 and name HG2*) (resid 23 and name HG12) 0.0 3.0 5.5 1s
assign (resid 17 and name HG*) (resid 23 and name HN ) 0.0 3.0 8.0 1w
assign (resid 17 and name HG2*) (resid 26 and name HG1*) 0.0 3.0 6.0 1vw
assign (resid 17 and name HG2*) (resid 26 and name HG2*) 0.0 3.0 6.5 1vw
assign (resid 17 and name HA) (resid 26 and name HG1*) 0.0 1.5 7.0 1vw
assign (resid 17 and name HB) (resid 26 and name HG2*) 0.0 1.5 7.0 1vw
assign (resid 17 and name HB) (resid 56 and name HD1*) 0.0 1.5 7.0 1vw
assign (resid 17 and name HB) (resid 56 and name HD2*) 0.0 0.3 0.8 5w
assign (resid 17 and name HG*) (resid 56 and name HB1) 0.0 0.3 0.8 5w
assign (resid 17 and name HG*) (resid 56 and name HB2) 0.0 0.3 0.8 5w
assign (resid 17 and name HG1*) (resid 56 and name HD1*) 0.0 1.5 4.0 1s
assign (resid 17 and name HG2*) (resid 56 and name HD1*) 0.0 1.5 4.0 1s
assign (resid 17 and name HG*) (resid 61 and name HG1*) 0.0 4.0 9.0 1tw
assign (resid 17 and name HG*) (resid 63 and name HN ) 0.0 3.0 8.0 1w
assign (resid 17 and name HG*) (resid 63 and name HB*) 0.0 4.0 9.0 1tw
assign (resid 18 and name HN) (resid 56 and name HD1*) 0.0 1.5 6.5 1w
assign (resid 18 and name HN) (resid 56 and name HD2*) 0.0 1.5 6.5 1w
assign (resid 19 and name HD1) (resid 56 and name HD1*) 0.0 1.0 6.0 1w
assign (resid 19 and name HD2) (resid 56 and name HD1*) 0.0 1.0 6.0 1w
assign (resid 19 and name HB2) (resid 57 and name HB1) 0.0 0.0 5.0 1w
assign (resid 19 and name HB2) (resid 57 and name HB2) 0.0 0.0 5.0 1w
assign (resid 19 and name HN) (resid 56 and name HD1*) 0.0 1.0 6.5 1vw
assign (resid 19 and name HB1) (resid 57 and name HB1) 0.0 0.0 5.0 1w
assign (resid 19 and name HB1) (resid 57 and name HB2) 0.0 0.0 5.0 1w
assign (resid 19 and name HB1) (resid 57 and name HB2) 0.0 0.0 5.0 1w
assign (resid 19 and name HB1) (resid 57 and name HB3) 0.0 0.0 5.0 1w
assign (resid 19 and name HB2) (resid 57 and name HB2) 0.0 0.0 5.0 1w
assign (resid 19 and name HB2) (resid 57 and name HB3) 0.0 0.0 5.0 1w
assign (resid 19 and name HB2) (resid 57 and name HB3) 0.0 0.0 5.0 1w
assign (resid 19 and name HB2) (resid 57 and name HB3) 0.0 0.0 5.0 1w
assign (resid 20 and name HN) (resid 55 and name HG1*) 0.0 1.5 4.0 1s
assign (resid 20 and name HB1) (resid 55 and name HG1*) 0.0 1.5 6.5 1w
assign (resid 20 and name HB2) (resid 55 and name HG1*) 0.0 1.5 7.0 1vw
assign (resid 21 and name HA) (resid 25 and name HB2) 0.0 0.0 5.3 1w
assign (resid 21 and name HB) (resid 25 and name HB2) 0.0 0.0 5.3 1w
assign (resid 21 and name HB) (resid 25 and name HB3) 0.0 0.0 5.3 1w
assign (resid 21 and name HB*) (resid 25 and name HB2) 0.0 2.0 6.0 1w
assign (resid 21 and name HB) (resid 25 and name HD22) 0.0 0.0 3.5 1m
assign (resid 21 and name HB) (resid 25 and name HD21) 0.0 0.0 5.0 1w
assign (resid 21 and name HB*) (resid 26 and name HG2*) 0.0 1.5 6.5 1w
assign (resid 21 and name HA) (resid 55 and name HG2*) 0.0 1.5 6.5 1w
assign (resid 22 and name HG1*) (resid 54 and name HA ) 0.0 1.5 6.5 1w
assign (resid 22 and name HG1*) (resid 55 and name HG2*) 0.0 1.5 6.5 1w
assign (resid 22 and name HG1*) (resid 56 and name HA ) 0.0 1.5 6.5 1w
assign (resid 22 and name HG1*) (resid 59 and name HD1*) 0.0 3.5 9.0 1vw
assign (resid 22 and name HG1*) (resid 59 and name HD1*) 0.0 3.5 9.0 1vw
assign (resid 22 and name HG2*) (resid 52 and name HA ) 0.0 1.5 5.0 1m
assign (resid 22 and name HG2*) (resid 52 and name HD1) 0.0 1.5 5.0 1m
assign (resid 23 and name HD2*) (resid 54 and name HB*) 0.0 2.5 8.0 1w
assign (resid 23 and name HD2) (resid 54 and name HD1) 0.0 1.5 6.5 1w
<table>
<thead>
<tr>
<th>Residue 1</th>
<th>Name 1</th>
<th>Residue 2</th>
<th>Name 2</th>
<th>dE (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>HA</td>
<td>4</td>
<td>HA</td>
<td>0.20</td>
</tr>
<tr>
<td>3</td>
<td>HD*</td>
<td>4</td>
<td>HB2</td>
<td>0.50</td>
</tr>
<tr>
<td>5</td>
<td>IN</td>
<td>5</td>
<td>IN</td>
<td>0.50</td>
</tr>
<tr>
<td>6</td>
<td>UN</td>
<td>7</td>
<td>UN</td>
<td>6.50</td>
</tr>
<tr>
<td>8</td>
<td>HG2</td>
<td>9</td>
<td>HG2</td>
<td>6.50</td>
</tr>
<tr>
<td>10</td>
<td>UN</td>
<td>11</td>
<td>UN</td>
<td>6.50</td>
</tr>
<tr>
<td>12</td>
<td>UN</td>
<td>13</td>
<td>UN</td>
<td>7.00</td>
</tr>
<tr>
<td>14</td>
<td>UN</td>
<td>15</td>
<td>UN</td>
<td>6.50</td>
</tr>
<tr>
<td>16</td>
<td>UN</td>
<td>17</td>
<td>UN</td>
<td>6.50</td>
</tr>
<tr>
<td>18</td>
<td>UN</td>
<td>19</td>
<td>UN</td>
<td>6.50</td>
</tr>
<tr>
<td>20</td>
<td>UN</td>
<td>21</td>
<td>UN</td>
<td>6.50</td>
</tr>
<tr>
<td>22</td>
<td>UN</td>
<td>23</td>
<td>UN</td>
<td>6.50</td>
</tr>
<tr>
<td>24</td>
<td>UN</td>
<td>25</td>
<td>UN</td>
<td>6.50</td>
</tr>
<tr>
<td>26</td>
<td>UN</td>
<td>27</td>
<td>UN</td>
<td>6.50</td>
</tr>
<tr>
<td>28</td>
<td>UN</td>
<td>29</td>
<td>UN</td>
<td>6.50</td>
</tr>
<tr>
<td>30</td>
<td>UN</td>
<td>31</td>
<td>UN</td>
<td>6.50</td>
</tr>
<tr>
<td>32</td>
<td>UN</td>
<td>33</td>
<td>UN</td>
<td>6.50</td>
</tr>
</tbody>
</table>
assign (resid 16 and name HA) (resid 17 and name HN2) 0.0 1.5 6.5 lw
assign (resid 17 and name HA) (resid 18 and name HN) 0.0 0.00 2.5 ls
assign (resid 17 and name HB) (resid 18 and name HN) 0.0 1.5 5.0 lm
assign (resid 17 and name HGL2) (resid 18 and name HN) 0.0 1.5 6.5 lw
assign (resid 18 and name HA) (resid 19 and name HD1) 0.0 0.00 3.5 ls
assign (resid 18 and name HA) (resid 19 and name HD2) 0.0 0.00 3.5 ls
assign (resid 18 and name HB) (resid 19 and name HG) 0.0 1.0 6.0 ls
assign (resid 18 and name HB) (resid 19 and name HN) 0.0 1.0 6.5 lw
assign (resid 19 and name HB) (resid 20 and name HN) 0.0 0.00 5.0 lw
assign (resid 19 and name HB) (resid 20 and name HN) 0.0 0.00 5.0 lm
assign (resid 19 and name HD1) (resid 20 and name HN) 0.0 0.00 2.5 ls
assign (resid 19 and name HD2) (resid 20 and name HN) 0.0 0.00 5.0 lw
assign (resid 19 and name HGL2) (resid 20 and name HN) 0.0 0.00 5.5 lw
assign (resid 19 and name HGL2) (resid 20 and name HB) 0.0 0.00 5.5 lw
assign (resid 19 and name HB) (resid 21 and name HB1) 0.0 0.00 5.5 lw
assign (resid 19 and name HB) (resid 21 and name HB2) 0.0 0.00 5.5 lw
assign (resid 20 and name HA) (resid 21 and name HN) 0.0 0.00 3.5 lm
assign (resid 20 and name HB) (resid 21 and name HN) 0.0 0.00 2.5 ls
assign (resid 20 and name HB) (resid 21 and name HN) 0.0 1.0 6.5 lw
assign (resid 21 and name HA) (resid 22 and name HN) 0.0 0.00 2.5 ls
assign (resid 21 and name HB1) (resid 22 and name HN) 0.0 0.00 3.5 lm
assign (resid 21 and name HB2) (resid 22 and name HN) 0.0 0.00 3.5 lm
assign (resid 21 and name HA) (resid 22 and name HG) 0.0 1.5 5.0 lm
assign (resid 22 and name HA) (resid 23 and name HN) 0.0 0.00 2.5 ls
assign (resid 22 and name HB) (resid 23 and name HN) 0.0 0.00 3.5 lm
assign (resid 22 and name HGL2) (resid 23 and name HN) 0.0 1.5 5.0 lm
assign (resid 22 and name HGL2) (resid 23 and name HN) 0.0 0.00 5.5 lw
assign (resid 22 and name HN) (resid 24 and name HN) 0.0 0.00 3.5 lw
assign (resid 24 and name HA) (resid 25 and name HN) 0.0 0.00 5.5 lw
assign (resid 24 and name HG) (resid 25 and name HN) 0.0 1.0 6.5 lw
assign (resid 25 and name HA) (resid 26 and name HN) 0.0 0.00 3.5 ls
assign (resid 25 and name HA) (resid 26 and name HN) 0.0 0.00 2.5 ls
assign (resid 25 and name HB1) (resid 26 and name HN) 0.0 0.00 5.0 lw
assign (resid 25 and name HB2) (resid 26 and name HN) 0.0 0.00 5.0 lw
assign (resid 25 and name HN) (resid 26 and name HN) 0.0 0.00 5.0 lw
assign (resid 26 and name HN2) (resid 27 and name HN) 0.0 0.00 3.5 ls
assign (resid 26 and name HA) (resid 27 and name HN) 0.0 0.00 5.0 lw
assign (resid 26 and name HB1) (resid 27 and name HN) 0.0 0.00 5.0 lw
assign (resid 27 and name HA) (resid 28 and name HN) 0.0 0.00 5.0 lw
assign (resid 27 and name HB) (resid 28 and name HN) 0.0 1.0 6.0 lw
assign (resid 28 and name HA) (resid 29 and name HN) 0.0 0.00 2.5 ls
assign (resid 28 and name HA) (resid 30 and name HN) 0.0 0.00 5.0 lw
assign (resid 29 and name HA) (resid 30 and name HN) 0.0 0.00 2.5 ls
assign (resid 29 and name HA) (resid 30 and name HN) 0.0 0.00 5.5 lw
assign (resid 29 and name HB1) (resid 30 and name HN) 0.0 0.00 5.5 lw
assign (resid 30 and name HA) (resid 31 and name HN) 0.0 0.00 2.5 ls
assign (resid 30 and name HA) (resid 31 and name HN) 0.0 0.00 3.5 lm
assign (resid 31 and name HB1) (resid 32 and name HN) 0.0 0.00 3.5 lm
assign (resid 31 and name HB2) (resid 32 and name HN) 0.0 0.00 3.5 lm
assign (resid 31 and name HB) (resid 32 and name HN) 0.0 0.00 2.5 ls
assign (resid 31 and name HN) (resid 32 and name HN) 0.0 0.00 3.5 lm
assign (resid 31 and name HB) (resid 32 and name HN) 0.0 0.00 5.0 lw
assign (resid 32 and name HB) (resid 33 and name HN) 0.0 0.00 3.5 lm
assign (resid 32 and name HA) (resid 33 and name HN) 0.0 0.00 3.5 lm
assign (resid 32 and name HB1) (resid 33 and name HN) 0.0 0.00 3.5 lm
assign (resid 33 and name HB) (resid 34 and name HN) 0.0 0.00 3.5 lm
assign (resid 33 and name HN) (resid 34 and name HN) 0.0 0.00 2.5 ls/m
assign (resid 33 and name HB1) (resid 34 and name HN) 0.0 0.00 3.5 lm
assign (resid 33 and name HB2) (resid 34 and name HN) 0.0 0.00 3.5 lm
assign (resid 33 and name HB) (resid 34 and name HN) 0.0 1.0 6.5 lw
assign (resid 34 and name HA) (resid 35 and name HN) 0.0 0.00 5.2 lw
assign (resid 34 and name HA) (resid 35 and name HN) 0.0 0.00 2.5 ls
assign (resid 35 and name HN) (resid 36 and name HN) 0.0 0.00 5.0 lw
assign (resid 35 and name HN) (resid 36 and name HN) 0.0 0.00 2.5 ls
assign (resid 36 and name HA) (resid 37 and name HN) 0.0 0.00 3.5 lm
assign (resid 36 and name HA) (resid 37 and name HD) 0.0 0.00 3.5 lm/s
assign (resid 36 and name HG*2) (resid 37 and name HD2) 0.1 1.5 6.5 lw
assign (resid 36 and name HG*2) (resid 37 and name HD1) 0.1 1.5 6.5 lw
assign (resid 36 and name HD*) (resid 37 and name HD) 0.1 1.5 7.0 lw
assign (resid 36 and name HN) (resid 37 and name HD) 0.1 1.5 7.0 lw
assign (resid 36 and name HA) (resid 39 and name HN) 0.0 0.00 5.0 lw
assign (resid 38 and name HA) (resid 39 and name HN) 0.0 0.00 3.5 lm
assign (resid 39 and name HA) (resid 40 and name HN) 0.0 0.00 5.0 lw
assign (resid 39 and name HA) (resid 40 and name HN) 0.0 0.00 2.5 ls
assign (resid 39 and name HN) (resid 40 and name HN) 0.0 0.00 5.0 lw
assign (resid 40 and name HA) (resid 41 and name HN) 0.0 0.00 2.5 ls
assign (resid 40 and name HN) (resid 41 and name HN) 0.0 1.0 6.0 lw
assign (resid 21 and name HA) (resid 26 and name HN) 0.0 0.0 5.5 !tw
assign (resid 21 and name HB2) (resid 26 and name HG1*) 0.0 1.5 6.5 !tw
assign (resid 21 and name HB1) (resid 26 and name HG1*) 0.0 1.5 7.0 !tw
assign (resid 21 and name HN) (resid 55 and name HDG2*) 0.0 1.5 6.5 !tw
assign (resid 21 and name HA) (resid 55 and name HDG2*) 0.0 1.5 6.5 !tw
assign (resid 21 and name HA) (resid 56 and name HDI*) 0.0 1.5 6.5 !tw
assign (resid 22 and name HA) (resid 26 and name HG1*) 0.0 1.5 6.5 !tw
assign (resid 22 and name HGI*) (resid 54 and name HN) 0.0 1.5 7.0 !tw
assign (resid 22 and name HGI*) (resid 54 and name HA) 0.0 1.5 7.0 !tw
assign (resid 22 and name HGI*) (resid 54 and name HN) 0.0 1.5 7.0 !tw
assign (resid 23 and name HGI1) (resid 50 and name HG2*) 0.0 1.5 7.0 !tw
assign (resid 23 and name HG12) (resid 50 and name HG2*) 0.0 1.5 7.0 !tw
assign (resid 23 and name HG2*) (resid 52 and name HA) 0.0 1.5 5.0 !m
assign (resid 23 and name HB) (resid 52 and name HN) 0.0 0.0 5.0 !w
assign (resid 23 and name HG2*) (resid 54 and name HN) 0.0 0.0 5.5 !w
assign (resid 23 and name HG12) (resid 54 and name HN) 0.0 0.0 5.5 !w
assign (resid 23 and name HG12) (resid 54 and name HB1) 0.0 0.0 5.5 !w
assign (resid 23 and name HG2*) (resid 54 and name HB2) 0.0 0.0 5.5 !w
assign (resid 23 and name HG1) (resid 54 and name HB1) 0.0 0.0 5.5 !w
assign (resid 23 and name HDI*) (resid 54 and name HB2) 0.0 0.0 5.5 !w
assign (resid 23 and name HDI*) (resid 54 and name HB1) 0.0 0.0 5.5 !w
assign (resid 23 and name HGI2) (resid 59 and name HE*) 0.0 0.0 5.5 !w
assign (resid 23 and name HDI*) (resid 59 and name HD*) 0.0 0.0 5.5 !w
assign (resid 23 and name HB) (resid 59 and name HE*) 0.0 0.0 5.5 !w
assign (resid 23 and name HGI2) (resid 59 and name HD*) 0.0 0.0 5.5 !w
assign (resid 23 and name HGI2) (resid 59 and name HE*) 0.0 0.0 5.5 !w
assign (resid 23 and name HGI1) (resid 59 and name HE*) 0.0 0.0 5.5 !w
assign (resid 23 and name HGI1) (resid 59 and name HD*) 0.0 0.0 5.5 !w
assign (resid 23 and name HN) (resid 69 and name HDI*) 0.0 0.0 5.5 !w
assign (resid 23 and name HN) (resid 69 and name HD2*) 0.0 0.0 5.5 !w
assign (resid 24 and name HA) (resid 28 and name HB*) 0.0 0.0 5.5 !w
assign (resid 25 and name HA) (resid 29 and name HN) 0.0 0.0 5.5 !w
assign (resid 26 and name HA) (resid 30 and name HN) 0.0 0.0 5.5 !w
assign (resid 26 and name HA) (resid 34 and name HN) 0.0 0.0 5.5 !w
assign (resid 26 and name HN) (resid 55 and name HDG2*) 0.0 1.5 6.5 !w
assign (resid 27 and name HA) (resid 31 and name HDI*) 0.0 1.5 7.0 !w
assign (resid 27 and name HA) (resid 31 and name HB1) 0.0 1.5 7.0 !w
assign (resid 27 and name HA) (resid 31 and name HB2) 0.0 1.5 7.0 !w
assign (resid 27 and name HA) (resid 43 and name HDI*) 0.0 1.5 7.0 !w
assign (resid 27 and name HA) (resid 43 and name HD2*) 0.0 1.5 7.0 !w
assign (resid 28 and name HA) (resid 32 and name HN) 0.0 0.0 5.0 !w
assign (resid 28 and name HA) (resid 38 and name HBI) 0.0 0.0 5.5 !w
assign (resid 28 and name HA) (resid 38 and name HB2) 0.0 0.0 5.5 !w
assign (resid 28 and name HA) (resid 38 and name HB) 0.0 1.5 6.5 !w/m
assign (resid 30 and name HA) (resid 34 and name HN) 0.0 0.0 5.0 !w
assign (resid 30 and name HA) (resid 34 and name HB1) 0.0 0.0 5.5 !w
assign (resid 30 and name HA) (resid 34 and name HB2) 0.0 0.0 5.5 !w
assign (resid 30 and name HB) (resid 41 and name HN) 0.0 1.0 5.5 !w
assign (resid 30 and name HB) (resid 43 and name HDI*) 0.0 1.0 7.0 !w
assign (resid 31 and name HA) (resid 35 and name HN) 0.0 0.0 5.0 !w
assign (resid 31 and name HB) (resid 41 and name HG*) 0.0 0.0 3.5 !s
assign (resid 31 and name HB) (resid 41 and name HG*) 0.0 2.0 7.0 !w
assign (resid 36 and name HB) (resid 41 and name HDI*) 0.0 1.0 6.5 !w/m
assign (resid 36 and name HA) (resid 69 and name HDI*) 0.0 1.5 6.5 !w
assign (resid 41 and name HB) (resid 69 and name HDI*) 0.0 1.5 7.0 !w
assign (resid 41 and name HB) (resid 69 and name HD2*) 0.0 1.5 7.0 !w
assign (resid 41 and name HA) (resid 69 and name HDI*) 0.0 0.0 5.0 !w
assign (resid 41 and name HA) (resid 71 and name HA) 0.0 0.0 5.0 !w
assign (resid 41 and name HA) (resid 71 and name HDI) 0.0 0.0 5.0 !w
assign (resid 42 and name HA) (resid 70 and name HN) 0.0 0.0 5.0 !w
assign (resid 42 and name HA) (resid 70 and name HN) 0.0 0.0 5.0 !w
assign (resid 43 and name HA) (resid 50 and name HDI*) 0.0 1.5 6.5 !w
assign (resid 43 and name HA) (resid 50 and name HD2*) 0.0 1.5 6.5 !w
assign (resid 43 and name HB1) (resid 50 and name HDI*) 0.0 1.5 4.0 !s
assign (resid 43 and name HB1) (resid 50 and name HD2*) 0.0 1.5 4.0 !s
assign (resid 43 and name HB2) (resid 50 and name HDI*) 0.0 1.5 7.0 !w
assign (resid 43 and name HB2) (resid 50 and name HD2*) 0.0 1.5 7.0 !w
assign (resid 43 and name HB) (resid 48 and name HN) 0.0 0.0 5.5 !w
assign (resid 44 and name HB) (resid 48 and name HN) 0.0 1.5 4.0 !s
assign (resid 54 and name HG2*) (resid 59 and name HE*) 0.0 2.0 7.5 \text{ lvw}
assign (resid 55 and name HA) (resid 61 and name HB) 0.0 0.0 5.5 \text{ lvw}
assign (resid 56 and name HD2*) (resid 61 and name HN) 0.0 1.5 7.0 \text{ lvw}
assign (resid 56 and name HD1*) (resid 61 and name HN) 0.0 1.5 7.0 \text{ lvw}
assign (resid 56 and name HD2*) (resid 61 and name HB) 0.0 1.5 5.0 \text{ lm}
assign (resid 56 and name HD2*) (resid 61 and name HG2*) 0.0 3.0 6.5 \text{ lm/s}
assign (resid 56 and name HD1*) (resid 61 and name HG2*) 0.0 3.0 6.5 \text{ lm/s}
assign (resid 56 and name HN) (resid 61 and name HD1*) 0.0 1.5 6.5 \text{ lw}
assign (resid 57 and name HA) (resid 61 and name HN) 0.0 0.0 5.0 \text{ lw}
assign (resid 61 and name HD1*) (resid 65 and name HB2) 0.0 1.5 5.0 \text{ lm}
assign (resid 61 and name HD1*) (resid 65 and name HB1) 0.0 1.5 4.0 \text{ ls}
assign (resid 61 and name HG2*) (resid 65 and name HN) 0.0 1.5 6.5 \text{ lw}
assign (resid 61 and name HG2*) (resid 67 and name HB*) 0.0 2.5 6.0 \text{ lm}
assign (resid 61 and name HG2*) (resid 67 and name HG) 0.0 1.5 7.0 \text{ lvw}
assign (resid 61 and name HD1*) (resid 67 and name HB*) 0.0 2.5 5.0 \text{ ls}
<table>
<thead>
<tr>
<th>Residue</th>
<th>HN</th>
<th>Hα</th>
<th>Hβ</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Met 1</td>
<td>4.080</td>
<td>1.933</td>
<td></td>
<td>CyH 2.095, NeH 7.394, 6.762</td>
</tr>
<tr>
<td>Gln 2</td>
<td>8.849</td>
<td>5.122</td>
<td>1.716</td>
<td>Cy2H 0.519, Cy1H 0.910, 0.682, C6H 0.445</td>
</tr>
<tr>
<td>Ile 3</td>
<td>8.203</td>
<td>4.032</td>
<td>1.577</td>
<td>CeH 6.936, C6H 7.104</td>
</tr>
<tr>
<td>Phe 4</td>
<td>8.510</td>
<td>5.522</td>
<td></td>
<td>CyH 0.586, 0.539</td>
</tr>
<tr>
<td>Val 5</td>
<td>9.202</td>
<td>4.590</td>
<td>1.765</td>
<td>CyH 1.422, 1.130, C6H 1.337, 2.770</td>
</tr>
<tr>
<td>Lys 6</td>
<td>8.677</td>
<td>5.217</td>
<td>1.523</td>
<td>CyH 1.038</td>
</tr>
<tr>
<td>Thr 7</td>
<td>8.648</td>
<td>4.773</td>
<td>4.691</td>
<td>CyH 1.740, C6H 0.911, 0.847</td>
</tr>
<tr>
<td>Leu 8</td>
<td>9.066</td>
<td>4.154</td>
<td>1.804</td>
<td>CyH 1.126</td>
</tr>
<tr>
<td>Thr 9</td>
<td>7.541</td>
<td>4.290</td>
<td>4.448</td>
<td></td>
</tr>
<tr>
<td>Gln10</td>
<td>7.711</td>
<td>3.471</td>
<td>1.207</td>
<td>CyH 1.282, 1.106, C6H 1.489, 2.786</td>
</tr>
<tr>
<td>Lys11</td>
<td>7.182</td>
<td>4.200</td>
<td>1.678</td>
<td>CyH 0.943</td>
</tr>
<tr>
<td>Thr12</td>
<td>8.578</td>
<td>4.917</td>
<td>3.810</td>
<td>Cy2H 0.737, Cy1H 1.295, 0.977, C6H 0.580</td>
</tr>
<tr>
<td>Ile13</td>
<td>9.370</td>
<td>4.395</td>
<td>1.738</td>
<td>CyH 0.988</td>
</tr>
<tr>
<td>Thr14</td>
<td>8.644</td>
<td>4.829</td>
<td>3.903</td>
<td>CyH 1.063, C6H 0.734, 0.738</td>
</tr>
<tr>
<td>Leu15</td>
<td>8.635</td>
<td>4.623</td>
<td>1.250</td>
<td>CyH 2.137, 1.979</td>
</tr>
<tr>
<td>Glu16</td>
<td>8.029</td>
<td>4.766</td>
<td>1.702</td>
<td>CyH 0.542, 0.289</td>
</tr>
<tr>
<td>Val17</td>
<td>8.813</td>
<td>4.571</td>
<td>2.186</td>
<td>CyH 2.032</td>
</tr>
<tr>
<td>Glu18</td>
<td>8.524</td>
<td>4.946</td>
<td>1.466</td>
<td>CyH 2.074, 1.926, C6H 3.870, 3.672</td>
</tr>
<tr>
<td>Pro19</td>
<td>3.996</td>
<td>2.304</td>
<td>1.887</td>
<td></td>
</tr>
<tr>
<td>Ser20</td>
<td>6.937</td>
<td>4.242</td>
<td>4.022</td>
<td>CyH 1.128</td>
</tr>
<tr>
<td>Asp21</td>
<td>7.954</td>
<td>4.554</td>
<td>2.823</td>
<td>Cy2H 0.649, Cy1H 1.985, 1.515</td>
</tr>
<tr>
<td>Thr22</td>
<td>7.819</td>
<td>4.762</td>
<td>4.679</td>
<td>CyH 2.318, 2.216</td>
</tr>
<tr>
<td>Ile23</td>
<td>8.455</td>
<td>3.521</td>
<td>2.379</td>
<td>C6H 6.797, 7.722</td>
</tr>
<tr>
<td>Glu24</td>
<td>9.793</td>
<td>3.771</td>
<td>1.898</td>
<td>CyH 0.860, 0.588</td>
</tr>
<tr>
<td>Asn25</td>
<td>7.838</td>
<td>4.401</td>
<td>3.080</td>
<td>CyH 1.462, Cy1H 1.655, C6H 3.061</td>
</tr>
<tr>
<td>Val26</td>
<td>7.987</td>
<td>3.254</td>
<td>2.226</td>
<td>Cy2H 0.560, Cy1H 1.878, 0.532, C6H 0.75</td>
</tr>
<tr>
<td>Lys27</td>
<td>8.473</td>
<td>4.442</td>
<td>1.884</td>
<td>CyH 2.143, NeH 6.396, 6.090</td>
</tr>
<tr>
<td>Ala28</td>
<td>7.953</td>
<td>4.027</td>
<td>1.496</td>
<td></td>
</tr>
<tr>
<td>Lys29</td>
<td>7.744</td>
<td>4.069</td>
<td>2.041</td>
<td>CyH 1.977, 1.884, C6H 1.552, 1.465, C6H 3.067</td>
</tr>
<tr>
<td>Ile30</td>
<td>8.201</td>
<td>3.363</td>
<td>2.231</td>
<td></td>
</tr>
<tr>
<td>Gln31</td>
<td>8.493</td>
<td>3.680</td>
<td>2.368</td>
<td></td>
</tr>
<tr>
<td>Asp32</td>
<td>7.920</td>
<td>4.214</td>
<td>2.737</td>
<td></td>
</tr>
<tr>
<td>Lys33</td>
<td>7.307</td>
<td>4.199</td>
<td>1.708</td>
<td></td>
</tr>
<tr>
<td>Residue</td>
<td>HN</td>
<td>$\alpha$</td>
<td>$\beta$</td>
<td>Other</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
<td>----------</td>
<td>---------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Thr66</td>
<td>8.67</td>
<td>5.15</td>
<td>3.94</td>
<td>CyH 0.789</td>
</tr>
<tr>
<td>F1eu67</td>
<td>9.34</td>
<td>4.99</td>
<td>1.40</td>
<td>CyH 1.751, C6H 3.869, 3.784, C6H 0.588</td>
</tr>
<tr>
<td>His68</td>
<td>9.14</td>
<td>5.04</td>
<td>3.06</td>
<td>CyH 7.051, C6H 8.332</td>
</tr>
<tr>
<td>Leu69</td>
<td>8.20</td>
<td>5.06</td>
<td>1.47</td>
<td>CyH 1.193, C6H 0.726, 0.609</td>
</tr>
<tr>
<td>Val70</td>
<td>9.11</td>
<td>4.24</td>
<td>1.89</td>
<td>CyH 0.789, 0.696</td>
</tr>
<tr>
<td>Leu71</td>
<td>8.06</td>
<td>4.89</td>
<td>1.58</td>
<td>CyH 1.630, C6H 0.837, 0.727</td>
</tr>
<tr>
<td>Arg72</td>
<td>8.51</td>
<td>4.12</td>
<td>1.61</td>
<td>CyH 1.515, C6H 3.014, NeH 6.996</td>
</tr>
<tr>
<td>Leu73</td>
<td>8.27</td>
<td>4.26</td>
<td>1.50</td>
<td>CyH 1.422, C6H 0.793, 0.737</td>
</tr>
<tr>
<td>Arg74</td>
<td>8.35</td>
<td>4.14</td>
<td>1.73</td>
<td>CyH 0.734, 1.488, C6H 2.945, 3.073, NeH 7.142</td>
</tr>
<tr>
<td>Gly75</td>
<td>8.40</td>
<td>3.85</td>
<td>3.82</td>
<td></td>
</tr>
<tr>
<td>Gly76</td>
<td>7.86</td>
<td>3.70</td>
<td>3.61</td>
<td></td>
</tr>
<tr>
<td>Residue</td>
<td>HN</td>
<td>HA</td>
<td>HB</td>
<td>Other</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
<td>------</td>
<td>--------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Ile36</td>
<td>6.351</td>
<td>4.454</td>
<td>1.454</td>
<td></td>
</tr>
<tr>
<td>Pro37</td>
<td>4.636</td>
<td>2.412</td>
<td>1.950</td>
<td></td>
</tr>
<tr>
<td>Pro38</td>
<td>4.099</td>
<td>2.222</td>
<td>2.037</td>
<td></td>
</tr>
<tr>
<td>Asp39</td>
<td>8.537</td>
<td>4.411</td>
<td>2.767</td>
<td>2.668</td>
</tr>
<tr>
<td>Gln40</td>
<td>7.842</td>
<td>4.461</td>
<td>1.825</td>
<td></td>
</tr>
<tr>
<td>Gln41</td>
<td>7.495</td>
<td>4.246</td>
<td>2.494</td>
<td></td>
</tr>
<tr>
<td>Arg42</td>
<td>8.487</td>
<td>4.474</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu43</td>
<td>8.755</td>
<td>5.326</td>
<td>1.567</td>
<td>1.164</td>
</tr>
<tr>
<td>Ile44</td>
<td>8.823</td>
<td>4.903</td>
<td>1.762</td>
<td></td>
</tr>
<tr>
<td>Phe45</td>
<td>8.848</td>
<td>5.236</td>
<td>3.001</td>
<td>2.836</td>
</tr>
<tr>
<td>Ala46</td>
<td>8.708</td>
<td>4.072</td>
<td>0.955</td>
<td></td>
</tr>
<tr>
<td>Gly47</td>
<td>8.041</td>
<td>4.104</td>
<td>3.431</td>
<td></td>
</tr>
<tr>
<td>Lys48</td>
<td>7.983</td>
<td>4.599</td>
<td>1.919</td>
<td></td>
</tr>
<tr>
<td>Gln49</td>
<td>8.669</td>
<td>4.479</td>
<td>1.982</td>
<td></td>
</tr>
<tr>
<td>Leu50</td>
<td>8.530</td>
<td>4.039</td>
<td>1.505</td>
<td>1.002</td>
</tr>
<tr>
<td>Gln51</td>
<td>8.379</td>
<td>4.480</td>
<td>2.336</td>
<td>1.972</td>
</tr>
<tr>
<td>Asp52</td>
<td>8.179</td>
<td>4.367</td>
<td>3.321</td>
<td>2.824</td>
</tr>
<tr>
<td>Gly53</td>
<td>8.059</td>
<td>3.906</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg54</td>
<td>7.452</td>
<td>4.721</td>
<td>2.245</td>
<td>2.147</td>
</tr>
<tr>
<td>Thr55</td>
<td>8.834</td>
<td>5.219</td>
<td>4.539</td>
<td></td>
</tr>
<tr>
<td>Leu56</td>
<td>8.246</td>
<td>4.046</td>
<td>2.053</td>
<td>1.222</td>
</tr>
<tr>
<td>Ser57</td>
<td>8.512</td>
<td>4.249</td>
<td>3.859</td>
<td>3.746</td>
</tr>
<tr>
<td>Asp58</td>
<td>7.946</td>
<td>4.295</td>
<td>3.012</td>
<td>2.286</td>
</tr>
<tr>
<td>Tyr59</td>
<td>7.294</td>
<td>4.641</td>
<td>3.443</td>
<td>2.584</td>
</tr>
<tr>
<td>Asn60</td>
<td>8.183</td>
<td>4.361</td>
<td>3.309</td>
<td>2.817</td>
</tr>
<tr>
<td>Ile61</td>
<td>7.162</td>
<td>3.333</td>
<td>1.355</td>
<td></td>
</tr>
<tr>
<td>Glu62</td>
<td>7.669</td>
<td>4.499</td>
<td>2.354</td>
<td>1.886</td>
</tr>
<tr>
<td>Lys63</td>
<td>8.349</td>
<td>3.944</td>
<td>2.019</td>
<td>1.865</td>
</tr>
<tr>
<td>Gln64</td>
<td>9.341</td>
<td>3.307</td>
<td>2.530</td>
<td>2.403</td>
</tr>
<tr>
<td>Ser65</td>
<td>7.787</td>
<td>4.616</td>
<td>3.922</td>
<td>3.654</td>
</tr>
<tr>
<td>Thr66</td>
<td>8.706</td>
<td>5.264</td>
<td>4.069</td>
<td></td>
</tr>
<tr>
<td>Leu67</td>
<td>9.478</td>
<td>5.119</td>
<td>1.639</td>
<td></td>
</tr>
<tr>
<td>His68</td>
<td>9.246</td>
<td>5.133</td>
<td>3.060</td>
<td>2.824</td>
</tr>
<tr>
<td>Leu69</td>
<td>8.225</td>
<td>5.176</td>
<td>1.597</td>
<td>1.114</td>
</tr>
<tr>
<td>Val70</td>
<td>9.124</td>
<td>4.413</td>
<td>2.066</td>
<td></td>
</tr>
<tr>
<td>Leu71</td>
<td>8.145</td>
<td>4.947</td>
<td>1.666</td>
<td></td>
</tr>
<tr>
<td>Arg72</td>
<td>8.596</td>
<td>4.248</td>
<td>1.742</td>
<td>1.520</td>
</tr>
</tbody>
</table>

APPENDIX E: UbAbu(26)Nva(30) assignment table (contd)