The Unusual Structure of Module 13 of Factor H, the Shortest Complement Control Protein Domain

Christopher J. Fenton

A thesis submitted for the degree of Doctor of Philosophy

The University of Edinburgh
September 3, 2007

Abstract

Factor H (fH) is a crucial regulator of the alternative pathway of the complement system, a part of the innate immune system in mammals. Structural information regarding factor H is of great value for investigating the various functions of fH, its interactions with the host's molecules and those of pathogens. As part of our ongoing efforts to solve a complete structure of factor H, the structure of module 13 (fH~13) has been solved by NMR spectroscopy.

To achieve this goal a recombinant fH~13 protein was produced in *Pichia pastoris* in our laboratory and we have prepared unlabelled, ^{15}N labelled and ^{15}N,^{13}C labelled samples of this module. A set of 2D homonuclear, 3D ^{15}N-edited and ^{13}C-edited NOESY spectra have been obtained forming an extensive set of spatial restraints for structure calculation. Adequate NMR experiments have been collected to obtain resonance assignments.

As a part of the familiarization with the procedures for the assignment of homonuclear NMR spectra of proteins, four short polypeptide sequences from the KefC ion channel were analyzed. These corresponded to variations in a loop region participating in intermolecular interactions and were therefore hypothesized to form structured elements. Our analysis did not support the existence of secondary or tertiary structural elements in these peptides.

Among the 88-individual CCP-modules of the regulators of complement activation, module sequence-lengths range from 51 to 67 amino acid residues and fH~13 possesses the shortest sequence. The solved solution structure of fH~13, reflects this short primary sequence and is unusual amongst the complement control proteins (CCP modules). fH~13 possess the expected disulfide-bonding pattern and consensus tryptophan, but lacks many overall 3D-structural features that characterise a "typical" CCP-module. fH~13 possesses only two β-strands out of a maximum of eight. The most similar structure to fH~13 is fH~15, while the most dissimilar CCP module is CR1~16. One side of the fH~13 domain reveals a highly localised positively charged patch composed of eight residues.

To recognize host from non-host cell membranes, factor H binds to polyanions such as sialic acid or heparan sulphate which are bound on to the surface of host cells. There are three putative polyanion binding sites located in modules 7, 13 and 20, whose involvement in this process is, to various extend, supported by experimental evidence. The one in module 13 is the most disputed of the three polyanion binding sites.

Binding studies using gel mobility shift assay were performed using a range of heparin derived oligosaccharides from disaccharide to dodecasaccharide with negative results. Similarly, NMR titrations using a fully sulphated heparin-derived tetrasaccharide yielded a negative result. This is despite the considerable accumulation of positive charge on one side of the fH~13 molecule. These results point to the importance of an adequate distribution of positively charged residues for the binding of polyanions. High charge density only is insufficient to initiate binding in this case.

Abbreviations and symbols

AMAC	aminoacridone
C1 – C9	members of complement cascade numbered in order of their activation
C3a, C4a	anaphylatoxin fragments of C3 and C4
C3b, C4b	proteolysis activated fragments of C3 and C4
C4BP	C4b binding protein
CA	cofactor activity
CCP	complement control protein
CE	Combinatorial Extension
CIE	Cation Exchange
CR1	complement receptor type 1
CR2	complement receptor type 2
DAA	decay accelerating activity
DAF	decay accelerating factor
DPFGSE	double pulsed field gradient spin-echo
EDTA	ethylenedinitrilotetraacetic acid
fH	factor H
FID	free induction decay
FPLC	fast protein liquid chromatography
γ_X	gyromagnetic ratio of spin X
GMSA	gel mobility shift assay
hetNOE	steady state heteronuclear NOE
HSQC	heteronuclear single quantum coherence
I	nuclear spin quantum number
KTN	K^+ transport, nucleotide binding
m	magnetic quantum number
M	bulk magnetisation
MAC	membrane attack complex
MASP	mannan-binding lectin associated serine protease
MBP	mannan-binding protein
MCP	membrane cofactor protein
MTF	molecular template file
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser enhancement
NOESY	nuclear Overhauser effect spectroscopy
PAGE	polyacrylamide gel electrophoresis
PDB	protein databank
PMSF	phenylmethylsulphonyl fluoride
ppm	parts per milion
PTM	post translational modification
RCA	regulator of complement activation
RECOORD	recalculated co-ordinate database
RMD	restrained molecular dynamics
RMSD	root mean square deviation
SA	simulated annealing
SCR	short consensus repeat

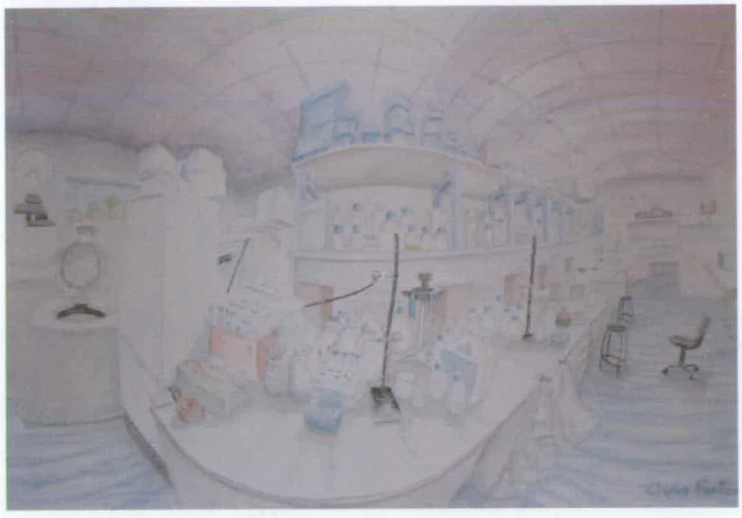
SDS	sodium dodecyl sulfate
SCR	short consensus repeat
T_1	longitudinal relaxation time
T_2	transverse relaxation time
TOCSY	total correlation spectroscopy
VCP	*Vaccinia* complement protein
YPD	yeast peptone dextrose
ω	Larmor frequency

Names and abbreviations for amino acids

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

Declaration

I hereby declare that I composed this thesis entirely myself and that unless otherwise stated, it describes my own research.



Acknowledgements

I would particularly like to thank:

Prof. Paul Barlow, Dr. Graeme Ball, Dr. Juraj Bella, Berbal Blaum, Dr. Krystyna Bromek, Claire Egan, Dr. Andrew Herbert, Henry Hocking, Dr. Lan Jin, Dr. Marc Lenior, Ursula Lodge, Dr. Tran Pham, Dr. Philip Robinson, Christoph Schmidt, Dr. Dinesh Soares

The rest of the BioNMR group and everyone I've encountered in lab 38 and lab 120 for all their help.

BBSRC for providing my funding

And most importantly, Dr. Dušan Uhrín.

Contents

List of Figures

Chapter 1

Introduction

1.1 The Complement System

The human immune system consists of two main systems, an adaptive immune system and the innate immune system, that work together to destroy invading pathogens. The adaptive immune system primarily relies on the production of antibodies to produce a targeted response to specific pathogenic antigens. However, in order to produce a full antibody response to a pathogenic invader the adaptive immune system has to have prior contact with the invader's antigens. As a first line of defense against novel pathogens, the adaptive immune system is of limited value. On the other hand, the complement system, as part of the innate immune system, is able to respond rapidly to destroy any foreign material which does not have suitable protection. Therefore the complement system provides a suitable first line of defense against novel pathogens [1].

1.1.1 The Roles of the Complement System

The complement system has three main strategies it uses to help destroy foriegn invaders.

The Membrane Attack Complex

All activation pathways of the complement system lead ultimately to the activation of the terminal pathway on cell membranes. To do this, one of the C5 convertases needs to catalyse the cleavage of complement protein C5 to C5b and C5a. C5b will

then associate with the fluid phase complement proteins C6, C7, C8 and also multiple copies of the complement protein C9, assembling on the cell membrane. This protein assembly, known as the membrane attack complex (MAC) is able to penetrate the cell membrane. Once the cell membrane has been penetrated, the MAC works as a transmembrane pore allowing cytoplasmic fluid to leak from the cell. Once an invading pathogen has triggered the complement system, the proteolytic cascade effect of the activation pathways will result in many MACs forming on its lipid membranes, leading to lysis of its cells [1].

Inflammation

Activation of the complement system involves the cleavage of proteins C3, C4 and C5. As a result, proteins C3a, C4a and C5a are produced, respectively. Although these proteins play no more role in the continuing proteolytic cascade, these are all anaphylotoxins that promote inflammation. They recruit and activate phagocytes to the site of infection, were the phagocytes will destroy the invading pathogen [2].

Opsonisation

C3b is also produced during the activation of the complement system when C3 is cleaved into C3a and C3b [3]. C3b contains a thio-ester group which it uses to bind to pathogenic cell membranes [4]. Thus, if an invading pathogen activates the complement system, its cell membranes rapidly become opsonised with C3b. Here, C3b plays an important role in the proteolytic cascade, recruiting other complement activators to the cell membrane, finally resulting in the formation of the MAC. However, C3b can aid the immune system in another way. C3b can use its thio-ester group to bind to antigen-antibody complexes. The C3b attached to the antigen-antibody complex then acts as a ligand for erythocyte-borne complement receptor type I (CR1). The opsonised pathogen is then carried on the erythocyte to the liver and spleen where phagocytes are found. Here, C3b allows phagocytes to recognise the pathogen as foreign, inducing the pathogen's destruction via phagocytosis [5].

1.1.2 Activation Pathways of the Complement System

There are three activation pathways for the complement system; the classical, lectin and alternative pathways (Ficure 1.1). Over thirty different proteins are involved in the complement system, either found in the blood serum or on the surface of cells exposed to the blood serum. The main proteins of the complement system are called C1-C9. C5-C9 are involved in the membrane attack complex, as discussed above, while C1-4 are involved in the activation pathways.

The Classical Pathway

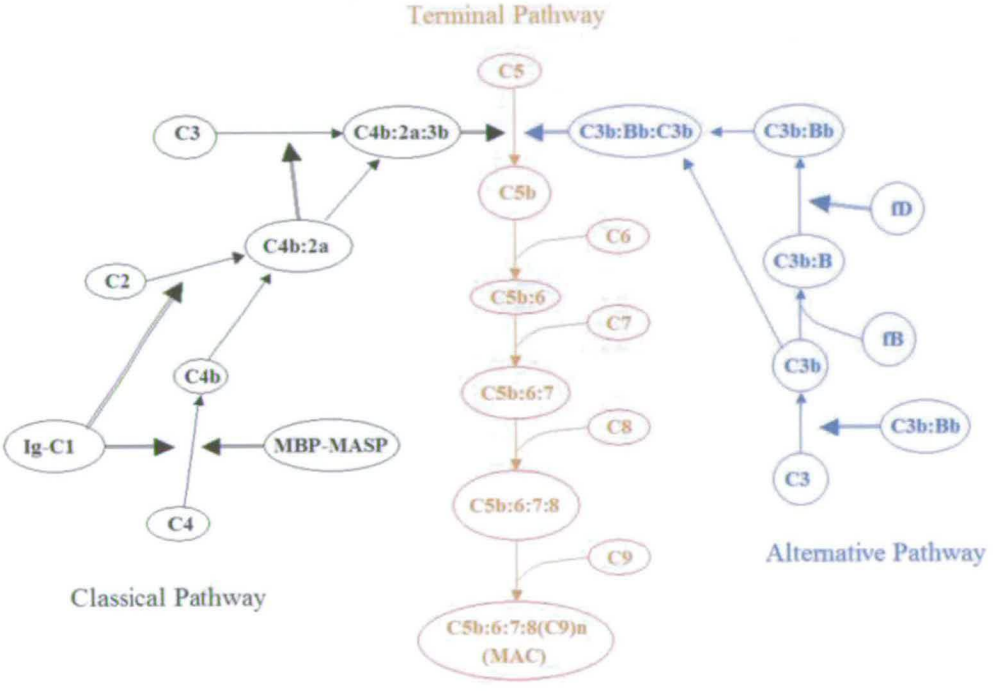


Figure 1.1: A diagram of the complement system, showing the interactions between the Classical (black), Alternative (blue) and Terminal (red) pathways. The complement proteins C1-9 are shown, as are MBP/MASP, factor B (fB and its breakdown product Bb) and factor D (fD). Single arrows depict binding interactions, while double arrows depict emzymatic reactions. Derived from [1]

Discovered first (hence the name), the classical pathway actually relies on the presence of antibody-antigen complexes for its activation. The complement protein C1q can bind to these complexes via the antibody's F_c domain. C1q in turn recruits the proteins

C1r and C1s to form a complex known as C1 or the C4 protease. As indicated by the name, the C4 protease cleaves the complement protein C4 into two fragments known as C4a and C4b. C4a plays a role in inflammation, as described above. Meanwhile, C4b binds to the complement protein C2, and C2 is subsequently cleaved to 2a forming the C3 protease (C4b2a). This cleaves C3 into C3a and C3b. Again, C3a plays its role in inflammation described above. C3b on the other hand has many roles, but in the classical pathway it associates with the already formed C3 protease to form the C5 protease (C4b3b2a). This complex cleaves C5 into C5b, initiating the terminal pathway and the formation of MAC [6].

The Lectin Pathway

The lectin pathway is activated by foreign carboydrates on pathogenic surfaces, such as those found on many types of bacterium, and unlike the classical pathway is independent of antibodies. The pathway is activated when a protein called the Mannose Binding Protein (MBP) binds to foreign carbohydrates. MBP has a similar structure to C1q and it readily forms a complex with MBP-associated serine proteases (MASPs), which are similar to C1r and C1s. Once bound to foreign carbohydrates, this protein complex will cleave C4 into C4a and C4b, and from this point on the proteolytic cascade is the same as that of the classical pathway [7].

The Alternative Pathway

The alternative pathway is also antibody independent, being activated by any surface membrane not protected by complement regulators. C3, which also plays a role in the other two pathways, is the initiator of the alternative pathway. C3 is present in high levels in the blood, where a small proportion of it reacts with water to form a molecule with similar properties to C3b. This hydrolysed C3 associates with a protein called factor B, and this process initiates the cleavage of Factor B into Factor Bb by the enzyme Factor D. The new complex, C3(H_2O)Bb can then cleave C3 into C3b, in a manner analagous to that of the protein complex C4b2a.

As is clear from the above description, a pathogenic organism is not required to initiate

the formation of C3b in the blood, and therefore C3b is constantly produced at low levels in blood sera. C3b binds to nucleophilic groups on surfaces where it can associate with factor B to form C3bB, which is the alternative pathway's C3 convertase. This in turn generates more molecules of C3b. The C3 convertase will also associate with further molecules of C3b to form the C5 convertase (C3bBbC3b), which initiates the terminal pathway, and the formation of the MAC on the lipid surface.

Therefore, with low levels of $C3(H_2O)$ already in the blood, the alternative pathway can respond rapidly to an invasion by foreign organisms without the need for an antibody response. To protect the bodies own surfaces from attack there are various membrane bound and plasma-borne regulators which promote the degradation of C3b or the decay of the C3 and C5 covertases if they become bound to host surfaces [6].

1.1.3 Regulators of Complement Activation (RCA) Proteins

As the complement system can become activated on any surface, regulators are required to protect host cells that are in contact with blood sera. These regulators inactivate the complement system at key places, preventing the opsonisation of host cells and the formation of the MAC. Some regulators are membrane-bound and specific to the cell type on which they are found. Other regulators are in the fluid-phase, and these protect surfaces not enclosed by a membrane. The fluid-phase regulators are also recruited by many different pathogenic organisms to protect these foreign cells from complement attack [8].

There is a family of protein regulators known as the regulators of complement activation (RCA). All the proteins in this family are characterised by having their genes in a cluster known at 1q32, which is found on chromosome 1 [9]. The family includes membrane bound regulators such as CR1, CR2, membrane cofactor protein (MCP) and decay acceleration factor (DAF), as well as the fluid-phase regulators C4 binding protein (C4BP) and factor H (fH) (see Figure 1.2). As they are fluid-phase, C4BP and fH are often the targets for recruitment by foreign organisms [8].

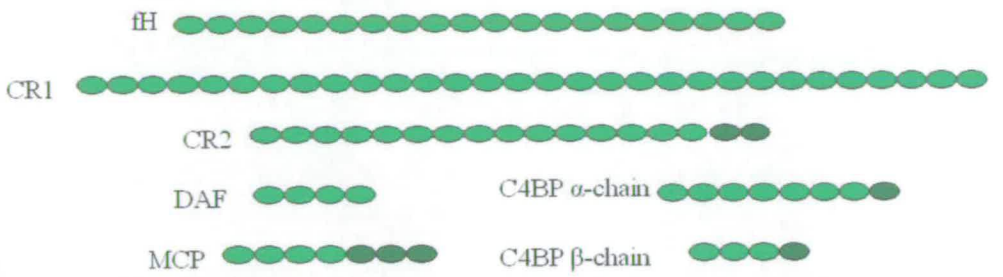


Figure 1.2: A schematic representation of several members of the RCA family. The pale green circles represent CCP module domains, while the dark green ovals represent non-CCP domains, mostly transmembrane or intracellular domains. Derived from [10]

The predominant structural motif for these proteins is the complement control protein (CCP) module. This protein domain is the main domain type (and in the case of factor H, the only domain type) within the RCA proteins. While the different proteins in this family regulate different points in the complement system, the extensive use of CCP modules in their makeup results in them all having related structure and function [11][12]. It is important to note that, while all RCA proteins contain CCP modules, not all CCP modules belong to RCA proteins.

1.1.4 Complement Control Protein (CCP) Modules

CCP modules consist of approximately 50-70 amino acids folded into a domain where the C and N-termini are at opposite ends of the molecule [12][13]. CCP modules are joined together by a series of linker sequences to form structures that resemble beads on a string [12][14]. There are generally two disulphide bridges holding the module together, made up of conserved cysteine residues joined in a 1-3, 2-4 formation [13]. The sequence of a single CCP module is generally quoted as starting from the first cysteine and ending at the last. The linker, measured between the last cysteine of the current module and the first cysteine of the subsequence module, can vary in length between three and eight residues. Also conserved is a tryptophan residue which is found in almost all CCP modules. This is found between the third and the last cysteine residue. There is also a hydrophobic core where many residues are often conserved or conservatively replaced with other hydrophobic residues. It is here that the hydrophobic ring

of the tryptophan is usually found [15][12][13].

As well as some sequence similarity, there is also much structural similarity between CCP modules. The CCP module is elongated, with the approximate dimensions being 40 Å by 15 Å by 15 Å [12][16]. This barrel shape is maintained by up to eight anti-parallel β-strands aligned along the long axis of the domain, forming a β barrel-like structure. While eight strands is the maximum number of strands used to make a CCP module (labeled 1-8 from N- to C- terminus [17]), six is the usual number found, with strands 1 and 3 often absent. Where the full complement of eight β-strands occur in a CCP module, two small β-sheets are found at the N- and C-termini, made up of strands 1 and 3 (N-terminus) and 5 and 8 (C-terminus). The other four strands make up a twisted β-sheet that covers one side of the molecule and results in the hydrophobic core. Within this core lies the two disulphide bridges and other alkyl and aromatic residues. Between β-strands 2 and 3 lies a hypervariable loop. This area shows low sequence similarity between CCP modules, and is also of variable length.

Although CCP modules all share a great deal of structural similarity they vary greatly in their functions [6][12]. This diverse functionality is presumably the result of differences in their primary sequences. Studying the structures of different CCP modules can help in understanding the causes of their different functionality and also help with the design of treatments of diseases where the complement system plays a role. The structures of many CCP modules from different members of the RCA family have been determined [13].

The remaining CCP module structures have been modelled based on their primary sequence and their homology to experimentally determined structures [18][13]. The sequences of 243 CCP modules from 48 proteins were classified using a clustering procedure. This involved sequence alignment to organize the sequences into 9 clusters labeled A-J. The structures of experimentally determined modules were then used as templates for generating models, the clustering procedure ensuring that the most appropriate set of templates would be used in each case.

After the first run of the modeling procedure had been completed the structures of DAF CCP modules 1-4 became available [19]. This allowed for comparison between the DAF models generated by the procedure and the experimentally determined structures. The C^{α} RMSDs between the models and the experimental structures were 1.7, 2.0, 1.2 and 1.9 Å respectively for DAF modules 1-4. This indicates a high level of structural consistency and therefore supports the modeling strategy employed [18].

However, 33 CCP module sequences could not be assigned to a cluster (fH~13 is one such CCP module) and therefore had no experimentally determined template. Also, for clusters D, E and I no experimentally determined structures currently exist, accounting for a further 50 structures that can not be modeled by this method. In total, 83 of the 243 CCP modules examined could not be modeled for lack of a suitable template. Increasing the number of experimentally determined structures increases the number of available templates for modeling. This allows more CCP modules to be modeled and can improve the accuracy of the homology models.

1.1.5 Known RCA Structures

Some of the experimentally determined CCP modules are detailed below. For many RCA proteins, the molecules themselves are too large or flexible to have the whole of their structures determined by X-ray crystallography or NMR. Instead recombinant proteins are prepared of between one and four sequential CCP modules from an RCA protein. These samples are then rigid or small enough to have their structures determined by X-ray crystallography or NMR techniques. This is known as the modular approach and a list of some of the CCP module structures solved by this method to date are detailed below [14][20].

The first CCP module to have it's structure solved was module 16 of factor H. This was done using proton homonuclear NMR spectroscopy [21]. The first structure of a CCP module pair was also part of factor H, that of modules 15-16 [22]. Other sections of factor H with published structures are module 5 [23], modules 19-20 [24] and module

7 [25].

DAF, a membrane bound complement regulator, has had all of it's CCP modules structurally determined. The first of these was the module pair 3-4 and this was by X-ray crystallography [26]. The structures of modules 2-3 were then solved by NMR spectroscopy [27]. Following this, the entire CCP module chain (domains 1-4) was structurally determined using X-ray crystalogrpahy [19]. There were, however, differences between the X-ray structures and those determined by NMR. The X-ray structure of all four modules suggested a rigid, rod-like shape for the CCP modules. However, the NMR structure of modules 2-3 implied a flexible junction between the two modules. The differences could have resulted from the difference in experimental conditions. By their nature, X-ray structures require the sample to be in a solid, crystalline state. Elements of the sample molecule that are otherwise flexible in solution could form rigid conformations that are stabilized by crystal packing forces. In contrast, NMR experiments for structural determination are carried out on samples in solution. This could explain why flexible regions identified by NMR are not present in the X-ray structures.

MCP, another membrane bound complement regulator, has had the structure of it's first two N-terminal CCP modules determined by X-ray crystallography [28]. Here a rigid, bend structure is suggested for the two modules.

CR1 modules 15, 16 and 17 have been structurally determined by NMR spectroscopy using double module constructs of 15-16 and 16-17 [16][29]. These modules represent one of the two copies of site 2 of CR1 (site 1 is found between modules 1-3, while the two copies of site 2 are found at modules 8-10 and at modules 15-17). A triple module construct of 15-17 was created but the side-chain assignment experiments were of poor quality. Once structures existed for the two double module constructs, molecular modeling was used to produce a model structure for all three modules. The structure of 15-17 was found to have an extended head-to-tail conformation with flexability between modules 16-17. Also, a large positively charged patch on one 'face' of module 15 was found. Mutagenesis studies revealled that this patch is required for the binding of

C4b and C3b to CR1.

CR2 has also had the structure of its first two CCP modules determined by X-ray crystalography [30]. The module pair was crystallized with its ligand, C3d, and the crystal strucutre shows how the two interact. In this case a hydrophobic interaction between the modules results in their folding up into a V-shape. This results in only module 1 being in contact with the ligand.

The structure of the first two modules of the α-chain of C4BP was determined by NMR spectroscopy. This revealed an elongated structure for the two modules. These two modules were known to be a binding site for the M4 protein of *Streptococus pyogenes*. Thus, chemical shift mapping was used to map the binding site of the M4 protein on these two modules. This revealed a reorientation of the two modules upon binding [31].

1.2 Factor H

The RCA protein factor H (fH) is an important protein for host surfaces that do not have their own surface bound complement regulators. The protein was discovered in 1965 and initially called β_{1F}H globulin [32]. Synthesized primarily in the liver but also locally during inflammation by, for example, fibroblasts, it is secreted into the blood in high concentrations (500 μg/ml) [33]. The molecule itself is a 155 kDa glycoprotein composed of 20 district CCP modular domains which, if stretched out, would give the protein a length of $\approx$80 nm. Once in the blood fH can bind to and protect host surfaces via polyanions found on their surface [34][35].

Factor H is is the largest member of the factor H protein family. Other members of the family include factor H-like Protein 1 (fHL-1), also known as reconectin. Produced by alternative splicing of the RNA transcript of fH, the 49 kDa protein fHL-1 consists of the first seven (N-terminal) CCP domains of fH with four addition amino-acids at the C-terminal end which are unique to fHL-1. There is also a groups of proteins known as factor H-related proteins (FHRs) which are also members of the factor H protein

family. These are composed of between 4 and 5 CCP modules from fH and are not produced by alternative splicing but are coded for by separate genes [34].

1.2.1 The Functions of Factor H

Factor H regulates the alternative pathway by preventing the build up of the pathways C3 and C5 convertases on host surfaces. To do this it needs to recognise and bind to host surfaces as opposed to foreign cell membranes. It does this by binding to polyanions such as sialoglycoproteins and glycosaminoglycans (GAGs). Human cell surfaces possess a high surface density of sialic acids. Endothial cells have been shown to have heparin-like GAGs on their surfaces too, such as heparan sulphate. In contrast, most bacteria lack molecules of these types on their surfaces, allowing factor H to distinguish host from non-host [35][36][37]. Once bound to host cells factor H can inhibit the proteolytic cascade in two ways. These functions of fH are known as its decay accelerating activity and it's cofactor activity.

Decay Accelerating Activity

Factor H shares its decay accelerating activity with other complement regulators, most notibly DAF. However, factor H is the main fluid-phase decay accelerating factor. The C3 and C5 convertases of the alternative pathway (C3bBb and C3bBb3b respectively) represent important points in the proteolytic cascade, producing increased C3b and C5b concentrations on the activated surface and leading to opsonisation of the surface and formation of the MAC. Factor H regulates the alternative pathway at these points by acceleratiing the decay of both of these convertases. The DAA of factor H is located within the first four *N*-terminal CCP domains of the protein (as discussed in section 1.2.2).

Factor B's cleavage product, factor Bb, is present in both of these convertases, and it is this that factor H causes to be displaced from C3b to decay the convertases. There is also competitive inhibition of factor H with either factor B (during formation of the

convertase) for C3b, with factor H binding to C3b with a 70-fold higher affinity then factor B [36]. With factor H bound to C3b in place of factor B (or its breakdown product, factor Bb), neither the C3 convertase nor the C5 convertase can form.

Cofactor Activity

While the decay accelerating activity can halt the formation of the convertases on host surfaces or hasten their decay, the cofactor activity facilitates the breakdown of C3b. With its opsonising properties combined with its high serum concentration, it is of utmost importance that C3b is not allowed to build up on host cells. An enzyme called factor I can breakdown C3b into inactive constituents, C3i and C3f, but it requires cofactors to become active. This reaction is irreversible, leading to the inactivation of C3b. Along with MCP, factor H acts as such a cofactor for factor I [38]. Thus factor H protects host cell membranes from opsonisation by C3b.

Furthermore, factor H also acts as a cofactor for factor I in the blood plasma. Thus factor H contributes to the continued turnover of C3b in the blood.

1.2.2 Structural and Binding Characteristics of Factor H

As stated above, factor H consists of 20 CCP domains joined by linkers of various lengths. The protein is believed to be heavily glycosylated. Single N-glycosylation sites are located on domains 4, 9, 12, 14, 17 and 18, and two N glycosylation sites on module 15. However, the potential glycosylation site on module 12 is not believed to be glycosylated [34][39].

Although when stretched out as an extended chain the protein would have an overall length of $\approx$80 nm, analysis by synchrotron X-ray scattering and neutron scattering (with additional molecular modelling) suggested a shorter length of $\approx$38 nm for the glycoprotein in solution. This required the protein to be folded back on itself, perhaps allowing the various binding sites to be in close proximity [39].

The decay accelerating activity of factor H is located within the first four N-terminal

CCP domains of the protein [40]. Co-factor activity has also been mapped to these first four domains [41][42]. However, the ability to discriminate between host and non-host cell surfaces requires domains 11-20 [36]. Also, removal of domains 11-15 or 16-20 results in a loss of 97% of the DAA expressed by domains 1-4 [36]. Although currently no structure exists for modules 1-4, work is being done within our group to rectify this.

There are three recognized C3/C3b binding regions of factor H, each binding to a distinct region of the C3b protein. The DAA region (domains 1-4) bind to native C3b [43]. Thus this is the site of competitive inhibition between factor H and factor B for C3b, confirming this region's status as the DAA region of the glycoprotein. The second region for C3 binding exists between domains 8-15, and this binds to the C3c fragment of C3 [43]. The third region is on domains 19-20, which binds to the C3d fragment [43].

There also are two confirmed polyanion binding sites identified for factor H. The first binding site is in domains 19-20 and this region has been well characterized. The structure of the 19-20 module pair has been solved by NMR [24] and X-ray crystalography [44]. Also, GMSA and chemical shift mapping were carried out on the module pair in solution with a fully sulphated heparin-derived tetrasaccharide [24]. The chemical shift mapping revealed a region on module 20 that was particularly perturbed by the addition of the heparin tetrasaccharide. This region contained, amongst other residues, a high density of lysine and arginine residues. Sialic acid is also known to bind fH in the region of domains 16-20 [45].

The second polyanion binding site was identified on domain 7 [46][47]. This domain is also implicated in age-related macular degeneration [48](see Section 1.2.3), suggesting that domain 7 has an important functional role in factor H.

There is some evidence for a third heparin binding site in factor H, although there is debate over the location of the site. Before the two heparin binding sites on domains 7 & 20 were identified there was indirect evidence for a site of heparin interation on domain 13. A photoaffinity-tagged heparin probe labeled the region between domains 12 and 15. Then, CNBr cleavage at the amino acid residue Met787 (located between

domains 13 and 14) split the photoaffinity tagged region [37]. Also, sequence analysis based on known heparin binding sites pointed to domain 13. Heparin binding sites identified in other proteins suggest a consensus sequence consisting of alternate arginine and lysine residues as a binding sequence for heparin. Factor H contains only one such sequence of basic residues, and that is located in the linker region between domains 12 and 13 at residues 731-736 (the sequence is KLKKCK). Domain 13 itself has more basic residues in its sequence then any other CCP domain (15 out of 57 residues). This results in two particular runs of basic residues at 746-750 (HLKAKK) and at 760-766 (RYRCRGK) [37][35]. Domain 13 itself has a high theoretical pI value of $\approx$8.8.

However, more recently a study suggested this third heparin binding site is actually located on domain 9[49]. An extensive series of fragments of factor H was generated in order to more accurately locate this third site of heparin interaction. This library of fH fragments were then passed individually through a heparin-agarose column and their elution profiles were recorded. Fragments of the domains 8-9; 9-11; 8-11; 8-13; 8-14; and 8-15 bound to the column. After ruling out the possible involvement of domain 8 in the binding using constructs of domains 1-7 and 1-8, domain 9 was identified as the prime suspect for heparin binding. Fragments containing the domains 11-14 and 12-15 did not bind to the heparin-agarose column, suggesting the domain 13 could not be a heparin binding site [49]. However, it is worth noting that module 9 has a theoretical pI value of $\approx$4.3 which is low for a heparin binding module. Also, the linker between domains 13 and 14 is unusually long, allowing for a high degree of flexibility that could lead to binding between the two modules that could interfere with the putative heparin binding site. This region of factor H between 12-14 is currently being investigated by our group.

1.2.3 Diseases Associated With Factor H

There are two main diseases associated with factor H; atypical hemolytic uremic syndrome (aHUS) and age related macular degeneration (AMD). Also, some pathogenic organisms recruit factor H onto their membrane surfaces to evade destruction by the

alternative pathway of complement.

Atypical Hemolytic Uremic Syndrome

Hemolytic uremic syndrome is a disease of the kidneys which ultimately leads to acute renal failure [50]. It is caused by platelet thrombi forming in the microcirculation of the kidney causing vessel wall thickening and detachment of the endothelial cells from the basement membrane. Cases of HUS can be placed in two categories; a diarrhea associated form (D+HUS) and a non-diarrhea associated form (D−HUS). D+HUS is caused by Shiga-toxin producing bacteria such as enterohemorrhagic strains of *Escherichia coli*. However, D−HUS, also known as atypical HUS (aHUS), can occur sporadically or is familial [50].

Genetic studies with some aHUS patients demonstrated the involvement of point mutations in the factor H gene in aHUS [51][52]. However, only one gene allele was affected. As factor H is codominantly expressed this resulted in either significantly reduced blood plasma levels of factor H or with expression of variant forms of factor H alongside expression of normal factor H [53].

Point mutations in factor H that cause aHUS predominantly result in single amino-acid substitutions, although a few result in stop codons and the truncation of the expressed molecule [54]. The mutations (including 2 of the known stop codons) are mainly found around domain 20 [55][54]. This domain is functionally important as it is part of a binding site for both polyanions and C3b [24]. Chemical shift mapping of the heparin binding region on the module pair 19-20 revealed that most of the aHUS mutations on module 20 coincided with this binding site [24][43]. Other domains with point mutations are modules 1, 8 (a stop codon), 15 (three mutations known, including one stop codon), 16 (again, three mutaions known, 17 (a single stop codon) and 19 [56][54].

Mutants that result in aHUS have also been found in the RCA protein MCP and factor I.

Age-Related Macular Degeneration (AMD)

Age-related macular degeneration is the leading cause of blindness in the elderly and a major public health issue in the developed world [57]. In AMD extracellular deposits called drusen (consisting of proteins, lipids and other cellular debris) concentrate in and around the macular between the retina and the epithelium [58][48]. As the disease progresses this causes photoreceptor dysfunction in the macular [48][59]. At present, no therapy for this disease has been shown to be broadly effective [59].

AMD is caused by a variety of enviromental and genetic risk factors [59]. Enviromental factors include age and smoking. A genome-wide screen of 96 cases and 50 controls looking for polymorphisms associate with AMD descovered a common variant of the factor H gene to be strongly associated with factor H [48]. The risk allele involved a tyrosine to histidine change at amino acid 402 in domain 7 [48][60][61].

As mentioned above, module 7 is a known heparin binding domain. It is hypothasised that the Y402H mutation causes a weakening of this binding interaction, leading to over activation of the compliment system in the eye, resulting in AMD [48]. Work in our group is being carried out to characterize the structures of both the disease-associated and the non-disease associated forms of domain 7 on its own [25] and in the context of 6-8 constructs.

1.3 Project Aims

Factor H is an important regulator of the alternative pathway, with many factor H mutations associated with serious diseases. It is the ultimate aim of our group to solve the complete structure of factor H for the benefit of both our understanding of the complement system and its associated diseases. To achieve this, constructs of single and multiple modules from factor H are being created and then their structures are being solved using NMR techniques. Therefore, as part of the larger group project, the first aim of this individual project was to purify a construct of the domain 13.

Before the start of this project a strain of *Pichia pastoris* was transformed to produce a construct of fH-13-14. Thus this was to be examined to determine a suitable purification method. Meanwhile, Claire Egan was working to produce and purify a construct containing just fH-13. Once an NMR sample containing module 13 became available (either fH-13-14 or just fH-13) a suite of NMR spectra was to be acquired on the sample to enable the protein to be assigned.

In order to familiarize myself with biomolecular NMR techniques, short peptides derived from the KefC potassium efflux system were to be fully assigned first. After assignment, the data on these short polypeptides was to be used to investigate for evidence of secondary structural elements.

As the shortest CCP module (50 residues from first cysteine to last cysteine), the structure of fH-13 would be of interest to those who study the structure of CCP modules, as it may demonstrate a minimum degree of secondary structural elements required to give a CCP module its structure. It may also help answer whether there is a specific reason for the module being so small, maybe to do with the overall structure of fH. Therefore, assignment of the fH-13 construct and its structural determination was the main aim of this project.

The protein also required assignment in order to answer whether fH-13 was indeed a third heparin binding site for fH. There is evidence both for and against fH-13 being a third heparin binding site, and it is important for all future studies of Factor H that this matter is resolved. Thus, the aim was to use tetrasaccharides derived from heparin to titrate the protein with the sugar. By acquiring appropriate NMR spectra at each titration point the porturbation of chemical shifts for each residue could be determined as a measure of the binding between the two molecules.

Chapter 2

Background

2.1 Backgroud to NMR

2.1.1 Nuclear spin

Some atomic nuclei possess the property of spin characterized by the spin quantum number, I. The values of I start at zero and then are multiples of $\frac{1}{2}$, and a nucleus with an I values greater then zero possesses a magnetic moment, μ. If this nucleus is placed in an external magnetic field then it will precess due to the interaction between the external magnetic field and the nculeus's magnetic moment. The frequency of this precession is known as the Larmor frequency, or the magnetic resonance frequency. It is this resonance frequency that is measured in NMR spectroscopy.

The spinning nucleus will align itself with respect to the external field in one of the available quantum states each labeled with a magnetic quantum number, m. The number of available quantum states can be calculated from the magnetic quantum number. The magnetic quantum number, m, has $2I + 1$ values in integer steps from $+I$ to $-I$: therefore $m = I, I-1, I-2... - I+1, -I+2, -I$. A nucleus with $I = \frac{1}{2}$ can have m values of $+\frac{1}{2}$ and $-\frac{1}{2}$, while a nucleus with a spin quantum number of $I = 1$ can have m values of 1, 0 and -1. Nuclei with spin quantum numbers $\frac{1}{2}$ produce the simplest NMR spectra as these nuclei only have two possible spin states. This group includes ^{1}H, which is found in 99% natural abundence and less abundant nuclei such as ^{13}C, ^{15}N and ^{31}P. Therefore NMR is well suited for investigating all organic molecules, including biomolecules, because predominant nuclei present are hydrogen, carbon and nitrogen.

2.1.2 Magentic Resonance

When a sample of nuclei with a non-zero spin quantum number is placed in a magnetic field the spins will precess at their resonance frequencies. Three factors determine the resonance frequency for a particular nucleus:

1. Different elements, and different isotopes of each element, that have the property of spin will resonate at their own frequencies (all other factors being equal). The constant that relates an isotope's resonance frequency directly to the magnetic field strength is expressed as the gyromagnetic ratio.

2. The precessional frequency of a nucleus is also directly proportional to the external magnetic field. Doubling the magnetic field strength will double the resonance frequency measured.

3. The local magnetic environment at the nucleus is usually slightly different to the external magnetic field. Surrounding electrons and neighbouring atoms contribute to the magnetic environment of the nucleus in question, producing the shielding constant. The relationship between these factors is expressed in the Larmor equation:

$$v = \frac{\gamma B}{2\pi} = \frac{\gamma B_0}{2\pi}(1 - \sigma) \tag{2.1}$$

Where v is the resonance frequency of the nucleus; γ is the gyromagnetic ratio for the nucleus; B is the magnetic environment of the nucleus; B_0 is the external magnetic field and σ is the shielding constant.

The two possible spin-states for a spin $\frac{1}{2}$ nucleus are orientated either with the external magnetic field or against it. The former, known as the α-state, has a lower energy state compared to the latter, known as the β-state. This energy difference is small. Therefore the population of precessing α-state nuclei will be only slightly larger then the number of β-state nuclei in accordance with the Boltzman distribution equation:

$$\frac{N_\alpha}{N_\beta} = e^{-\frac{\Delta E}{kT}} \tag{2.2}$$

Where N_α is the number of nuclei in the α-state, N_β is the numebr of nuclei in the β-state, $\Delta - E$ is the energy difference between the two states, and K is the Boltzman constant.

The nuclei in the sample can be induced to change their spin quantum state by irradiating the sample with electromagnetic radiation with a frequency equal to the resonance frequency of the nuclei. This induced change in the sample leads to a non-equilibrium state, and as the sample returns to its equilibrium state energy is released. Some of this energy will take the form of electromagnetic radiation. It is the frequency of this radiation which is recorded in NMR spectroscopy as it is the same as the Larmor frequency.

Because the energy difference between the α- and β-states is small, the difference in the distribution of the two states is also small. NMR spectroscopy is therefore relatively insensitive. One way to increase the sensitivity of NMR spectroscopy is to increase the external magnetic field strength. This has the effect of increasing the energy difference between the α-state and the β-state, thus increasing the difference in distribution between the two states. As the energy difference between the two states is related to the resonance frequency of the precessing nucleus, increasing the external field strength also increases the resonance frequency as described below.

NMR experiments give information on the specific magnetic environment of a resonating nucleus, with generic chemical groups resonating at characteristic NMR frequencies. Spins can also interact with each other either directly or indirectly in a variety of ways which can be detected, some of which will be discussed later.

2.1.3 Fourier Transform NMR

To measure the resonance frequency of the nuclei in a sample, the sample is placed into an external magnetic field (B_0) which is generally described as being applied along the z-axis. The nuclei will start precessing around the z-axis. As slightly more protons will precess in the α-state state a very small magnetic moment is orientated along the z-axis in the direction of B_0 (Fig 2.1 [A]). This is known as the bulk magnetization

vector, **M**. The spins themselves are randomly distributed along the z-axis so that the x-y components of the spins interfere destructively. There is therefore no net magnetization in the x-y plane at this point.

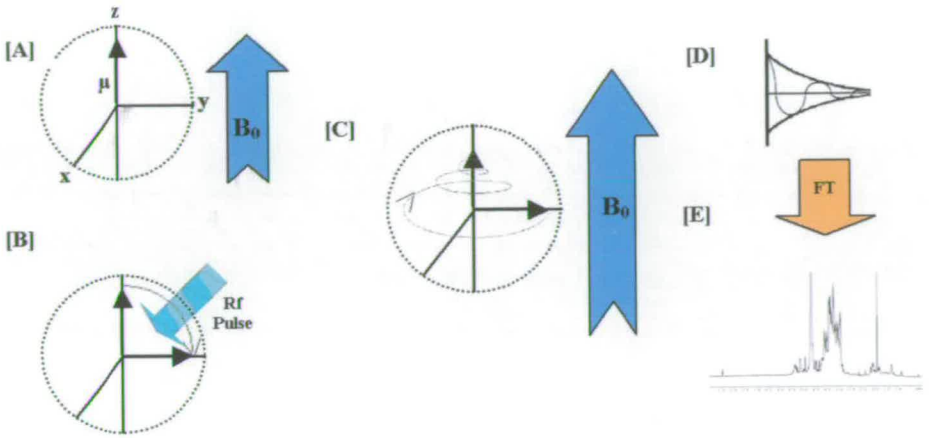


Figure 2.1: Diagram of a simple NMR experiment. The nucleus is placed in a magnetic field (B_0) and begins to precess along the z-axis [A]. An electromagnetic pulse moves the magnetization to the y-axis [B]. Once the pulse has ended the magnetization relaxes back to the z-axis again in a sinusoidal decay in the xy-plane [C]. This decay induces in the receiver coil an oscillating current which is represented graphically as a free induction decay (FID) [D]. The FID is converted into an NMR spectrum by Fourier transformation [E].

A pulse of electromagnetic radiation (rf pulse) is applied to the sample at an angle perpendicular to B_0, in this instance along the y-axis. The magnetic component of this pulse, B_1, will causes the bulk magnetization, **M**, to precess around the y-axis. The strength of the rf pulse is in excess of the external magnetic field and so B_0 can be neglected during this pulse. The rf pulse is switched off once enough time has elapsed for the magnetic moment to precess through the required angle. In NMR experiments, the bulk magnetization is usually flipped through either 90° (as in Fig 2.1 [B]) or 180°. A 180° flip will invert the populations of the two spin states, so now the excess of spins is in the β-state. A 90° flip will equalize the net magnetization with respect to the z-axis, while the bulk magnetization is now along the y-axis.

To represent these motions the vector model is traditionally used. Although the vector model is superceeded by product operator formalism when describing the pulse

sequences used in NMR experiments, the vector model can adequately describe the simple NMR experiment used here. In the vector model a rotating frame of reference is used to visualize the movement of the bulk magnetization. Once **M** has been moved onto the y-axis it will begin to precess around the z-axis at the Larmor frequency for the nuclei. To simplify the visualization of the model, the viewpoint for the model is described as rotating around the z-axis at the Larmor frequency as well, in the same direction as the precessing **M**. This is the rotating frame. Thus in this example, the bulk magnetization will appear static on the y-axis until the sample returns to equilibrium.

Once the rf pulse has ended, the external magnetic field again becomes the dominant field. Therefore the nuclei find themselves in a high energy state and start returning to their equilibrium state along the z-axis. To loose the potential energy of this high energy state the nuclei must be exposed to a fluctuating magnetic field, allowing the dissipation of energy. There are two processes of relaxation by which this can occur: spin-lattice relaxation (with time constant T_1) and spin-spin relaxation (with time constant T_2). In spin-lattice relaxation the rotation of the molecule in solution causes the magnetic field of each spin to fluctuate with respect to its neighbour. Spin-lattice relaxation is characterized as fluctuations in the magnetic field along the x-y plane, causing the bulk magnetization to return to the z-axis. In spin-spin relaxation, the fluctuation of the magnetic fields of nearby spins causes a fluctuation in the z-axis magnetization experienced by individual spins. This in turn leads to a loss of phase coherence of spins in the x-y plane until eventually the spins interfere destructively again. Thus spin-spin relaxation can be characterized as the loss of bulk magnetization in the x-y plane.

The trajectory of the relaxation of the bulk magnetization is shown in Fig 2.1 [C]. In the xy-plane of the sample is a receiver coil. The rotation of the bulk magnetization around this plane will induce in its coil a very small electric current that is then amplified. The sinusoidal decay of this rotation in the xy-plane means the oscillating current will take the form of a free induction decay (Fig 2.1 [D]). This primary NMR signal is converted into an NMR spectrum by Fourier transformation (Fig 2.1 [E]).

2.1.4 Chemical Shift

A bare proton in a 2.35 Tesla magnetic field will resonate at approximately 100 MHz. However, the magnetic field experienced by the nucleus (B) of an isolated atom will be slightly lower then external magnetic field (B_0), and this affects the resonance frequency of the nucleus. This is due to the shielding effect of the hydrogen atom's electron. B_0 induces the electron to circulate in its orbital. In turn this motion of a charged body produces a magnetic field that is opposed to that of the external magnetic field. This opposing field acts to lower the strength of the magnetic field experienced by the nucleus (B) compared to the external field (B_0). This effect is known as nuclear shielding and the difference in B from B_0 can be accounted for using the shielding constant, σ.

In molecules the variation in electron density, as well as the motion of each electron induced by the external field can together produce localized magnetic fields that either oppose or augment the external field. Thus the electronic structure in the immediate environment of a nucleus has a small but measurable effect on the magnetic environment (and therefore the resonance frequency) of a nucleus. Each separate magnetic environment in a molecule produces a separate resonance frequency, and this variation is known as chemical shift.

The shielding constant is an impractical measurement of chemical shift, therefore the term δ is used. This is a dimensionless parameter which is essentially the difference of v of the nucleus from the standard reference frequency, v_{ref}. For protons the standard is usually the resonance frequency of the protons in tetramethyl silane (TMS). To calculate δ the frequency difference, $v - v_{ref}$, is divided by v_{ref} to make it dimensionless but also frequency independent (thus allowing direct comparisons of chemical shifts measured at different signal strengths). The chemical shift range for protons in a 100 MHz spectrometer will be in the order of 1000 Hz. Therefore δ is scaled up by 10^6 and expressed in parts per million or ppm. ppm can easily be converted into Hertz:

$$\delta = \frac{v - v_0}{v_0} \times 10^6 \tag{2.3}$$

Where, v is the resonance frequency of the nucleus of interest; v_0 is the resonance frequency of the standard nucleus; and δ is the chemical shift expressed in parts per

million.

Each proton will produce a peak at its chemical shift. The signal from the standard will produce a peak at 0 ppm, while all the other resonating protons will give rise to peaks relative to this. The area under a peak is directly proportional to the number of resonating protons. This can be seen in a typical 1D proton spectrum of a protein, as shown in Figure 2.2. While 1D stands for one dimensional, 1D spectra actually contain two dimensions: resonance and intensity. In the 1D proton spectrum of a protein there are four main regions.

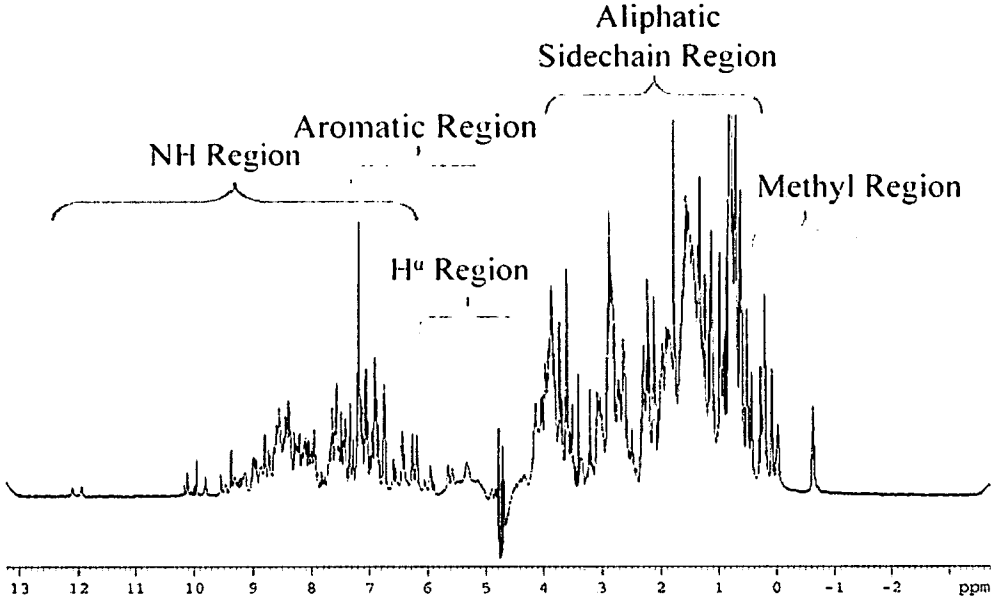


Figure 2.2: A 1D protein spectrum of fH-13 showing the five main regions of interest in protein 1D spectra.

In large molecules such as proteins, there will be various contributions to the shielding of an individual nucleus. The electrons orbiting the nucleus themselves can circulate opposing B_0 producing diamagnetic shifts, or complementing B_0, producing paramagnetic shifts. The greater the electron density of the atom, the greater the shift, and therefore these local contributions are very small for protons. Hydrogen bonds and unpaired electrons produce strong deshielding effects. Charged or polar groups produce local electric fields, thus pH can effect the chemical shift of groups. Neighboring groups

can also produce diamagnetic or paramagnetic shifts when subjected to the motion of the molecule in solution, produced by the shielding anisotropy of the group. Along with through-bond effects, this is particularly important in folded proteins as a cause of deviation of chemical shifts from random coil values.

2.2 Multidimensional NMR Experiments for Proteins

In large molecules such as proteins, where there are many magnetically distinct proton nuclei with similar chemical shifts, a simple 1D NMR spectrum can become crowded. Multidimensional NMR experiments are used to resolve the individual nuclei as well as to investigate the relationships between them. The first to be discussed here are 2D NMR experiments. The pulse sequences needed for 1D spectra are is usually composed of 5 generic parts; relaxation, preparation, evolution, mixing and acquisition. This same sequence is repeated with different evolution times over the course of multidimensional experiment.

During relaxation the nuclei in the sample are either at (as in the very start of the experiment), or returning to (as in when the sequence is being repeated with a different evolution time), their equilibrium state. The preparation phase generally involves a 90^o pulse so that the bulk magnetization is in the xy-plane. Here it is left to precess freely during the evolution period. The length of time, t_1, over which this free precession is allowed to occur determines how much magnetization is used during the mixing period. Varying t_1 will also label the chemical shift of nuclei
Then comes the mixing period, where phase coherences are transfered between spins. Depending on the experiment, the transfer can be between protons (as in homonuclear experiemnts) or from carbon or nitrogen to protons (as in heteronuclear experiments). For example, in a 2D COSY experiment the phase coherences are transfered between nuclei through scalar interactions. This interaction is mediated through covalent bonds. Thus a 2D COSY spectrum can be used to identify nuclei that are J-coupled, as will be explained below in more detail.

After the mixing period comes the acquisition phase, where the signal is recorded

in the spectrometer's receiver coils. This also has a variable time parameter, t_2. The value of t_1 is varied during the progression of the experiment. Once all the FIDs have been collected a Fourier transformation of data along the t_1 and t_2 dimensions, results in a 2D spectrum.

There are three main 2D homonuclear spectra: COSY; TOCSY and NOESY. These spectra have two ^{1}H resonance axis. They each have a 'diagonal': Across the spectrum, from the bottom left corner to the top right corner, will lie a peak for each peak to be found in a 1D spectrum at its correct resonance in both resonance dimensions. Homonuclear spectra have crosspeaks on both sides of the diagonal, producing symmetry-related peaks. Crosspeaks connect two resonances and represent the interactions between the two nuclei.

A 2D heteronuclear spectrum will not have a diagonal for obvious reasons. However, by adding additional evolution and mixing periods into the pulse sequence, 3D and 4D NMR experiments can be made. These experiments are very useful for resolving resonances in large proteins, and for generating additional distance restraints for structural determination, as will be discussed later. Meanwhile, the three main 2D NMR experiments will be discussed in more detail, starting with the COSY experiment.

2.2.1 Scalar Coupling and Correlated Spectroscopy

Nuclei that are connected through covalent bonds can interact through the shared electrons of their bonds in a phenomenon known as scalar coupling (or spin-spin coupling). Take, for example, two covalently bonded spin-$\frac{1}{2}$ nuclei called D and E. In an individual molecule the magnetic state of D (α- or β-state) will affect, in a small way, the magnetic field experienced by E, and vica versa.

Thus, with D precessing in the α-state (D_α), nucleus E experiences extra shielding, resulting in a slightly lower resonance frequency for E. With D precessing in the β-state the opposite is true. As the populations of the α- and β-states are approximately equal across the whole sample this results in the characteristic splitting of the reso-

nance peak for E. The signal peak for E splits into two peaks each half the intensity of what the single peak would have been. The difference in the two peaks, referred to as the J-coupling, is measured in Hertz and (as it is a scalar value) is field independent. Instead the magnitude of the J-coupling is dependent on the number of bonds that seperate D from E, and also the dihedral and torsion angles of those bonds. D will also exhibit the same phenomenon as a result of E, and the splitting that occurs will have the same value of J.

In a COSY spectrum the crosspeaks represent interactions through covalent bonds between scalar coupled nuclei. The couplings decay rapidly with increased separation between coupled nuclei. The result of this decay is that usually only 2- and 3-bond coupling constants are large enough to be significant. The COSY spectrum therefore contain important local structural information. Bonded nuclei in biomolecules form isolated spin systems, for example peptide monomers or saccharide monomers. For example, the backbone NH proton in an amino-acid residue will show a crosspeak to α-protons in the same residue, but not to any β-protons as they are not within 3 bonds of the NH-protons. These networks of scalar coupled spins can be mapped using the COSY spectrum allowing the assignment of resonance frequencies to individual nuclei.

In a TOCSY spectrum, like a COSY spectrum, the crosspeaks represent interactions through covalent bonds between the nuclei, again mediated by scalar coupling. Unlike in a COSY spectrum, the two interacting nuclei do not have to be within 3 covalent bonds of each other to produce a crosspeak. This is because in the mixing phase of the experiment there is a special spin lock pulse. This pulse forces the bulk magnetization to precess around the axis from which the rf pulse was applied. Thus all the precessing nuclei have equal Larmor frequencies. This allows the transfer of magnetization between all the precessing nuclei in the same spin system. In a TOCSY spectrum the magnetization can be transfered between nuclei up to six covalent bonds away from each other, depending on the efficiency of the transfer and the mixing time. This means that, for example, from the resonance frequency of a backbone NH proton it is possible to see most of the side-chain proton resonances, if not all, for that residue as crosspeaks to the NH resonance.

2.2.2 The Nuclear Overhauser Effect and NOE Correlated Spectroscopy:

While the phenomenon of scalar coupling can be used to investigate nuclei that are covalently bonded together, the nuclear Overhauser effect (NOE) can be used to investigate nuclei which are close in space. The distance information obtained from NOE experiments is the main source of structural information used in NMR structural determination. The NOE is the result of dipolar interactions between nuclei and the relaxation that results from these interactions. Unlike scalar coupling, the dipolar coupling does not produce signal splitting in isotropic media, but instead affects the peak intensity of the interacting nuclei. The NOE is affected by molecular tumbling but can be observed in isotropic samples.

To explain the NOE effect, consider a 1D NOESY experiment on two non-equivalent nuclei A and B which are close in space but not necessarily connected through bonds. Nuclei A and B both produce their own magnetic fields that (because the two nuclei are close in space) are capable of influencing the magnetic enviroment of the other nucleus. Because both A and B are part of a rotating molecule, the magnetic field produced by A is experienced as a fluctuating magnetic field by B and vice versa, in the same manner as occurs in spin-lattice relaxation.

This interaction allows the transfer of magnetization between spins in a process known as cross-relaxation. During the steady-state 1D NOESY experiment, a weak radiofrequency pulse is used to saturate nucleus A. The saturation of proton A will cause a non-equilibrium state to arise where the α- and β-states of the proton are no longer in a Boltzman distribution. This is a high-energy state and so nucleus A relaxes to equilibrium As it does so it will, through cross-relaxation, affect the distribution of the α- and β-spin states of B. The nucleus B signal can thus become weaker or stronger depending on the new distribution of these two states. Large molecules which tumble slowly tend to produce an increase in the β-spin state of the population of nucleus. For small molecules, which tend to tumble fast, the opposite is true.

1D NOESY experiments are of little practical use in NMR spectroscopy of large pro-

teins. There is the problem of signal overlap, which in a 1D NOE experiment would prevent the selective saturation of individual spin resonances. There is also the problem of spin-diffusion whereby magnetization is transfered throughout the molecule. While this phenomenon can be utilized for the examination of protein-ligand interactions, in a NOE experiment it would lead to the undesirable transfer of magnetization between remote nuclei.

Thus 2D NOESY spectroscopy is employed, whereby a non-equilibrium state is created for all nuclei at the same time by the mixing pulse. During the following mixing period the magnetization is transfered between spins via the NOE. In the resulting spectra the crosspeaks therefore represent nuclei that are relatively close in space, with the intensity of the crosspeak related to the distance. The intensity of NOE decays rapidly as the distance between nuclei increases at the rate of r^{-6}. Peaks due to NOE will only be observed if nuclei A and B are within 5-6 Å of each other. NOE crosspeaks can therefore be used as short-range distance restraints.

During assignment, homonuclear COSY and TOCSY spectra can't be used to join individual spin systems together within the primary structure of a polypeptide. This is because even the closest protons in adjacent peptide residues are always more then 3 bonds from each other, or are separated by a non-protonated carbon. Therefore homonuclear scalar coupling experiments can only be used to identify protons of the same spin system. With NOESY spectra it is possible to identify crosspeaks between residues that are adjacent in the primary sequence, and so assign each spin system to its place in the primary sequence. Crosspeaks will also be seen between proton resonances where the protons themselves are not close in the primary sequence but are close in the tertiary structure, and these NOE restraints are the basis for structural determination by NMR.

2.3 Heteronuclear Spectra

A suite of good quality homonuclear proton spectra are suficient to derive the structures of proteins up to about 8 kDa. For larger proteins many protons will have similar

chemical shifts leading to peak overlap, even in 2D spectra. This means that a high percentage of the protein's proton nuclei can't be unambiguously assigned chemical shifts. Isotopic labeling can be used to overcome problems of overlap or poor quality spectra.

As mentioned above, in proteins, the hydrogen, carbon and nitrogen are the nuclei of principle importance. ^{1}H, ^{15}N and ^{13}C all have spin $\frac{1}{2}$ and thus are suitable for protein NMR. However, ^{15}N and ^{13}C do not occur in high abundance in nature (0.13 and 1.1% repectively). To overcome this, special 'labelled' protein can be engineered by expressing proteins in host organisms fed on media containing only these isotopes (the method for labelling is described in section 3.1).

Labeling allows the effective use of heteronuclear experiments. Here the magnetization is transfered between heteronuclei, allowing the spread of the signals through another frequency range (represented through another dimension on the spectra). Protons can thus be differentiated not just by their chemical shift, but also by the atom-type they are bonded to. Also, the chemical shift of the heteronucleus can be used to help identify a chemical group or spin system.

The most common heteronuclear experiment is the HSQC (Heteronuclear Single Quantum Coherence), a 2D experiment where one dimension represents proton resonances and the other either ^{15}N or ^{13}C resonances. In this experiment, magnetisation will start on ^{1}H then be transfered via scalar coupling to ^{15}N or ^{13}C for chemical shift labelling, before being transfered back to the ^{1}H for detection. This way, only protons covalently linked to ^{15}N or ^{13}C will be visible in the spectrum. An example of an ^{15}N,^{1}H HSQC is shown in Figure 2.3.

The ^{15}N,^{1}H HSQC is a very common heteronuclear experiment and is used as a test of quality of isotopically labelled samples. The experiment is very sensitive and so can be run quickly using very few scans, or on diluted samples. As it correlates ^{1}H nuclei with ^{15}N the resulting spectrum will have a peak for each backbone amide group giving a simple fingerprint for each protein sample. A poorly folded protein can be easily iden-

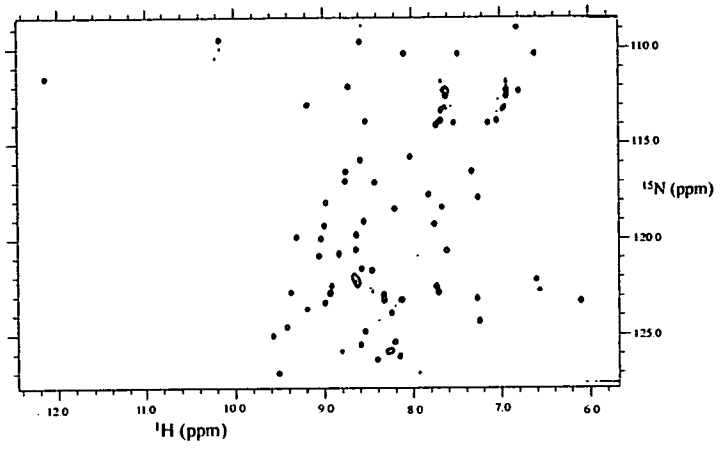


Figure 2.3: An ^{15}N HSQC of factor H, Module 13

tified as the crosspeaks tend to cluster in one region of the spectrum. Sidechain amide groups are also visible, usually resonating at significantly different nitrogen chemical shifts from their backbone counterparts.

Isotopic labeling is also used to produce ^{15}N- or ^{13}C-edited versions of TOCSY or NOESY spectra. Here, the isotopic labeling is used to spread the signals through an additional dimension. Thus, a 3D ^{15}N-edited TOCSY spectrum would spread the TOCSY peaks through the frequency range of nitrogen-15. ^{15}N and ^{13}C edited spectra are used to correctly assign nuclei in larger proteins, or small proteins that produce poor quality spectra. This can help produce additional unambiguous distance restraints, and improve the precision or accuracy of the derived structures. A more detailed explaination of the types of heteronuclear spectra applicable to proteins, and their uses, is given in section 2.6.

The chemical shift range for ^{15}N and ^{13}C is much greater then for ^{1}H. When running ^{15}N- or ^{13}C- edited experiments it is common not to sample the entire frequency range for these heteronclear nuclei. This saves time, and as hundreds of increments may be required for each experiment, this can represent a significant saving of time. The nuclei lying outside the frequency range or sweep width of the experiment are still present in the FID. However, the waveform of these signals will be insufficiently

sampled for their correct frequency to be determined. Thus their peaks are still present in the resulting spectrum but at incorrect chemical shifts. The peaks are known as aliased peaks, and are said to be folded into the spectrum. It is possible to calculate the correct frequency of an aliased peak from its phase and its chemical shift in the spectrum, provided one knows how many times the peak has been folded. Thus before carrying out a heteronucler experiment a sweep width is chosen for the heteronucleus that will fold the aliased peaks into empty regions of the spectrum.

2.4 Residual Dipolar Couplings

Dipolar coupling (D-coupling) is a direct through-space interaction between NMR-active nuclei. Like scalar coupling, D-coupling produces splitting of a signal. D-coupling is a through space interaction which depends on the immediate orientation of the vector that connects two nuclei with respect to the external magnetic field and not just the distance between them. Unlike J-coupling it is not scalar but vectorial in nature, and is decribed only by the dipolar tensor.

In solid-state NMR where there is very little molecular motion in the sample, any dipolar interactions produce large signal splitting. In a sample of micro-crystalline solid there are many crystals at different orientations relative to each other and to B_0. This will lead to dipolar couplings of different magnitudes superimposed to yield a broad NMR signal. However, in liquid-state NMR the molecules are constantly tumbling rapidly, changing speed and direction due to collisions with other molecules. This tumbling is often at a rate that far exceeds the coupling frequency for dipolar interactions and so in isotropic liquids dipolar interactions don't produce signal splitting in the NMR spectrum. Molecular tumbling can be restricted using special aligning media introduced into the sample.

The type of alignment media used for fH-13 was the liquid crystaline media consisting of cetylpyridinium bromide (CPBr)/hexanol/sodium bromide solution. There are other types of media. Liquid crystaline media based on phospholipids was first used for alignment to measure RDCs [62]. Suspensions of charged rod-shaped viruses and phages

can also be used [63]. However, these biomolecular alignment media are expensive compared to the chemical media CPBr/hexanol/NaBr [64]. Because CPBr/hexanol/NaBr proved effective in aligning fH-13 it was chosen.

The various types of alignment media all restrict the isotropic molecular tumbling so that some splitting due to D-coupling can be seen in NMR spectra, albeit a fraction of the magnitude of couplings seen in solid-state. These reduced splitting are known as residual dipolar couplings (RDCs). The observed signal is split by the sum of $J + D$ couplings. In order to obtain the dipolar coupling constant the J-coupling constant must also be measured to high accuracy. Because a dipolar coupling has a tensorial character, RDCs can give long range structure information and so are complementary to NOE restraints and are important tools in protein structure determination.

2.5 Assignment of NMR Spectra of Unlabeled Protein Samples

The assignment of the unlabeled and ^{15}N-labeled protein samples was carried out using the assignment program ANSIG [65]. This method for assigning unlabeled polypeptides using COSY, TOCSY and NOESY 2D spectra was adapted from 'Resonance Assignment Strategies for Small Proteins' by Christina Redfield [66]. The method was used in the assignment of the KefC polypeptides.

Assignment of polypeptides in unlabeled samples was carried out in three basic steps:

- Using the COSY and TOCSY spectra, individual spin systems were identified and all the peaks were picked.
- Using a table of average resonances for the amino acids, each spin system was labeled with a degenerate marker corresponding to a group of amino acids of with the spin system is most probably a member.
- Using the NOESY spectrum, the position of the spin systems in the primary sequence is ascertained.

Once the position in the primary sequence of each spin system has been identified, the type of amino acid should be known. The amino acid type should fall within the group identified in the second step. If not then all three steps need to be repeated for that spin system.

2.5.1 Identifying Spin Systems and Picking Peaks

Identifying individual spin systems is done mainly through the COSY spectrum. First, crosspeaks are picked in the fingerprint region of the spectrum, between about 6.5-10.0 ppm in the directly detected dimension (the H^N proton region) and about 3-6 ppm (the H^α proton region) in the indirectly detected dimension. Although some of these crosspeaks may correspond to interactions with side chain N-H protons, most will be J-coupled interactions between backbone H^N protons and H^α protons. The number of resonance peaks in this region should therefore correlate fairly well with the number of amino acids in the polypeptide with some exceptions; N-terminal H^N resonances may not appear in the spectrum due to exchange and this will be pH dependent; Proline residues have no backbone H^N proton; and Glycine residues may have two distinct peaks in the fingerprint region representing the two H^α protons.

To identify all the resonances that belong to a single spin system the TOCSY spectrum can be used. The TOCSY spectrum has the potential to show crosspeaks to all resonances in a spin system from a single resonance, although this is not the case for all amino acids. However, by looking at all the TOCSY crosspeaks from the original H^N proton, the H^α, H^β, H^γ etc. proton resonances can often be found, although it is not always possible to unambiguously distinguish between them. In order to do this the COSY spectrum is again used. Protons attached to adjacent carbon atoms will produce COSY crosspeaks, and from this it is possible in most cases to distinguish between H^α, H^β and H^γ protons within a spin system.

As stated before, some NH resonances with crosspeaks in the fingerprint region do not belong to the backbone but to the sidechain of amino acids. Often these peaks are not J-coupled through protons to the rest of their residue, but can be traced to the

residue of origin through NOESY crosspeaks (see section 2.5.3).

2.5.2 Labeling Spin Systems

Some residue types, namely glycine, alanine and threonine, can be identified easily if there is little peak overlap in the regions where their peaks lie:

Glycine

Glycine will most likely have two separate COSY crosspeaks from the backbone H^N proton, representing the interaction through bonds with the two H^α-protons. As glycine is unique in this respect, this is an instant way of identifying it. If the H^α-proton resonances are too close together such that two distinct crosspeaks cannot be seen, then glycine can be identified as a spin system with no H^β protons.

Alanine and Threonine

Alanine and threonine residues will produce large single H^α-H^β J-coupled crosspeaks. For alanine the H^β resonances fall in the methyl region of COSY and TOCSY spectra. For threonine the H^β resonances are usually to be found at a slightly higher chemical shift then the H^α resonances. The threonine H^β proton will also have a large COSY crosspeak to the H^γ methyl protons, producing a large single crosspeak to the methyl region of the spectrum.

Valine, Leucine and Isoleucine

Valine, leucine and isoleucine all have crosspeaks to methyl groups in COSY and TOCSY spectra. The position of these methyl groups in the amino acid, and thus the identity of the amino acid, can be easy determined from the COSY and TOCSY spectra (see Figure 2.4). In valine, the easiest to identify, crosspeaks from the H^β proton to two large peaks in the methyl region of the COSY spectrum can be seen, representing H^β-H^γ interactions. That these peaks belong to the same residue can be confirmed through the TOCSY spectrum by looking for H^α-H^γ or H^N-H^γ crosspeaks.

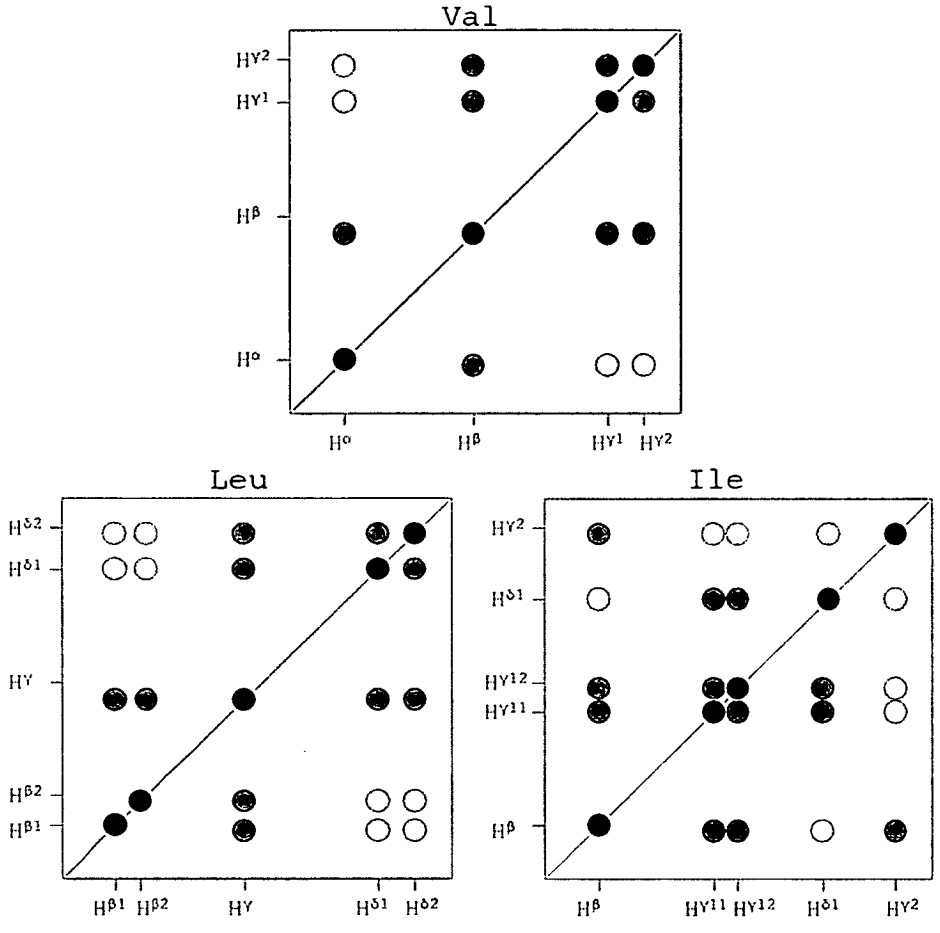


Figure 2.4: Valine, leucine and isoleucine make characteristic patterns in the aliphatic and methyl regions in COSY and TOCSY spectra. The gray circles represent crosspeaks in COSY spectra. In addition, the open circles represent crosspeaks in TOCSY spectra.

The pattern of crosspeaks in leucine is very similar to that of valine. Here a crosspeak from the H^{γ} proton to two methyl groups can be seen in the COSY spectrum. Again, this can be confirmed through the TOCSY spectrum by looking for H^{β}-H^{δ}, H^{α}-H^{δ}, and even H^{N}-H^{δ} crosspeaks can be seen if there is a high degree of TOCSY transfer.

Isoleucine residues produce a complex pattern in the methyl region of TOCSY and COSY spectra. The following crosspeaks can be found in the COSY spectrum of an isoleucine residue (not including H^{N} crosspeaks): between the H^{α} to a single H^{β}; be-

tween the H^{β} to the H^{γ} protons in the methyl region; between the H^{β} to two H^{γ} not in the methyl region; and between the two H^{γ} not in the methyl region to a single H^{δ} in the methyl region.

J Residues

For the residues not described above identification becomes harder. However, these residues can be placed into to broad categories, primarily by whether the residues in question have H^{γ} protons or not. Those designated J residues have no H^{γ} protons. This category therefore includes cysteine, serine, asparagine, aspartate, histidine, phenylalanine, tyrosine and tryptophan. Aside from the fact that neither COSY nor TOCSY crosspeaks can be found to protons beyond the H^{β} protons, these residues can usually be identified because their H^{β} protons will generally be found at less then 2.5 ppm.

While cysteine, serine and aspartate can't be reliably distinguished using 2D homonuclear NMR spectra, the aromatic residues and asparagine can be individually identified using COSY, TOCSY and NOESY spectra if the conditions are favorable. If groups beyond the H^{γ} position produce NOE crosspeaks to the H^{β} protons, or even H^{α} or H^{N} protons, then this is evidence that these protons belong to the same residue, although the NOE interaction could be inter-residue and this ambiguity has to be taken into account during assignment.

U Residues

The final group of amino-acids is designated type U residues and includes the residues lysine, arginine, methionine, glutamine and glutamic acid (proline would also be in this group but as it has no H^{N} proton it is treated separately). These amino-acids are distinguished from the other types by having two H^{γ} protons J-coupled to the H^{β} protons, with the latter protons usually resonating at more then 2.2 ppm. Because of this definition, any unidentified leucine residues are also likely to be placed in this group. As there is no methionine in any of the proteins assigned as part of this project, it is not discussed.

For glutamine and glutamic acid, the H^{γ} proton resonances are generally found at a higher ppm then the H^{β} resonances. Although distinguishing between these two amino-acids is hard, sometimes the protons on the H^{δ}-amide group of glutamine can be seen with a faint NOE-coupled crosspeak to the H^{γ} resonances.

Arginine, lysine, leucine and proline, on the other hand, generally have H^{γ} proton resonanting at a lower ppm then the H^{β} resonances. The protons of the extended chains of arginine and lysine are all J-coupled to each other. For arginine, the $H^{N\epsilon}$ proton will produce a crosspeak to the H^{δ} protons. For lysine the $H^{N\zeta}$ protons will produce a crosspeak to the H^{ϵ} protons.

Proline

Proline residues are hard to identify as they have no H^{N} backbone protons. Their H^{α}-H^{β}, H^{β}-H^{γ} and H^{γ}-H^{δ}, protons are strongly J-coupled though, and are usually quite easy to find in spectra with little overlap once most other side chains are assigned.

If this is not the case, the best chance of assigning a proline residue is through NOE interactions with the subsequent residue. NOESY crosspeaks from the proline H^{α} and H^{β} protons to the H^{N} backbone proton of the subsequent residue should be seen.

2.5.3 Assignment of Aromatic Side Chains

Although the aromatic amino acids have no H^{γ} protons to propagate the magnetization in COSY and TOCSY spectra, it is still possible to assign their aromatic groups prior to sequential assignment. The ring protons of aromatic groups tend to be found between 6.0-8.0 ppm. NOE crosspeaks in NOESY spectra can often be seen between the H^{δ} protons of the aromatic group and the H^{β} protons of its own residue. The specific aromatic group can be identified by the pattern it makes in 2D TOCSY and COSY spectra as shown in Fig 2.5.

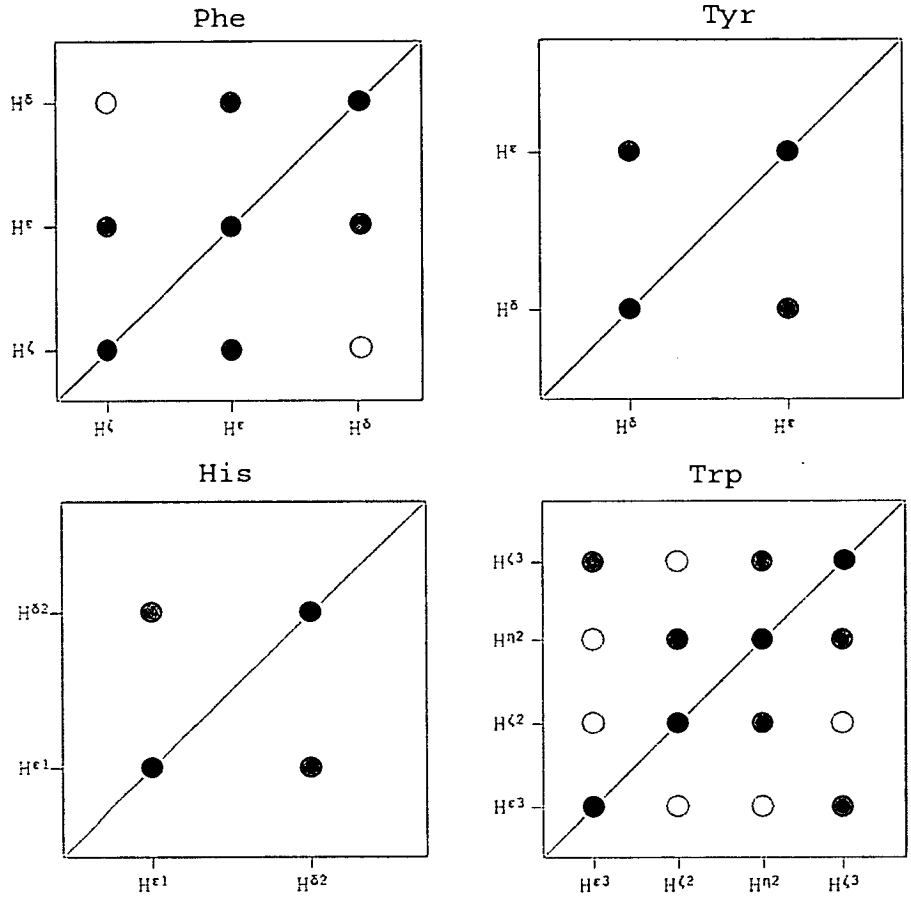


Figure 2.5: Aromatic side chains can be identified by the symmetrical patterns they make in COSY and TOCSY spectra between 6 ppm and 8 ppm. The gray circles represent crosspeaks in COSY spectra. In addition, the open circles represent crosspeaks in TOCSY spectra. Diagram adapted from 'NMR of Macromolecules', page 85, Figure 6 [66]

Phenylalanine

Provided the polypeptide structure allows the phenylalanine ring to spin freely the H^{δ} protons will be magnetically equivalent and appear at the same resonance frequency. The same will also be true of the H^{ϵ} protons. Therefore, along the H^{ζ}, there should be three resonances giving rise to four COSY crosspeaks, or six TOCSY crosspeaks. All these aromatic protons should be found around 6-7 ppm.

Tyrosine

Like Phenylalanine, Tyrosine's H^{δ} and H^{ϵ} protons form two magnetically equivalent pairs, and thus have two resonance frequencies (provided the aromatic ring is free to rotate). The H^{δ} and H^{ϵ} resonances are J-coupled together, and so produce two symmetrically related crosspeaks in both TOCSY and COSY spectra. The H^{ζ} proton is rarely seen. As with phenylalanine, all these aromatic protons should be found around 6-7 ppm.

Histidine

The $H^{\delta 2}$ (at around 10 ppm) and $H^{\epsilon 1}$ (at around 7 ppm) protons are J-coupled and so will produce a pair of symentrically related crosspeaks in both TOCSY and COSY spectra. The $H^{\delta 1}$ and $H^{\epsilon 2}$ protons are in exchange with water, often broadening their resonance frequencies beyond definition.

Tryptophan

The $H^{N\epsilon 1}$ proton is often found at a high chemical shift, $\approx$10 ppm, making its assignment easier. It is usually possible to see a COSY (and therefore also TOCSY) crosspeak to the $H^{\delta 1}$ proton. Also, there should be a strong NOESY crosspeak to the $H^{\zeta 1}$ proton. This $H^{\zeta 1}$ proton is then part of the same spin system as the $H^{\eta 2}$, $H^{\zeta 3}$ and $H^{\epsilon 3}$ protons, forming the complex pattern shown in Figure 2.5.

2.5.4 Identifying Sequential Residues

Sequential assignment of polypeptides using homonuclear spectra is entirely reliant on short range NOE restraints. Once individual spin systems have been identified and assigned to degenerate groupings, NOE restraints between adjacent spin systems need to be found. The most common connections for spin system *(i)* to spin system *(i-1)* are summarized in Table 2.1:

Therefore, in the NOESY H^N strip of any given spin system there will be peaks not only belonging to that spin system *(i)* but also belonging to the previous spin system

(i)	*(i-1)*
H^N	H^α
H^N	H^N
H^N	H^β

Table 2.1: The most common NOE restraint connections between residue *(i)* to residue *(i-1)*.

(i-1). By overlaying the NOESY and TOCSY/COSY spectrum these peaks can be found (see Figure 2.6).

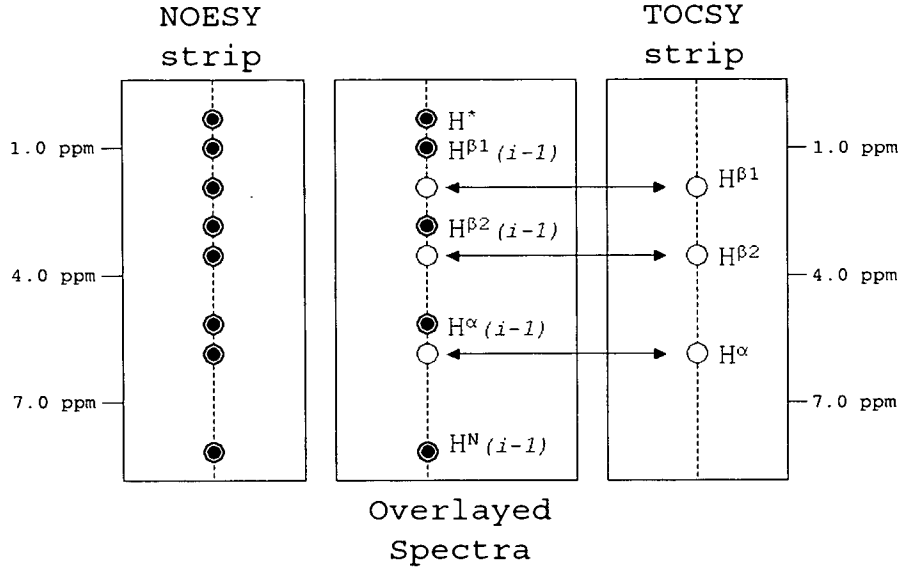


Figure 2.6: By overlaying H^N strips of spin system *(i)* from NOESY and TOCSY spectra it is possible to find NOESY peaks belonging to the previous spin system *(i-1)*. The peak marked H* belongs to a non-sequential residue and can be used as a distance restraint for structure calculations (see section 3.5).

Although it is possible to find sequential residues using this method, it is not always straightforward for a number of reasons. Proline residues, not having an H^N proton, can't be sequentially assigned in this way. Different spin systems may have similar H^α or H^β resonances so that a specific adjacent spin system can't be found. The larger the

polypeptide, the more likely this is to be a problem. Also, the H^N strip will not just contain NOE crosspeaks to the *(i-1)* residue but to other residues in close proximity to *(i)*. These restraints are the basis of structural determination by NMR but make sequential assignment harder.

However, with the spins systems labeled with degenerate markers, a group of 3 or more sequentially assigned spin systems may have a unique place in the primary sequence. For example, a group of three amino-acids, U-J-ALA, could well be unique within the primary sequence of a small polypeptide.

The larger the protein the less likely it is that small groups of spin systems with degenerate markers will be unique in the primary sequence. With labeled samples, other NMR experiments can be used to overcome some of the problems associated with the assignment strategy available to unlabeled samples.

2.6 Assignment of NMR Spectra in ^{15}N and ^{13}C labeled Protein Samples

The assignment of ^{15}N-labeled protein samples was carried out using the assignment program ANSIG [65], while the assignment of ^{15}N, ^{13}C-labeled protein samples was carried out using the assignment program Analysis [67][68] With samples labeled with ^{15}N and ^{13}C nuclei it is possible to carry out experiments looking at more then just protons. As mentioned previously, the most basic heteronuclear experiment is the HSQC (heteronuclear single quantum coherence) experiment an example of which is seen in Figure 2.3.

As mentioned above, with large proteins there can be a problem with overlapping resonances in 2D homonuclear spectra. With labeled samples it becomes possible to separate out the proton signals further using three dimensional spectra. So, for example, in an ^{15}N edited TOCSY spectrum the F2, F3 plane resembles a normal ^1H,^{15}N HSQC, with ^1H-^1H TOCSY transfer in the F1 dimension. Therefore, for each residue

the TOCSY transfer from each backbone H^N can now be assigned directly to the ^{15}N chemical shifts. Peaks that would normally overlap in the H^N region of homonuclear 2D TOCSY spectra are spread out in the F2 dimension. 2D planes can then be extracted from the 3D spectrum and analyzed separately (see Fig 2.7).

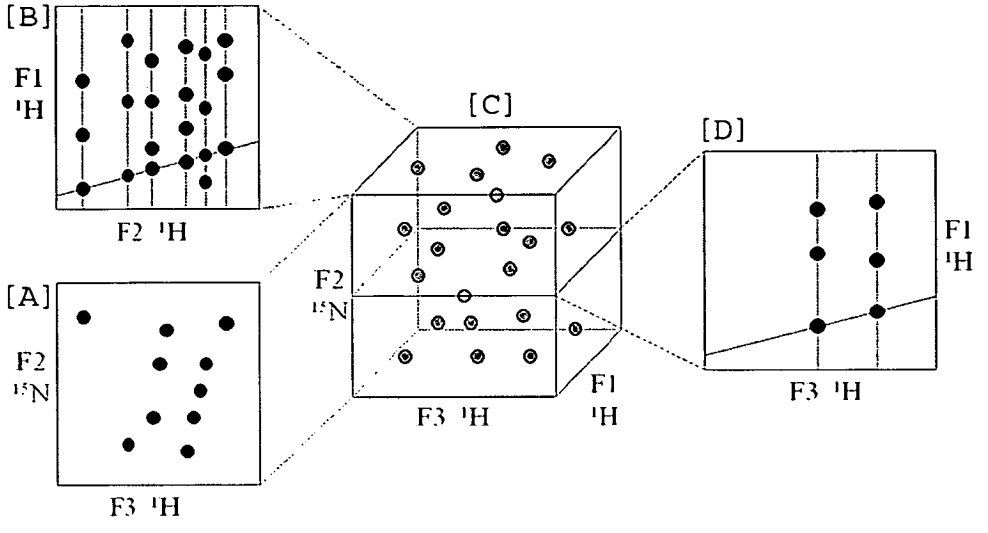


Figure 2.7: In this example of a 3D ^{15}N edited TOCSY spectrum [C], an ^1H,^{15}N HSQC [A] shows TOCSY transfer in the F1 dimension [B]. Individual ^1H-^1H planes can then be extracted from the 3D spectrum based on the ^{15}N chemical shifts [D].

For an ^{15}N labeled sample the assignment method is essentially the same as that for an unlabeled sample and also leads to the assignment of N^H nuclei. Each N^H crosspeak is assigned by studying its TOCSY and NOESY dimensions in ^{15}N edited TOCSY and ^{15}N edited NOESY spectra respectively. This is done in the same manner as H^N strips are assigned in homonuclear TOCSY and NOESY spectra, but with the added advantage of signal separation due to the ^{15}N chemical shifts.

2.6.1 Assignment of ^{13}C Labeled Protein Samples

Backbone Assignment: CBCANH and CBCA(CO)NH Spectra

In both these spectra, magnetisation is transfered from the C^{β} and C^{α} nuclei of a residue *(i - 1)* to the N-H group of the subsequent residue *(i)*. By taking an H^N strip of residue *(i)* in the ^{13}C-^{1}H plane it is possible to assign the C^{β} and C^{α} nuclei of residue *(i - 1)*.

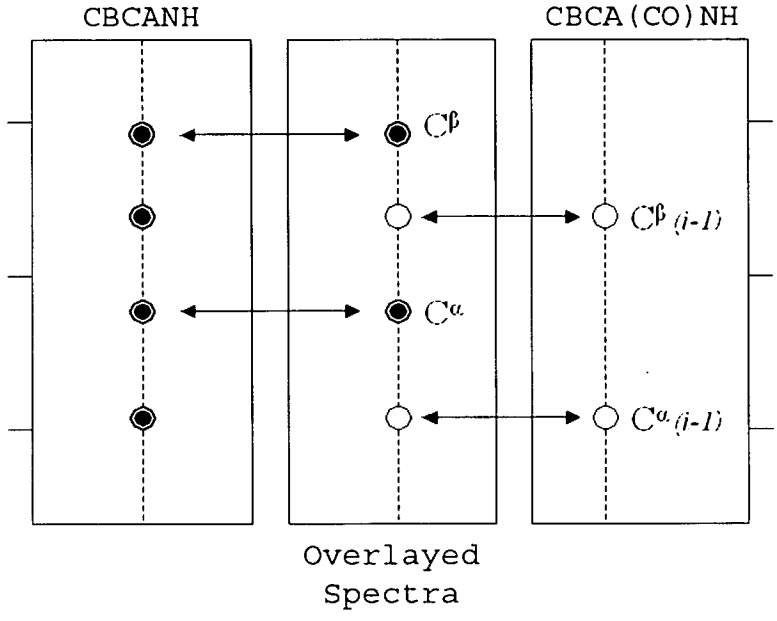


Figure 2.8: The use of CBCANH and CBCA(CO)NH spectra to assign backbone carbon nuclei. For any given H^N strip in the CBCA(CO)NH spectrum, the C^{β} and C^{α} nuclei of residue *(i -1)* can be assigned. By overlaying this strip with the complementary strip in the CBCANH spectrum it is possible to assign the C^{β} and C^{α} nuclei of residue *(i)*

However, in CBCANH spectra the magnetisation of the C^{β} and C^{α} nuclei of *(i)* is also transfered to it's own N-H group. As a consequence, by overlaying the ^{13}C-^{1}H planes of these spectra it is possible to assign the C^{β} and C^{α} nuclei of both residue *(i)* and *(i -1)* for any given plane (Fig 2.8). This process allows sequential walking along the protein's backbone and leads to their assignment.

Sidechain Assignment: H(C)(CO)NH-TOCSY, (H)C(CO)NH-TOCSY and HCCH-TOCSY Spectra

With the H(C)(CO)NH-TOCSY and the (H)C(CO)NH-TOCSY spectra, magnetisation moves along a residue's sidechain *(i -1)* to the N-H group of the subsequent residue *(i)*. In the H(C)(CO)NH-TOCSY, the sidechains protons magnetisation is labeled and so the F1 dimension is an ^{1}H dimension. With the (H)C(CO)NH-TOCSY spectrum it is the sidechain carbons chemical shift that become labeled, and so the F1 dimension is a ^{13}C dimension (see Figure 2.9).

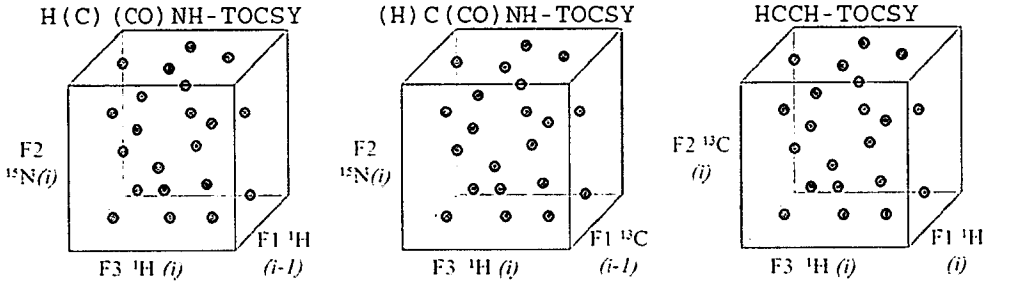


Figure 2.9: Schematic representations of H(C)(CO)NH-TOCSY, (H)C(CO)NH-TOCSY and HCCH-TOCSY spectra.

In the HCCH-TOCSY spectrum magnetisation is transfered from protons to their attached carbons, then through the carbon chain before being transfered back onto the attached protons for acquisition. This spectrum was used to verify ^{1}H and ^{13}C sidechain assignments made elsewhere.

Assignment of Aromatic Resonances

Four different experiments were acquired for the assignment of aromatic sidechains. The first two were ^{1}H-^{13}C HSQCs optimized for the aromatic region. These were used initially to deduce the carbon assignments of the aromatic groups using assignments already made for their attached protons.

The second two experiments, (HB)CB(CGCD)HD and (HB)CB(CGCDCE)HE could be used to directly link C^{β} and H^{β} assignments with those of aromatic side chains. In

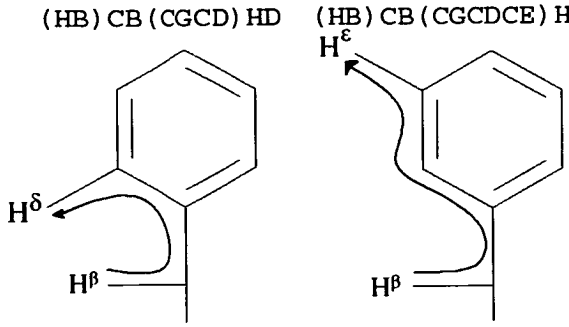


Figure 2.10: Schematic representations of the transfer of magnetisation through phenylalanine in (HB)CB(CGCD)HD and (HB)CB(CGCDCE)HE experiments.

the (HB)CB(CGCD)HD the magnetisation is transfered from the H^{β} protons to their attached carbon nuclei. The magnetisation then passes through the carbon chain of aromatic residues until it reaches the C^{δ} carbon. Magnetisation is then transfered to the H^{δ} proton for acquisition (see Figure 2.10). The resulting 2D spectrum links C^{β} resonances on one dimension with H^{δ} resonances on the other. Practically the same magnetisation transfer happens in the (HB)CB(CGCDCE)HE spectrum. Here, however, one more step is added and the magnetisaton is partially transfered to the C^{ϵ} and acquisition occurs on the H^{ϵ} proton. In this case the 2D spectrum links C^{β} resonances on one dimension with H^{δ}/H^{ϵ} resonances on the other.

Once the protein's resonances have been completely, or nearly completely, assigned, the NOESY specta can be analysed. This process starts with picking the peaks in all the NOESY spectra and assigning them to known resonances when possible. If the resonance assignment is not obvious, both the ANSIG [65] and Analysis [67][68] assignment programs can suggest possibilities based upon current assignments. Some of these possibilities can be eliminated upon closer inspection. Often there are a number of possible assignments for a given dimension of a peak. In this case the dimension is left unassigned and the peak assignment is ambiguous.

Chapter 3

Materials and Methods

3.1 Expression and Purification of fH~13 Constructs

Before a clone of module 13 became available Ursula Lodge (University of Edinburgh) had already produced a clone of the module pair 13-14. This module pair has an unusually long linker region (seven residues, SMAQIQL) and so would have made a good sample for structural analysis. Attempts to produce a sample of this protein were undone by what appeared to be protein degradation. This is described in more detail in Section 5.1 below.

Collaborator Claire Egan completed the cloning of fH module 13 and expressed and purified both the unlabeled and the ^{15}N-labeled samples. The construct was expressed in the yeast *Pichia pastoris* using the plasmid KM71 pPICzα. The fermentation method described below refers to the expression of the ^{13}C-^{15}N-labeled sample. An outline of the method of purification of fH~13 as developed by Claire Egan is given in section 3.1.2.

3.1.1 Fermentation

To prepare the inoculant for fermentation, the transformed yeast were streaked onto yeast peptone dextrose (YPD)-agar plates with added Zeocin to 100 ug/ml. The plates were incubated at 30 °C for approximately four days. A single large, isolated colony was then used to incubate 10 ml of BMG (the recipe for BMG is given in Table 3.1).

This was then incubated at 30 °C for approximately two days. The 10 ml of culture was then used to inoculate a further 200 ml of BMG. This was incubated for one day. It was then centrifuged at 1,500 rpm (390 g) for 10 min. The supernatant was removed and the pellet was then resuspended in ≈70 ml of 0.1 M potassium phosphate buffer, pH 6.0. This was to remove any unlabeled glycerol remaining from the BMG prior to fermentation.

1 M KPO_4 Buffer, pH 6.0	50 ml
10x Glycerol Solution	50 ml
Distilled Water	350 ml
Added after Autoclaving:	
Yeast Nitrogen Base	50 ml
500x Biotin	1 ml

Table 3.1: The recipe for BMG

Calcium Sulphate	$CaSO_4.2H_2O$	0.5625 g
Magnesium Sulphate	$MgSO_4.7H_2O$	9 g
Potassium Sulphate	K_2SO_4	6 g
Distilled Water		500 ml
1 M KPO_4 Buffer, pH 5.0		50 ml

Table 3.2: The recipe for the minimal media for $^{13}C^{15}N$-labelled growth (550 ml)

Before inoculating, the fermentation chamber needed to be autoclaved. Prior to autoclaving, 550 ml of minimal media was added to the fermentation chamber (BioFlo 3000) (the recipe for the minimal media is given in Table 3.2). Once autoclaved, PTM 1 post translational modification trace salts medium (4.35 $ml.l^{-1}$; 2.2 ml in this case, see Table 3.3) as well as antifoam (0.5 ml) were added to the chamber. The prepared inoculant was then added to the fermentation chamber. Also added were 8 g of ^{15}N-labeled ammonium sulphate dissolved in 20 ml of distilled water and 10 g of ^{13}C-labeled glucose dissolved in 25 ml of distilled water. The fermentation run was then started, with the pH regulated to pH 5.0 (the base feed was KOH). The yeast were then fed when the agitation was high and the dissolved oxygen was low, the deatails of which are given in Table 3.4

Concentration (mM)	Salt
24.0 mM	cupric sulfate ($CuSO_4.5H_2O$)
0.534 mM	sodium iodide (NaI)
17.8 mM	manganese sulfate ($MnSO_4.5H_2O$)
0.827 mM	sodium molybdate ($NaMoO_4.2H_2O$)
0.323 mM	boric acid (H_3BO_3)
2.1 mM	cobalt chloride ($CoCl_2.6H_2O$)
147.0 mM	zinc chloride ($ZnCl_2$)
234.0 mM	ferrous sulfate ($FeSO_4.7H_2O$)
1.64 mM	biotin
188.0 mM	sulfuric acid (H_2SO_4)

Table 3.3: Concentrations of salts in the PTM 1 trace salts medium.

Time After Inoculation	Food	Amount
≈17h	^{13}C-labeled Glucose	10 g in ≈75ml
≈21h	^{13}C-labeled Gycerol	1 g
≈24h	^{13}C-labeled Methanol	4.0 ml of 78%(v/v) solution
≈41h	"	"
≈48h	"	"
≈65h	"	"
≈72h	"	"
≈90h	"	"
≈97h	"	"

Table 3.4: The details of feeding for the yeast during fermentation

After fermentation the chamber was emptied and its contents centrifuged at 1,500 rpm (390 g) for 10 min at 4 °C, with a slow deceleration time to prevent cells from lysing. The supernatant was then poured off and to it was added PMSF (phenylmethylsulphonyl fluoride) and EDTA (ethylenedinitrilotetraacetic acid) to 0.5 mM and 5.0 mM respectively. In total, ≈700 ml of supernatant was collected and its OD_{280} reading was determined to be 1.897. A sample of the raw supernatant was run on SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis), the results of which are shown in Figure 3.2. (As the same protein molecular mass ladder was used to reference all the gels, the ladder and it's corresponding molecular masses are shown in Figure 3.1.)

3.1.2 Purification Protocol for fH~13

Claire Egan developed the method for the purification of fH~13 from the supernatant. First the raw supernatant was concentrated to approximately 50 ml through a 1 kDa

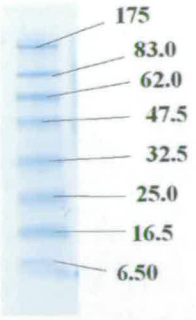


Figure 3.1: The protein marker New England Biolabs Broad Range Prestained Protein Marker P7708S, showing the molecular masses for the different bands (in kDa). The bands have the same masses in the P7708G mixture.

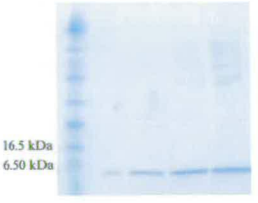


Figure 3.2: SDS-PAGE showing different volumes of the raw supernatant. The order of samples on the gel is (l-r) marker; 2μl of supernatant; 4 μl of supernatant; 8 μl of supernatant; 16 μl of supernatant. The protein marker used was the New England Biolabs Broad Range Prestained Protein Marker P7708S

cut-off filter.

The concentrated supernatant was then buffer-exchanged in to a 20 mM potassium phosphate buffer before being passed through a gravity flow chromatography SP-Sepharose column. The protein was eluted from this column using a 20 mM potassium phosphate buffer pH 5.0 with 1 M sodium chloride.

The elutant from this column was then buffer-exchanged into a 20 mM potassium phosphate buffer again to remove the salt. The solution was then put through a 1 ml porous heparin column (to which the protein bound) using fast protein liquid chromatography (FPLC). The protein was then eluted from the column using a gradient of 20 mM potassium phosphate buffer pH 5.0 with 1 M sodium chloride.

3.2 Aquired Spectra

3.2.1 Spectra Aquired on the KefC Derived Synthetic Polypeptides

The four polypeptides were individually dissolved in 10 mM deuterated sodium acetate, pH 3.0, with 10% D_2O and 0.01% Sodium Azide. Spectra were aquired on an 800 MHz NMR spectrometer at 298 K. Proton 1D spectra with water presaturation and proton 2D COSY, TOCSY and NOESY spectra were aquired for all four polypeptides using the parameters given in Table 3.5).

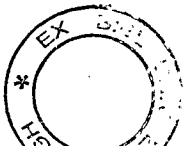
Spectrum	Dims	Number of complex points	Spectral widths (Hz)
NOESY	2	2048 x 1280	8802.8 x 6999.7
TOCSY	2	2048 x 1280	8802.8 x 7000.4
COSY	2	2048 x 1600	8802.8 x 700.4

Table 3.5: Parameters for the NMR spectra acquired on the four polypeptides. All three experiments were acquired on all four polypeptides using these parameters. The exception was the Hel polypeptide which had 4096 x 2048 complex points for each experiment.

The 2D spectra were processed in AZARA [69] (see section 3.4) and then imported into individual ANSIG [65] projects for the assignment of each polypeptide.

3.2.2 NMR Experiments on the Unlabelled and ^{15}N-labeled fH-13

Six NMR spectra were acquired at 288 K for the assignment of the ^{1}H and ^{15}N resonances of fH-13. These were acquired on a Bruker 600 MHz AVANCE spectrometer using a triple-resonance cryoprobe. The sample buffer was 20 mM sodium phosphate buffer at pH 6.0 with 10% D_2O and 0.01% sodium azide, with the expection of one homonuclear 2D NOESY experiment that was taken in 99.9% D_2O. The latter experiment was useful in uncovering NOESY crosspeaks that had been obscured in the H_2O sample by the water signal. These could then be used as additional distance restraints. Details of the experiments are given in Table 3.6.



Spectrum	Dims	Number of complex points	Spectral widths (Hz)
COSY	2	2048 x 1536	10162.6 x 6800.4
TOCSY	2	2048 x 1344	10162.6 x 6800.4
NOESY	2	2048 x 1010	10162.6 x 6800.4
NOESY in D_2O	2	2048 x 1024	7183.9 x 6599.6
^{15}N,^{1}H-HSQC	2	2048 x 256	10162.6 x 1191.6
^{15}N TOCSY-HSQC	3	2048 x 64 x 192	10162.6 x 1191.6 x 6898.6
^{15}N NOESY-HSQC	3	2048 x 72 x 256	10162.6 x 1191.6 x 6898.6

Table 3.6: Spectra acquired on fH-13 for the assignment of it's ^{1}H and ^{15}N resonances.

3.2.3 NMR Experiments on the ^{13}C,^{15}N-labeled fH-13

The homonuclear and ^{15}N-heteronuclear NMR spectra for fH-13 were already assigned before the ^{13}C, ^{15}N labeled sample became available. It was therefore only necessary to acquire the minimum set of ^{13}C spectra used for assignment. Assignment was made using the Analysis program from the CCPN suite [67][68] The experiments used for assignment are described in Section 2.6.1. These were acquired on a Bruker 600 MHz AVANCE spectrometer using a triple-resonance cryoprobe.

Details of the sample preparation is given in section 3.1. Details of the experiments is given in Table 3.7. Two ^{13}C-^{1}H HSQC spectra were taken of the aromatic region of the spectrum.

Spectrum	Nuclei labeled in F_3, F_2, and F_1	Number of complex points	Spectral widths (Hz)
^{13}C NOESY-HSQC	^{1}H;^{13}C;^{1}H	1024 x 48 x 144	10162.6 x 5838.6 x 7199.4
HCCH-TOCSY	^{1}H;^{13}C;^{1}H	1024 x 52 x 112	10162.6 x 5838.6 x 4559.4
CBCA(CO)NH	^{1}H;^{15}N;^{13}C	2048 x 54 x 128	10162.6 x 1191.4 x 11312.2
CBCANH	^{1}H;^{15}N;^{13}C	1024 x 54 x 112	10162.6 x 1191.6 x 11312.2
(H)C(CO)NH-TOCSY	^{1}H;^{15}N;^{13}C	1024 x 56 x 112	10162.6 x 1191.6 x 11315.4
H(C)(CO)NH-TOCSY	^{1}H;^{15}N;^{1}H	1024 x 56 x 144	10162.6 x 1191.6 x 4559.4
^{13}C-^{1}H HSQC	^{1}H;^{13}C	1024 x 80	7183.9 x 6035.0
^{13}C-^{1}H CT-HSQC	^{1}H;^{13}C	1024 x 184	7183.9 x 6035.0
(HB)CB(CGCD)HD	^{1}H;^{13}C	1024 x 64	8389.3 x 4071.7
(HB)CB(CGCDCE)HE	^{1}H;^{13}C	1024 x 64	8389.3 x 4071.7

Table 3.7: Spectra acquired on a double labeled sample of fH-13.

Often (in particular with larger proteins), the ambiguity of ^{13}C chemical shifts means that the assignments can't be completed based solely on one pair of triple-resonance experiments. However, because the complete assignment of the ^{1}H,^{15}N HSQC obtained from the analysis of ^{15}N-edited 3D spectra it was not necessary to acquire more triple resonance experiments for the backbone assignment. The (H)C(CO)NH-TOCSY experiment was used to assign much of the carbon chemical shifts in fH~13. The H(C)(CO)NH-TOCSY experiment was used to confirm the ^{1}H resonance assignments obtained previously from the analysis of the homonuclear and ^{15}N-edited spectra.

3.3 Relaxation Data

T_1 and T_2 rates of relaxation were measured for the backbone ^{15}N nuclei for each residue. Also, the heteronuclear NOE cross-relaxation rate was measured for the backbone NH for each residue. The relaxation data can give information on the mobility of individual residues on a nano-second timescale.

3.3.1 T_1 and T_2 Relaxation Experiments

The T_1 and T_2 relaxation time for the ^{15}N nuclei were measured by acquiring ^{1}H,^{15}N HSQC type spectra. These spectra have varying delays before acquisition, which allows relaxation to occur. The relaxation produces a drop in peak intensity as the relaxation delay increases, and so the measurement of the peak intensity in the different spectra enables the calculation of T_1 and T_2 for individual residues. The delays used are shown in Table 3.8. These spectra were acquired on a Bruker 600 MHz AVANCE spectrometer using a triple-resonance cryoprobe at 288 K.

3.3.2 Determination of T_1 and T_2

The method for calculating T_1 and T_2 relaxation times from the acquired spectra is the same. The intensity of each peak was measured as it is for the generation of NOE distance restraints (as described in section 3.5). Overlapping peaks can't be used as

Experiment	Number collected	Relaxation Delay (ms)
T_1	8	50 (x 2), 250, 350, 450, 650, 750, 800
T_2	8	10 (x 2), 20, 50, 70, 90, 100, 110

Table 3.8: The relaxation delays used in the acquisition of T_1 and T_2 relaxation spectra. The spectra with the shortest delay were acquire both before and after acquisition of the other spectra

their intensity can not be measured precisely, and residues whose signals were not present in the ^1H,^{15}N HSQC (such as proline) could also not have their T_1 and T_2 values determined.

The error in peak intensity also needs to be determined for each spectrum. This is done by first picking peaks in about twenty empty regions of the spectrum contain only noise. These regions were processed using the same integration boxes as were used for the regular peaks and their intensities were averaged. This produced an average noise intensity for each spectrum which could be used as the associated error for each peak in that spectrum.

To calculate the T_1 or T_2 values, a file is produced for each residue listing the ^1H,^{15}N HSQC peak intensity for each relevant spectrum together with its associated error, together with a sequence file. A program called `fitt_gauss` (written by Dr Kyrstyna Bromek-Burnside from our group) was then used to fit the data from each residue to the following exponential decay:

$$I(t) = I(0)xe^{-t/T_1} \tag{3.1}$$

Where $I(t)$ is the crosspeak intensity at the time t and $I(0)$ is the crosspeak intensity at time 0.

The program produces an output file listing the relaxation time (T_1 or T_2 depending on the input data) and an error value.

3.3.3 Heteronuclear NOE Relaxation Experiments

To determine the rate of heteronuclear NOE relaxation (HetNOE) between the backbone ^{15}N nuclei and their bonded protons two types of ^{1}H,^{15}N HSQC spectra were recorded. A saturated spectrum was recorded within which was a 5 s delay. During this delay there was 3 s of proton spin saturation, allowing an NOE to develop between the proton and its ^{15}N. The other spectrum contained a 5 s delay but no saturation. This is the reference spectrum. These spectra were acquired on a Bruker 600 MHz AVANCE spectrometer using a triple-resonance cryoprobe at 288 K.

The peaks in both spectra were then integrated and the HetNOE was calculated using the ratio:

$$\mathrm{NOE} = \frac{\mathrm{I}}{\mathrm{I(0)}} \tag{3.2}$$

Where I(0) is the reference intensity and I is the saturated intensity.

The ratio of the saturated intensity to reference intensity was thus calculated for each residue where it was possible. As with the determination of T_1 and T_2, residues with overlapping peaks or those that were unassigned could not have their HetNOE determined.

3.4 Processing Bruker Spectra

3.4.1 Scripts used for Processing Bruker Data

NMR experiments were process using AZARA software [69]. Processing involves Fourier transformation of the data and additional modifications to improve the quality of the final spectrum. In Table 3.9 is an example of a script used by AZARA to process the spectrum, in this case a ^{15}N-edited 3D NOESY spectrum. The input file for the script is a file detailing the basic parameters used in acquiring the spectrum (ser.par) and a link to the FID. When run by AZARA the script will output the Fourier transformed spectrum (spc) and a file that again details the basic parameters used (in this

case it would be spc.par).

Commands for processing each dimension of the spectrum are grouped in separate sections of the script (script_com 1, script_com 2, script_com 3). Their meaning is explained in detail in the AZARA manual.

In brief, in each dimension the data is multiplied by a sine-bell function (to remove sinc-wiggles due to truncation of the FID) prior to zerofilling and Fourier transformation. The spectrum is then phase corrected and only the real points are kept for displaying. A baseline correction is often applied to the directly detected dimension to 'smiling' baselines. Water removal can be done after the phase correction by reverse Fourier transformation of the data prior to its removal. This is a more effective method then applying the water deconvolution to the unphased data.

3.4.2 Contouring a Spectrum for ANSIG

The assignment of the ^{15}N-labeled sample was undertaken using the ANSIG software package [65]. For each 2D plane of a spectrum ANSIG requires a contour file to describe the topography of the plane. Therefore, a contour file must be created for each 2D spectrum. For a 3D spectrum a minimum of 2 contour files are required to fully describe the spectrum, one contour file describing a set of planes which are perpendicular to those described by the other contour file. For example, for an ^{15}N edited 3D NOESY spectrum one contour file would describe all the planes in the indirectly detected proton dimension. These planes would have as their axis the directly detected proton dimension and the ^{15}N dimension. Another contour file would describe all the planes in the ^{15}N dimension, whose axis were the two proton dimensions, while each of these planes would have a specific ^{15}N frequency. See Figure 2.7 for a description of a 3D ^{15}N-edited spectrum.

To create a contour file, an AZARA program called `contours` is used, and this in turn requires a script file. An example of a contour script is shown in Table 3.10. The contour script needs; a link to the parameter file (in this case spc.par, see above); the

```
input                ser.par
output               spc

interlace 2

script_com 1
    complex
    avance 12 16
    sinebell2 90
    zerofill 1
    fft
    avance_phase
    phase 182.3 0
end_script

script_com 1
    complex
    ifftn
    conv_box 32
    fftn
    fft
    reduce
    upper 1024
end_script

script_com 2
    mask_ppm
    complex
    sinebell2 90
    zerofill 1
    fft
    phase 90 -180
    reduce
end_script

script_com 3
    mask_ppm
    complex
    sinebell2 90
    zerofill 1
    fft
    phase 90 -180
    reduce
end_script

script_com 1
    base_poly 8 0
end_script
```

Table 3.9: AZARA processing script for Bruker NMR data. The script was used for processing ^{15}N-edited 3D NOESY data.

name of the output contour file; and a list of contour levels to be displayed for the plane (or for the set of planes if it is a 3D spectrum). The contour levels can be ascertained using the AZARA program. Both positive and negative contour levels are required for most spectra.

```
input                    ./spc.par
output                      ./cnt
ppm_range 1              5.0 11.3
dims 1 2

levels          135085    -135085
levels          189119    -189119
levels          264767    -264767
levels          370674    -370674
levels          518943    -518943
levels          726520    -726520
levels         1017128   -1017128
levels         1423979   -1423979
levels         1993570   -1993570
levels         2790998   -2790998
```

Table 3.10: An example of a script used by AZARA to create a contour file for ANSIG. This script was used to create a contour file for a ^{15}N,^{1}H-HSQC.

3.4.3 Residual Dipolar Couplings

The medium used for the aligned sample was a solution of cetylpyridinium bromide (CPBr)/hexanol/sodium bromide solution [70]. 1.5 ml of a 6.5% stock solution of the alignment media was prepared by mixing 0.0418 g of CPBr with buffer containing 25 mM sodium bromide; 20 mM sodium phosphate buffer at pH 6.0 and 4.5% hexanol by volume. This was used to test the splitting of the ^{2}H signal due to the alignment media prior to the addition of the protein. A 5.0% solution of the alignment media produced a ^{2}H signal splitting of 12 Hz.

Thus a sample of 0.24 mg of ^{15}N labeled fH~13 was freeze-dried and added to a 5.0% solution of the alignment medium. To reach the desired level of splitting, this sample was then diluted by added 88 μl of 20 mM sodium phosphate buffer at pH 6.0 to bring

the amount of alignment media to 4.1%. This gave splitting in the ^{2}H signal of 11 Hz in a 600 MHz magnet.

Residual dipolar couplings were then collected for the backbone N-H bonds only. To measure these, two IPAP-HSQC (In-Phase Anti-Phase-HSQC) experiments were recorded [71], one with the protein sample aligned and one in isotropic medium.

Spectrum	Dims	Number of complex points	Spectral widths (Hz)
Unaligned IPAP-HSQC	2	2048 x 1024	10162.6 x 1191.6
Aligned IPAP-HSQC	2	2048 x 328	10162.6 x 1191.6

Table 3.11: Spectra acquired on fH~13 for the determination of NH RDCs

3.5 Structure Calculation

Once the NOESY spectra have been assigned the crosspeak intensities need to be converted into distance restraints for structure calculation. The process is different for both ANSIG and Analysis data, but the principles are the same.

Distance Restraint Generation

Before the distance restraints were generated the NOESY peaks were first integrated. Set linewidths for the integration box were specified for each spectrum. These were chosen to give not the whole peak but a reasonable relative intensity for each peak. If linewidths were chosen to capture the whole peak in every case this would result in many peak intensities being larger than is actually the case. This is because many peaks overlap sightly (for example, $H^{\beta 1}$ and $H^{\beta 2}$ in the same spin system overlap in many cases). A more rigorous method would be an integration box of tailored size for each peak, but because of the large number of peaks this is impractical.

Peaks that could not be assigned to a specific nucleus were then used as ambiguous restraints. Before they were used, these restraints needed to be assigned to all the nuclei with the same chemical shift, based on a standard chemical shift list for

the protein in question. This list was generated automatically as part of the CCPN Analysis suite of programs, but required separate scripts for generation in ANSIG. Unambiguous assignments from all spectra (including NOESY spectra) were used to generate the average chemical shift table. At this point chemical shift assignments with large standard deviations were discovered and checked for errors in their assignment.

The peaks from each NOESY spectrum were then compared to the chemical shift list. For ANSIG data this was done using the AZARA program `connect`, requiring a connect script. Ambiguously assigned peaks were automatically assigned matching chemical shifts to their ambiguous dimensions using the chemical shift table. Any peaks for whom no match could be found were checked. A resonance not finding a match was usually the result of unassigned nuclei.

The intensity of a NOESY crosspeak is directly proportional to 'r^{-6}', where 'r' is the distance between the two nuclei that generated the crosspeak. However, in protein NMR structure calculations the NOE restraints are treated in a more qualitative manner. NOE restraints are put into separate distance classes based on their relative intensities. The distance classes used are shown in Table 3.12. The bands are quite broad in accord with established procedures. This allows more flexibility in structure calculation, as a greater number of possible structures would fit a given restraint. This is necessary to take into account spin effects and possible spin diffusion which would alter the intensity of the NOE peak. Therefore, a large number of these restraints was required to get an accurate structure.

Intensity	**Distance**
>3.0	≤2.7Å
1.5-3.0	≤3.3Å
0.3-1.5	≤5.0Å
<0.3	≤6.0Å

Table 3.12: Distance classes for the NOE distance restraints.

3.5.1 Structure Calculation

Structure calculation was carried out using CNS [72]. As well as the NOE restraint files, two other files were required as input for structure calculation; the force field and the MTF (molecular template file).

The force field used was an updated version of the PARALLHDG force field [73]. The force field describes the atoms present in each L-amino acid in a number of ways: their element; their charge; what other atoms they are bonded to and whether their bonds are part of dihedral or improper angles with other bonds. The force field also described the typical bond lengths and angles found in proteins; the changes that take place during disulphide bridge formation; and the strength of any van der Waals (non-bonding) interactions.

The force field in turn was used in conjunction with a sequence file (simply containing the primary sequence of fH~13) to generate the MTF. It is in the MTF that the disulphide bridges are defined (as confirmed by NOE data). The MTF file was used as an initial template for the structure calculation.

The structure calculation itself consisted of two main parts which are run in parallel; restrained molecular dynamics (RMD) and simulated annealing (SA). In SA, the protein is first heated to 2000 K, giving the atoms kinetic energy. The protein is then cooled, where the atoms explore possible conformations and the potential energy is measured for each atom. The atoms are then moved and their potential energy is measured again. A move is accepted if the potential energy is lower. The overall process is repeated until the potential energy doesn't drop.

In RMD, the NOE restraints are applied to a model of the structure. From here a potential energy function is calculated for each atom. The atoms are given mass and velocity (calculated from the kinetic energy used in SA). A force is then applied to each atom, based on its potential energy. After a given time the position of each atom is recorded. The process is then repeated iteratively so that a global (as opposed to

local) energy minimum is found.

During the structure calculations, prochiral swapping occurs. Protons which share the same carbon can be chirally distinct. However, they can't be differentiated by assignment alone. To counter this, their assignments are swapped during the calculation. The lower energy state after swapping is selected.

Three periods of simulated annealing were carried out in series, with each set of structures generated from one period being used as the templates for the subsequent period. After this the structures needed to be analysed and the NOE restraints filtered and checked prior to another round of structure calculation.

3.5.2 Analysing the Calculated Structures

After a single round of structure calculation the ambiguous NOE restraints were further refined to improve the structure. Refinement was done based on the most accurate structures from the last round of calculations. To ascertain a structure's accuracy a number of energy terms were used to describe the structure. This is based on the principle that the higher the potential energy of a structure the less likely it is to be correct. The total energy of a structure is calculated as a sum of all the energy terms:

$$Etotal = \sum(E_{bond} + E_{angle} + E_{dihe} + E_{impr} + E_{vdw} + E_{noe}) \tag{3.3}$$

Where, E_{bond} is the bond energy; E_{angle} is the energy of the bond angles; E_{dihe} is the energy of the dihedral angles; E_{impr} is the energy of the improper angles; E_{vdw} is the energy of the van der Waals interactions; and E_{noe} is the energy of the NOE interactions. For refinement with RDC data there is also an E_{tenso} term for the energy of the RDC restraints (see section 3.5.4).

The NOE energy was calculated by comparing the original distance restraint to the actual distance within the structure, with an energy penalty for deviations. The equation

for calculating this is:

$$E_{noe} = \sum_k \begin{cases} (L_k - D_k)^2 & \text{if } D_k < L_k \\ 0 & \text{if } L_k \leq D_k \leq U_k \\ (D_k - U_k)^2 & \text{if } U_k < D_k \end{cases} \quad (3.4)$$

Where U and L are the upper and lower bounds respectively from the original distance restraint; and D is the actual distance between the nuclei in the calculated structure that the restraint describes [74].

If the distance from the original restraint was more then 0.3 Å greater than the upper distance band then the restraint was said to be violated in that structure. Violations from each structure were recorded in a separate file. If a restraint was violating in many structures it was examined to see what the problem was. Often the restraint was incorrectly assigned, and either changing the assignment or making the peak ambiguous in one or more dimensions would solve the problem. Overlapping of peaks could often lead to incorrect intensities and violating restraints. For example, small peaks that lay close to the diagonal could give incorrectly high intensities as their integration box would overlap with the large diagonal peaks. The restraint would then record the distance between the nuclei to be closer than it actually is. These NOE restraints would have to be adjusted or removed.

The ensemble of calculated structures should converge on a low energy for both E_{total} and E_{noe} where >99% of the NOE restraints are satisfied for a number of structures. On a plot of E_{total} or E_{noe} for all the calculated structures convergence can be seen as a plateau of low energy structures (Figure 6.8). It is these converged structures that are used to filter the NOE restraints for the next round of calculations.

3.5.3 Filtering and Checking the Structures Using ARIA

Filtering and checking of the NOE restraints was undertaken by the ARIA suite of programs. An ambiguous NOE restraint can have many matching assignments attached to its ambiguous dimension. Not all of these matching assignments may be correct. In the filtering step the NOE restraints are examined against the converged structures.

The contribution of each matching assignment to the intensity of an NOE peak is calculated. If assignments contributed less then a specified amount, their contribution was removed. The initial cutoff for the contribution was set to 0.99 in the filtering script, which means that if an assignment contributed less then 1% to the intensity of a crosspeak its contribution was removed.

After filtering, the restraints are checked. Here, any diagonal peaks that were picked in the NOE spectrum are removed as they do not represent interactions between nuclei. Also, if there are duplicate restraints the duplicates are removed. The highest intensity peak is always kept, while the lower intensity multiples are removed.

The new set of NOE restraints are used in a new round of structure calculations. The process of calculation, analysis, filtering and checking is repeated until the calculated structures between one round and the next barely change. The threshold for contributing assignments during filtering can be raised, from for example 1% for the first two rounds to 2% for the next two roundsd.

3.5.4 Refinement of the Calculated Structures

The final calculated structures can be further refined using the RDC restraints and also by recalculation in water.

The RDC restraints are used to refine the final structures in a single structure calculation step using the TENSO module for CNS [75]. Here simulated annealing takes place, with the introduction of a Tenso force constant for each atom. The Tenso force constant is increased geometrically from 0.001 kcal.(mol.Hz^2)$^{-1}$ to 0.6 during the first cooling step and fixed throughout subsequent cooling steps. After calculation the structures were analysed as before, except now there is an E_{tenso} term for the energy of the RDC restraints.

The final set of structures can be further refined by recalculating the structures in explicit solvent. This has been shown to improve the quality of the final structures,

helping to orientate the amino acid side chains and improve the statistics (such as the Ramachandran statistics) of the final ensemble [76][77]. A maximum of 50 structures with the lowest E_{total} from the RDC refinement could be selected at a time for water refinement. This is because water refinement in CNS can only be carried out on a maximum of 50 structures.

3.6 Chemical Shift Mapping of fH~13 with Heparin

For the titration at pH 6.0 two stock solutions were prepared. The first consisted of 0.777 ml of a 0.5 M stock solution of the heparin tetrasaccharide in distilled water with 0.02 % (w/v) sodium azide (an explaination of the heparin tetrasaccharide used is given in section 7.1). The second solution consisted of 1.00 ml of a 20 μM of ^{15}N-labeled fH~13 in 20 mM sodium phosphate buffer at pH 6.0, with 10 % (v/v) D_2O and 0.05 % (w/v)sodium azide. The stock solution of fH~13 was the source of an initial NMR sample that was used to optimize conditions and to obtain NMR data for fH~13 in the absence of the tetrasaccharide.

Ratio of fH~13 to Tetrasaccharide	Concentration of fH~13 (μM)	Concentration of tetrasaccharide (μM)
1:0	20	0
1:21.7	10	217
1:10	20	200
1:7.5	"	150
1:5	"	100
1:4	"	80
1:3	"	60
1:2	"	40
1:1	"	20
1:0.65	"	13

Table 3.13: Samples prepared for the titration between fH~13 and a heparan sulphate derived tetrasaccharide at pH 6.0.

Subsequently ^{1}H, ^{15}N HSQC experiments were acquired for nine different ratios of protein to tetrasaccharide (Table 3.13). For the first titration point, 140 μl of the first solution was mixed with: 14 μl of a 0.250 mM NMR sample of ^{15}N labeled fH~13; 160 μl of buffer (40 mM sodium phosphate buffer at pH 6.0, with 10 % (v/v) D_2O and 0.1

% (w/v) sodium azide); and finally 20 μl of D_2O to give a ratio of 1:21.7 of fH~13 to tetrasaccharide. For the second titration point, 14 μl of the 0.250 mM sample of fH~13 was then added to the 1:21.7 sample to make the 1:10 sample. Subsequent samples were made by adding aliquots of the 20 μM stock solution of ^{15}N labeled fH~13 to the previous sample, thus incrementally diluting the tetrasaccharide whilst maintaining the concentration of the protein.

Titrations between other CCP modules and heparin-derived oligosaccharides previously carried out in our group had shown improved binding at lower pH. The titration between fH~13 and the heparin-sulphate derived tetrasaccharide was therefore repeated at pH 5.5. Stock solutions of tetrasaccharide and fH~13 in 20 mM sodium phosphate buffer at pH 5.5 were prepared as before. ^{1}H, ^{15}N HSQC spectra were acquired at five titration points as shown in Table 3.14.

Ratio of fH~13 to Tetrasaccharide	Concentration of fH~13 (μM)	Concentration of tetrasaccharide (μM)
1:0	18.8	0
1:1	18.8	18.8
1:3.3	16.8	43.5
1:11	7.67	86.4
1:22	4.9	108

Table 3.14: Samples prepared for the titration between fH~13 and a heparin derived tetrasaccharide at pH 5.5.

3.7 Structure Analysis Method

This details the method for the structure analysis carried out on the closest to mean structure of fH~13 that was generated using the ^{13}C,^{15}N-labeled sample. This was carried out in collaboration with Dinesh Soares.

3.7.1 Structural Comparisons and Superposition

Combinatorial Extension (CE) was employed to compare all 34 experimentally-determined CCP module structures, (including both GABA conformers [78]) against the closest-to-mean structure of fH~13. CE is a method for the calculation of pairwise structure

alignments using characteristics of their local geometry as defined by vectors between C^{α}-positions [79].

The individual domain boundaries for each CCP module were considered from one residue before the first cysteine untill three residues after the fourth cysteine (both inclusive, where available). In cases where structures were solved by both NMR and X-ray diffraction, the highest resolution X-ray structure was used for comparison. Where both liganded and unliganded structures were available, the highest resolved unliganded X-ray structure was used. A few residues were missing in the crystal structure of C1r-2. Therefore, in this case the structure with the most determined residues was employed i.e. Protein Data Bank (PDB) [80] ID: 1GPZ (for both modules) [81]. The two interleukin-2 receptor alpha (IL2Rα) CCP modules, which exhibit atypical "strand-swapped" individual CCP module structures [82][83][84], were excluded from this analysis.

Superposition (overlay) of all 35 individual CCP-modules was undertaken using the multiple protein structure alignment program server, MAMMOTH-multi [85].

Surface Analysis

The solvent accessible surface area (ASA) for fH-13 was calculated using the GETAREA version 1.1 server [86][87] using default settings. Therefore the radius of water probe in Angstroms was 1.4 Å; the Select output level was equal to only the total area). ASA is defined as the surface mapped out by the centre of a probe sphere, of radius 1.4 Å, as if it were rolled around the van der Waals surface of the protein. The electrostatic surface representations were computed and displayed using GRASP [88]. The lipophilic surface renditions were created using MOLCAD [89] under SYBYL version 6.9 (Tripos Associates, MO, USA). Electrostatic and lipophilic surface potential scales are shown in Figure 6.21.

Chapter 4

Structural Assessment of KefC Derived Synthetic Polypeptides

4.1 Project Aims

Four synthetic polypeptides were supplied to our lab by Prof Ian Booth from the University of Aberdeen. The aim of this project was to use NMR spectroscopy to determine whether any secondary structural elements could be found within each of the four polypeptides. This was achieved by the acquisition of homonuclear 2D COSY, TOCSY and NOESY spectra on each of four synthetic polypepides. Proton resonances were assigned by working with these spectra.

4.2 The Four Synthetic Polypeptides

The four polypeptides that were investigated are all based on a sequence of residues found in the KefC potassium efflux protein system. This membrane protein is found in *Escherichia coli* cells where it protects the cell from toxic concentrations of electrophiles by removing potassium from the cell. The sequences for the four polypeptides used in this project are found in Table 4.1 They are all based around a strongly conserved region of the protein, residues 252-271, with the sequence LASSEYRHALESDIEPFGL. This region is found in the amino-terminal hydrophobic domain between two putative transmembrane helices. Thus Hal20 represents the native sequence, while Hal22 was supposed to be an elongated version of Hal20, but the sequence was shown to be erro-

neous upon NMR assignment.

Label	Amino Acid Sequence
Hal22	LAASSEYR**HAL**ESDIEPFKGLL
Hal20	LASSEYR**HAL**ESDIEPFKGL
Hel	LASSEYR**HEL**ETAIDPFKGL
Dik	LASSEYRHALES**DIK**PFKGL

Table 4.1: The sequence of the four polypeptides analysed. The names of polypeptides were derived from their amino acid sequences, as shown in bold.

Within this sequence, the region HELE*X*DIEPFK (where *X* is either alanine, serine or threonine) has been found to be important for the regulatory function of the protein. Mutagenesis analysis showed that mutations E262K (found in Dik) and D264A (found in Hel) resulted in spontaneous activation of the KefC channel. This resulted in a loss of K^+ from the bacterial cells. After suspension of the cells in K^+ media, the native cells were observed to have a K^+ concentration of 430 K^+g^{-1}, while for the D264A and E262K mutants these were 310 μmol K^+/g of dry cells and 274 μmol K^+/g of dry cells respectively [90].

Also important for regulation of the KefC ion channel is a carboxy-terminal hydrophobic domain. This is a KTN (K^+ transport, nucleotide binding) domain, with a highly conserved structure. These are found in a variety of prokaryotic and eukaryotic K^+ ion channels. In the KefC channel a tetramer of these KTN domains assemble near the cytoplasmic vestibule of the ion pore. As with the HELE*X*DIEPFK loop, a number of mutations in the KTN domain result in spontaneous activation of the KefC channel (R416S, V427A and R516C). Futhermore, introduction of a R527E mutations alone reduces the K^+ effux from the bacterial cells. However, the addition of E262R and D264R mutations alongside the R527E mutation restores the protein's activity [91]. This suggests a direct interaction between E262 and D264 with R527 which mediates the channel's activation. It was thus hypothesised that the HELE*X*DIEPFK loop and the KTN domain interact to regulate the protein through salt bridge connections. An X-ray crystalography structure structure of the KTN domain exists already [91]. A structure of the HELE*X*DIEPFK loop could be used to model the binding of these two regulatory features.

4.3 Resonance Assignment

1D proton spectra of each polypeptide are shown in Figure 4.1. These revealed impurities in each sample which can be more easily seen in Figure 4.2. Although the impurities did not affect the assignment, they did impede the search for non-sequential NOE crosspeaks, as explained below. The 1D spectra were also very similar. Indeed, the chemical shifts for equivalent residues were generally within 0.05 ppm of each other between the polypeptides, with the exeption of the proline and the N-terminal residues.

Assignment was carried out according to the method described in section 2.5. The assignment of each polypeptide was carried out to completion, with a few exceptions described below. Because the chemical shifts for equivalent residues were so similar assignment was relatively easy once one of the polypeptides had been fully assigned. An assigned amide region of a 1D spectrum for each polypeptide, showing the positions of each backbone H^N chemical shift, is shown in Figure 4.2.

All the serine and tyrosine residues were fully assigned apart from their OH groups (the H^{γ} for serine and the $H^{\gamma 1}$ for threonine) which are rarely assigned. All the lysine residues were missing their H^{ζ} nuclei chemical sifts. For all the arginines the guanidinium resonances were unassigned. Finally, the histidines H259 were all missing their $H^{\delta 1}$ and $H^{\epsilon 2}$ assignments.

4.4 Analysis of the NMR Data

After the four polypeptides had been assigned, the NMR data was analysed in a search for evidence of secondary structure elements within them. This was done in three different ways: analysis of the chemical shifts; analysis of $J_{H^N H^{\alpha}}$ coupling constants; and examination of 2D NOESY spectra.

The ramdom coil values were taken from a table in 'NMR of Proteins and Nucleic Acids', page 17 [92]. These values were recorded using a tetrapeptide GGXA, where X was the residue whose shifts were to be measured. The samples of tetrapeptide were

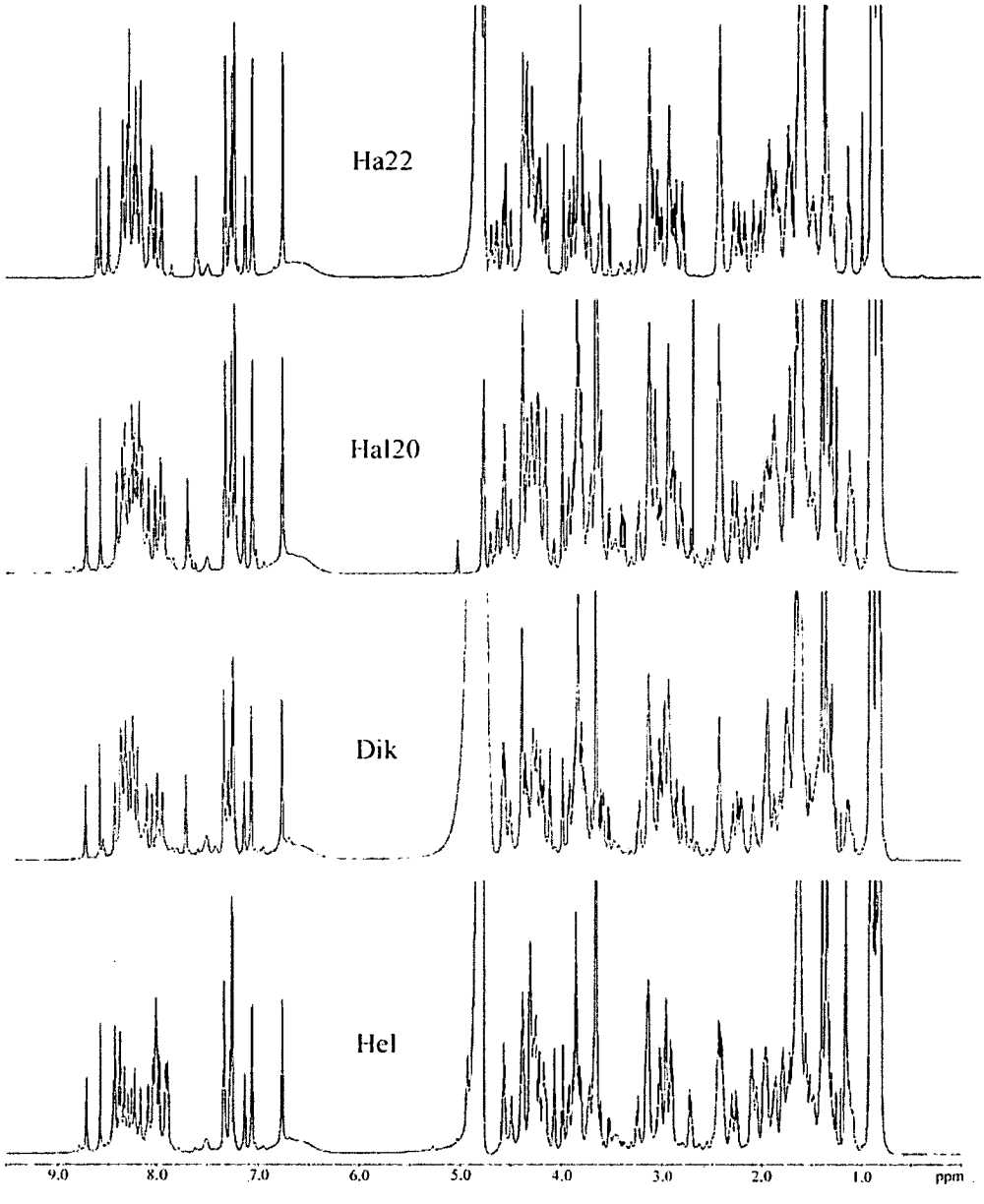
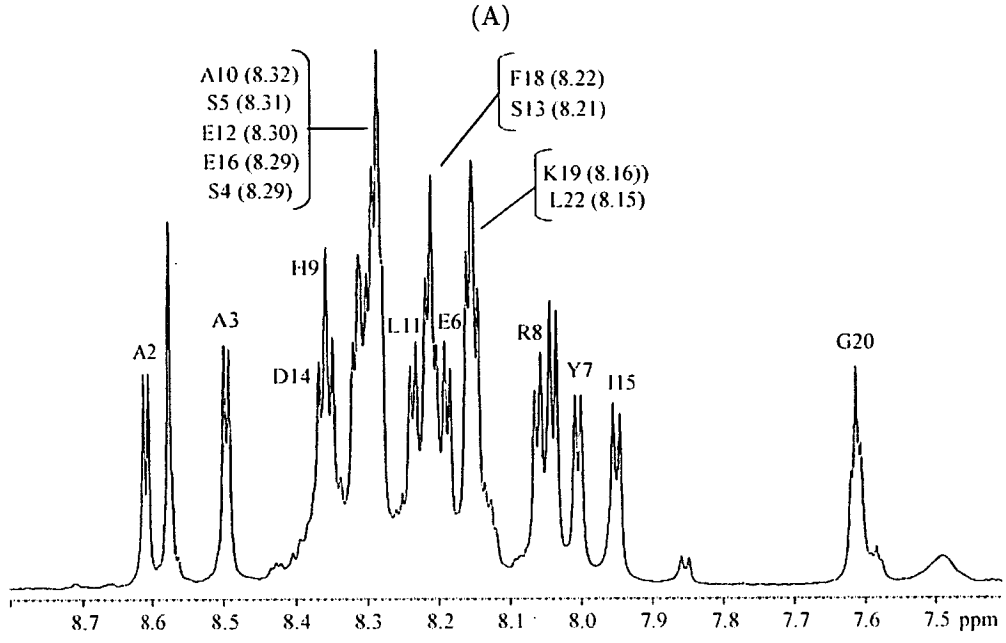
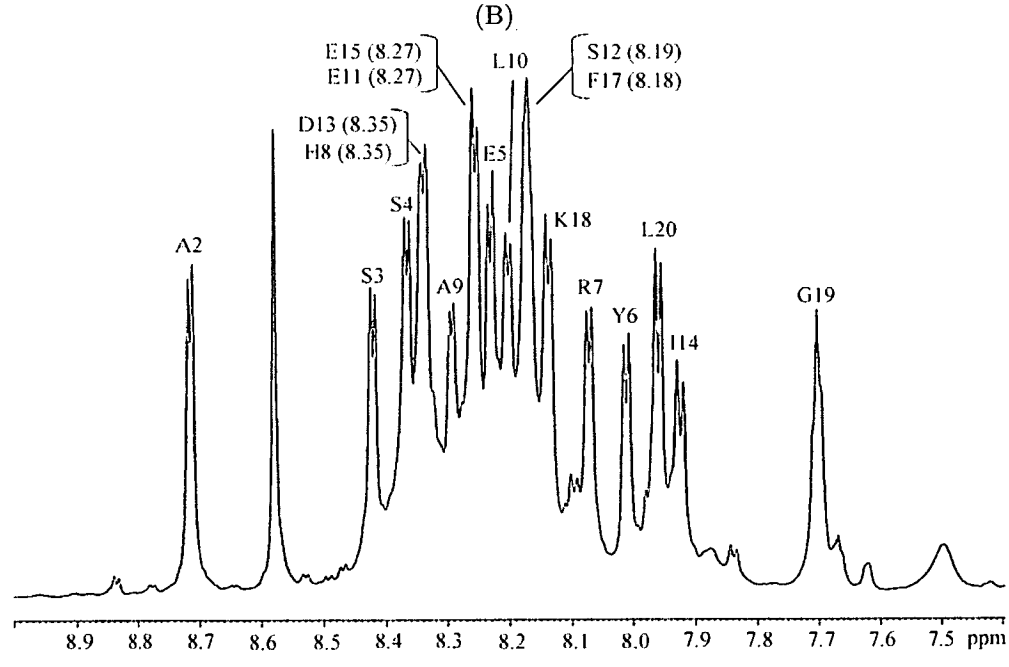


Figure 4.1: Comparison of the 800 MHz 1D proton spectra from each of the four polypeptides.


(A)
A10 (8.32)
S5 (8.31)
E12 (8.30)
E16 (8.29)
S4 (8.29)
F18 (8.22)
S13 (8.21)
K19 (8.16))
L22 (8.15)
A2
A3
H9
D14
L11
E6
R8
Y7
I15
G20
8.7 8.6 8.5 8.4 8.3 8.2 8.1 8.0 7.9 7.8 7.7 7.6 7.5 ppm


(B)
E15 (8.27)
E11 (8.27)
L10
S12 (8.19)
F17 (8.18)
D13 (8.35)
H8 (8.35)
E5
S4
K18
L20
A2
S3
A9
R7
Y6
I14
G19
8.9 8.8 8.7 8.6 8.5 8.4 8.3 8.2 8.1 8.0 7.9 7.8 7.7 7.6 7.5 ppm

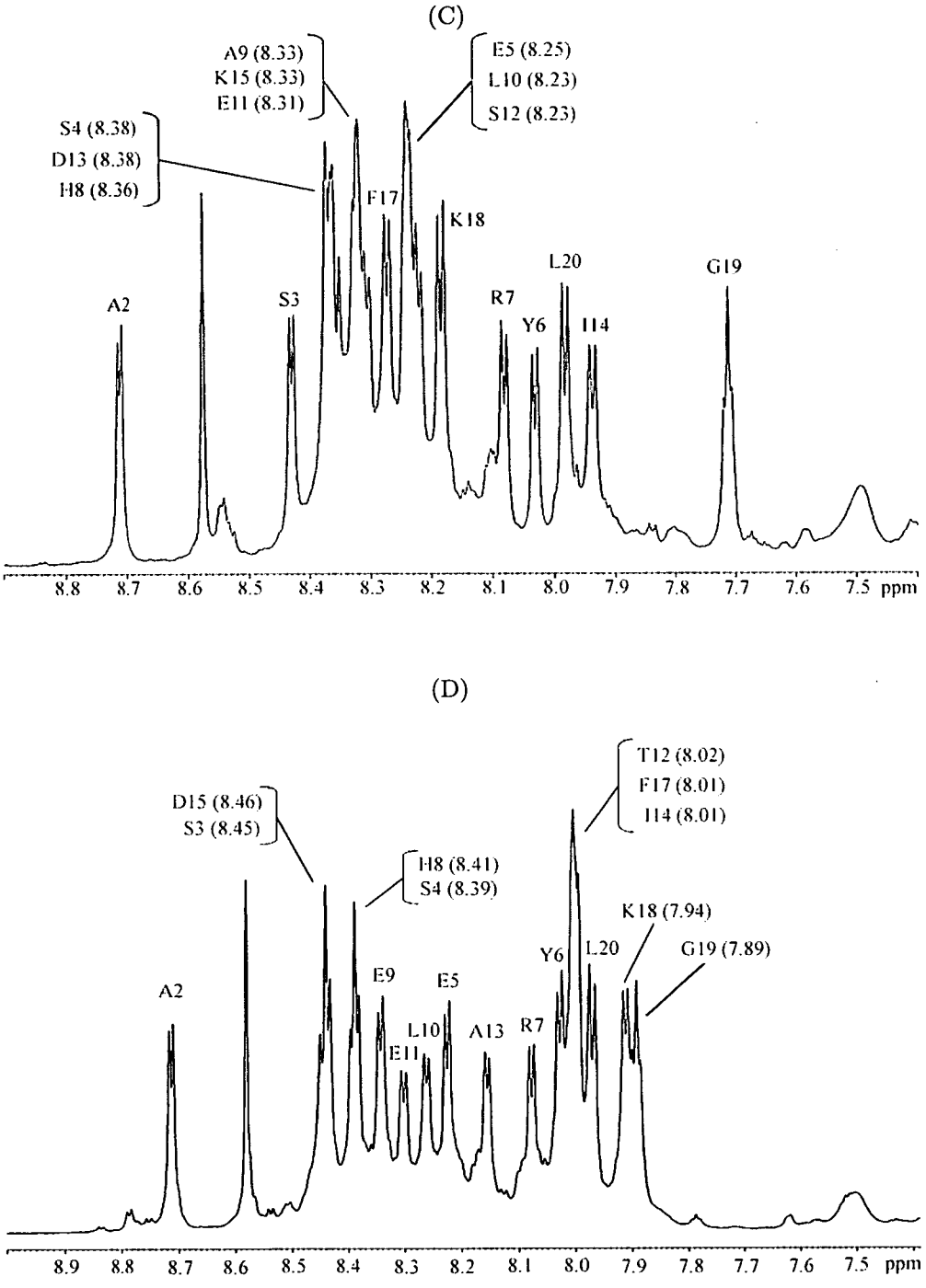


Figure 4.2: Comparison of the amide region of 1D proton spectra from each of the polypeptides, showing the assignments for the backbone H^N resonances. Impurities of $\leq$ 10% of the main signals can be seen in all the samples. (A) Hal22; (B) Hal20; (C) Dik; (D) Hel.

recorded at pH 7.0, 37 °C. The results of the comparison are shown in Figure 4.4. The backbone amide chemical shifts were not used for comparison as these are strongly affected by pH differences.

The 1D spectra collected for each sample were of sufficient quality to see splitting of the backbone H^N peaks. This splitting is due to J-coupling with the backbone H^α nuclei. The magnitude of the $J_{H^N H^\alpha}$ coupling is related to the ϕ dihedral angle of the peptide bond. In α-helices and β-sheets the dihedral angles adopt certain orientations producing characteristic $J_{H^N H^\alpha}$ coupling values, as summarized in Table 4.2.

Secondary Structure	ϕ/**degrees**	$^3J_{HnH\alpha}$
Antiparrallel β-sheet	-139	8.9
Parrallel β-sheet	-119	9.7
α-helix	-57	3.9

Table 4.2: Parameters for regular secondary structures of proteins. Adapted from NMR of Macromolecules, p.68 [66]

Therefore the $J_{H^N H^\alpha}$ coupling constants were measured for those residues which were free of overlap. Glycine residues, where the extra H^α produces additional splitting, were excluded. The results are summaries in Table 4.3.

The NOESY spectra were also used to search for evidence of secondary or tertiary structure in the polypeptides (Fig 4.3. If two resonances in non-sequential residues produced an NOESY crosspeak this would mean that the residues were within 6 Å of each other. This would not be the case if the residues were part of an unstructured section of the polypeptide. Therefore the NOESY spectrum for each polypeptide was examined to check for non-sequential NOE crosspeaks. The impurities present in each sample complicated this analysis, but careful examination did not yield any non-sequential NOESY crosspeaks.

The chemical shifts were also compared between the four polypeptides to see if the differences between the native and mutant sequences could lead to differences in any

Residue	$J_{H^N H^\alpha}$ **(Hz)**	**Residue**	$J_{H^N H^\alpha}$ **(Hz)**	**Residue**	$J_{H^N H^\alpha}$ **(Hz)**	**Residue**	$J_{H^N H^\alpha}$ **(Hz)**
Hal22		Hal20		Dik		Hel	
252Ala	6.6	-	-	-	-	-	-
253Ala	5.1	253Ala	5.1	253Ala	5.1	253Ala	5.3
254Ser	-	254Ser	5.9	254Ser	5.9	254Ser	6.3
255Ser	-	255Ser	5.9	255Ser	-	255Ser	6.5
256Glu	6.2	256Glu	6.2	256Glu	-	256Glu	6.5
257Tyr	7.3	257Tyr	6.2	257Tyr	7.0	257Tyr	6.7
258Arg	7.0	258Arg	6.2	258Arg	7.0	258Arg	6.2
259His	-	259His	-	259His	-	259His	-
260Ala	-	260Ala	6.2	260Ala	-	260Glu	6.2
261Leu	6.2	261Leu	6.2	261Leu	-	261Leu	6.2
262Glu	-	262Glu	-	262Glu	7.3	262Glu	6.7
263Ser	-	263Ser	-	263Ser	6.2	263Ser	-
264Asp	-	264Asp	-	264Asp	-	264Ala	5.9
265Ile	6.6	265Ile	8.1	265Ile	8.1	265Ile	-
266Glu	-	266Glu	-	266Ala	-	266Asp	7.5
268Phe	-	268Phe	-	268Phe	6.2	18Phe	-
269Lys	-	269Lys	7.0	269Lys	7.7	269Lys	6.7
271Leu	-	271Leu	7.7	271Leu	8.1	271Leu	7.5
272Leu	-	-	-	-	-	-	-

Table 4.3: $J_{H^N H^\alpha}$ for each polypeptide as measured from 1D proton spectra were possible. The first residue for each polypeptide, as well as the proline and glycine residues, are not included as their couplings could not be measured.

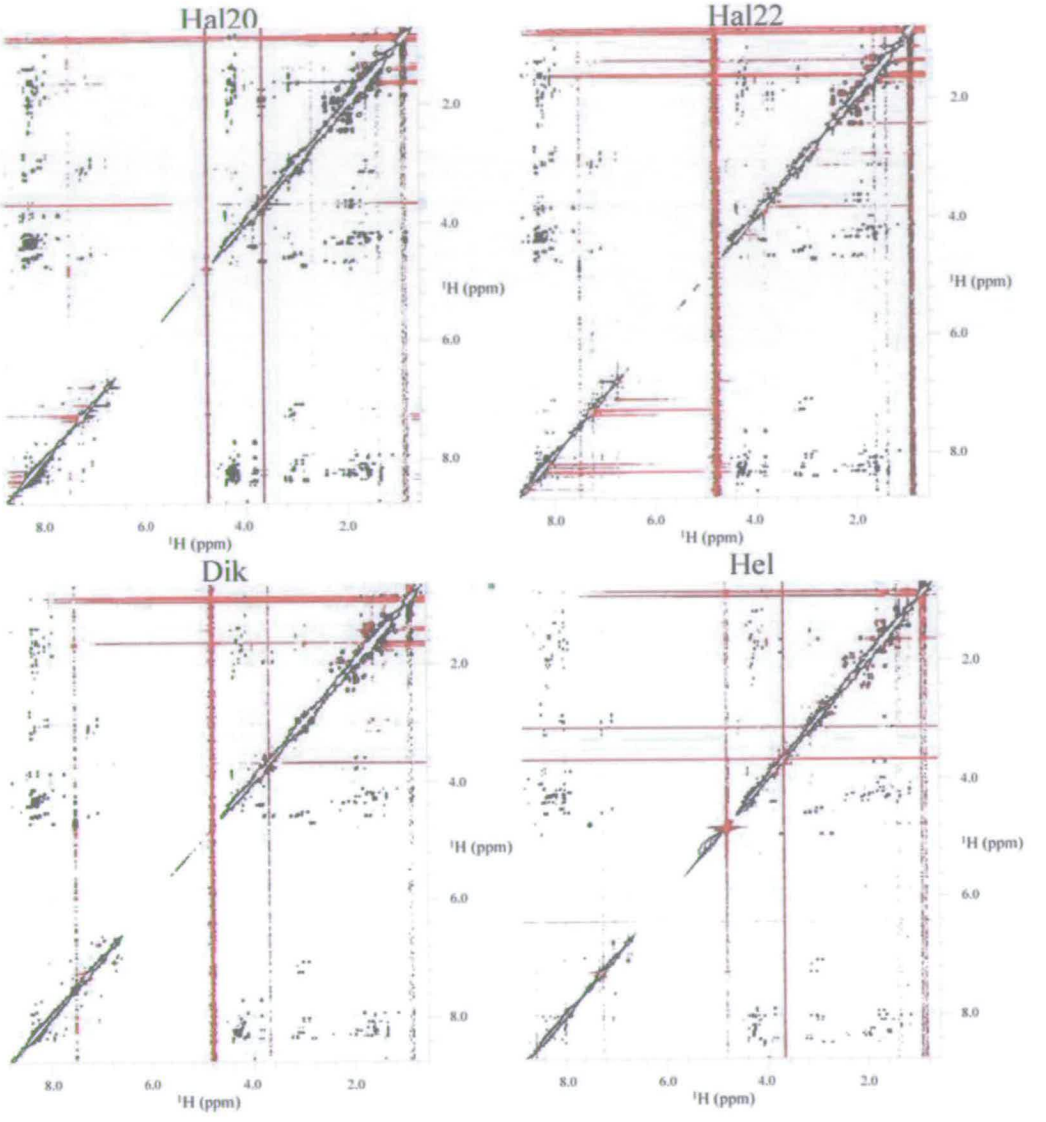
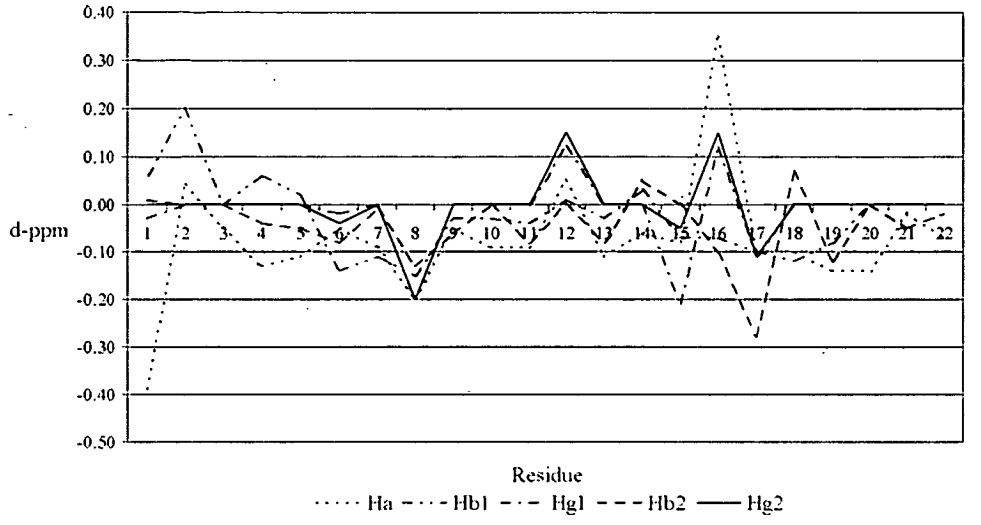
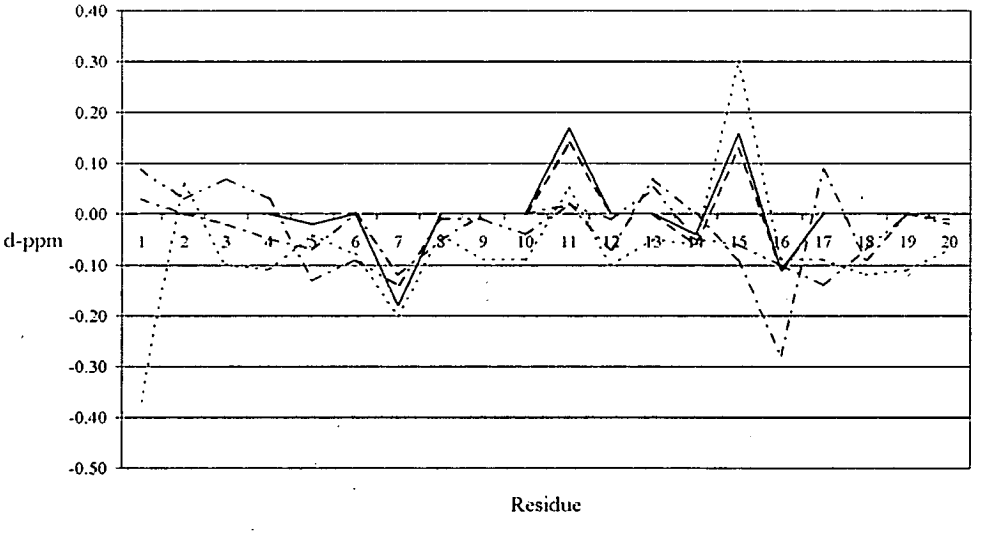


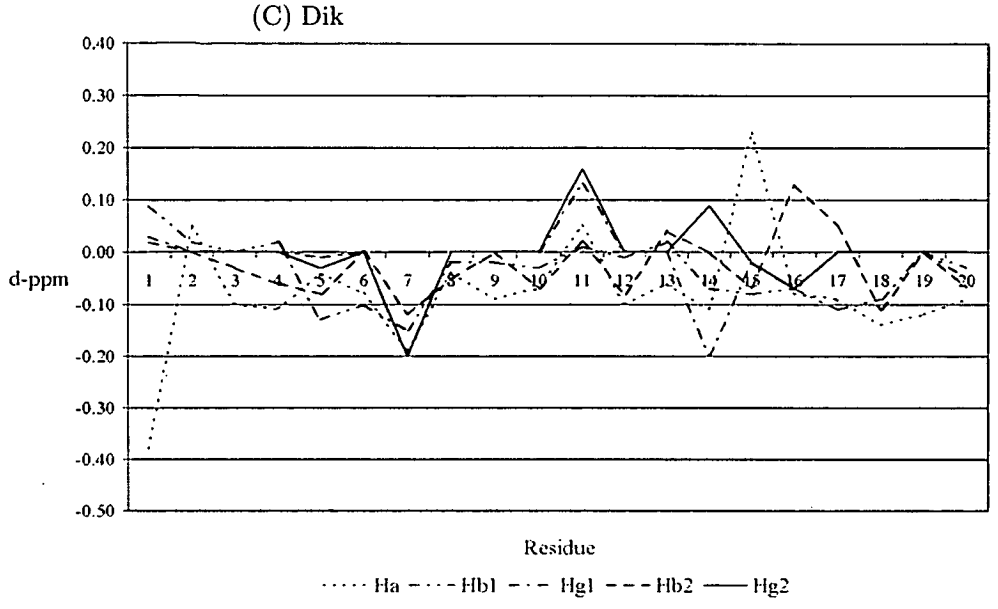
Figure 4.3: NOESY spectra for all four polypeptides.

(A) Hal22



(B) Hal20





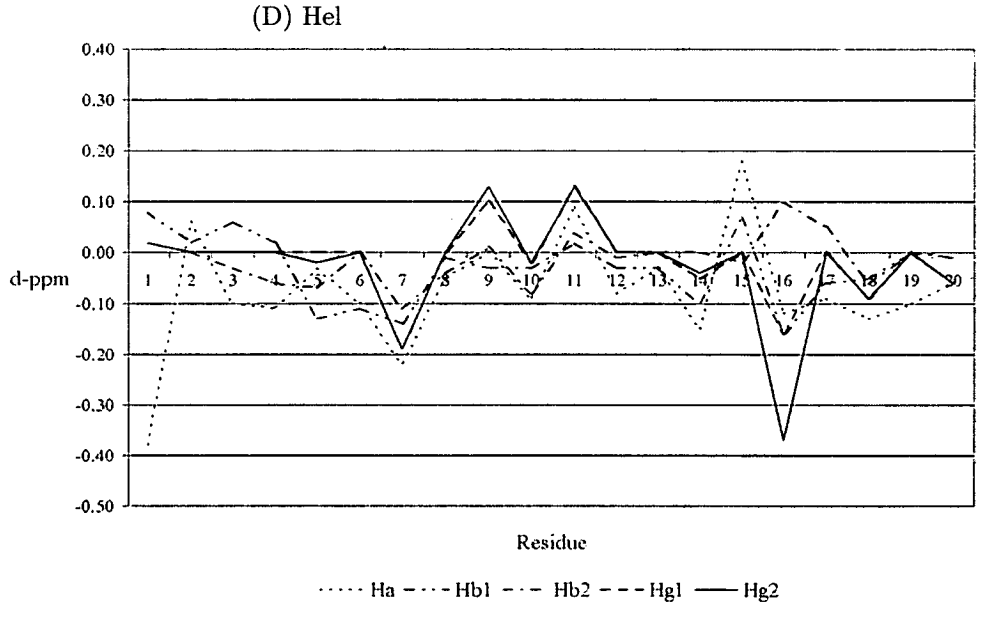
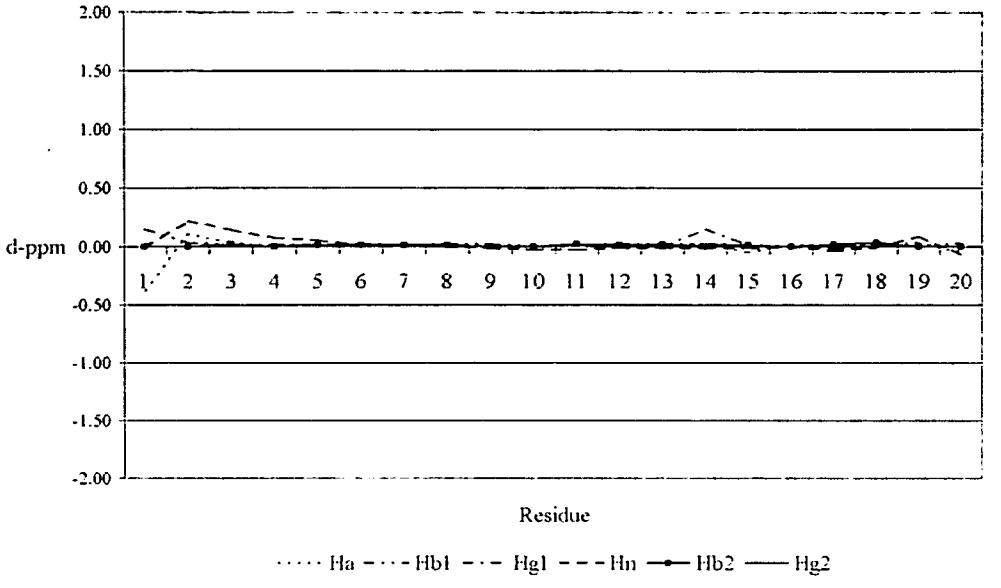
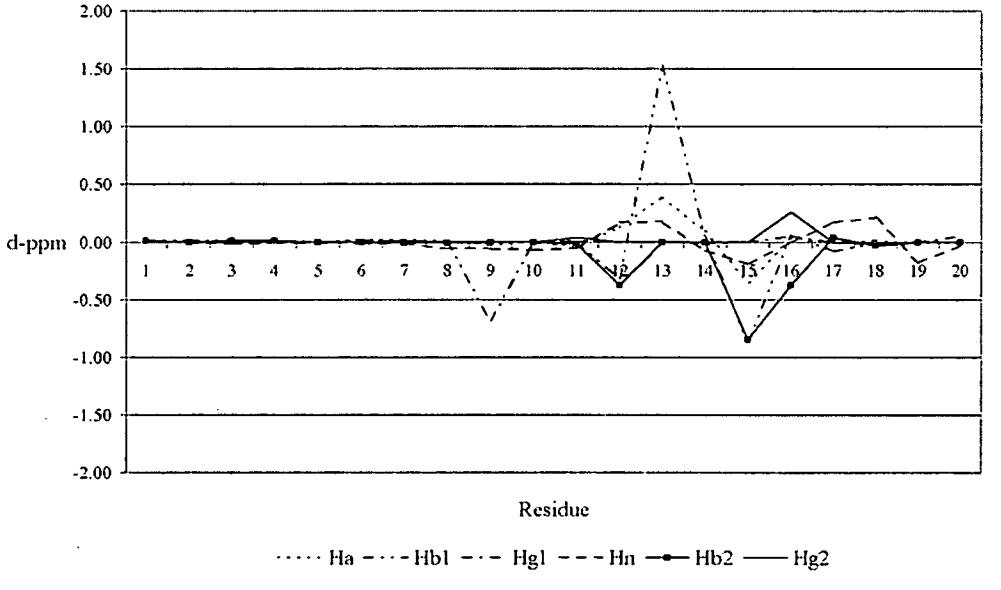


Figure 4.4: Chemical shift comparison between the polypeptides and random coil chemical shifts. For each polypeptide, the equivalent random coil chemical shifts were subtracted from the Hal polypeptide chemical shifts. (A) Hal22; (B) Hal20; (C) Dik; (D) Hel.

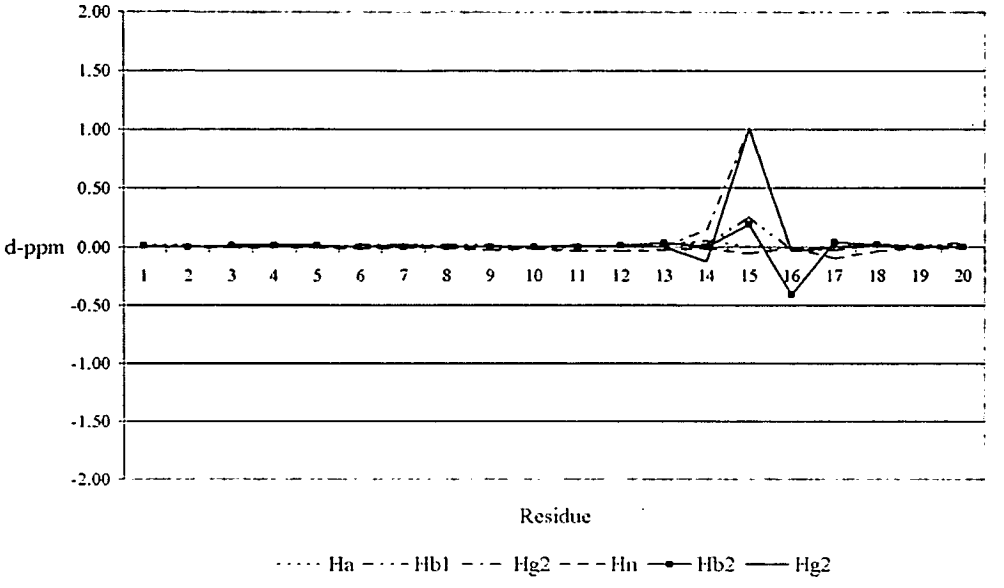
(A) Hal20-Hal22



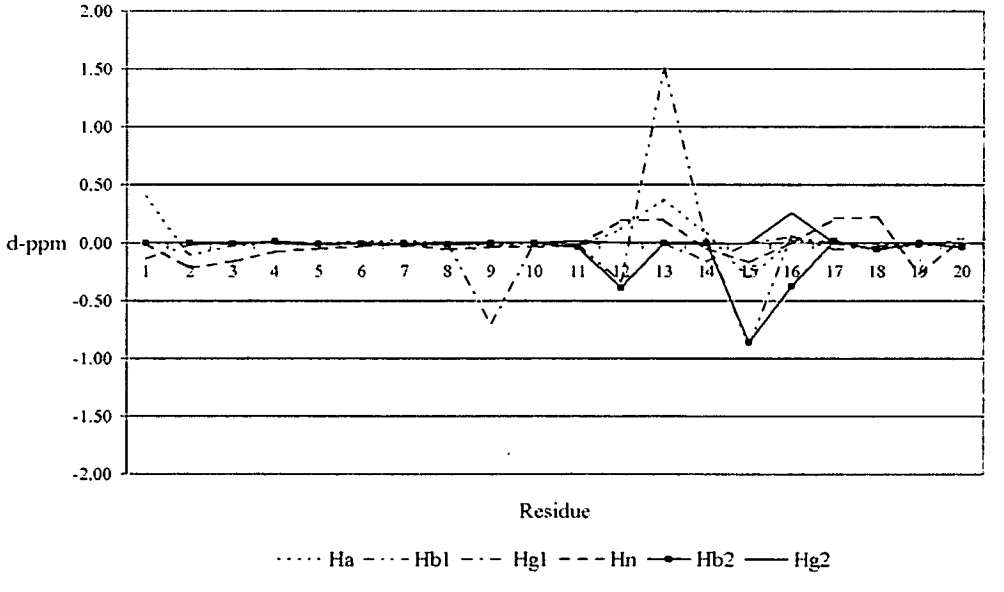
(B) Hal20-Hel

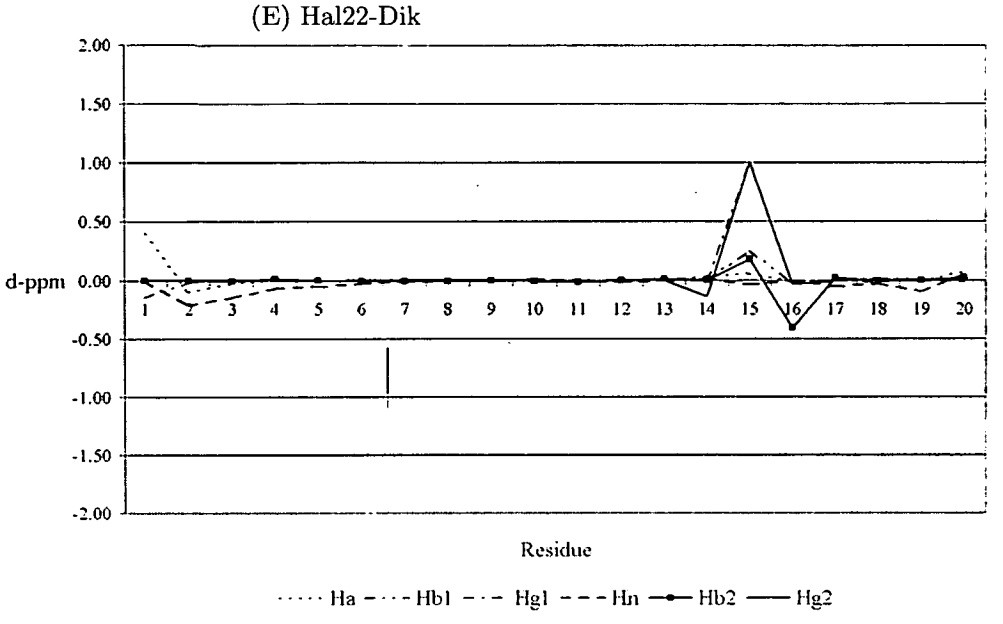


(C) Hal20-Dik



(D) Hal22-Hel





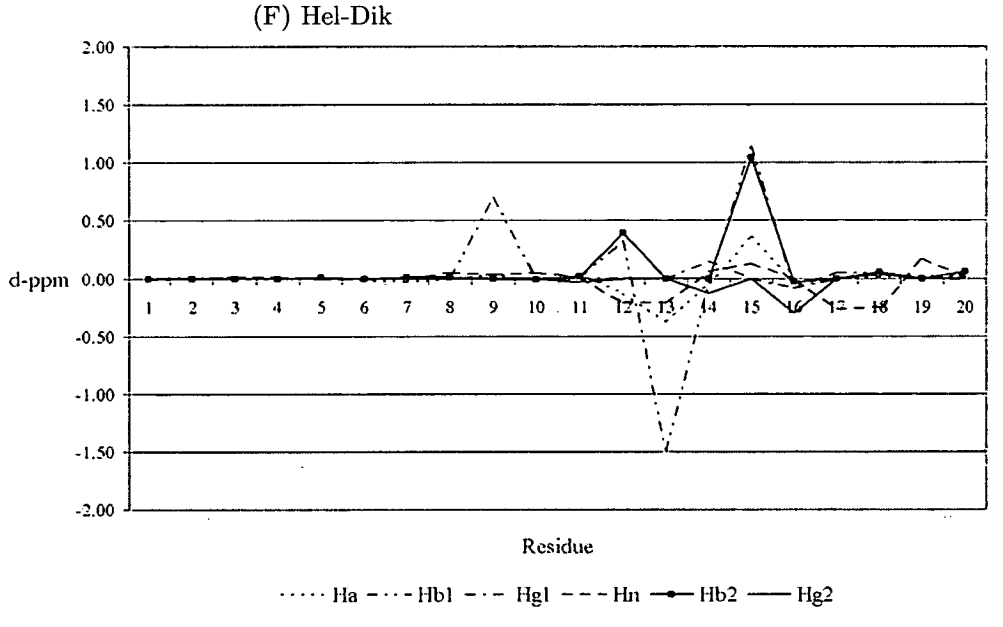


Figure 4.5: Chemical shift comparison between the four polypeptides. (A) Hal20-Hal22; (B) Hal20-Hel; (C) Hal20-Dik; (D) Hal22-Hel (E) Hal22-Dik (F) Hel-Dik.

secondary structure. Comparison of the chemical shifts between the four polypeptides is given in Figure 4.5.

4.5 Discussion

With few exceptions, no differences larger then 0.2 ppm could be seen between any of the chemical shifts for the polypeptides and the random coil values. The exceptions were the proline residues and the N-terminal residues. However, overall there was not enough of a difference to suggest any secondary structural elements in the polypeptides. The small differences observed were most likely due to the difference in the pH of the sample and the reference peptides.

The measured $J_{H^N H^\alpha}$ lay in between 5.1 and 8.1 Hz. These values indicate an abscence of α-helices and β-sheets and are typical for random coil peptides. Along with the lack of non-sequential NOESY crosspeaks, this provided further evidence that the polypeptides were unstructured.

Comparisons of the chemical shifts between the different polypeptides yielded chemical shift changes only at, or immeadiately next to, the mutated residues. In Figure 4.5, Hal20 and Hal22 show no significant chemical shift changes. This is not surprising as the two polypeptides have almost exactly the same structure. For comparisons between the native Hal sequence and both Hel and Dik, significant differences in chemical shifts are restricted to the mutations present in Hel and Dik and their surrounding residues.

In conclusion, the NMR data provided no evidence that the polypeptides contained any secondary or tertiary structural elements. In fact, the lack of non-sequential NOESY peaks, the recorded $J_{H^N H^\alpha}$ and the lack of significant deviation from random coil chemical shift values strongly indicated that the polypeptides were unstructured. This is not completely unexpected as the sequences where part of the loop in the native KefC protein. Without the support of the rest of the protein these did not fold by themselves as suggested by molecular modeling [93].

The HELE*X*DIEPFK loop and the KTN domain are hypothesized to interact via salt bridges. These salt bridges, if they exist, would be enough to hold the HELE*X*DIEPFK loop in a specific conformation just as hydrogen bonds hold elements of secondary structure together. To examine this, ^{15}N or ^{13}C edited NMR spectra could be acquired on ^{15}N or ^{13} labeled polypeptides in the presence of an unlabeled sample of the KTN domain. This would allow the HELE*X*DIEPFK loop to be examined on its own through the edited spectra, without the additional resonances of the KTN domain crowding the spectra.

Chapter 5

Structural Determinaton of fH~13: The Unlabelled and ^{15}N-labeled Samples

5.1 Attempts at the Purification of fH~13-14

Prior to the work on the single fH~13 module, the focus of the project was on a *Pichia pastoris* clone overexpressing fH~13-14 (K753-V865), produced by Ursula Lodge (University of Edinburgh). The fH~13-14 clone was based on the plasmid KM71 pPICzα, however fH~13-14 had yet to be successfully purified.

In order to determine the most suitable method of purification, the clone was cultured in a fermenter as described in Section 3.1.1. The protein was not visible on an SDS-PAGE gel before incubation at 37 °C for two hours in the presence of the enzyme Endo Hf (New England Biolabs, product code P0703L). This enzyme removes all but one of the carbohydrate residues from the protein, leaving a GluNac residue. Once this had been carried out, samples of the raw supernatant could be run on SDS-PAGE. This produced bands of protein that migrated to between the 6.5 and 16.5 kDa markers. This was evidence that the fH~13-14 protein was N-glycosylated prior to treatment with Endo Hf. However, not one but three protein bands were visible after Comassie-blue staining, suggesting the fH~13-14 protein construct was undergoing degradation.

5.1.1 Determining the Point of Degradaton of fH-13-14

In order to assess the stage during production at which this apparent degradation occurred an expression study was undertaken. The following procedure was carried out on both the transformed strain of *Pichia pastoris* and on an untransformed control strain. The yeast were streaked onto YPD-Agar plates with added Zeocin to 100 μg/ml. The plates were incubated at 30 °C for approximately four days. A single large, isolated colony was then used to inoculate 10 ml of BMG (the recipe for BMG is given in Table 3.1). This was then incubated at 30 °C for approximately two days. The 10 ml of BMG was then used to inoculate a further 90 ml of BMG. This was incubated for one day at 30 °C. After incubation, 25 ml of the inoculated BMG was added to a further 225 ml of fresh BMG and this was incubated for a further two days at 30 °C. The inoculated BMG was then centrifuged at 1,500 rpm (390 g) for 5 min. The supernatant was discarded, and the pellet of cells was resuspended in 200 ml of BMM. This media contained methanol instead of glycerol, and since protein production is under the control of a methanol-inducible promoter, the resuspension in BMM resulted in induction of the cells for protein expression.

The inoculated BMM was then incubated at 28 °C. Approximately every 24 hours after induction a 15 ml sample of supernatant was taken from the inoculated BMM. These samples were frozen in liquid nitrogen after extraction and kept at -80 °C until they could all be loaded on to a SDS-PAGE gel together. Samples from the untransformed control strain (data not shown) were collected and treated using the same method as those from the transformed strain. The control gel served to show the positions of protein bands arising from the host cell.

Samples were collected at 23, 49, 71 and 95 hours after induction. The supernatant collected at each time-point was split into four samples of 0.5 ml each for individual preparation. These are labeled on the gel in Figure 5.1 as samples 1-4 for each time-point. The supernatant in sample 1 was kept at 4 °C for 2.5 hours. The supernatant in sample 2 was incubated at 37 °C for two hours to simulate the effects of Endo Hf treatment in the absence of Endo Hf. For sample 3, Endo Hf was added to the super-

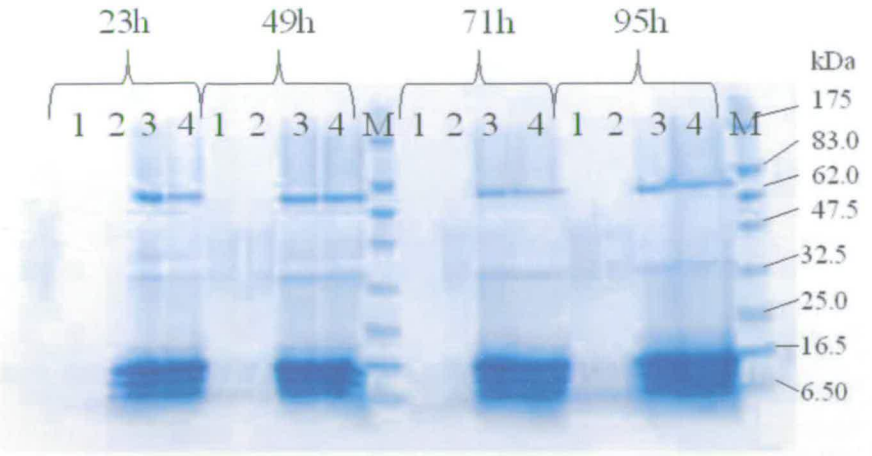


Figure 5.1: SDS-PAGE showing the results of the experiment to monitor the degradation of the fH-13-14 protein. Four samples of supernatant were taken during the course of the experiment at 23, 49, 71 and 95 hours after induction. The supernatant was filtered and kept at -80 °C prior to preparation for the gel; 0.5 ml of filtered supernatant was used per lane. For each time point there are in turn four lanes (1) supernatant kept at 4 °C for 2.5 hours (2) supernatant incubated at 37 °C for 2 hours (3) Endo Hf added to supernatant then incubated at 37 °C for two hours (4) Endo Hf, PMSF and EDTA added to supernatant then incubated at 37 °C for two hours. There are also two lanes of protein marker, labeled (M). The protein marker used was the New England Biolabs Broad Range Prestained Protein Marker P7708S

natant then incubated at 37 °C for two hours. An aliquot of this Endo Hf solution was used for samples at time points 23-71 h, while 3 μl was used for samples at 95 h as considerably more protein was found in this sample. For sample 4, Endo Hf, PMSF (to 0.5 mM) and EDTA (to 5.0 mM) were all added to the supernatant then incubated at 37 °C for two hours. The inclusion of PMSF and EDTA in sample 4 was designed to prevent protein degradation by proteases that could occur due to the raised temperature during Endo Hf treatment.

Thus a comparison between the different time-points should indicate whether protein degradation was occurring during protein expression, and if so at what point after induction. Furthermore, a comparison amongst samples 3-4 should determine whether protein degradation was occurring during Endo Hf incubation instead of or as well as during protein expression.

The results shown in Figure 5.1 demonstrate that Endo Hf treatment is required for

the expressed protein to be visible on the gel at the expected point, as no protein can be seen in columns 1 and 2. This suggests that the protein is heterogeneously glycosylated, producing a range of glycoproteins of various molecular weights. However, three distinct bands can be seen between the 6.5 and 16.5 kDa markers at all points after Endo Hf treatment. Therefore it was concluded that the protein was undergoing significant degrading during the entire expression phase, and no significant degradation was occuring during Endo Hf treatment.

Various protein purification columns were tried to separate out the three bands, as outlined below. As an initial purification step (to clean the sample before FLPC) various methods were tried. A hydrophobic interaction column (HIC) under gravity flow was tried at binding buffer $(NH_4)_2SO_4$ concentrations from 1-2 M (Fig 5.2A). Also a cation exchange column under gravity flow between pH 4.5-6.0 was also tried (Fig 5.3A). However, the protein still appeared on SDS-PAGE as three separate bands.

After the initial gravity flow purification step, the protein needed to be deglycosylated using Endo Hf before FPLC. A concanavalin A column was used to remove sugar residues from the solution. The solution was then tested through a variety of FPLC columns.

For the products of the HIC gravity flow column, cation exchange at pH 5 (Fig 5.2B) and pH 6 (Fig 5.2C) aswell as reversed phase (Fig 5.2D) were tried but none achieved band separation. For the products of the CIE gravity flow column, the following were tried (in order of increasing hydrophobicity): ethyl HIC (Fig 5.3B); isopropanol HIC (Fig 5.3C); and a phenyl HIC(Fig 5.3D) were tried. However, none of these methods achieved band separation. It was thus decided to focus the project on the newly available single module fH-13 sample produced by Claire Egan.

5.2 Sample Optimization

Claire Egan performed the cloning, fermentation and purification of both unlabeled and ^{15}N-labeled samples of fH-13, details of which can be found in section 3.1. Claire

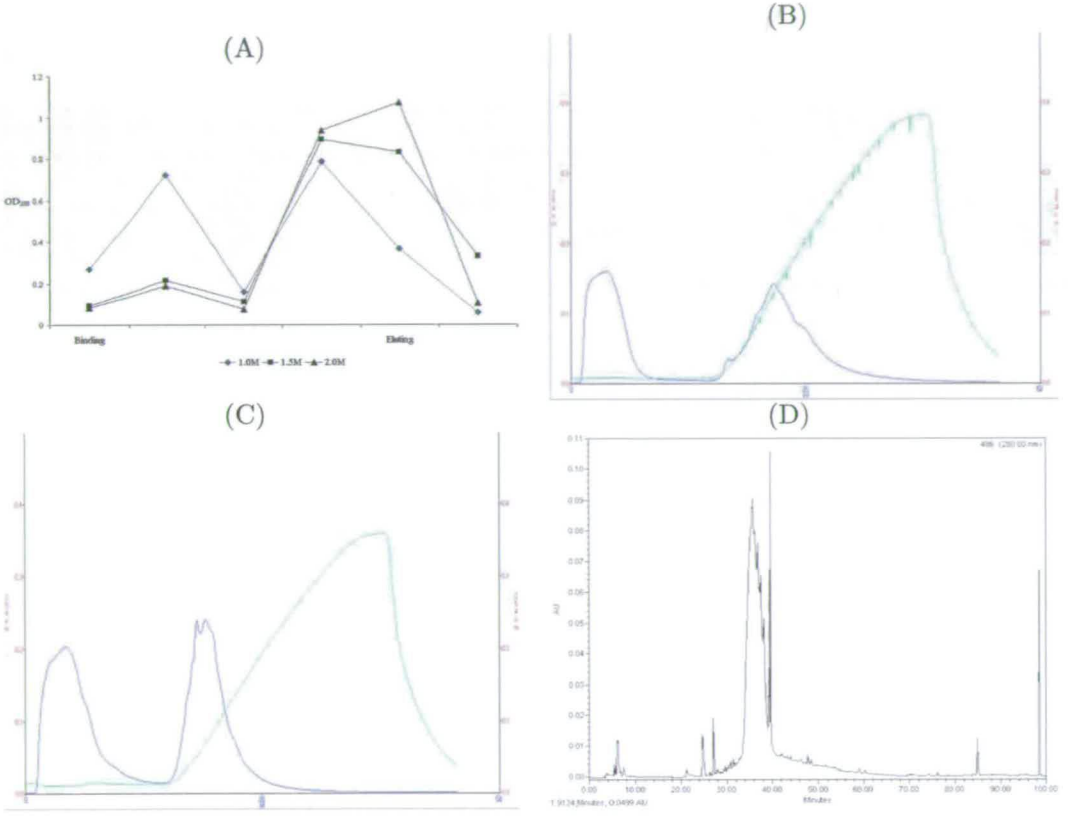


Figure 5.2: Chromatograms from the attempts to purify fH~13-14. (A) Results from the gravity flow HIC column at binding buffer $(NH_4)_2SO_4$ concentrations from 1-2 M. (B) Products from the gravity flow HIC column put through an FPLC CIE column at pH 5.0 after deglycosylation. (C) Products from the gravity flow HIC column put through an FPLC CIE column at pH 6.0 after deglycosylation. (D) Products from the gravity flow HIC column put through an FPLC reversed phase column after deglycosylation.

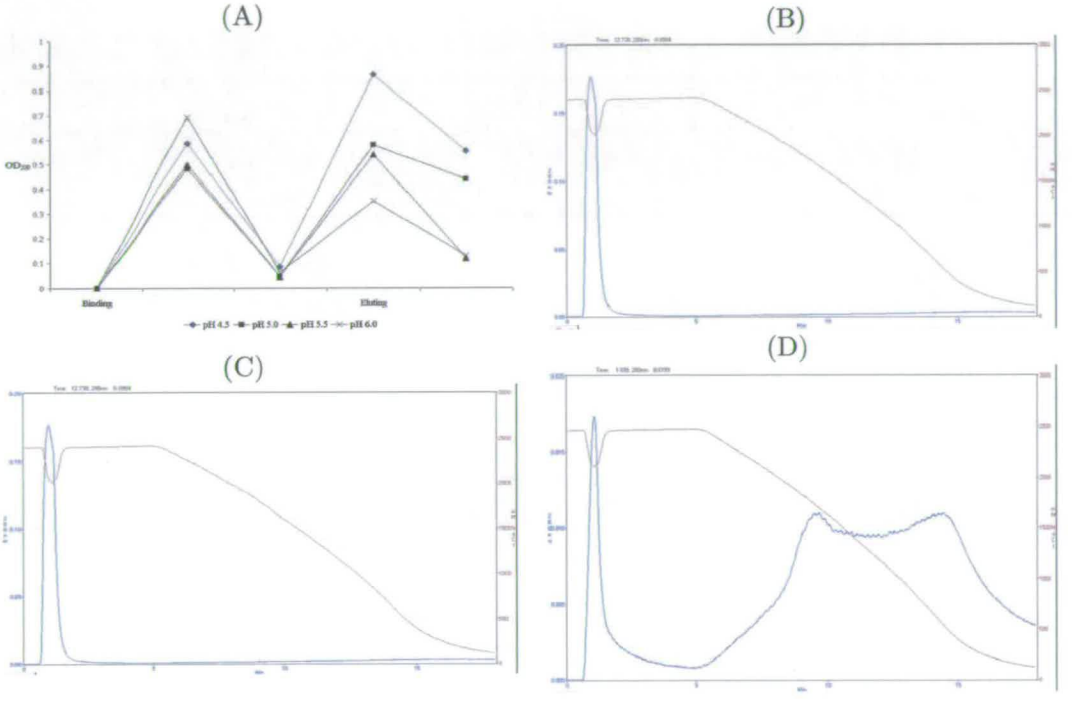


Figure 5.3: Chromatograms from the attempts to purify fH~13-14. (A) Results from the gravity flow CIE column at buffer pH 4.5-6.0. (B) Products from the gravity flow CIE column put through an FPLC ethyl HIC column after deglycosylation. (C) Products from the gravity flow CIE column put through an FPLC isopropanol HIC column after deglycosylation. (D) Products from the gravity flow CIE column put through an FPLC phenyl HIC column after deglycosylation.

Egan also validated the sample was fH-13 by mass spectrometry (data not shown). It had a sequence of AEAAG KCKSSNLII LEEHLKNKKE FDHNSNIRYR CRGKEGWIHT VCINGRWDPE VNCS (determined from the NMR assignment) corresponding to residues K752-S804 of fH with a five residue secretion signal peptide at the N-terminus. The samples were then handed to me for the NMR studies.

The initial unlabeled fH-13 NMR sample was used to determine the best conditions for running the NMR experiments. The pH, salt concentration and temperature were optimized by recording 1D proton spectra. The recorded spectra were examined to see under which conditions the protien was gave the best quality spectra. The efficiency of TOCSY transfer was also checked under various sample conditions. A poor transfer efficiency would indicate dynamics between folded and unfolded states or the aggregation of the protein sample. 1D TOCSY spectra were therefore acquired to see which conditions gave the best TOCSY transfer.

To determine how temperature would affect the sample of fH-13, 1D proton spectra were recorded at two temperatures; 310 K and 288 K. For these experiments the sample conditions were 200 mM salt, pH 5.9 in 20 mM sodium phosphate buffer. The results of changing the sample temperature on the unlabeled sample of fH-13 can be seen from Figure 5.4. There was no significant difference in the 1D proton spectra. 1D proton TOCSY spectra were also acquired at the two temperatures, with the magnetization transfer starting in the methyl region. Again, there was no significant difference in TOCSY transfer between the two different temperatures.

To determine the optimum pH for fH-13, 1D proton spectra were recorded between pH 4.3 and pH 6.5. The sample conditions were 310 K and 10 mM sodium phosphate buffer. Although the lower pH samples were slightly outside the sodium phosphate buffer range, the sample pH were measured before acquisition of the spectra. The resulting spectra can be seen in Figure 5.5. The peak dispersion is seen to increase with increasing pH. This is particularly noticeable in the methyl region (Figure 5.5 (B)) where CH_3 peaks in the 0.4 to 0.6 ppm region moved to a higher ppm at pH 3.5 indicating protein unfolding. The TOCSY transfer also improved dramatically with

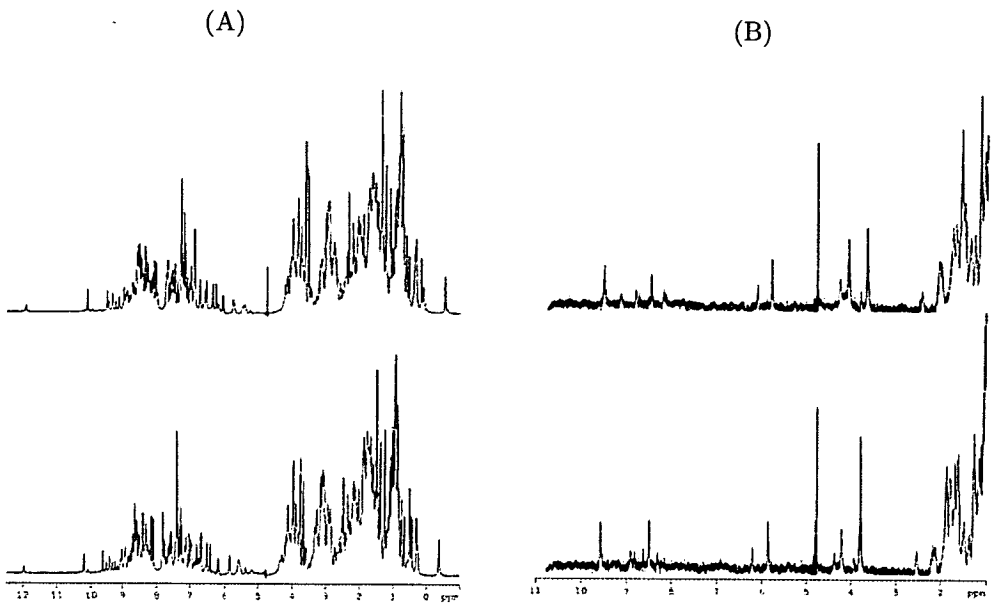
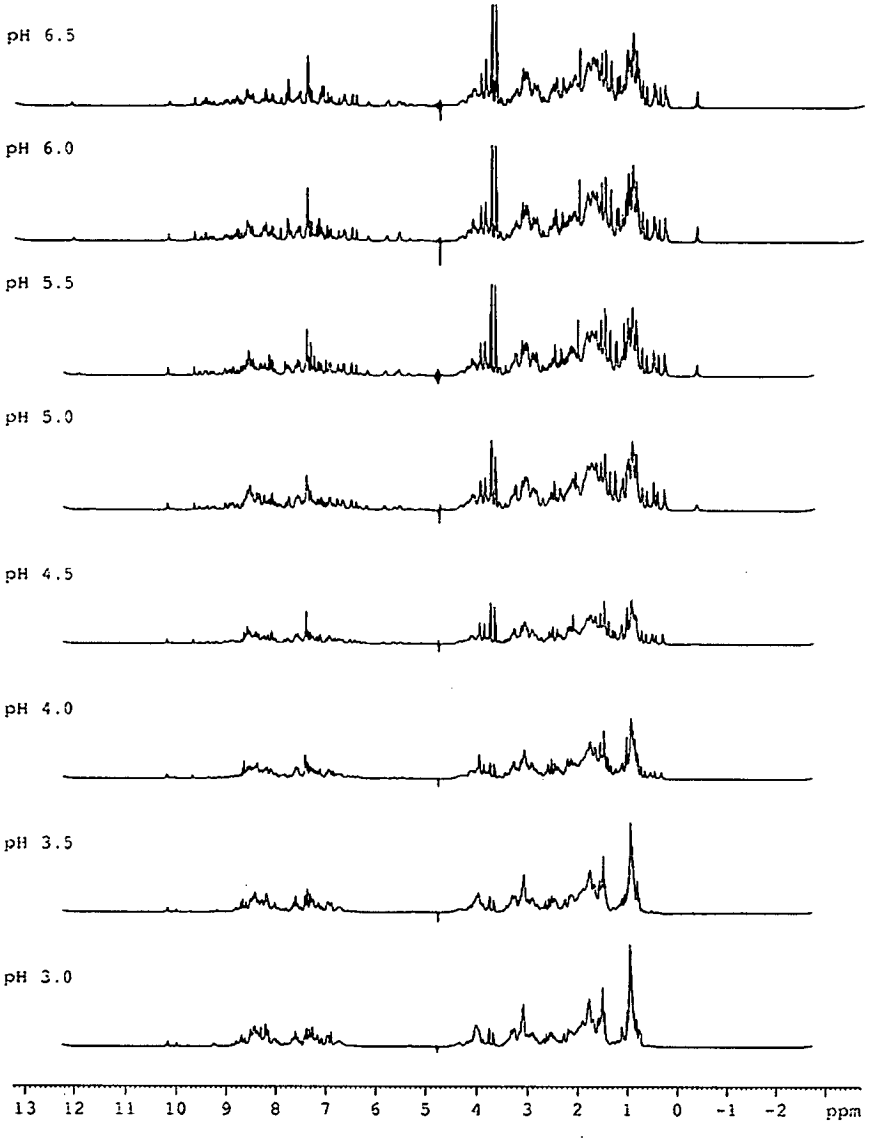


Figure 5.4: Effects of changes in temperature on a sample of unlabeled fH~13. The top spectra were taken at 298 K and the lower spectra were taken at 310 K. (A) 1D proton NMR spectra with a water suppression using the double pulsed field gradient spin-echo (DPFGSE). (B) 1D proton TOCSY spectra with the selective TOCSY spin lock pulse for protons in the methyl region (0.5 ppm±1.5 ppm). There appeared to be no significant difference between the two temperatures.

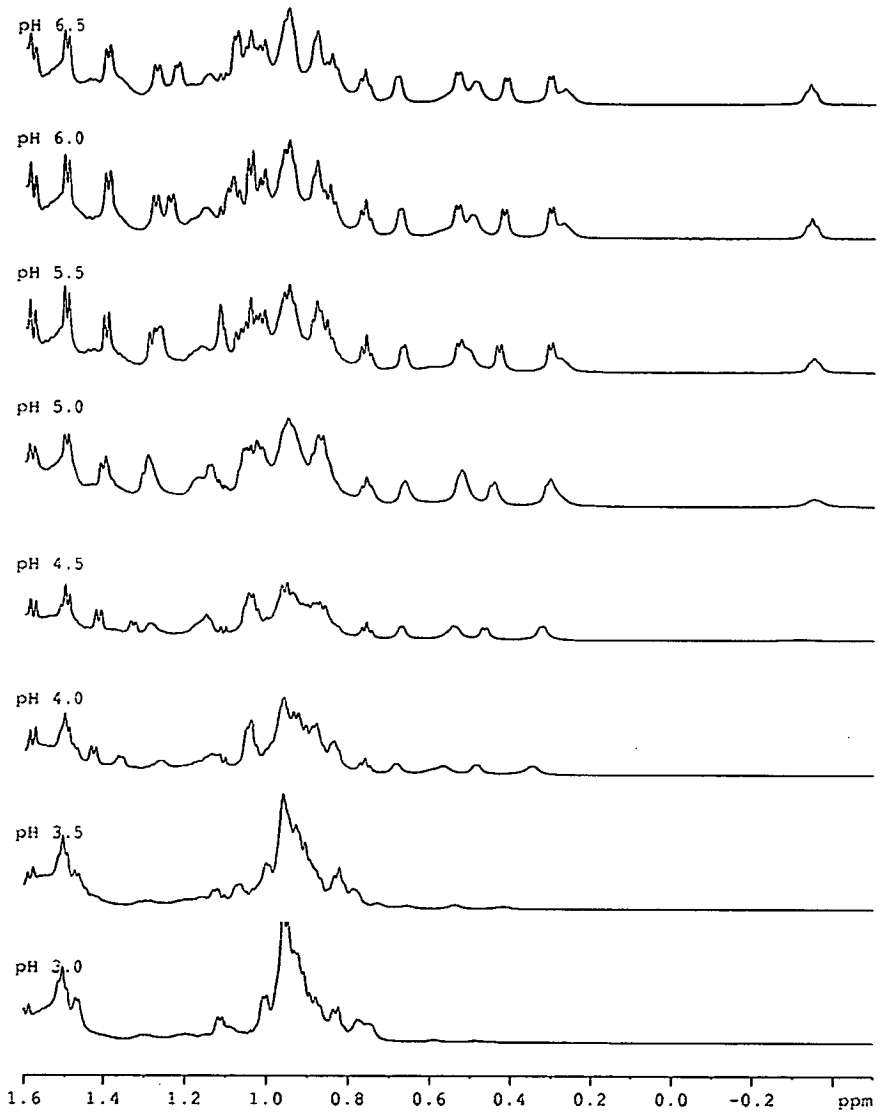
increasing pH (Figure 5.5 (C)): below pH 4.0 very little magnetization reaches the backbone NH protons. It was therefore decided to determine the structure of fH~13 at pH 6.0 and a ^{15}N labeled sample was used for assignment.

Samples of protein can aggregate and also precipitate in low salt solutions. Increasing the salt concentration can help prevent aggregation and improve the quality of the NMR spectra. To determine whether increasing the salt concentration would improve the quality of the sample, 1D proton spectra were recorded between 0 mM and 200 mM sodium chloride. The sample buffer used was 20 mM sodium phosphate buffer at pH 6.1. The pH changed gradually from pH 6.1 to pH 5.9 with increasing additions of sodium chloride, while the experiments were all recorded at 310 K. The recorded spectra can been seen in Figure 5.7. Varying the salt concentration made no significant difference to the quality of the spectra, and so no additional salt was added to the ^{15}N sample which was used for these studies. This maximizes the sensitivity of the NMR ex-

(A)


pH 6.5
pH 6.0
pH 5.5
pH 5.0
pH 4.5
pH 4.0
pH 3.5
pH 3.0
13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1 -2 ppm

(B)


pH 6.5
pH 6.0
pH 5.5
pH 5.0
pH 4.5
pH 4.0
pH 3.5
pH 3.0
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
ppm

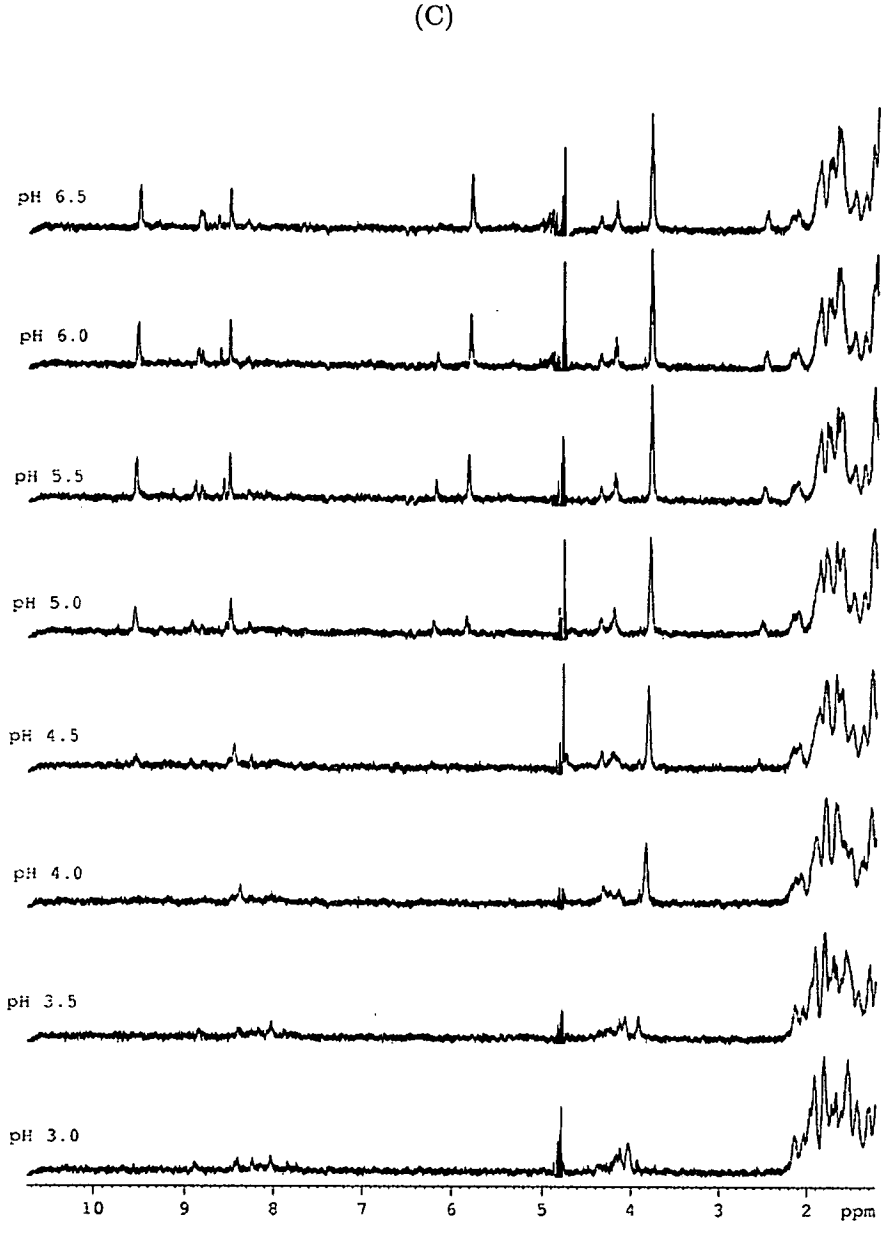


Figure 5.5: Effects of changes in pH on a sample of unlabeled fH-13. (A) A series of 1D proton NMR spectra with a DPFGSE used for water suppression. (B) Expansion of the methyl region of the spectra. (C) A series of 1D proton TOCSY spectra with the selective TOCSY transfer from protons at 0.5 ppm ± 1.5 ppm (the methyl region of the spectrum)

periments as the addition of salt is detremental, particularly with regard to cryoprobes.

Experiments for ^{1}H and^{15}N assignment were recorded at 310 K and assignment was carried out at this temperature initially. However, it became apparent during assignment that certain backbone NH peaks were missing from the ^{15}N, ^{1}H HSQC. Therefore a series of ^{15}N, ^{1}H HSQCs were taken between 310 K and 288 K. It was at 288 K that the most additional peaks were observed (see Fig 5.6).

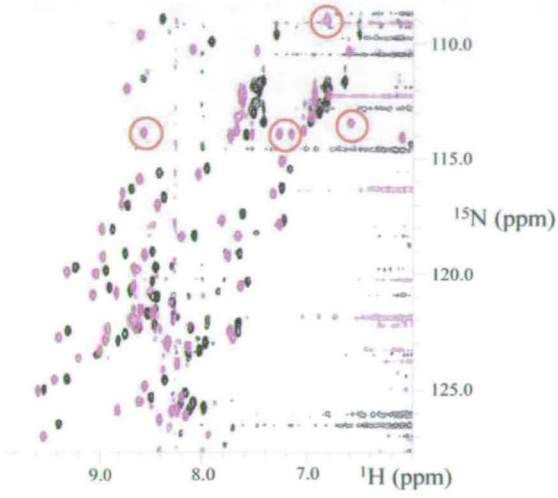
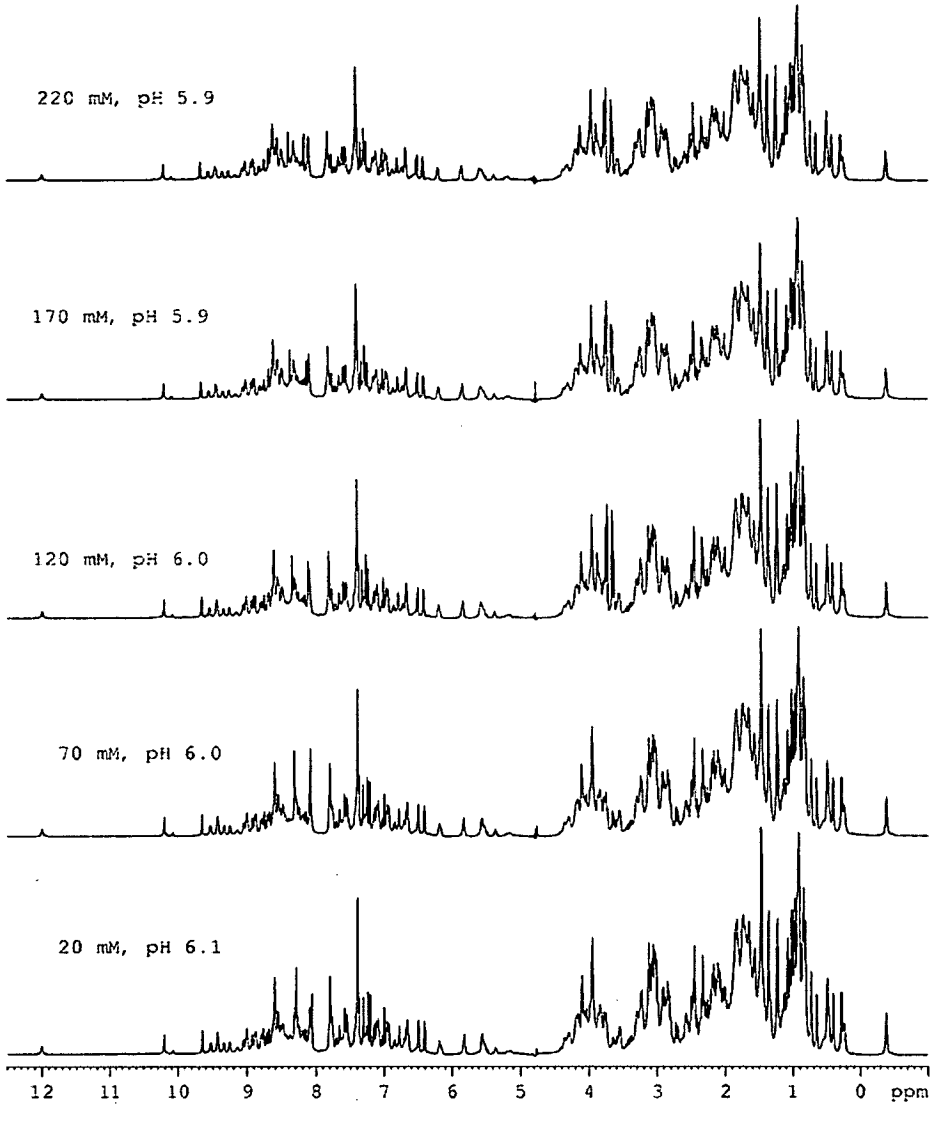


Figure 5.6: Effects of changes in temperature on a sample of ^{15}N labeled fH~13. Two ^{15}N, ^{1}H HSQCs were taken at different temperatures; 310 K (black) and 288 K (purple). Additional peaks (circled in red) were seen at 288 K that were not seen at 310 K.

As assignments had already been made at 310 K, a series of ^{15}N, ^{1}H HSQCs at temperatures between 310 K and 288 K were recorded. These were used to transfer assignments from the spectra recorded at 310 K to the new spectra recorded at 288 K. Following the assignment of NH resonances at this new temperature, experiments for ^{1}H assignment were re-recorded at 288 K. The appearence of new peaks at the lower temperature could be attributed to the slowing down of ms timescale dynamic processes in the protein. Two backbone NH crosspeaks became visible at the lower temperature (I759 and H764).

(A)


220 mM, pH 5.9
170 mM, pH 5.9
120 mM, pH 6.0
70 mM, pH 6.0
20 mM, pH 6.1
12
11
10
9
8
7
6
5
4
3
2
1
0
ppm

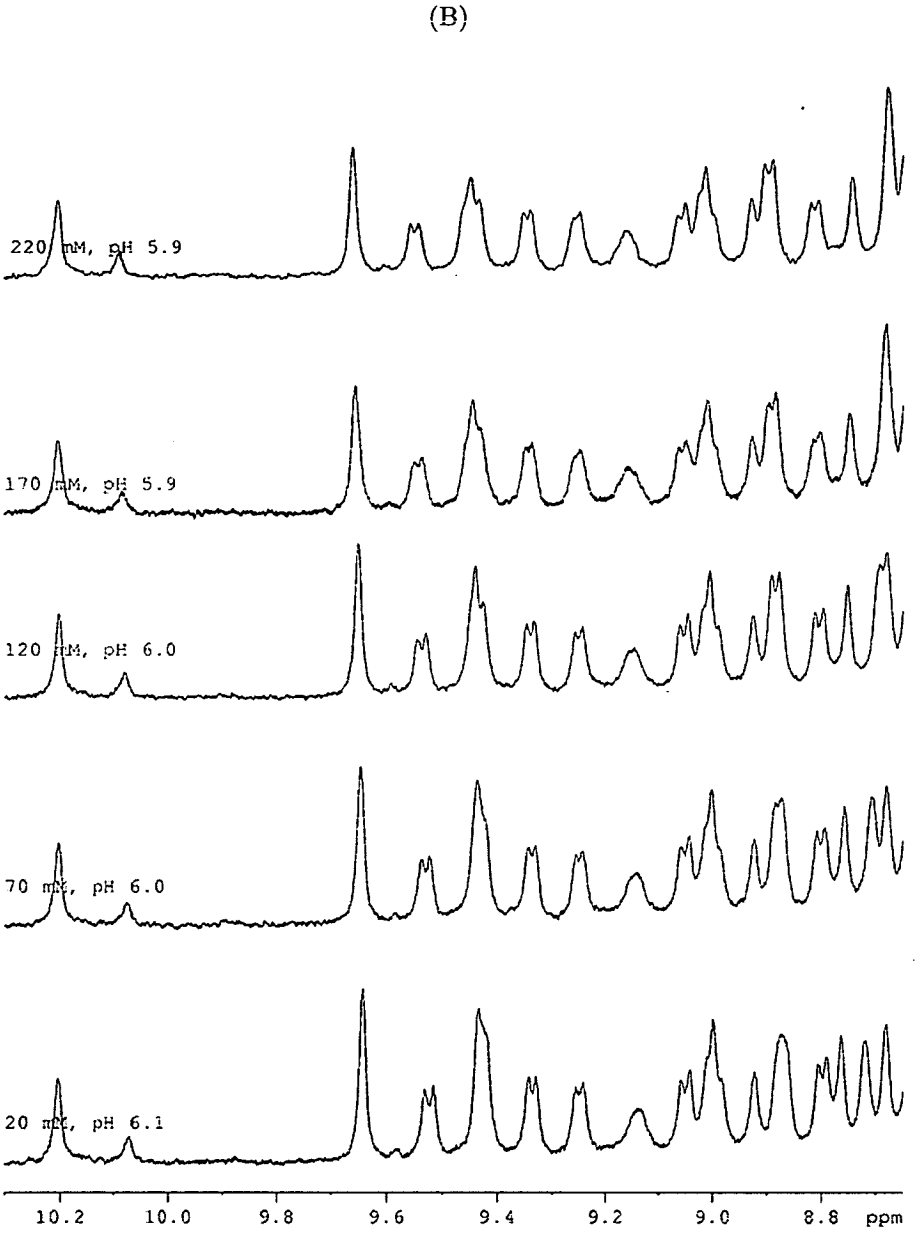


Figure 5.7: Effects of changes in salt concentration (NaCl) on a sample of unlabeled fH~13. (A) A series of 1D proton NMR spectra with a DPFGSE for water suppression. (B) Expansion of the NH region of the spectra shown in A.

5.3 Resonance Assignment

A full list of resonance assignments from the unlabelled and ^{15}N-labeled fH~13 samples can be found in Appendix B.

5.3.1 Backbone Assignment

The backbone assignment was carried out using the method outlined in section 2.5.4. Where they existed, high intensity NOESY crosspeaks from H^{α} of residue $(i-1)$ to the H^{N} of residue (i) were used to link sequential residues. An example of five residues linked by this method is shown in Figure 5.8. Using this method the backbone of fH~13 was assigned to near completion.

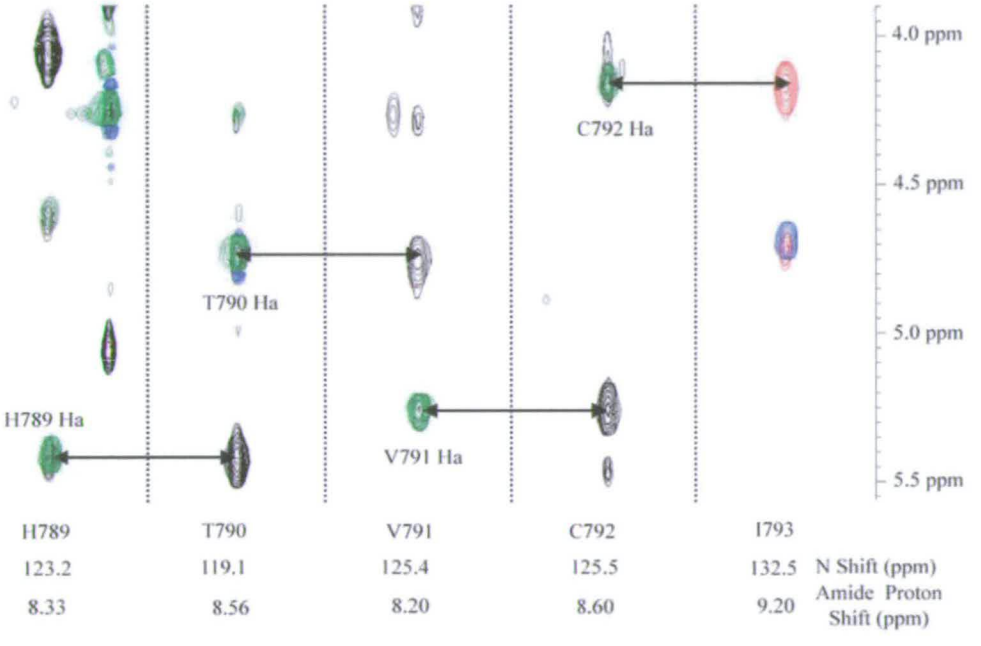


Figure 5.8: Five strips from the ^{15}N-edited 3D TOCSY overlayed with the corresponding ^{15}N-edited 3D NOESY strips for the residues H789-I793. ^{15}N-edited 3D TOCSY peaks are coloured green (positive) and blue (negative). ^{15}N-edited 3D NOESY crosspeaks are coloured black (positive) and red (negative). Strong NOESY crosspeaks between the H^{N} of residue of residue *(i)* and the H^{α} of residue *(i-1)* suggests these are sequential residues. The N-H^{N} strips of I793 were aliased in the ^{15}N edited spectra and so its crosspeaks appear negative in these spectra. An explanation of aliasing is given in section 5.3.1.

Figure 5.9 shows a ^{15}N, ^{1}H-HSQC with the assigned NH crosspeaks. Peaks with resonance frequencies beyond the sampled ^{15}N region are shown as aliased peaks. The spectra were aqcuired by setting the first sample point in the ^{15}N dimension to one-half of the dwell time. For the folded peaks, each folding changes the phase of the peak by 180°. Therefore, if a peak is folded once, it has a negative shift; if a peak is folded twice, it has a positive shift. The true chemical shifts of aliased peaks can been found in Appendix A.

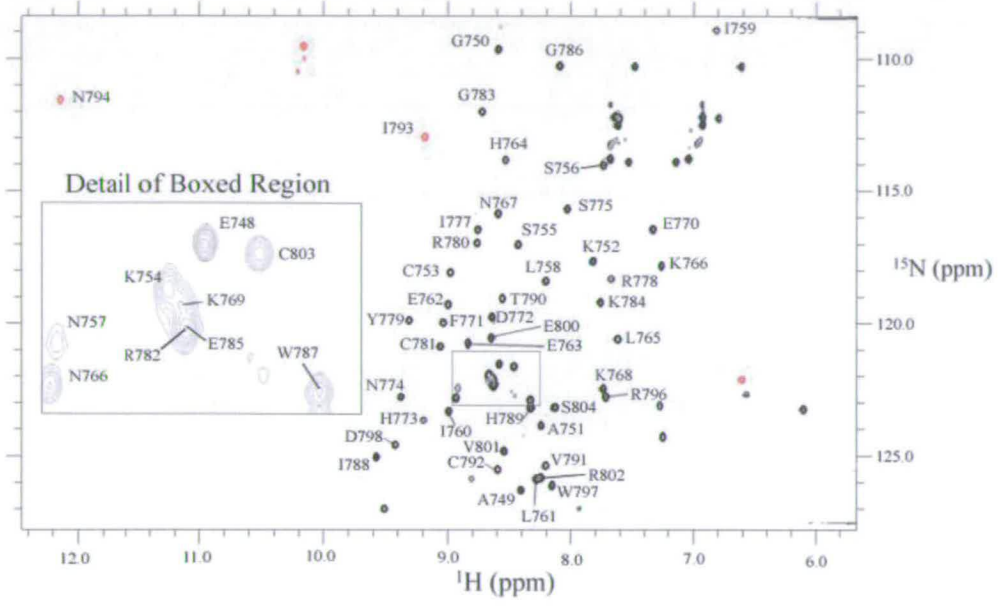


Figure 5.9: ^{15}N, ^{1}H-HSQC of fH~13. This was aquired using a 600 MHz Bruker NMR spectrometer with a cryoprobe. Positive peaks are black. Negative peaks are red. There is an unexplained peak at ≈123 ppm by ≈9.6 ppm in the ^{15}N,^{1}H HSQC. This peak produced no TOCSY, COSY or NOESY crosspeaks in any other spectra and therefore could not be assigned.

The first two N-terminal residues of the cloning artifact (believed to be EAEAGA) did not yield crosspeaks in the ^{15}N,^{1}H HSQC spectra. These two residues are part of the secretion signal peptide sequence a cloning artifact that ensured that the protein was secreted into the supernatant by *Pichia pastoris* after production in the cell. As this sequence was not part of the native fH~13 it was thought these two residues could be unstructured and highly flexible. This was later proved not to be the case when

the ^{13}C, ^{15}N labeled fH-13 sample became available. The real reason why there are missing assignments for the signal secretion peptide is explained in section 6.2.1

The backbone H^{α} of D798 was also missing from the assignment. For this residue, no crosspeak for the H^{N} to H^{α} interaction can be seen in the fingerprint region of any spectrum, nor can there be seen *J*-coupled crosspeaks from the H^{α} of D798 to its H^{β} protons. The missing crosspeaks to H^{α} to could be hidden under the water signal if the H^{α} resonance overlaps with that of water. There is also an unexplained peak at $\approx$123 ppm by $\approx$9.6 ppm in the ^{15}N,^{1}H HSQC. This peak produced no TOCSY, COSY or NOESY crosspeaks in any other spectra and therefore could not be assigned.

5.3.2 Sidechain Assignment

Using the method outlined in section 2.5 the assignment of fH-13 was carried out to almost completion. Out of the 59 residues that made up the construct, 40 were fully assigned. The residues that had their sidechains fully assigned were E748-A751; C753; S755-E763; L765; N767; E770-D772; N774-N776; Y779; C781; G783; E785-G786; I788; T790-G795; P799-S804. This list includes all the serine, threoine and tyrosine residues, which were fully assigned apart form their OH groups (the H^{γ} for serine and the $H^{\gamma 1}$ for threonine) which are rarely assigned.

All the lysine and arginine groups had unassigned terminal sidechain NH groups. These groups are normally exposed to water and the groups are ionisable so the protons are rarely observed. For the lysines (K752; K754; K766; K768; K769; and K784), this meant the N^{ζ} and H^{ζ} nuclei were unassigned. For the arginines (R778; R780; R782; and R796), the N^{ϵ} and H^{ϵ} nuclei were assigned in all residues except for in R782. However, the guanidinium protons were unassigned for all arginines.

K769 also had its H^{δ} and H^{ϵ} protons were unassigned in the ^{15}N spectra. This was because its backbone N-H peak in the ^{15}N HSQC overlapped with that of K754, E785 and R782 (see the boxed region of Figure 5.9). This meant that the sidechain resonances of K769 overlapped with those of these residues in both the homonuclear and ^{15}N-edited spectra to some extent. The K769, K754, E785 and R782 resonances were finally as-

signed almost completely using a double labeled sample and the H(C)(CO)NH-TOCSY spcectra (see sections 2.6.1 for the method and 6.2.2 for the assignment).

The methyl group $H^{\gamma 2}$ protons in I777 were unassigned in the ^{15}N spectra. There were no crosspeaks that could definitely belong to this group, suggesting their chemical shift overlapped with another resonance. This was confirmed using ^{13}C-edited spectra (see section 6.2.2).

The $H^{\epsilon 3}$ nuclei for both W787 and W797 were unassigned. Both H^{ζ} and the H^{η} nuclei were also unassigned for W797 in the ^{15}N spectra. These missing assignments were found using the $^{13}C^{15}N$-labeled sample of fH~13 using the carbon spectra. The H^{ζ} and the H^{η} assignments were missing as their resonances were incorrectly assigned to F771 aromatic resonances, as described in section 6.2.2.

The histidines H764, H773, and H789 were all missing their $H^{\delta 1}$ and $H^{\epsilon 2}$ assignments in the ^{15}N spectra. H764 was also missing its $H^{\epsilon 1}$ assignment in the ^{15}N spectra. The H764 $H^{\epsilon 1}$ was later found using the ^{13}C-^{1}H HSQCs of the aromatic region. Finally, the sidechain of E746 could also not be assigned.

In total, 92.6% of the ^{15}N and ^{1}H nuclei in fH~13 were assigned using the ^{15}N spectra. These assignments are given in Appendix B.

5.4 Non-NOE based Restraints

5.4.1 Hydrogen Bonds

To identify residues involved in hydrogen bonds a sample of fH~13 that had previously been dissolved in H_2O was freeze dried before being dissolved in 99.9% D_2O. A series of ^{1}H-^{15}N HSQCs were then acquired over the course of three hours.

Thus those residues whose NH backbone protons were exposed to the new solvent exchanged their ^{1}H with ^{2}D and disappeared from successive spectra over the range of time points taken. On the other hand, those NH groups which were protected from sol-

vent exchange by being engaged in hydrogen bonds retained their ^{1}H and their protons were be present in the spectrum after the three hours. The assigned ^{1}H-^{15}N HSQC D_2O spectrum acquired after three hours is shown in Figure 5.10.

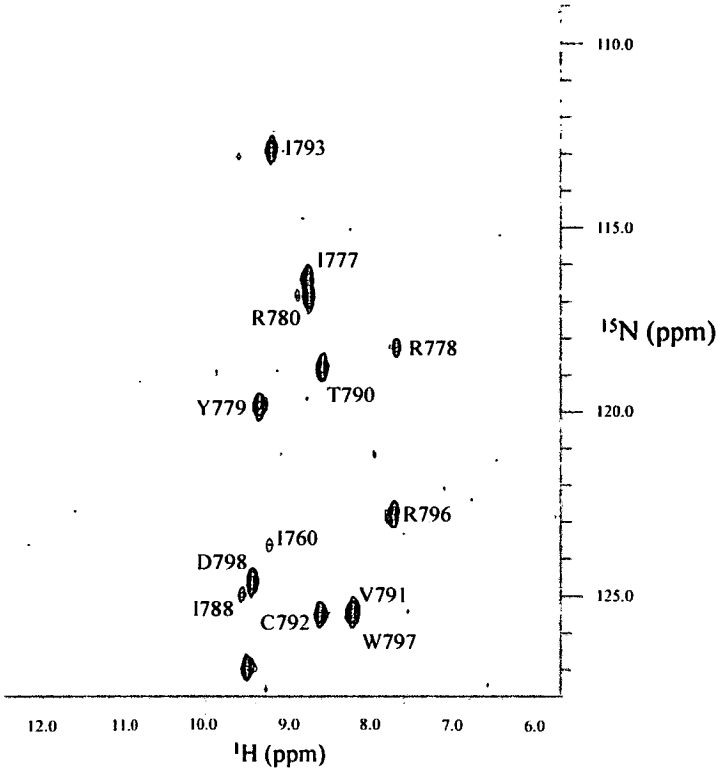


Figure 5.10: The ^{1}H-^{15}N HSQC acquired after 3 hours of exposure to D_2O. The peaks are assigned.

Once the hydrogen atoms involved in hydrogen bonding had been identified, suitable binding partners were found by examining structures of fH~13 determined using NOE distance restraints alone. It was not always possible to find suitable binding partners, or the binding partner was ambiguous (ie. it could have been one of a number of other atoms). For those hydrogen atoms to which partners can be assigned the hydrogen bond was inputted as two distance restraints into CNS. These restraints are given in Table 5.1.

The following H^N atoms could not be matched with suitable hydrogen bonding part-

H^N Partner	CO Partner
I777	T790
I788	Y779
T790	I777
C792	S775
I793	R796
R796	I793
D798	V791

Table 5.1: The hydrogen bond restraints

ners, although shown to be protected from deuterium exchange: R778, Y779, R780, V791 and W797.

5.4.2 Residual Dipolar Couplings

Residual dipolar couplings were collected - for the backbone NH bonds only - by acquiring two IPAP-HSQC experiments, one with the protein sample aligned and one in isotropic medium. In the IPAP experiment, two ^{1}H-^{15}N HSQC spectra are acquired in an interleaved manner. The difference between the two data sets is that one produces ^{1}H coupled crosspeaks that are in phase in the ^{15}N dimension while the other shows antiphase coupled peaks. Both spectra thus contain twice as many peaks as a regular ^{1}H-^{15}N HSQC spectrum. This could lead to excessive peak overlap that would make the spectra difficult to analyse unambiguously. By adding or subtracting the two IPAP-HSQC spectra the peak numbers are halved, thus restoring the simplicity of the spectra to that of a regular ^{1}H-^{15}N HSQC spectrum. The ^{1}H-^{15}N splitting can then be deduced from a comparison of the resonance frequencies of corresponding crosspeaks in the two edited spectra. This can be seen in Figure 5.11, where the difference between (A) and (E) represents the $J + D$ coupling, and the difference between (B) and (D) represents the J coupling alone.

After acquiring an IPAP experiment, the two interleaved ^{1}H-^{15}N HSQC data sets were separated prior to processing in AZARA. The spectra were assigned in CCPN Analysis and the extracted J coupling and D coupling for the backbone HN bond are shown in

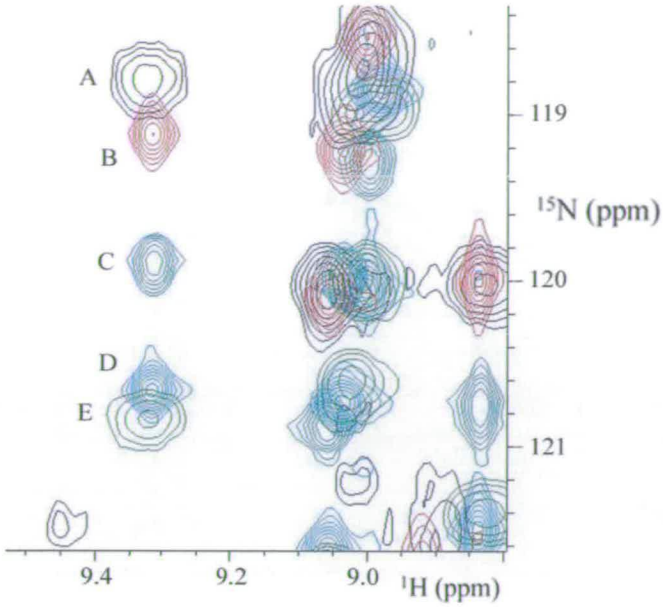


Figure 5.11: Illustration of the IPAP ^{1}H-^{15}N HSQC spectra of the isotropic and aligned samples of fH-13. (A) Peak of the sum of the aligned IPAP spectra (B) Peak of the sum of the unaligned IPAP spectra (C) Peak from a regular ^{1}H-^{15}N HSQC (D) Peak of the difference of the unaligned IPAP spectra (E) Peak of the difference of the aligned IPAP spectra. The peak is for the backbone NH of Y779.

Table 5.2.

Once the D couplings were calculated the error was estimated for both the J couplings from the unaligned sample and the $J + D$ couplings from the aligned sample by dividing the average linewidth at half height by the signal-to-noise ratio for the spectrum. For the unaligned IPAP spectrum the value was calculated to be 0.2 Hz, while for the aligned sample (which had broader linewidths and a considerably worse signal-to-noise ratio) it was 2.0. An error of 2.0 Hz was therefore used for the RDC restraints.

5.5 Structure Calculation and Analysis

Structures were calculated using the method described in section 3.5. In total, 607 unambiguous and 478 ambiguous NOE restraints were used in the calculation. It is worth noting that some of the ambiguous restraints will be represented more then once because there were ambiguous NOE restraints from 3 different spectra (homonuclear NOESY; ^{15}N-edited 3D NOESY; and a homonuclear NOESY in D_2O). The unambiguous NOE restraints were put into classes based on the distance between the two

Residue	$J + D$ coupling	J coupling	D coupling
C753	-96.4	-93.9	-2.5
S755	-123.5	-93.3	-30.2
S756	-92.8	-92.7	-0.1
L758	-67.7	-93.1	25.4
I760	-122.3	-92.3	-30.0
E763	-84.0	-92.1	8.1
H764	-110.2	-93.9	-16.3
L765	-86.2	-92.3	6.0
K766	-82.9	-93.0	10.0
N767	-94.1	-93.6	-0.5
K768	-96.1	-94.4	-1.7
E770	-108.8	-92.4	-16.4
D772	-97.6	-94.5	-3.1
N774	-103.1	-92.9	-10.1
S775	-92.8	-94.3	1.5
I777	-118.3	-94.1	-24.2
R778	-125.1	-91.5	-33.6
Y779	-123.9	-93.9	-30.0
R780	-123.2	-93.0	-30.2
C781	-107.3	-93.3	-14.0
G783	-110.7	-93.8	-16.9
K784	-93.6	-93.9	0.4
I788	-115.2	-92.5	-22.7
T790	-122.5	-94.1	-28.4
V791	-114.9	-93.6	-21.3
I793	-85.5	-94.4	8.8
N794	-94.8	-90.4	-4.3
G795	-99.5	-93.1	-6.4
R796	-80.5	-93.7	13.2
W797	-88.4	-93.9	5.5
D798	-108.7	-94.1	-14.6
E800	-116.5	-93.3	-23.2
V801	-88.7	-92.8	4.1
C803	-90.6	-93.1	2.5
S804	-76.5	-93.1	16.6

Table 5.2: The values of ^{15}N splittings for the backbone HN bond as calculated from the acquired IPAP spectra.

restrained atoms in the primary sequence. Intra-residue NOE restraints occur between nuclei in the same residue; sequential NOE restraints between nuclei in adjacent residues; short-range NOE restraints between nuclei which are 2-4 residues apart; and long-range NOE restraints between nuclei which are more then 4 residues apart. Table 5.3 shows the distribution of unambiguous NOE restraints within these classes.

Restraint Type	**Number**
Total NOE restraints	1085
Total Unambiguous NOEs	607
Total Ambiguous NOEs	478
For Unabiguous NOEs	
Intraresidue	308
Sequential	195
Medium Range, $2 \leq (i-j) \leq 4$	54
Long Range, $5 \leq (i-j)$	113
Total	607

Table 5.3: NOE statistics for the ^{15}N labeled sample of fH-13. The numbers for unambiguous restraints include only unique restraints.

Two rounds of structure calculation were carried out in total, with 100 structures calculated per round. Both unambiguous and ambiguous restraints were used from the beginning. The NOE restraints were analysed, filtered and checked after each round (as described in section 3.5.2). A plot of E_{noe} and E_{total} for each structure from the first round is shown in Figure 5.12. As can be seen from this plot, there was no clear sign of the structures having converged yet. The lack of convergence at this stage suggests that there may be many conformations of the protein in the lower energy structures. Therefore a broad selection of conformations needed to be taken for filtering and checking. As can be seen in Figure 5.12, after the 63^{rd} structure the E_{noe} and E_{total} increased at a steeper gradient then before indicating that conformations beyond this point were incorrect, and so this was chosen as a cut-off point. Therefore, after the first round of calculation, the 62 structures with the lowest E_{noe} were used for further analysis.

A plot of E_{noe} and E_{total} for each structure from the second round is shown in Figure

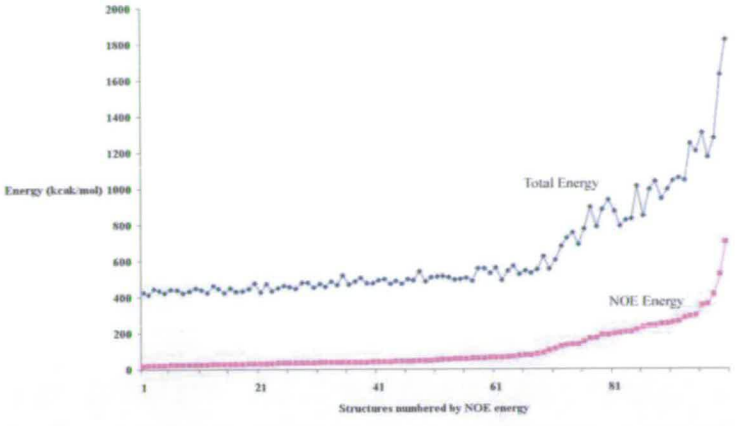


Figure 5.12: Total and NOE energy of structures of fH~13 after one round of structure calculation, ordered by their NOE energy. The structures show little sign of convergence.

5.13. The first 86 structures appear to have converged, although there in an abrupt increase in E_{noe} and E_{total} after the 70^{th} structure. Therefore, the 70 lowest energy structures by E_{noe} were used for furtehr analysis. There was no significant improvement in E_{noe} or E_{total} after this round. Additional rounds of structure calculation were tried. However no improvement was made from these additional rounds of structure calculations with NOE restraints and H bonds alone.

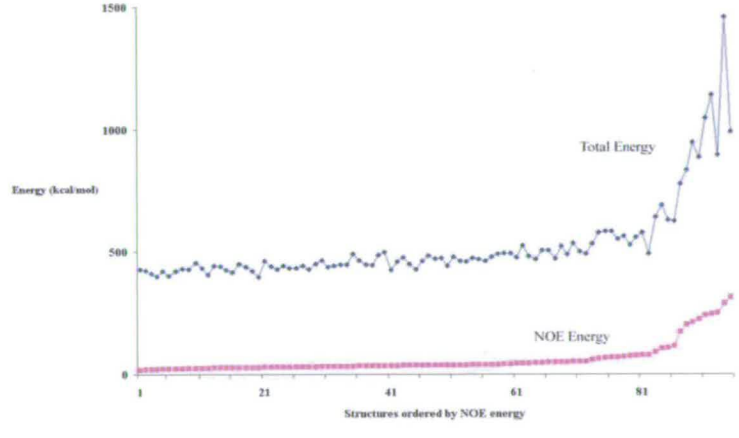


Figure 5.13: Total and NOE energy of structures of fH~13 after two round of structure calculation, ordered by their NOE energy. There is a step increase in E_{NOE} after the 70^{th} structure.

After this, the structures were refined using the RDC restraints before the final water refinement was undertaken (see section 3.5.4 for an explanation of these refinement steps). A plot of the E_{noe} for each structure after water refinement is shown in Figure 5.14. The first 20 structures appear to be converged, as after this the E_{total} increases at a steeper gradient as can be seen in Figure 5.14. Therefore the 20 structures with the lowest overall energy were selected. These are shown in Figure 5.15

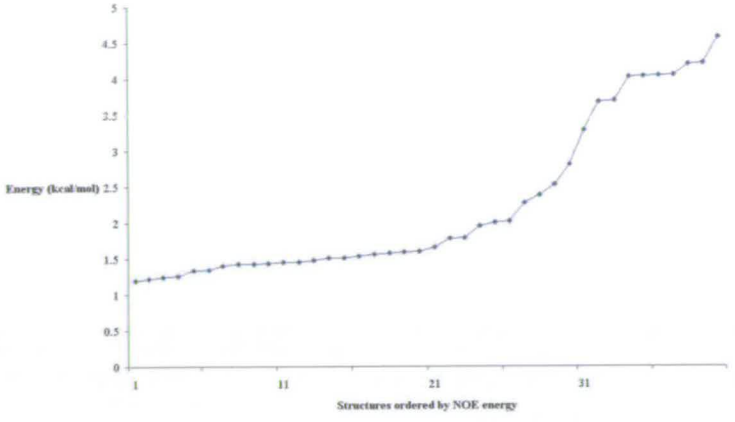


Figure 5.14: NOE energy of fH~13 structures after water refinement, ordered by their NOE energy. After the 20^{th} structure the E_{noe} increases dramatically.

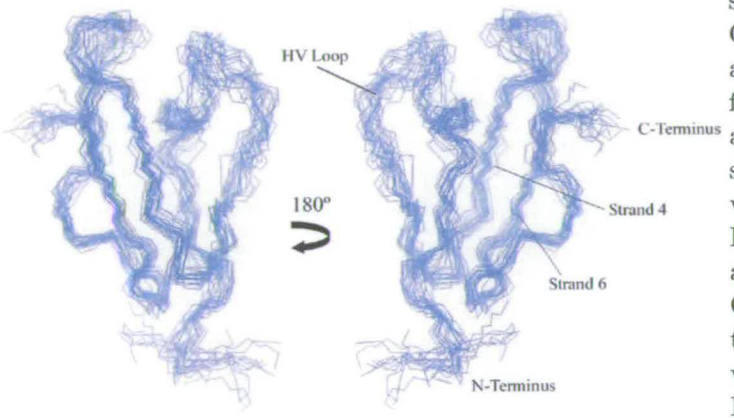


Figure 5.15: Overlay of 20 lowest E_{noe} structures of fH~13. Only the backbone atom traces are shown, fitted to the coordinate averaged structure (not shown). The hypervariable loop is labeled HV. β-strands 4 and 6 are labeled using the CCP module convention. The structures were generated using MOLMOL [94][95], which was also used to identify the secondary structure elements.

Various RMSD values (root mean square deviations) were calculated for this ensemble of structures. These are shown in Table 5.4 and Figure 5.16, the latter showing the RMSD of each residue for the ensemble. The RMSD values give an indication of how similar each structure in the ensemble is to each other. Backbone and all heavy atom RMSD values were calculated with and without the hypervariable loop (see section 1.1.4 for information on the hypervariable loop). The hypervariable loop was determined to be between residues I760-K769 since the residues in this region had higher then average RMSD values.

Residues	Backbone RMSD	Heavy atom RMSD
K752-S804	1.060	1.887
K752-E759, E770-S804	0.562	1.077

Table 5.4: Backbone atom and all heavy atom RMSD of fH~13. The values are given for the entire native sequence and for the native sequence without the hypervariable loop.

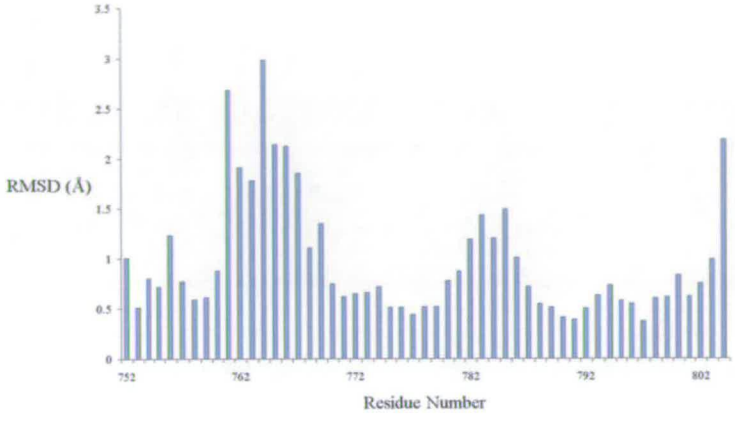


Figure 5.16: Graph showing backbone RMSD of each residue in the ensemble of structures.

The backbone RMSD was low once the hypervariable loop was removed from the calculation, suggesting good convergence between structures. The Ramachandran statistics for these structures also suggest that the backbone structure stereochemistry was reasonable once the hypervaribale loop was removed from the calculation. The results of the Ramachandran analysis of the ensemble is given in Table 5.5 (generated using PROCHECK [96]) and a plot is shown in Figure 5.17 (generated using MOLMOL [94][95]).

A good quality ensemble of structure would be expected to have over 90% of it's residues in the most favoured and additionally allowed regions of the plot. These structures just fall sort of this, even with the hypervariable loop removed from the calculations.

The RMSD for the entire ensemble including all heavy atoms (not just the backbone atoms) is twice as high. This means that, while the backbone of the molecule shows good convergence throughout the ensemble, the residue side-chains do not converge as well. Therefore an attempt was made to improve the structure of fH~13. A double labeled (^{13}C, ^{15}N) sample was made in order to generate more NOE restraints with

Residue Type	Number	Percentage
Without Reisudes 761-769 nor the precursor		
Residues in most favoured regions	428	56.3%
Residues in additional allowed regions	241	31.7%
Residues in generously allowed regions	64	8.4%
Residues in disallowed regions	27	3.6%
With residues 752-804		
Residues in most favoured regions	490	51.0%
Residues in additional allowed regions	341	35.5%
Residues in generously allowed regions	90	9.4%
Residues in disallowed regions	39	4.1%

Table 5.5: Ramachandran analysis of the residues in the ensemble of 20 lowest E_{noe} energy structures of fH~13. Calculated using PROCHECK [96].

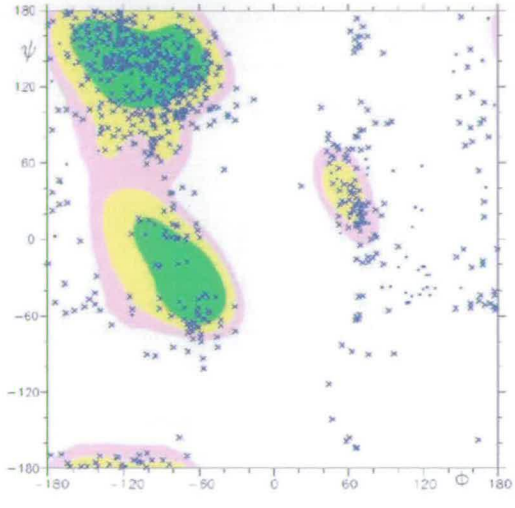


Figure 5.17: Ramachandran plot of the 20 lowest E_{noe} energy structures of fH~13. Dots indicate a Glycine residue, while crosses indicate each other residue. The colour changes from green, yellow, pink and white indicate decreasing favourability of regions. The diagram was made using MOLMOL [94][95].

the hope that these additional restraints would improve both the side chain convergence and the Ramachandran statistics. With the double-labeled sample it would be possible to record an ^{13}C-edited 3D NOESY spectrum, which can be used to collect NOE restraints between sidechain atoms with a greater resolution and less ambiguity than is possible with a homonuclear 2D NOESY spectrum. It is not possible to collect sidechain to sidechain NOEs with an ^{15}N-edited spectrum beyond those residue whose side chains contain nitrogen.

Chapter 6

Structural Determinaton of fH~13: The ^{13}C,^{15}N-labeled fH~13 Sample

6.1 Purification of The ^{13}C,^{15}N-labeled fH~13 Protein

While the clone for the ^{13}C,^{15}N-labeled fH~13 protein was prepared by Claire Egan, I undertook the fermentation and purification of this sample. The method of fermentation used is explained in section 3.1.1. The method for purification was broadly similar to that used for the previous samples, with changes outlined below.

For the unlabeled and ^{15}N-labeled samples, Claire Egan had first concentrated and buffer-exchanged the raw supernatant prior to the first stage of purification; gravity flow chromatography through SP-Sepharose at pH 5.0.

In an attempt to speed up the purification process a new method was tested; that is to dilute the raw supernatant in binding buffer in order to lower the salt concentration prior to the first stage of purification. Previously, salt was removed from the supernatant by first concentrating it to less then 100 ml, which is a lengthy process. This is a small enough volume to allow the concentrated supernatant to be buffer-exchanged into the appropriate binding buffer. By diluting the supernatant, the salt concentration is lowered without the need for concentration or buffer-exchange, although the volume of the sample to be loaded onto the SP-Sepharose column is considerably bigger.

To test whether diluted supernatant would bind to the SP-Sepharose beads three samples were prepared and labeled A, B and C.

- Sample A consisted of 1 ml of raw supernatant; 5 ml of 20 mM potassium phosphate buffer, pH 5.0 (the binding buffer for the SP-Sepharose column); and 1 ml of washed SP-Sepharose beads.
- Sample B consisted of 1 ml of raw supernatant; 5 ml of 20 mM potassium phosphate buffer, pH 5.0, with 1 M sodium chloride (the eluting buffer for the SP-Sephaorse column); and 1 ml of washed SP-Sepharose beads.
- Sample C consisted of 1 ml of raw supernatant and 5 ml of distilled water, thus acting as a control and a reference.

The samples A, B and C were all run on SDS-PAGE, with samples A and B run both with and without SP-Sepharose beads. The results are shown in Figure 6.1. For sample A, a protein around 6.5 kDa has stuck to the beads in the presence of binding buffer (Figure 6.1 (ii) lane A) while it is virtually absent in the surounding buffer (Figure 6.1 (i) lane A). For sample B, a protein around 6.5 kDa is present in the eluting buffer (Figure 6.1 (i) lane B) but is no longer stuck to the SP-Sepharose beads (Figure 6.1 (ii) lane B). Thus the gels demonstrated that protein of mass $\approx$6.5 kDa stuck to the SP-Sepharose beads in the presence of the binding buffer, and is eluted in the presence of eluting buffer.

Thus, the 0.5 liters raw supernatant was diluted with 3.5 liters of distilled water. This was then passed through (under gravity flow) a column of $\approx$3.5 ml of SP-Sepharose beads in a 1.13 cm diameter column, with a slow flow rate of $\approx$1ml/min. The through-flow from the column had an average OD_{280} reading of 0.332, as very little protein had stuck to the column. After passing the diluted supernatant through the column, 10 ml of eluting buffer was passed through the column and an OD_{280} reading of the eluant was determined to be 0.718. Samples of the through-flow and the eluant were both run on a SDS-PAGE, the results of which are shown in Figure 6.2.

Figure 6.1: SDS-PAGE showing the results of the test of the dilution method for the SP-Sepharose gravity flow column. For (i), 300 μl of each sample was TCA precipitated for the gel. Samples A and B did not contain SP-Sepharose beads from their samples as the beads had been spin-out. Therefore only the surrounding buffer, and any protein not stuck to the beads, was loaded onto these lanes. For gel (ii), 15 μl of SP-Sepharose beads from both samples A and B were washed in their respective dilution buffers to ensure only protein that had stuck to the SP-Sepharose beads would be present on the gel. Again, 300 μl of Sample C was TCA precipitated for this gel. On both gels (i) and (ii), the samples on the gel are marked as follows:(M) Marker; (A) sample A; (B) sample B; (C) sample C. The protein molecular weight marker used was the New England Biolabs Broad Range Prestained Protein Marker P7708G

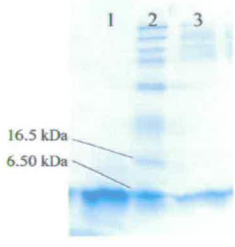


Figure 6.2: SDS-PAGE showing the products of the gravity flow SP-Sepharose column. The order of samples on the gel are found in lanes: (1) 1ml of column flow through (TCA precipitated); (2) protein marker; (3) 50μl of column elution. The protein marker used was the New England Biolabs Broad Range Prestained Protein Marker P7708S

As the OD_{280} readings demonstrate, only a small amount of protein had bound to the column, leaving a lot of protein still in the flow through. This was unfortunate but could not have been predicted from the gels in Figure 6.2. SP-Sepharose has a binding capacity of 120 mg of BSA per ml of gel (quoted from the Sigma Aldrich information booklet for SP-Sepharose), therefore unless there were large amounts of yeast protein in the supernatant, the column was unlikely to be overloaded. Diluting the flow through further and lowering the pH were both measures that could be taken to try and increase binding of the protein to the column.

Therefore, the 4.2 liters of flow through from the SP-Sepharose column was further

diluted with 1666 ml of distilled water and 320 ml of 0.2 M potassium phosphate. The flow through passed back through the gravity-flow SP-Sepharose column twice more. Protein was eluted from the column via 8 ml of eluting buffer after each passing. The OD_{280} readings of each 8ml batch was 0.404 and 0.355. This enabled enough protein to be extracted from the diluted supernatant for an NMR sample.

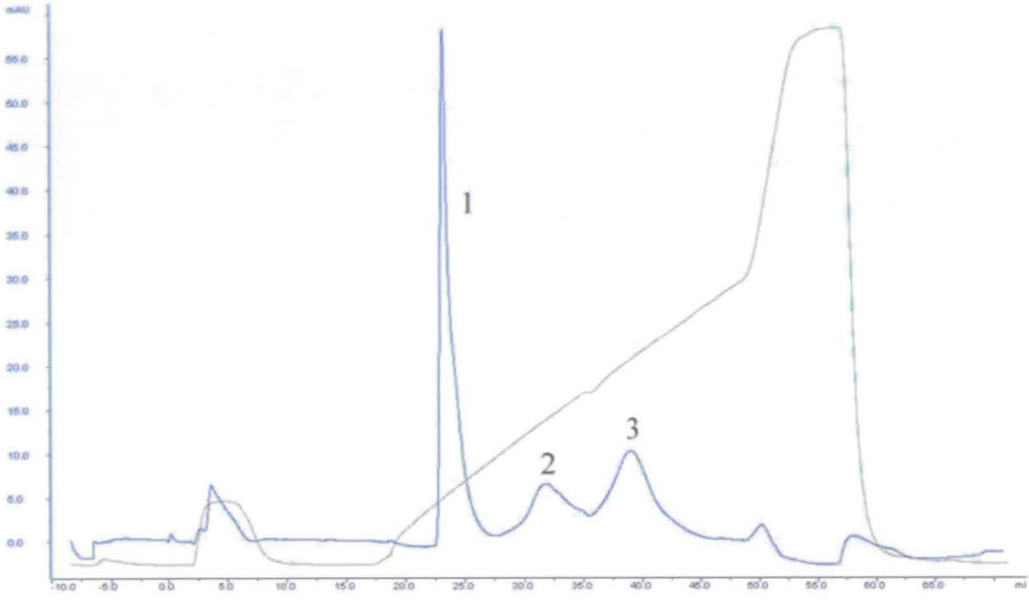


Figure 6.3: An FPLC trace of the ^{13}C labeled sample of fH~13 injected onto a porous heparin column. A subsequent salt gradient was used to elute the protein from the column, reaching a maximum salt concentration of 1 M. The OD_{280} trace is shown in blue and the conductivity (directly proportional to the salt concentration) is shown in brown. The sample was injected at 0 ml. Three separate species can be seen eluting from the column. These fractions were labeled 1, 2 and 3 in the order of their elution. The first fraction was confirmed to be fH~13.

For the final purification step, the elutions from the gravity-flow SP-sepharose column were diluted with 20 mM potassium phosphate buffer pH 5.0, with one part eluant to five parts buffer (to dilute the sodium chloride in the eluant). The diluted eluant was then put through a 1 ml porous heparin column using FPLC. 4.5 ml of diluted eluant was injected onto the column at a time as this was the size of the largest loop available. This required 29 injections for all the diluted eluant to pass through the column. The protein was then eluted from the column using a gradient of 20 mM potassium phosphate buffer pH 5.0 with 0-1 M sodium chloride. Three separate species were observed

eluting from the column (see Figure 6.3. These were collected separately and labeled fractions 1, 2 and 3 in the order they eluted from the column. The three fractions were run on SDS-PAGE, the results of which can be seen in Figure 6.4. From this it was determined that fraction 1 was most likely to be fH~13. In total, the three batchs of eluent from the SP Sepharose column (130 ml after dilution) were put through the porous heparin column using FPLC, generating 143 ml of faction 1.

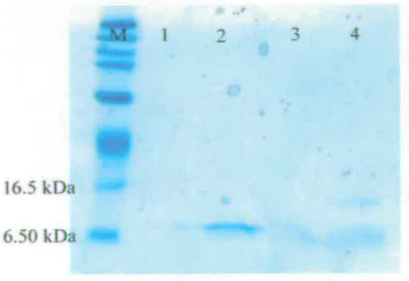


Figure 6.4: SDS-PAGE showing products of the FPLC porous heparin column. The samples on the column are found in lanes; (M) protein marker; (1) 1 ml of column through-flow; (2) 1 ml of fraction 1; (3) 1 ml of fraction 2; (4) 1 ml of fraction 3. The protein marker used was the New England Biolabs Broad Range Prestained Protein Marker P7708S

For the initial NMR sample, the 143 ml of fraction 1 (with an average OD_{280} of 0.020) were buffer exchanged into 20 mM sodium phosphate buffer, pH 6.0 with 10% (w/v) D_2O, and concentrated to $\approx$350 μl. Sodium Azide was added to 0.01%. The OD_{280} reading for the final sample was 1.643, and from this the concentration of fH~13 was determined to be 0.130 mM. The sample was placed in a Shigami tube. The sample purity was confirmed by SDS-PAGE (Figure 6.5) and a 1D ^1H NMR spectrum (data not shown).

Figure 6.5: SDS-PAGE showing the initial NMR sample used, the left lane being the protein marker and the right being 10μl of NMR sample. The protein molecular weight marker used was the New England Biolabs Broad Range Prestained Protein Marker P7708S

6.2 Assignment

Ten NMR experiments were acquired in total for the assignment of the ^{13}C,^{15}N-labeled sample of fH~13. These were: CBCANH and CBCA(CO)NH experiments

for backbone assignment; H(C)(CO)NH-TOCSY, (H)C(CO)NH-TOCSY and HCCH-TOCSY experiments for side-chain assignment; ^{1}H-^{13}C HSQC, ^{1}H-^{13}C CT-HSQC, (HB)CB(CGCD)HD and (HB)CB(CGCDCE)HE experiments for aromatic residue assignment; and an ^{13}C NOESY-HSQC for assignment assistance and restraint generation. The uses of these experiments is described in detail in section 2.6.1.

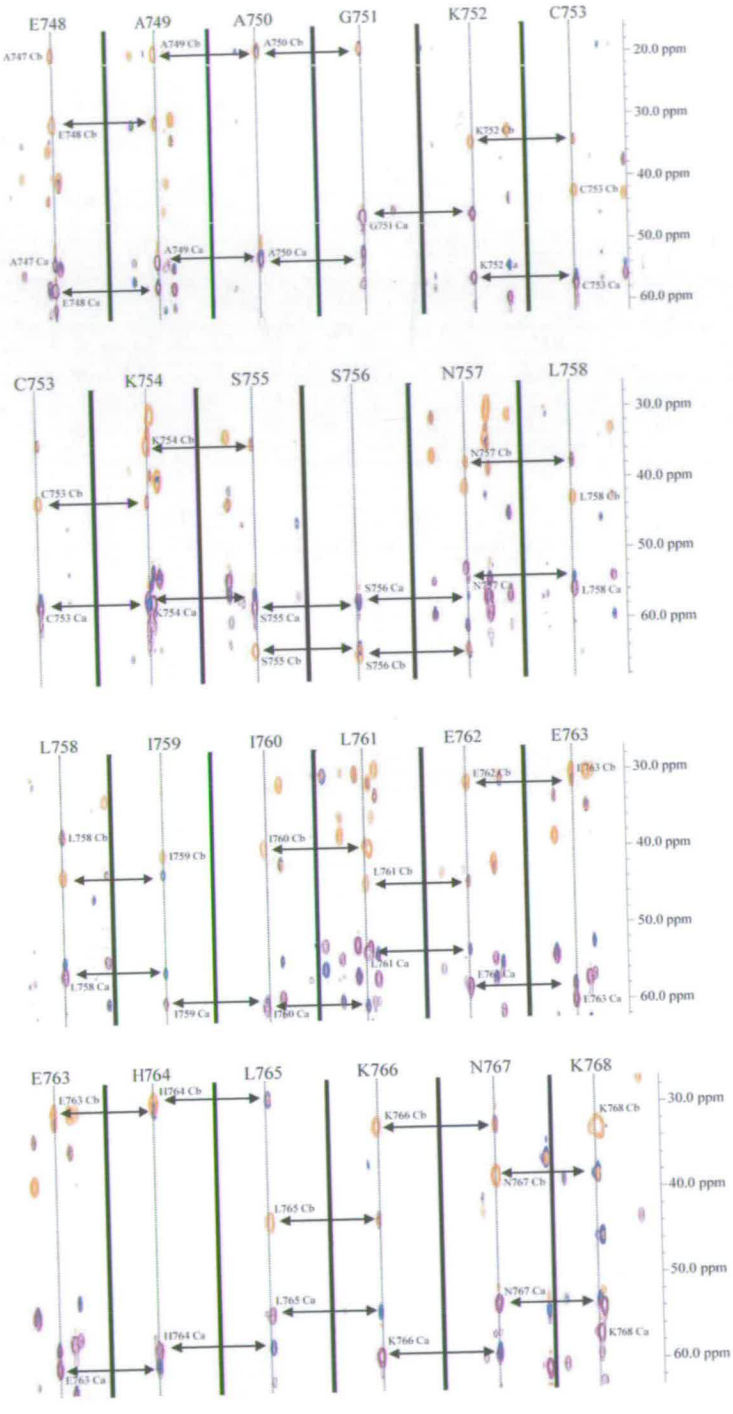
These spectra were also used to find previously unassigned ^{1}H and ^{15}N resonances and identify any assignments which may have been incorrectly assigned using only the unlabeled and the ^{15}N-labeled samples. The sample conditions used for the double-labeled sample were the same as those used for the labeled and ^{15}N-labeled samples. A full list of resonance assignments from the ^{13}C,^{15}N-labeled fH~13 samples can be found in Appendix C.

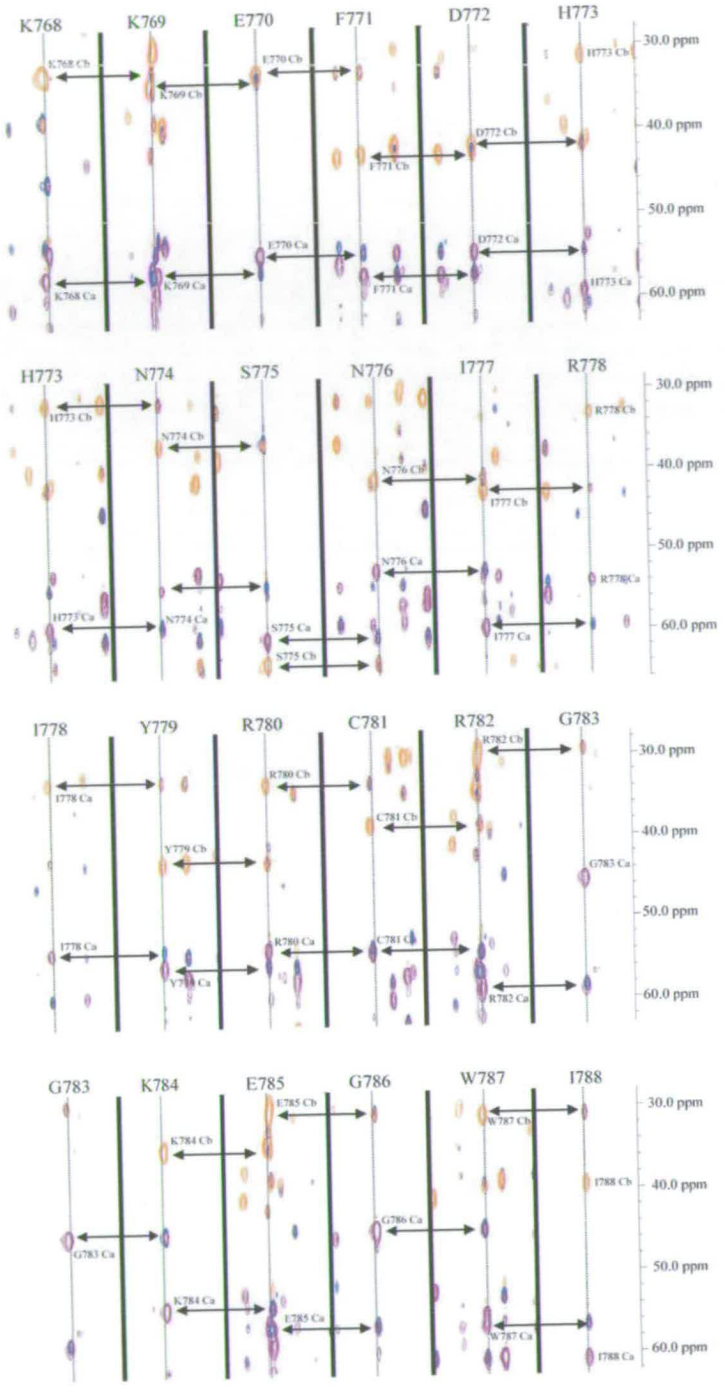
6.2.1 Backbone Assignments

As the H^N and N^H nuclei had already been assigned for most residues, only the backbone carbon nuclei needed assignment. This could be achieved by sequential walking using only the CBCANH and CBCA(CO)NH spectra, whereas without the previous assignments more spectra would have been required. Assignments were further checked using the (H)C(CO)NH-TOCSY and HCCH-TOCSY spectra. The number of side chain resonances identified in these experiments served to confirm the identity of the assigned residues. This method is described in detail in section 2.6.1. Figure 6.6 shows the assignment trail of the fH~13 backbone through the CBCANH and CBCA(CO)NH spectra. By this method the C^α and C^β nuclei were assigned for all residues with no exceptions. The backbone CO resonances could not be assigned as they have no attached proton.

As proline has no H^N proton there was a break in the sequential walk where the proline residue occurred. However, just as it was possible to assign the S804 C^α and C^β nuclei with no subsequent residue, it was possible to assign D798. The process of sequential walking is described in section 2.6.1.

As the backbone assignment approached the N-terminus discrepancies arose between


E748
A749
A750
G751
K752
C753
A747 Cb
A749 Cb
A750 Cb
E748 Cb
K752 Cb
C753 Cb
G751 Ca
A747 Ca
A749 Ca
A750 Ca
K752 Ca
C753 Ca
E748 Ca
20.0 ppm
30.0 ppm
40.0 ppm
50.0 ppm
60.0 ppm
C753
K754
S755
S756
N757
L758
K754 Cb
N757 Cb
C753 Cb
L758 Cb
S756 Ca
C753 Ca
K754 Ca
S755 Ca
L758 Ca
S755 Cb
S756 Cb
30.0 ppm
40.0 ppm
50.0 ppm
60.0 ppm
L758
I759
I760
L761
E762
E763
E762 Cb
I763 Cb
L758 Cb
I759 Cb
I760 Cb
L761 Cb
L761 Ca
E762 Ca
L758 Ca
I759 Ca
I760 Ca
E763 Ca
30.0 ppm
40.0 ppm
50.0 ppm
60.0 ppm
E763
H764
L765
K766
N767
K768
I763 Cb
H764 Cb
K766 Cb
K768 Cb
N767 Cb
L765 Cb
N767 Ca
L765 Ca
K768 Ca
H764 Ca
K766 Ca
E763 Ca
30.0 ppm
40.0 ppm
50.0 ppm
60.0 ppm


K768
K769
E770
F771
D772
H773
30.0 ppm
40.0 ppm
50.0 ppm
60.0 ppm
H773
N774
S775
N776
I777
R778
I778
Y779
R780
C781
R782
G783
G783
K784
E785
G786
W787
I788

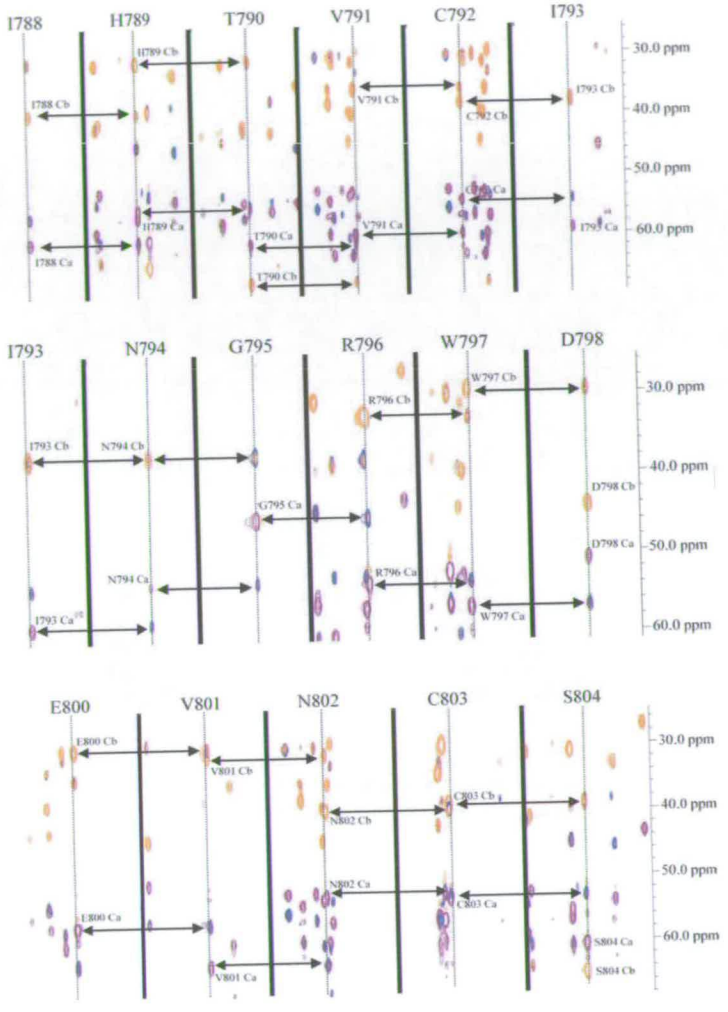


Figure 6.6: CBCANH and CBCA(CO)NH strips for fH~13 with sequential walk through the backbone for all residues except P799, which has no H^N proton and so no strip in these spectra. In the CBCA(CO)NH spectrum (blue peaks), the H^N strip shows the C^β and C^α nuclei of residue *(i -1)*. The CBCANH spectrum (purple positive peaks; orange negative peaks) shows these peaks and also the C^β and C^α nuclei of residue *(i)*.

the DNA sequence as supplied by Claire Egan and the NMR data. The sequence of residues for the secretion signal peptide is generally EAGA, sometimes preceded by an additional EA. Using the ^{15}N labeled sample, this signal sequence had been assigned AEAGA. The NOESY data from both homonuclear and ^{15}N edited spectra did not disagree with this assignment. NOE crosspeaks could be see from the H^N of residue *(i)* to the H^α of residue *(i-1)* for all residues in this sequence. However, the ^{13}C-edited spectra did disagree with this assignment. Sequential walk using the CBCANH and CBCA(CO)NH spectra proved the sequence to be AEAAG. The spin system assigned G750 became G751 and the spin system assigned A751 became A750. This explained why no assignments could be found for E748, nor for the H^N of A749 (as E748 never existed and A749 in fact was the N-terminus of the molecule).

6.2.2 Side-chain Assignments

The H(C)(CO)NH-TOCSY, (H)C(CO)NH-TOCSY and HCCH-TOCSY spectra were used to extend side-chain ^{13}C assignments beyond those obtained from the sequential walk. This was done according to the method described in section 2.6.1. Various non-protonated side-chain carbon nuclei could not be assigned. These nuclei are listed in Table 6.1 by their amino acid residue.

Amino Acid	*Unassigned Nucleus*	*Residues Affected*
Arginine	C^ζ	778 780 782 796
Asparagine	C^γ	757 767 774 776 794 802
Aspartate	C^γ	772 798
Glutamate	C^δ	748 762 763 770 785 800
Histidine	C^γ	764 773 789
Phenylalanine	C^γ	771
Tryptophan	C^γ, $C^{\delta 2}$ & $C^{\epsilon 2}$	787 797
Tyrosine	C^γ & C^ζ	779

Table 6.1: Carbon nuclei that could not be assigned by the method described in section 2.6.1. These nuclei have no attached protons and so do not show crosspeaks in the acquired spectra.

There were some other carbon nuclei that could not be assigned. The C^ϵ chemical shift

of K754 was not assigned. No corresponding peak was present in the (H)C(CO)NH-TOCSY. It is possible that the TOCSY transfer was not sufficiently efficient for the magnetization from this nucleus to reach the N^H nucleus of S755.

The three leucine residues were each missing one carbon assignment, where no distinct peak could be found in the (H)C(CO)NH-TOCSY spectrum. L759 and L765 were both missing assignments for their C^γ nuclei. The average chemical shift for this nucleus is similar to that of the C^δ nuclei. It is possible that the chemical shifts of the missing assignments are similar to those of their own C^δ nuclei. This would result in their peaks becoming overlapped in the (H)C(CO)NH-TOCSY spectrum. In the case of L765, the crosspeak for the $C^{\delta 1}$ nucleus had an elongated shape, suggesting that the C^γ chemical shift overlapped with the $C^{\delta 1}$ shift. For L759 there was no obvious candidate for overlap. The leucine residue L761 was missing its $C^{\delta 2}$ assignment. In this case the corresponding peak in the (H)C(CO)NH-TOCSY spectrum overlapped with that of the C^γ nucleus, making accurate assignment impossible.

In a few cases, various proton assignments were found that had been missing from the assignment of the ^{15}N sample. The missing histidine assignments, $H^{\delta 1}$ and $H^{\epsilon 2}$ were not found for either of the two residues. However, the missing $H^{\epsilon 1}$ of H764 was found. This were identified using the ^{13}C-^1H HSQCs of the aromatic region, where there was a crosspeak between it and its attached carbon.

The lysine K769 was missing it's H^δ and H^ϵ assignments due to overlap in both homonuclear and ^{15}N edited spectra. In the H(C)(CO)NH-TOCSY experiment magnetization from these proton resonances is recorded from the N^H of the subsequent residue, E770, which is not overlapped in this spectra. The missing assignments for K769 could therefore be assigned.

The tryptophans, W787 and W797, were both missing assignments for their $H^{\epsilon 3}$ protons. For both residues, this proton was identified using the ^{13}C-^1H HSQCs of the aromatic region. No crosspeaks were found in the homonuclear or the ^{15}N edited spectra for these protons.

The isoleucine I777 was missing the assignment of it's $H^{\gamma 2}$ protons. Once the $C^{\gamma 2}$ chemical shift was found using the (H)C(CO)NH-TOCSY spectrum, the HCCH-TOCSY spectrum was then used to assign the missing $H^{\gamma 2}$ protons. The chemical shift of these methyl protons was found to be overlapped with that of the $H^{\gamma 11}$ protons from the same residue, preventing assignment in the homonuclear and ^{15}N edited spectra.

A number of nuclei had been incorrectly assigned in the homonuclear and ^{15}N edited spectra. It was now possible to make less ambiguous assignments using the ^{13}C edited spectra. Proton chemical shifts could be directly associated with the carbon nuclei they are bound to using the HCCH-TOCSY spectrum. This helped to remove ambiguities in the proton assignments that originated from the ^{15}N sample.

This was the case for both I788 and I793. For the ^{15}N sample, the $H^{\gamma 11}$ nucleus of H788 was assigned to the same chemical shift as that of $H^{\gamma 12}$ as no separate peak could be found. However, in the HCCH-TOCSY spectrum the nucleus was associated with a different chemical shift and it was reassigned. This new shift was similar to that of the $H^{\gamma 2}$ group of the same residue. This resulted in overlap in the homonuclear and ^{15}N edited spectra and the incorrect assignment of this nucleus. In I793 the assignments for H^{δ} and $H^{\gamma 2}$ groups were the wrong way round and needed to be swapped. These assignments were in a heavily overlapped region of the homonuclear spectra and so no COSY crosspeak could be seen between the H^{δ} group and the $H^{\gamma 1}$ nuclei.

Residue W797 was previously missing assignments for its H^{ζ} and the H^{η} nuclei. These were found when examining the ^{13}C-^{1}H HSQCs of the aromatic region. The crosspeaks between the F771 aromatic carbons and their attached protons did not have the expected carbon chemical shifts. This pointed to the fact that these peaks had been incorrectly assigned. Initially these shifts had been assigned to the aromatic protons of F771 because their spin system produced the expected pattern in the homonuclear COSY and TOCSY spectra (see Figure 6.7(A) and Figure 6.7(B)). Also, there were NOESY crosspeaks from this aromatic spin system to the H^{β} protons of F771. However, there was also a single NOESY crosspeak from the aromatic spin system to the

$H^{\beta 2}$ of W797. This spin system was now reassigned to the missing W797 nuclei based on its carbon resonances. The F771 resonances H^{δ}, H^{ϵ} and H^{ζ} were now missing their assignments. From the ^{13}C-^{1}H HSQCs of the aromatic region a candidate peak was found corresponding to the expected carbon chemical shift for this group. The proton and carbon chemical shifts for the F771 H^{δ}, H^{ϵ} and H^{ζ} were all very similar and produced overlapping peaks in the homonuclear and carbon spectra (see Figure 6.7(C)). There was no recognizable pattern produced in the homonuclear COSY and TOCSY spectrum because of the overlap. However, these protons did collectively produce NOESY crosspeaks to the H^{β} protons of F771 (which remain unassigned due to overlap with neighboring peaks). In total, 97.9% of the ^{13}C, ^{15}N and ^{1}H nuclei in fH~13 were assigned.

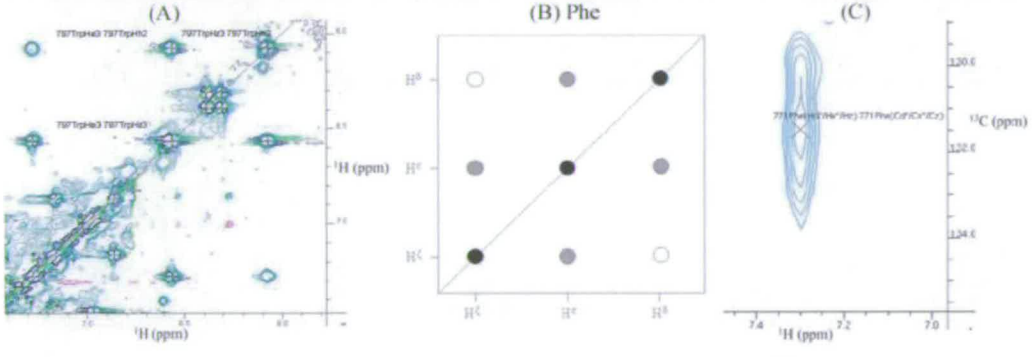


Figure 6.7: W797 was previously missing assignments for its H^{ζ} and the H^{η} nuclei. Initially these shifts had been assigned to the aromatic protons of F771 because their spin system produced the expected pattern (B) in the homonuclear COSY (green and blue peaks) and TOCSY (turquoise) spectra (A). The proton and carbon chemical shifts for the F771 H^{δ}, H^{ϵ} and H^{ζ} were all very similar and produced overlapping peaks in the ^{13}C-^{1}H HSQC (C).

6.3 Structure Calculation

6.3.1 Structure Calculation and Validation

Structures were calculated using the method described in section 3.5, with both ambiguous and unambiguous restraints included in the first round of structure calculations, as well as the hydrogen bond restraints (incorporated as distance restraints in their own file). In total, 677 unambiguous and 963 ambiguous NOE restraints were used for the calculation, comprising of restraints collected from the ^{13}C-edited experiments

as well as the restraints collected from the ^{15}N-edited and homonuclear experiments. This represents an increase of almost 50% in the number of restraints used compared to the ^{15}N labelled sample. Crosspeaks in the ^{13}C-edited 3D NOESY spectrum were only assigned in the F1 and F3 dimensions, therefore leading to the large increase in ambiguous NOE restraints. This meant that, as with the ^{15}N-labeled sample, many NOE restraints in the ambiguous data set were be repeated. The assignments for these restraints were refined during the filtering steps which took place after each round of calculation, removing these duplicate restraints. As with the ^{15}N-labeled sample, the unambiguous NOE restraints could be put into classes based on the distance between the two restrained atoms in the primary sequence (see section 5.5). Table 6.2 shows the distribution of unambiguous NOE restraints within these classes.

Restraint Type	**Number**
Total NOE restraints	1640
Total Unambiguous NOEs	677
Total Ambiguous NOEs	963
Intraresidue	308
Sequential	200
Medium Range, $2 \leq (i-j) \leq 4$	50
Long Range, $5 \leq (i-j)$	119
Total	677

Table 6.2: NOE statistics for the double labelled sample of fH-13. The numbers for unambiguous restraints include only unique restraints.

Six rounds of structure calculation were carried out in total with 100 structures calculated per round. A plot of E_{noe} and E_{total} for each structure from the first round is shown in Figure 6.8.

As can be seen from this plot, the structures show convergence in the first 39 structures. Therefore, the first 39 structures were chosen for further analysis. During filtering after the first round of calculation, assignments that contributed less then 1% to the volume of an ambiguously assigned crosspeak were rejected.

For analysing the structures from the second and third round steps the filtering level

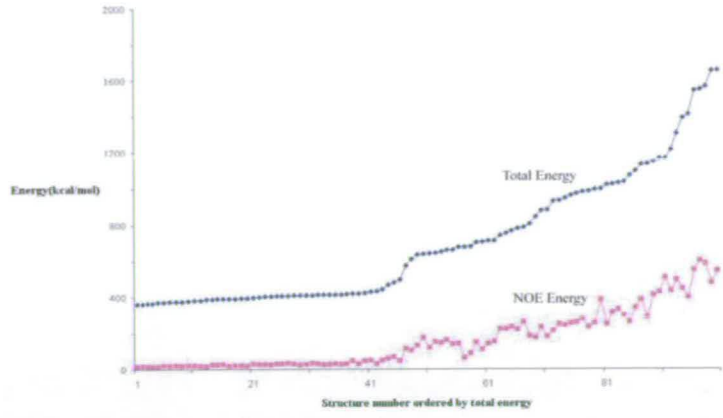


Figure 6.8: Total and NOE energy of structures of fH~13 after one round of structure calculation, ordered by their total energy. The first 39 structures show convergence.

was increased to 2%. Convergence was seen in the plot from round 2 after the first 70 structures, as can be seen in Figure 6.9.

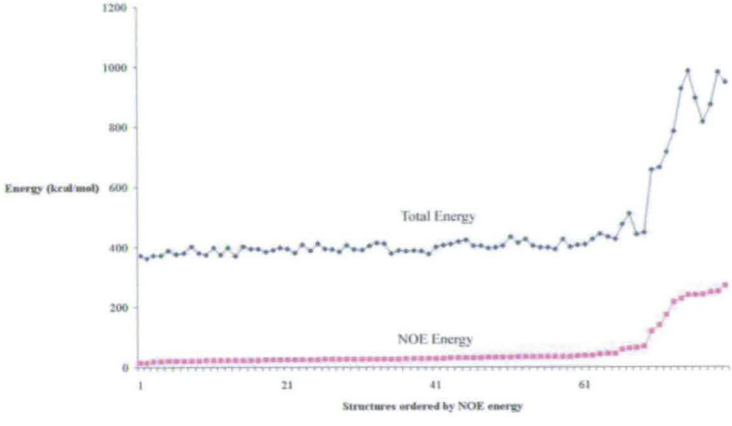


Figure 6.9: Total and NOE energy of structures of fH~13 after the second round of structure calculation, ordered by their NOE energy. The first 70 structures show convergence, while there is a jump in E_{noe} between the 67^{th} and the 68^{th} structures.

However there was a significant difference in the E_{noe} between the 67^{th} and the 68^{th} structures, and so the first 67 structures with the lowest E_{noe} were chosen for further analysis.

After each subsequent round of calculation the plots of E_{noe} and E_{total} for each structure showed signs of convergence, although generally a large number of structures had converged (59-65 structures). For the forth and fifth rounds of calculations the value for filtering was increase to 5% for filtering. There was little improvement in the E_{noe} and E_{total} plots between the fifth and the sixth round of calculations. Therefore after the sixth round the structures were refined using the RDC restraints before the final

water refinement was undertaken (see section 3.5.4 for an explanation of these refinement steps). As with the water refinement of the ^{15}N-labeled sample structures, the maximum of 50 structures with the lowest E_{total} were selected for water refinement. The final plot of E_{noe} for the water-refined structures is shown in Figure 6.10.

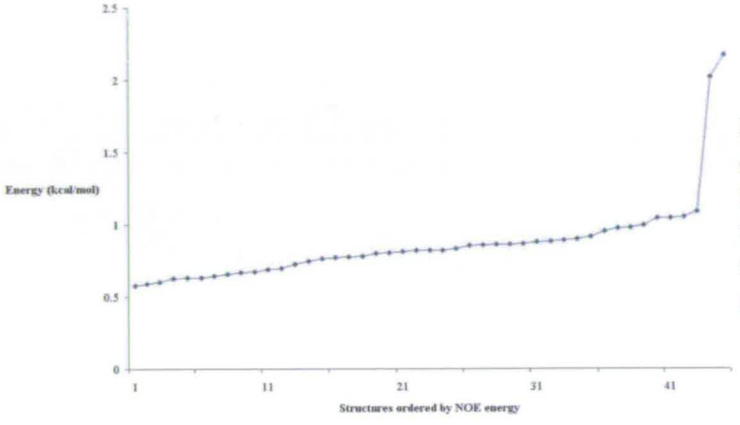


Figure 6.10: NOE energy of structures of fH~13 after water refinement. The first 43 structures show convergence.

The first 43 structures show convergence, but there was a large difference in E_{total} from the first to the last of these converged structures. Therefore, in a effort to select a smaller, more accurate group to represent the ensemble, the 20 structures with the lowest NOE energy were selected arbitrarily. These structures are shown in Figure 6.11

Various RMSD values for these structures are shown in Table 6.3 and Figure 6.12, the latter showing the RMSD of each residue for the ensemble. In this ensemble of structures the hypervariable loop appeared to be between E763-K769. As with the ^{15}N-labeled sample structures, identification of the hypervarible loop was based on the ensemble's backbone RMSD for this section and also by examining the backbone structures of the ensemble.

The shortening of the hypervarible loop between the two ensembles of structures could be the result of the increased number of NOE restraints connecting residue L761 to other residues for the double labeled sample.

There is also a significant improvement in the heavy atom and backbone RMSD (in-

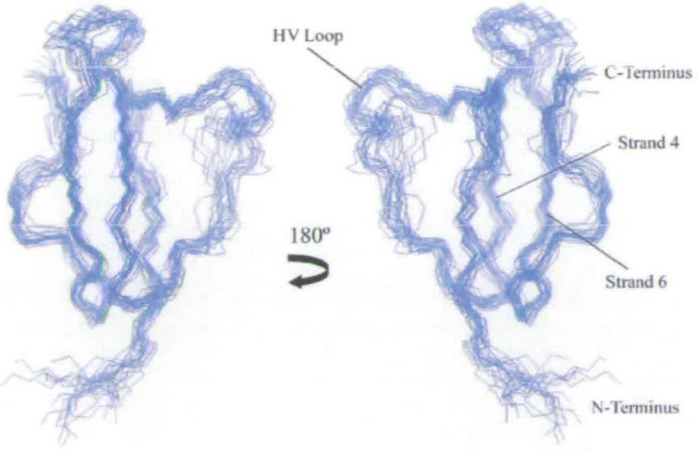


Figure 6.11: Overlay of 20 lowest E_{noe} structures of fH~13 for the doubled labeled sample. Only the backbone atom traces are shown, fitted to the coordinate averaged structure (not shown). The hypervariable loop is labeled HV. β-strands 4 and 6 are labeled using the CCP module convention. The structures were generated using MOLMOL [94][95], which was also used to identify the secondary structure elements.

Residues	**Backbone RMSD**	**Heavy atom RMSD**
K752-S804	0.837	1.512
K752-E762, E770-S804	0.577	1.140
K752-E759, E770-S804	0.528	1.061

Table 6.3: Backbone atom and all heavy atom RMSD of fH~13. The values are given for the entire native sequence; for the native sequence without the hypervariable loop; and for the native sequence with out residues 760-769, in order to make a comparison with the ^{15}N-labeled sample structures

cluding the hypervarible loop) between this ensemble and that calculated using the ^{15}N-labeled sample data (compare Table 6.3with Table 5.4). This suggests that the sidechain and backbone atoms are showing improved convergence throughout the whole molecule. In order to make a more direct comparison between the RMSD values for this ensemble and the previous ensemble (calculated from the unlabelled and ^{15}N-labeled sample data), RMSD values were calculated with the residues I760-K769 excluded.

These show a decrease in RMSD from those of the ^{15}N-labeled sample structures, but it is only small, suggesting that the improvement in convergence is mainly confined to the hypervaribale loop. Procheck statistics for these structures can be seen in Figure 6.13 (generated using MOLMOL [94][95]) and Table 6.4 (generated using MOLMOL

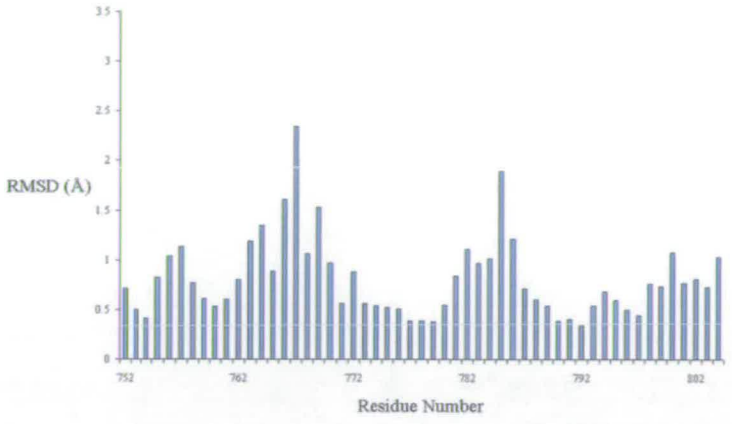


Figure 6.12: Graph showing RMSD of each residue in the ensemble of structures.

[94][95]). They appear worse then for the previous ensemble, with more then 5% of residues in the disallowed region. This suggests that while the overall convergence has improved slightly, this is at the expense of the backbone stereochemistry.

Residue Type	**Number**	**Percentage**
Without Reisudes 763-769 nor the precursor		
Residues in most favoured regions	471	57.4%
Residues in additional allowed regions	261	31.8%
Residues in generously allowed regions	45	5.5%
Residues in disallowed regions	43	5.2%
With residues 752-804		
Residues in most favoured regions	538	56.0%
Residues in additional allowed regions	312	32.5%
Residues in generously allowed regions	56	5.8%
Residues in disallowed regions	54	5.6%

Table 6.4: Ramachandran analysis of the residues in the ensemble of 20 lowest E_{noe} energy structures of fH~13. Calculated using PROCHECK [96].

A program called WHAT IF was also used to assess the quality for various geometric structural properties [97]. The ensemble of the 20 lowest energy structures was used for input, and WHAT IF produces a series of so-called Z-scores and RMS Z-scores as quality indicators. The Z-score relates the value of a parameter (eg. bond length) to a normalized Gaussian distribution derived from a database of structures. Z-scores smaller then -4 or larger then +4 are considered outliers. This means that the parameter value in the input structure is unlikely, not that it is necessarily in error. The

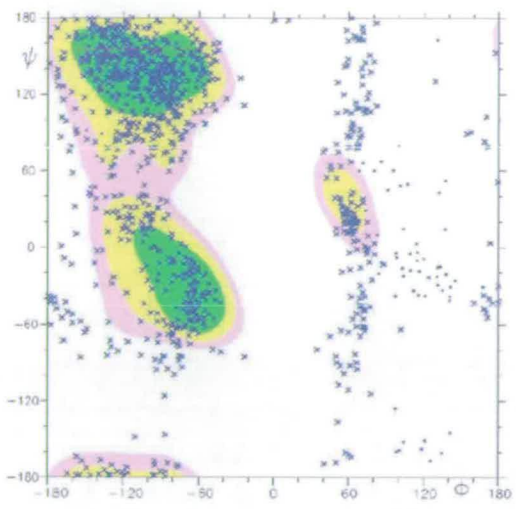


Figure 6.13: Ramachandran plot of the 20 lowest E_{noe} energy structures of fH~13. Dots indicate a Glycine residue, while crosses indicate each other residue found in the ensemble. The colour changes from green, yellow, pink and white indicate decreasing favourability of regions. The diagram was made using MOLMOL [94][95].

results from the WHAT IF checks are shown in Table 6.5.

Parameter	Value
Ramachandran Z-score	-6.327
RMS Z-score for bond lenths	1.015
RMSD in bond distance	0.020
RMS Z-score for bond angles	1.013
RMSD in bond angle	1.813
Standard deviation for Omega angle restraints	0.974
RMS Z-score for Improper dihedral distribution	1.260
Backbone Conformation Z-score	-8.453

Table 6.5: Structure quality results for the ensemble from the WHAT IF program.

These results show that for many geometric criteria the ensemble has values within excepted ranges. However, both the Ramachandran Z-score and the backbone conformation Z-score are very low. This is more evidence that the backbone stereochemistry is unusual in the ensemble.

Figure 6.14 shows the positions of the inferred hydrogen bonds in fH~13. Of the protected amide groups for which no binding partners were found (R778, Y779, R780, V791 and W797), R778, V791 and W797 have no obvious partners in the final structure ensemble. For residues Y779 and I788 there are potential binding partners. Y779

may form a hydrogen bond with the I788 CO group, while R780 may form a hydrogen bond with the I760 CO group.

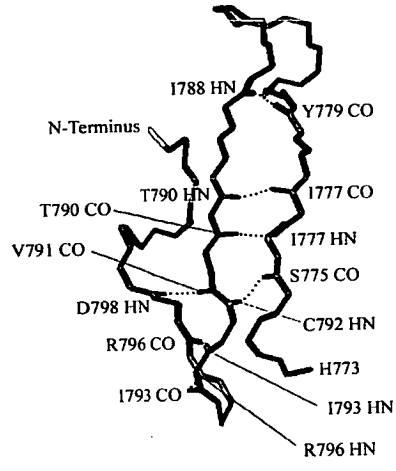


Figure 6.14: Position of the inferred hydrogen bond partners in fH~13. The HN partner is shown in red, while the CO partner is shown in blue. The diagram was made using MOLMOL [94][95].

6.3.2 Accuracy and Precision of The Final Structures

The number and magnitude of NOE violations that an ensemble of structures produces is one way to judge how well the ensemble fit the experimental data. NOE violations occur in a structure when the distance between two nuclei dictated by the NOE restraint is shorter then the actual distance in the calculated structure. Ideally, an ensemble of structures would have no NOE violations, and thus fit with the experimental data. However, in reality violations can occur in accurate ensembles where, for example, natural multiple stuctural conformations result in restraints that can't fit all structures in the ensemble [76].

After each round of structure calculations the number of violations occurring above a threshold is measured for the converged structures. A threshold of 0.3 Å was used for violations in the case of fH~13. Each violating restraint is then re-examined, usually by reverting back to the original spectrum from which the restraint was taken. Restraints may then be altered or removed if there is a rational reason for doing so. In the case of fH~13 there were two main reasons for altering or removing restraints. The crosspeaks for violating restraints were often found close to the diagonal in the

2D NOESY spectra, or were overlapping with a neighbouring peak in other ways. As mentioned in section 3.5.2 if a crosspeak significantly overlaps with another, especially a high intensity diagonal peak, this can result in the resulting restraint being placed in the wrong NOE class, and potentially producing a violation. Secondly, the protein was more fully assigned using the double-labeled sample. Therefore, due to missing assignments in the ^{15}N-labeled sample, some peaks which were unambiguously assigned for the ^{15}N-labeled sample now had ambiguities in their assignment for the double-labeled sample. These crosspeaks and other mis-assignments could potentially result in structure violations.

Three restraints produced violations in the final ensemble, although no single structure in the ensemble had more then one violation. The details of each violating restraint are summarized in Table 6.6.

Violating Restraint (nuclei-nuclei)	**Number of Structures Violation Occurs In**	**Restraint Distance (Å)**	**Violating Distance (Å)**
E762 H^N-(7.77ppm)	1	6.0	0.51
V801 H^N-V801 H^β	2	3.3	0.65
K752 $H^{\delta 1}$-(4.58ppm)	4	3.3	0.58

Table 6.6: Violations occurring in the final ensemble of 20 structures for fH-13. Two of the violating restraints were ambiguous in one of their dimensions. In these cases, a ppm value is quoted rather than an assigned nucleus. The violating distance is given as the RMS.

For each of these violations there was no rational reason to either remove or alter the original NOE crosspeak from which the violation was derived, with no obvious sign of crosspeak overlap or mis-assignment. The violations themselves are all above 0.5 Å above the restraint distance. Also, the RMSDs for the residue corresponding to the violating restraint are not above the average. However, the number of violating NOE restraints in the final ensemble of structures is low, indicating that the ensemble agrees reasonably well with the experiemntal data. Thus the ensemble of structures can be assumed to be reasonably accurate.

The RMSD values indicate that the ensemble is of relatively good precision in com-

parison with other CCP module structures.

Attempts were made to calibrate the restraints with the aim of improving the precision of the final structures. Usually the restraints are placed into four separate distance classes based on their intensities. However, the restraints can also be ordered via their intesity into a larger spectrum of distance classes based on the r^{-6} relationship between distance and peak intensity. This can be done through ARIA during the analysis of the ensemble after each round. ARIA scales each NOE intensity to one of many maximum distances from 2.2, 2.3, 2.4, 2.5 $\AA$ etc. up to 5.8 $\AA$. The CCPN Analysis program suite can also be used to place restraints into a broader spectrum of distance classes from 2.4, 2.5, 2.6 $\AA$ etc. up to 5.8 $\AA$. Both methods were tried (ARIA calibration for the ^{15}N-labeled sample data and CCPN Analysis for the double labeled sample data). However, both methods considerably increased the number of violations occuring per structure.

6.4 Relaxation Data

6.4.1 T_1 Relaxation Data

The T_1 relaxation values for fH~13 are shown in Figure 6.15. The residues that could not have their T_1 accurately determined, due to overlap, were 754, 761, 769, 782 and 802.

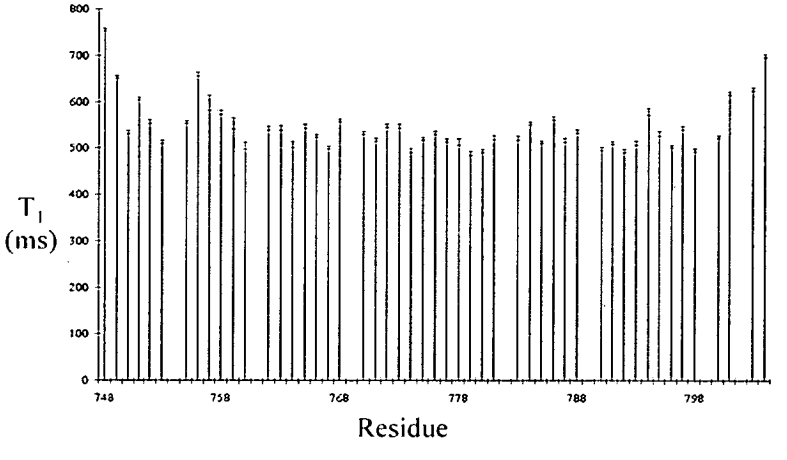


Figure 6.15: T_1 values for fH~13. Error bars are shown for each relaxation time.

The mean T_1 was 547 ms, with a standard deviation of 55 ms. Eight residues fall outside this range: 748, 749, 751, 756, 779, 801, 803, and 804. Of these, all but 756 and 779 are at the N-termius or C-termius of the molecule.

6.4.2 T_2 Relaxation Data

The T_2 relaxation values for fH~13 are shown in Figure 6.16. The residues that could not have their T_2 accurately determined, due to overlap, were 754, 761, 769, 782 and 802, the same residues as for the T_1 relaxation data.

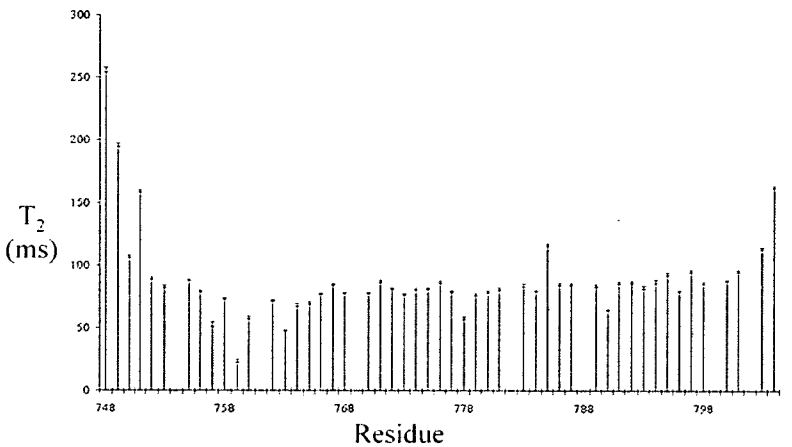


Figure 6.16: T_2 values for fH~13. Error bars are shown for each relaxation time.

The mean was 89 ms, with a standard deviation of 36 ms. Eight residues fall outside this range: 748, 749, 751, 757, 759, 763 and 804. Of these, all but 757, 759 and 763 are at the N-termius or C-termius of the molecule. 763 is the only residue with a particularly low T_2 value. This residue is at the start of the hypervariable loop.

6.4.3 HetNOE Relaxation Data

The HetNOE relaxation values for fH~13 are shown in Figure 6.16. The residues that could not have their T_1 accurately determined, due to overlap, were 761, 769, 782, 785 and 802.

The mean was 0.63, with a standard deviation of 0.23. Seven residues fall outside this range: 748, 749, 751, 756, 801, 803 and 804. Of these, all but 756 are at the N-termius or C-termius of the molecule. Unexpectedly, the residues that form the

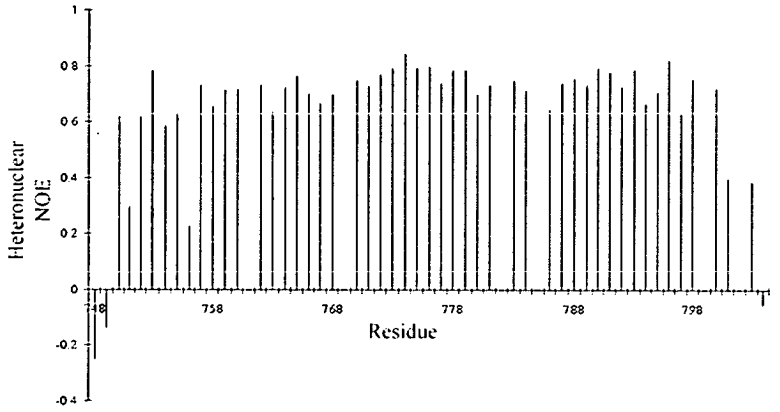


Figure 6.17: Heteronuclear NOE values for fH-13.

hypervariable loop in fH-13 show no significant variation in HetNOE from the rest of the protein. This suggests that the motions of the hypervariable loop are on the μs-ms timescale. This may explain why there are few NOE restraints defining this region, and is consistant with the relaxation data.

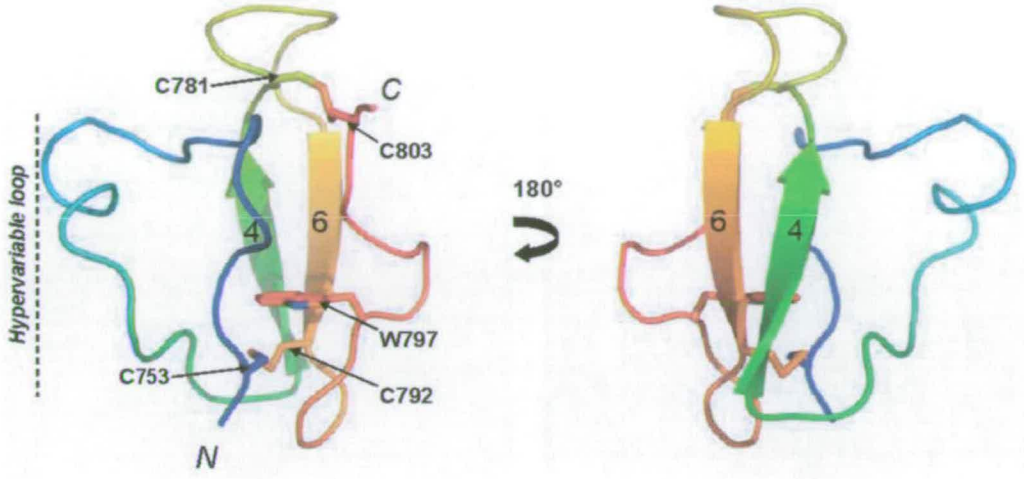
6.5 Structure Analysis

The analysis of the closest to mean structure of fH-13 described in this section, and it's comparison to other known CCP module structures, were done in collaboration with Dinesh Soares from our group. All Figures in this section (Figures 6.18, 6.19, 6.20, and 6.21) were also created in collaboration with Dinesh Soares. The method for the structure analysis is given in section 3.7.

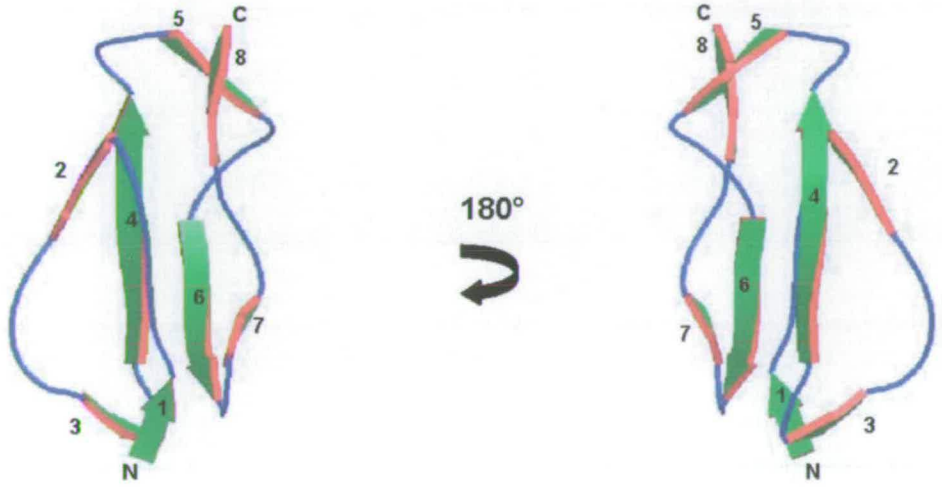
6.5.1 Results and Discussion

Factor H CCP-13: Sequence-Structure Correlation

CCP modules are characterised by having a compact hydrophobic core containing an almost invariant tryptophan residue. The core is enclosed in a framework of five extended segments that form β-strands for all or part of their lengths. The extended segments are aligned with the long axis of the module and are held together by two strictly conserved disulfide bridges. In almost all CCP-modules, a region that is highly variable in length, sequence and conformation, is inserted within the second extended region, projecting laterally, and is commonly referred to as the hypervariable loop [12]. As a representation of a typical CCP module structure, the structure of fH-5 is shown



(A)



(B)

Figure 6.18: The structure of fH~13 compared to fH~5 (not to scale). (A) A cartoon schematic representation of the structure of fH~13, coloured from blue to red (N- to C-terminus). The secondary structure elements (β-strands) have been assigned by default settings in PyMol [98] and labelled 4 and 6 [17]. The invariant cysteine and consensus tryptophan residues that characterise the CCP module fold are also labeled. (B) A Molscript [99] cartoon representation of fH~5, depicting a typical CCP-module structure. Again, the β-strands are numbered according to Wiles et al. [17]. The Figure 6.18B is taken from the Chapter 'Complement Control Protein Modules in the Regulators of Complement Activation' by D.C. Soares and P.N. Barlow [100].

in Figure 6.18B.

Among the 88-individual CCP-modules of the regulators of complement activation (including CR2 and VCP), module sequence-lengths range from 51 to 67 amino acid residues, when domain boundaries are considered from cysteine 1 to cysteine 4 (both inclusive) (http://www.bru.ed.ac.uk/~dinesh/Residue_lengths.pdf). fH~13 possesses the shortest known sequence (51 residues), while eight modules in CR1 and two in CR2 attain the longest sequence-length (67 residues per module) [100].

It can be seen in Figure 6.18A that fH~13 possess the expected disulfide-bonding pattern (Cys I to Cys III; Cys II to Cys IV) and consensus tryptophan, but lacks many overall 3D-structural features that characterise a "typical" CCP-module [12]. This is consistent with its short sequence-length and low sequence similarity to any other known CCP module.

Given that module 13 of factor H did not fall within any assigned sub-family-specific sequence cluster [18], it is not surprising that its structure exhibits unique features - only two β-strands are observed in the closest-to-mean structure (Figure 6.18A), corresponding to typically occurring β-strands 4 and 6 [17].

Figure 6.19 shows the unusual features of its sequence. These include the lack of consensus proline residues that typically occur at the third and/or fourth position in the stretch that follows the first cysteine, and at the second position before the fourth cysteine; the absence of any of the commonly occurring glycine residues between cysteines 1 and 2; and the lack of hydrophobic residues that predominate in β-strand 5 among other CCP module structures [100].

From the CE analysis, the most similar structure to fH~13 is fH~15 (RMSD = 3.09 Å over 56 residues), while the most dissimilar module is CR1~16 (RMSD = 5.05 Å over 56 residues). Overall, each of the other 34 structures compared with the fH~13 module have RMSD values > 3 Å. Ten structures have RMSD values in the range between 3.5 and 4 Å, and twelve greater than or equal to 4 Å. To place this in perspective, DAF~4

Compared CCP Module	fH-13
β2GPI-1	4.15 (55)
β2GPI-2	4.12 (51)
β2GPI-3	4.38 (48)
β2GPI-4	3.88 (59)
C1r-1	3.46 (55)
C1r-2	5.39 (38)
C1s-2	4.30 (66)
C4BPa-1	3.69 (57)
C4BPa-2	4.10 (52)
CR1-15	3.13 (47)
CR1-16	5.05 (56)
CR1-17	3.69 (57)
CR2-1	4.33 (59)
CR2-2	3.66 (49)
DAF-1	3.50 (58)
DAF-2	3.28 (56)
DAF-3	4.16 (61)
DAF-4	3.30 (51)
fH-5	3.47 (47)
fH-15	3.09 (56)
fH-16	3.69 (52)
fH-19	3.41 (45)
fH-20	3.91 (59)
Cis-GABA-2	4.15 (55)
Trans-GABA-02	4.00 (55)
IL15Rα-1	3.60 (56)
MASP2-1	3.57 (53)
MASP2-2	4.74 (56)
MCP-1	3.86 (50)
MCP-2	4.52 (59)
VCP-1	3.92 (59)
VCP-2	3.45 (53)
VCP-3	3.37 (56)
VCP-4	3.44 (56)

Table 6.7: Pairwise structural comparison of fH-13 versus all other individual CCP module structures based upon C^{α} RMSD values using the structural alignment program CE. All values in Angstroms (Å). Figures in brackets indicate the "Alignment length (gaps included)". The abbreviations used in the table and text are: β2GPI = β2-glycosylphosphatidylinositol-anchored protein (also called apolipoprotein H); C4BP = C4b-binding protein; CR = complement receptor; DAF = decay-accelerating factor; fH = factor H; GABA = gamma-aminobutyric-acid type B receptor, subunit 1 α; IL15Rα = interleukin-15 receptor α; MASP = mannan-binding lectin-associated serine protease; MCP = membrane cofactor protein; VCP = Vaccinia virus complement control protein. The CCP-module numbers follows the abbreviation. Some residues were not present (solved) in the electron density map for the C1r-02 module crystal structure, and this explains the short structural alignment length.

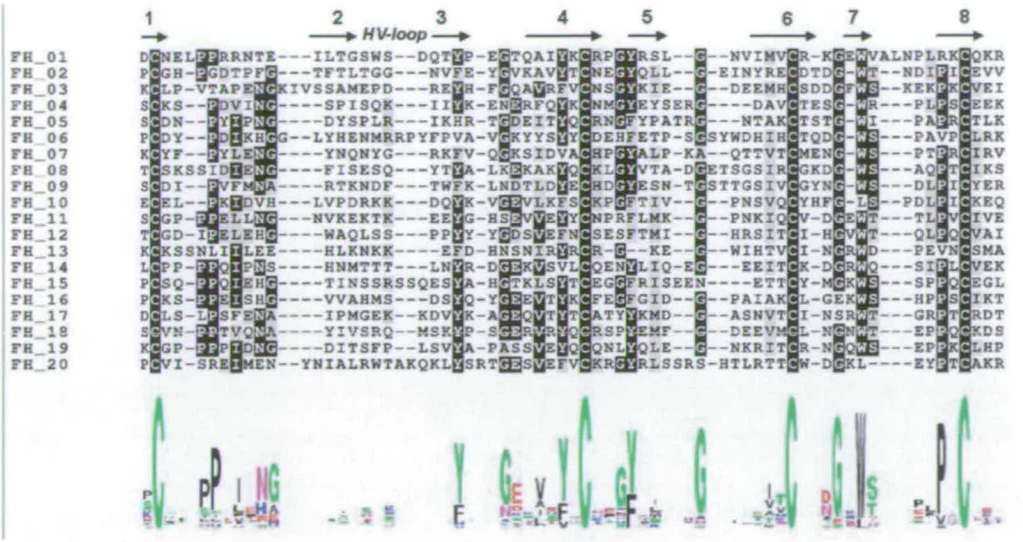
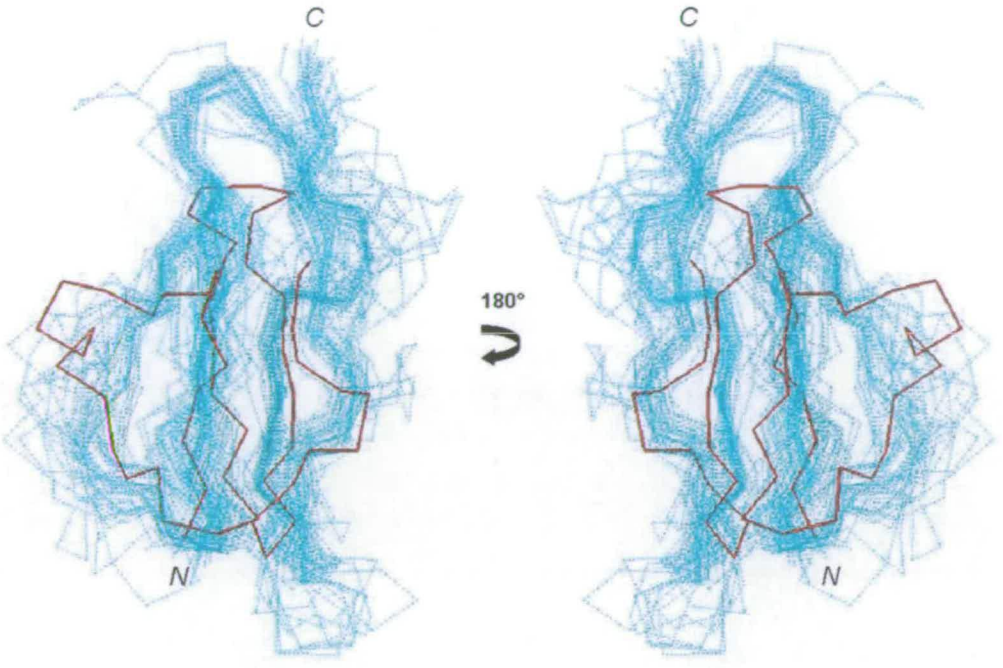


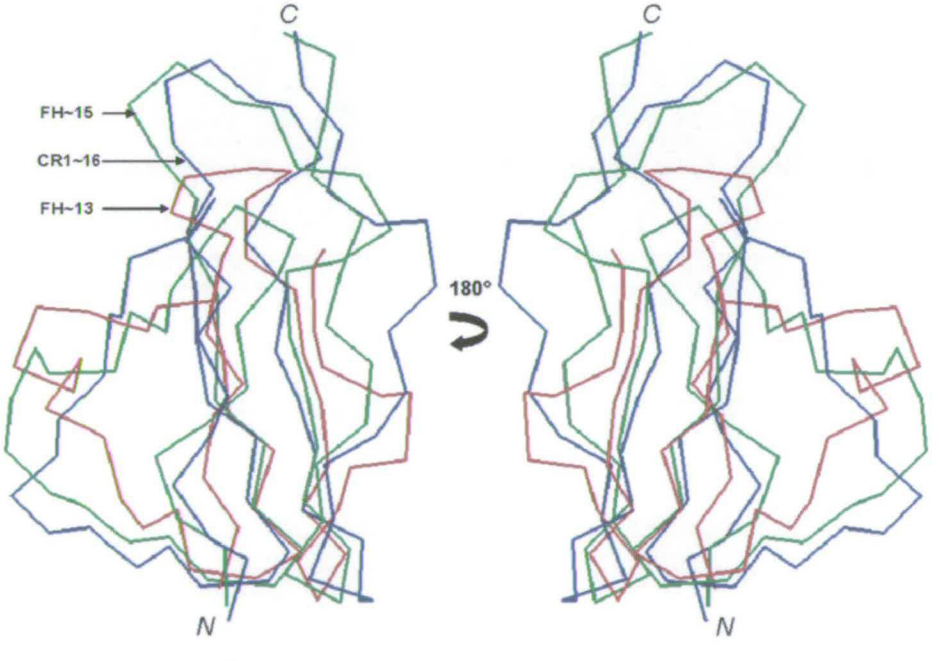
Figure 6.19: Multiple sequence alignment for factor H. The 20 factor H CCP modules were aligned using ClustalX [101], and shaded according to conservation using BOXSHADE version 3.21 [102]. The sequences have been rearranged according to module number. The typically occurring β-strands are indicated by arrows above the sequences, along with hypervariable loop (HV-loop). Consensus residues for each position in the alignment are shown below using WebLogo version 2.8.2 [103][104], with the greater the height of the residue, the greater the conservation. The strongly conserved residues compare well with the overall consensus for the entire RCA family reported in Soares and Barlow [100].

(which can be considered the most 'typical' CCP-module in such a comparison, with 33 out of 34 modules compared) produces RMSD values < 3 Å, with fH~13 being the only structure producing an RMSD value above 3 Å.

Examples of other 'divergent' modules include, fH~20, C4BPα~1 and VCP~1, which have RMSD values > 3 Å for each of 30, 20, and 17 other modules, respectively. Only three other pairs of modules for which comparisons have been made have RMSD values greater than 4 Å (fH~20 vs. C4BPa~1; fH~20 vs. CR1~15; and fH~20 vs. MASP2~2) [see: (Soares and Barlow 2005) and the updated table of values in DCS, Ph.D. thesis]. The two most dissimilar module structures, to date are fH~13 and CR1~16 (the only module-pair compared with RMSD > 5 Å, excluding C1r~2 from the analysis). Given these comparison results and the calculated overlay, structurally, fH~13 appears to be the most divergent CCP-module solved to date (Figure 6.20, Table 6.7).



(A)



(B)

Figure 6.20: The superposition of all CCP-module structures to data. (A) Two views rotated by 180 along the y-axis showing the C^{α}-trace overlay calculated by MAMMOTH-mult [85] for all the solved CCP module structures. fH~13 is shown as a red solid line, while the 34 other CCPs are traced with dashed lines in cyan. (B) The most similar (fH~15, green) and dissimilar (CR1~16, blue) CCP-module structures are compared to fH~13 (red). The images were created using RasMol version 2.7.3 [105].

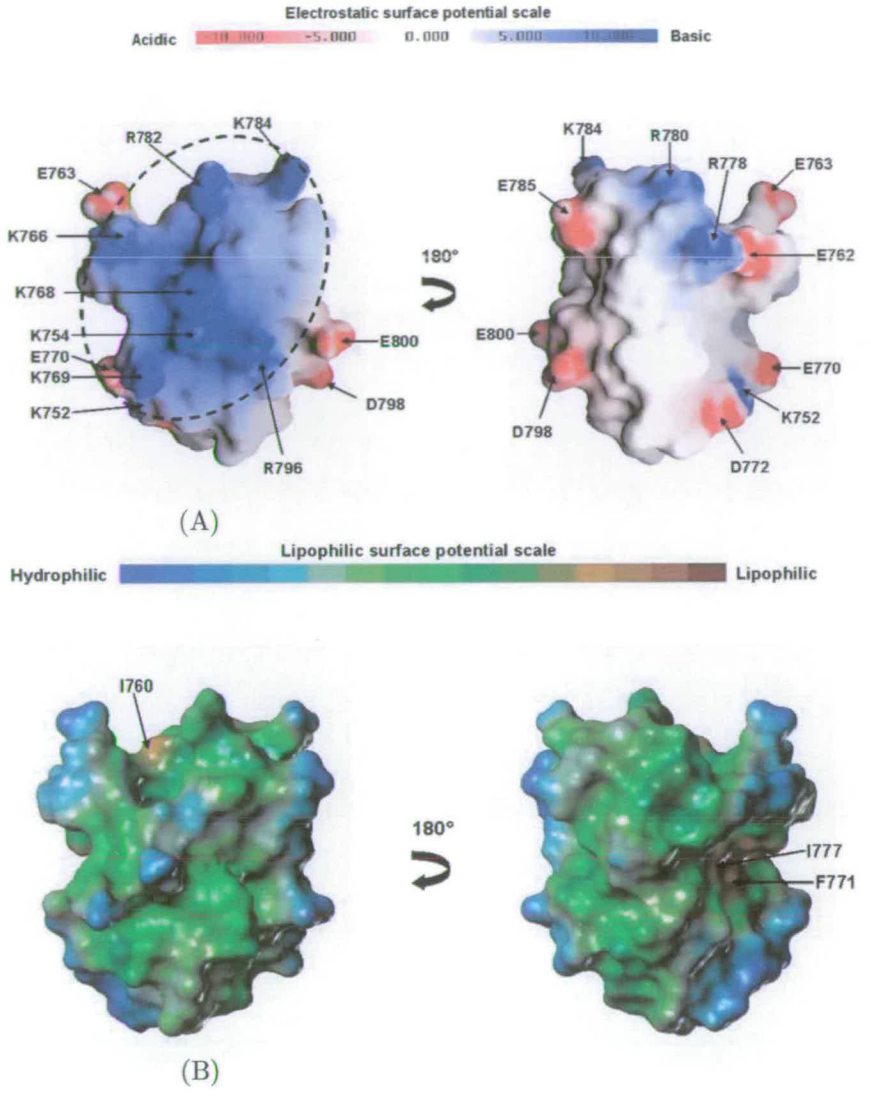


Figure 6.21: Surface representations of fH~13. (A) Two views rotated by 180 along the y-axis of a GRASP [88] electrostatic surface representation of fH~13. Negatively charged surface residues are coloured red and positively charged residues coloured blue and labelled; the potential ranges from -10 kT to $+10$ kT (k = Boltzmann's constant; T = temperature in Kelvin). Many basic residues appear to be exposed, and this is localised to one side of the protein (encircled region). (B) Two views rotated by 180 along the y-axis of a MOLCAD [89] lipophilic surface rendition of the molecule. Regions of high lipophilicity (hydrophobicity) are coloured brown and labelled and regions of high hydrophilicity are coloured blue. The surface representations in (A) and (B) are in approximately equivalent orientations to each other, and labels can be viewed in conjunction with each other.

Factor H CCP-13 has an accessible surface area of 3691.46 Å^2. Expectedly, given its small sequence-length, its ASA value is at the lower-end among individual CCP modules whose values have been calculated, with only module 3 from the NMR structures of VCP, i.e. PDB ID: 1VVC [17] and PDB ID: 1E5G [106] possessing a lower ASA.

However, Figure 6.21A shows many exposed charged residues occur on its surface. The 'front' side of the molecule (left-hand frame) reveals a positively charged patch, which include residues K752, K754, K766, K768, K769, R782, K784 and R796. However, the 'back' side is largely neutral, exposing a few negatively charged residues scattered on its surface. The R778 and R780 residues can be considered to be part of the same patch that extends onto the back of the molecule. fH-13 has previously been implicated in a number of protein-ligand interactions; with complement component fragment C3c, which also involves the neighbouring modules 12 and 14 [43]; as part of a microfilariae-binding site of the nematode parasite Onchocerca volvulus, which involve modules 8-20 [107]; and also heparin-binding localised at module 13 [37]. Heparin-protein interactions require electrostatic interactions between negatively charged heparin sulfate (and carboxyl groups), and suitably placed lysine and arginine residues in proteins. Thus the basic region on the module (shown in Figure 6.21A) could potentially be involved in heparin-binding.

Chapter 7

Binding Studies Between Heparin and fH~13

7.1 Introduction

As is discussed in section 1.2.2, the second putative heparin binding site is believed to be located in, or contain, fH~13. Our collaborator, Dr. Malcolm Lyon from the Department of Medical Oncology, Manchester University, had produced a fully sulphated haparin tetrasaccharide derived from heparin sulphate. It is advantageous to use smaller oligosaccharides such as tetrasaccharides as large, more easily and commercially available oligosaccharides can bind multiple protein molecules leading to precipitation. To make the sample of tetrasaccharide, heparin was enzymatically cleaved and then purified so produce a homogeneous sample of sulphated heparin. The structure of the tetrasaccharide is given in Figure 7.1.

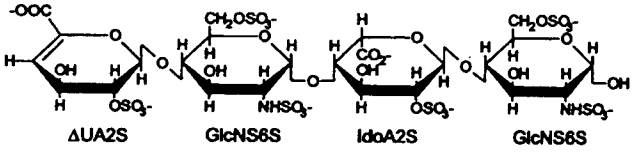


Figure 7.1: The structure of a heparin derived tetrasaccharide.

Dr. Lyon's group carried out an agarose gel mobility shift assay with fH~13 and the tetrasaccharide. The tetrasaccharide was labeled with a fluorescence tag, aminoacridone

(AMAC). While tetrasaccharide on its own would migrate through the gel, on addition of the recombinant fH~13, the sugar's progress was inhibited. This suggested that a proportion of the tetrasaccharide had bound to the protein and the resulting complex was now partially neutralized and it's migration towards the anode was significantly compromised. The results of this are shown in Figure 7.2.

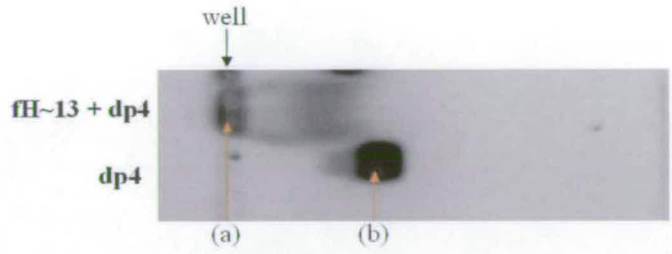


Figure 7.2: The first agarose GMSA experiment carried out on fH~13 by Dr. Lyon's group. The tetrasaccharide (dp4) was labeled with a fluorescence tag, AMAC. With only the tetrasaccharide loaded onto the well, the sugar is free to migrate through the gel (b). However, when the tetrasaccharide is loaded in conjunction with the recombinant fH~13 protein into the well, its progress through the gel is inhibited and much of the sugar remains in the well (a).

Along with the fact that fH~13 bound to a immobilized heparin column, we now had evidence that our recombinant fH~13 protein bound to the fully sulphated heparin-derived tetrasaccharide. The aim was then to map the binding site by titrating the protein with the tetrasaccharide and recording a series of NMR experiments (particularly ^{1}H, ^{15}N HSQCs) at each point in the titration. If the protein interacted with the tetrasaccharide during the titration, backbone NH crosspeaks might be seen to migrate during the course of the titration as a result of interactions between these residues in the protein and the sugar. These residues could be mapped onto the surface of the protein structure, producing a map of the site of interaction between the protein and the sugar.

7.2 Results

The titration of fH~13 with the heparin derived tetrassacharide was performed at two different pH values, pH 6.0 and 5.5. Previous titrations of this type carried out in our group showed improved binding at lower pH, which is why the titration was carried

out at pH 5.5. The molar ratios of protein to sugar ranged from 1:0.065 to 1: 22, a large excess of the tetrasscharide was used for the final point.

The first NMR sample of fH-13 with tetrasaccharide contained 10 μM ^{15}N labeled fH-13 with 214 μM tetrasaccharide in 20 mM sodium phosphate buffer at pH 6.0, with 10% D_2O and 0.05% sodium azide. This was used in conjunction with the fH-13 sample mentioned above to optimize the temperature at which the experiments would be recorded. ^{1}H, ^{15}N HSQC spectra were acquired at 37 °C and 25 °C for each sample. The HSQCs of the two samples were compared to see in which spectrum the most crosspeak displacement could be seen upon addition of the tetrasaccharide. The spectra at 25 °C seemed to show the most movement in crosspeaks and so this was chosen as the temperature at which the titration would be carried out.

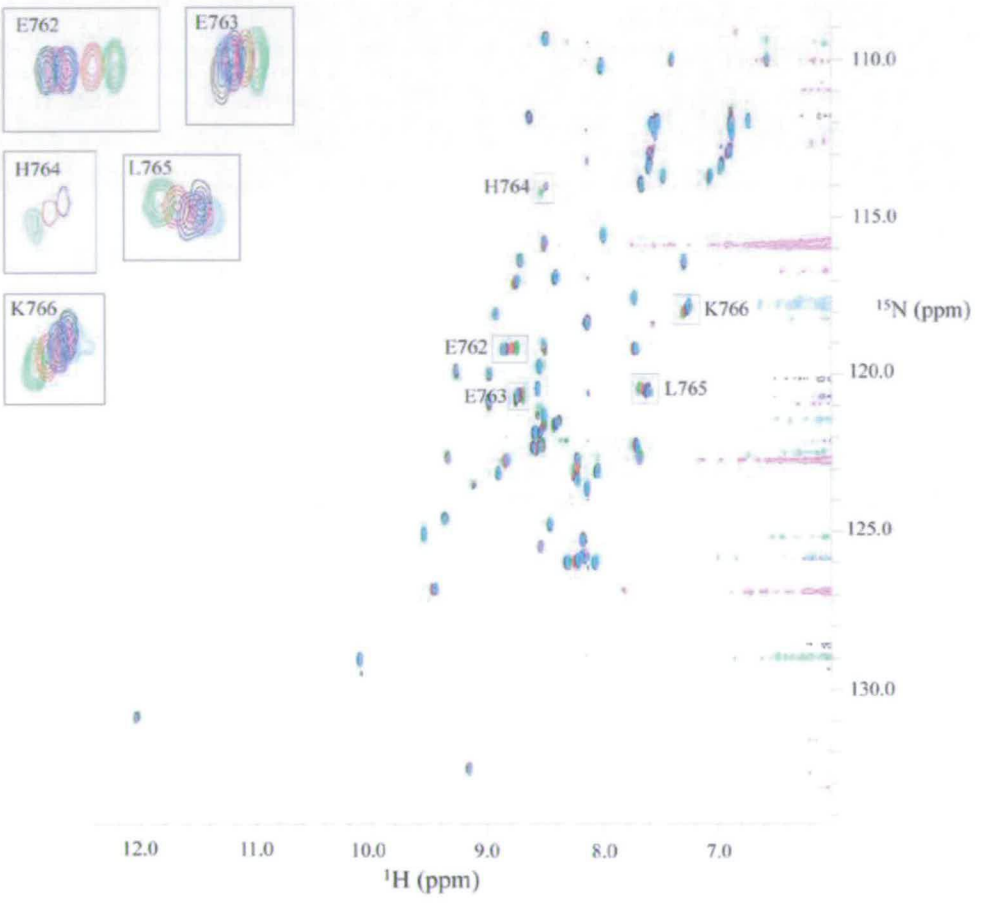
It was also in the acquired HSQCs that it was first noticed that additional peaks were present at the lower temperature. At a later date additional ^{1}H, ^{15}N HSQC were aquired on a sample of fH-13 at various temperatures ranging from 37 °C to 15 °C. This was discussed in Section 5.2.

In order to determine which residues in fH-13 were affected upon addition of the tetrasaccharide, the ^{1}H, ^{15}N HSQCs collected from each titration point are overlayed. The overlay of the ^{1}H, ^{15}N HSQCs from the titrations at pH 6.0 and pH 5.5 are given in Figure 7.3.

To determine which residues exhibited the most movement upon addition of the tetrasacharide graphs of the chemical shift difference between the free protein and the last titration point for each residue were produced, shown in Figure 7.4. Shift differences were calculated using the equation:

$$ShiftDifference = \sqrt{(\Delta^1 H)^2 + (\Delta^{15} N)^2} \tag{7.1}$$

(A)


E762
E763
H764
L765
K766
H764
K766
E762
E763
L765
110.0
115.0
120.0
125.0
130.0
15N (ppm)
12.0
11.0
10.0
9.0
8.0
7.0
1H (ppm)

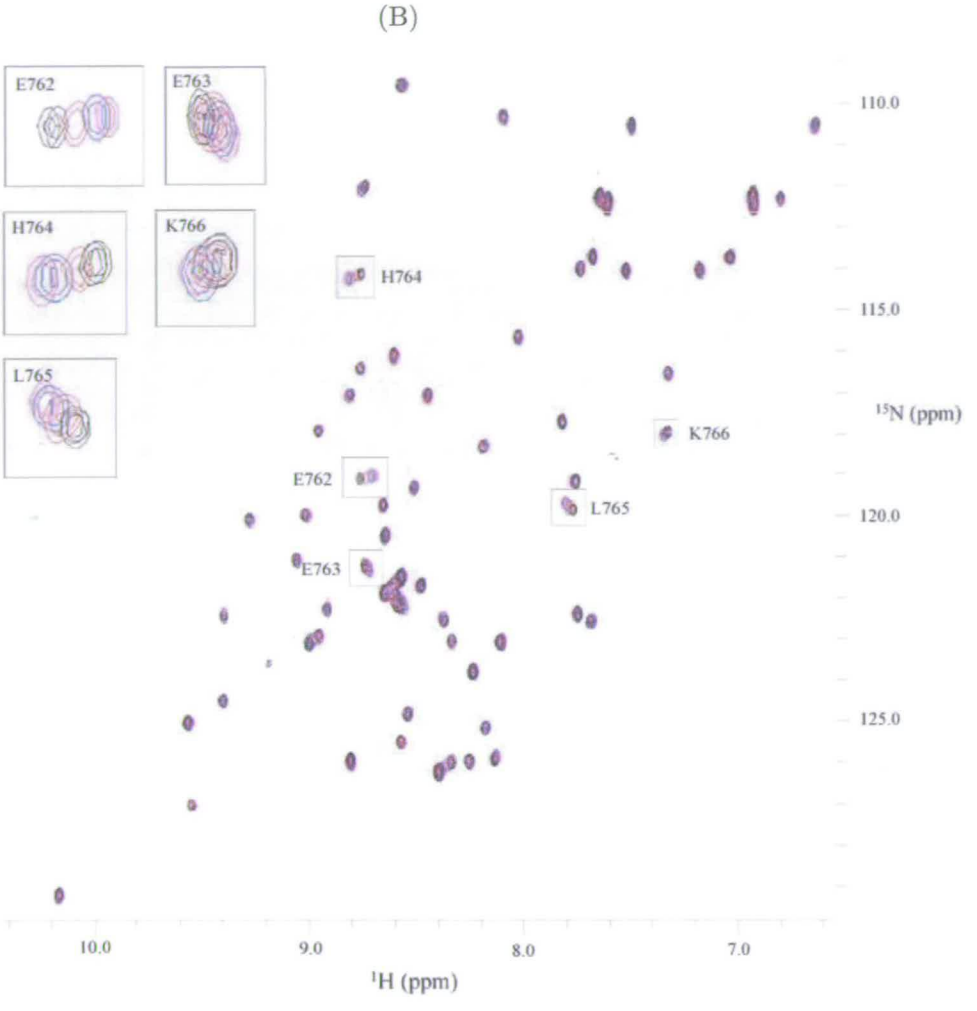


Figure 7.3: The overlay of the ^{1}H, ^{15}N HSQCs from the titrations at pH 6.0 and pH 5.5 between fH~13 and a heparin derived tetrasaccharide. The different ratios of fH~13 to tetrasaccharide are represented by different colours of crosspeak (negative peaks are not included). For pH 6.0 (A): (black) Free fH~13; (turquoise) 1:4 fH~13 to sugar; (magenta) 1:5 fH~13 to sugar; (blue) 1:7.5 fH~13 to sugar; (red) 1:10 fH~13 to sugar; (green) 1:20 fH~13 to sugar. The HSQCs from the 1:1 and 1:0.65 titration points are not shown as these two HSQCs showed very little deviation from that of fH~13 on its own. For pH 5.5 (B):(black) Free fH~13; (green) 1:1 fH~13 to sugar; (red) 1:3.3 fH~13 to sugar; (blue) 1:11 fH~13 to sugar; (magenta) 1:22 fH~13 to sugar.

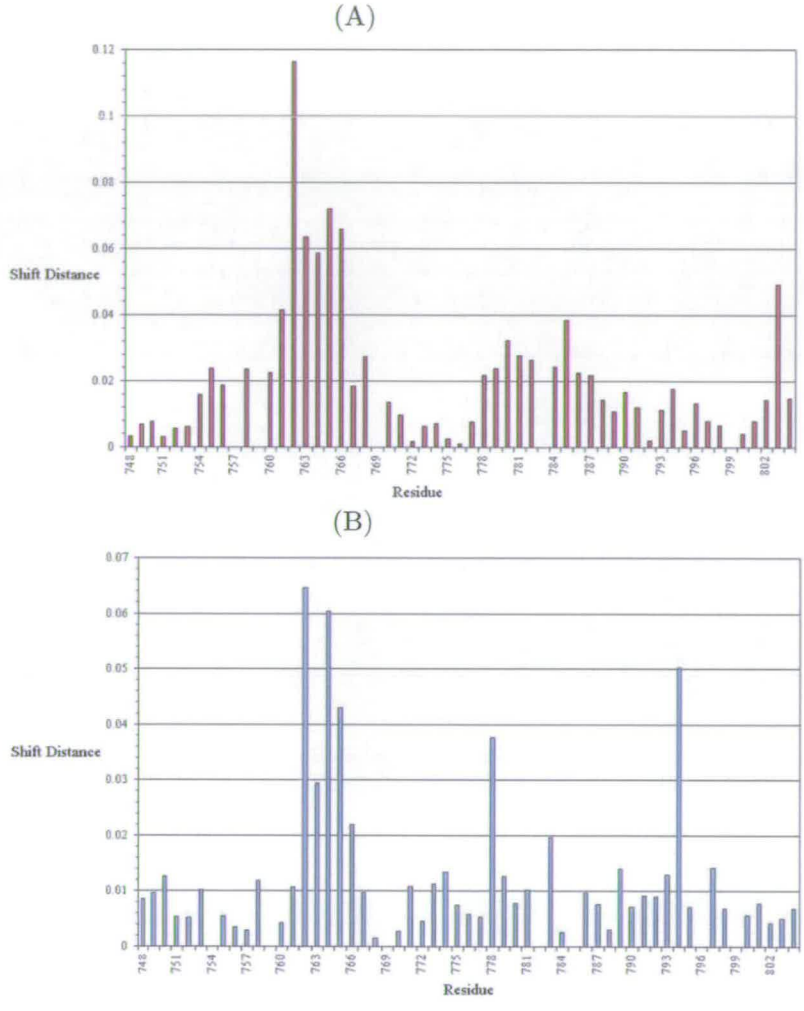


Figure 7.4: Graph of the chemical shift difference at pH 6.0 and pH 5.5 between the free fH~13 and fH~13 in the presence of excess sugar. (A) The chemical shift difference at pH 6.0 between the free fH~13 and 1:20 fH~13 to sugar samples for each residue for which it could be meassured. (B) The chemical shift difference at pH 5.5 between the free fH~1 and 1:22 fH~13 to sugar samples.

Where $\Delta^1 H$ is the difference in ^{1}H chemical shift; and $\Delta^{15} N$ is the difference in ^{15}N shift.

The chemical shift differences for some residues at pH 6.0 could not be measured; These were residues N757, I759, K769, R782, G783 and P799. K769 and R782 are all found clumped together in the ^{1}H, ^{15}N HSQC and so their individual chemical shifts can't be precisely determined. None of N757, I759 or G783 are visible at the temperature that these spectra were acquired (25 °C). Finally, P799 does not produce crosspeaks in ^{1}H, ^{15}N HSQC.

The ^{15}N-^{1}H HSQCs for the repeat of the titration at pH 5.5 are shown in Figure 7.3(B). Chemical shift differences for various residues could not be measured; These were K754, I759, K769, R782, E785, R796 and P799. Of these, K754, K769, R782 and E785 are all found clumped together in the ^{1}H, ^{15}N HSQC and so their individual chemical shifts can't be precisely determined. Neither I759 nor R796 were present in all titrations.

For the titrations at pH 6.0 and 5.5 the chemical shift differences are summarized in Figure 7.4. Most peaks move very little, while a few showed considerable movement. It was surprising to observe that the residues that showed the most pronounced movement in the overlayed spectra were the glutamic acid residues (E). These are negatively charged and known to be sensitive to pH changes.

It was therefore decided to also carry out a pH titration on fH-13. This was to determine whether the movement of peaks upon addition of the tetrasaccharide could be attributed to binding or more trivially to adventitious small changes in pH. Samples of the 20 μM stock solution of ^{15}N labeled fH-13 at pH 6.0 were used. The sample pH was changed using 20 mM dibasic or monobasic sodium phosphate solutions. As with the above titration, three NMR spectra were acquired at each titration point; a 1D proton spectrum with water presaturation; a 1D proton spectrum with water suppression gradients; and an ^{1}H, ^{15}N HSQC. There were four titration points in total; pH 6.5; pH 6.0; pH 5.5 and pH 5.3. The results of the titration are shown in Figures

7.5.

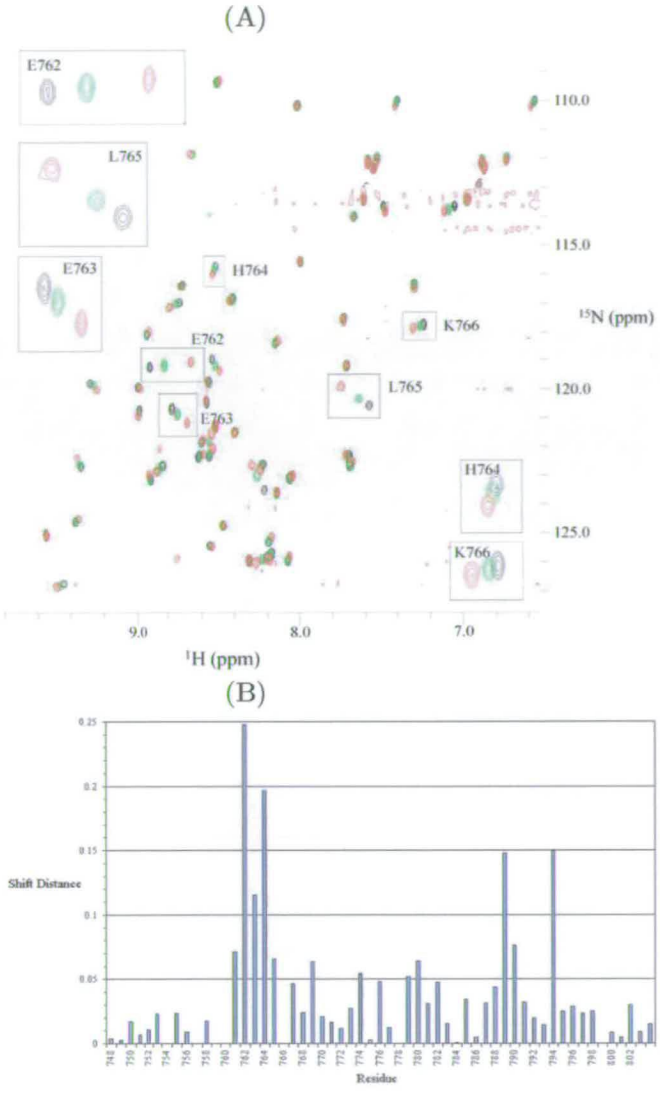


Figure 7.5: (A) Results of the pH titration of fH-13. The ^{1}H, ^{15}N HSQCs at pH 6.5, 6.0 and 5.5 are overlayed with their crosspeaks in black, green and red respectively (negative crosspeaks are not shown). (B) Graph of the chemical shift difference between the free fH-13 at pH 6.0 and free fH-13 at pH 5.5 for each residue for which it could be meassured.

7.3 Discussion

From Figures 7.3(A) and 7.4(A) it can be see that the residues whose chemical shifts showed the most pronounced perturbation upon titration with dp4 at pH 6.0 were; E762; E763; H764, L765; and K766. From Figures 7.3(B) and 7.4(B) it can be seen that in the titration pH 5.5 the residues that showed the most pronounced movement were E762; H764; L765; R778 and N794. Thus, with the exception of residues 778 and 794, the residues that show the most pronounced movement upon introduction of the

tetrasaccharide belong to the hypervariable loop.

The results of the pH titration (which can be seen in Figures 7.5 and 7.5) show that the peaks that were displaced most were also those of the hypervariable loop. In fact, from comparisons of the ^{15}N-^{1}H HSQCs it can be seen that with increased concentrations of sugar compared to protein the susceptible peaks shifted in similar directions as when they responded to a lowering of the pH. By placing the suseptible peaks onto the structure of fH-13 (Figure 7.7) it can be seen that it is the same regions of the protein affected by both pH and sugar titrations.

Taken together, this evidence strongly suggested that the movement of the crosspeaks was due to a lowering of the pH of the sample as the sugar concentration increased.

Based on our results, Dr. Lyon repeated the GMSA experiments on fH-13 using not just the tetrasaccharide but also longer lengths of heparin derived oligosaccharides. The results of this follow up experiment can be seen in Figure 7.6.

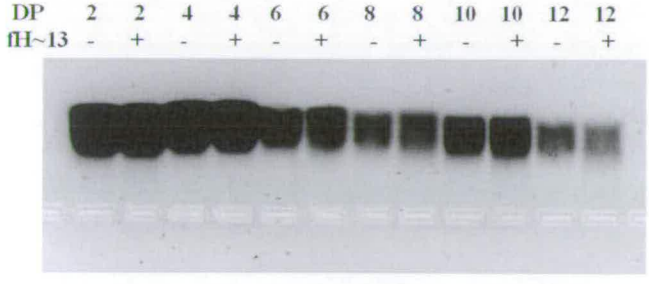


Figure 7.6: The second agarose GMSA experiment carried out on fH-13 by Dr. Lyon's group. This time the heparin derived oligosaccharides of degrees of polymerisation (dp) from 2 to 12 (all labeled with the fluorescence tag, AMAC) were used. Each oligosaccharide was run on the gel twice; with (+) the recombinant fH-13 protein; and without (-) the protein. None of the oligosacharides had their progress through the gel inhibited by the prescence of the protein

This time the results showed no evidence of binding of fH-13 to any of the oilgosaccharides ranging from a disaccharide to a dodecasaccharide. This was in contradiction to the previous GMSA result. It was Dr Lyon's opinion that the result to give the

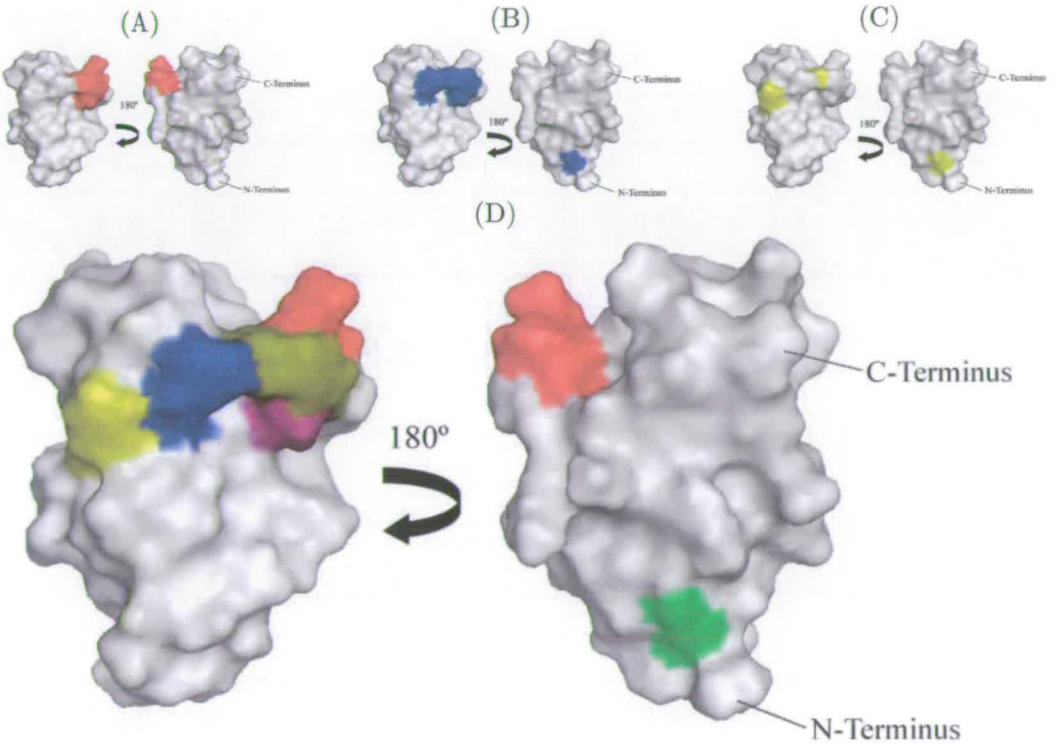


Figure 7.7: Surface representations of fH~13 showing residues whose chemical shift was most affected during the titrations. Two views rotated by 180° along the y-axis showing residues most affected during heparin titration: (A) at pH 6.0 (red); (B) at pH 5.0 (blue); (C) during the pH titration (yellow); (D) Two views rotated by 180° along the y-axis showing overlay of the most affected residues, colour coded as for (A), (B) and (C) with those residues shown in both (A) and (B) coloured purple; those residues shown in both (B) and (C) coloured green and those shown in (A), (B) and (C) coloured brown.

most credence to is that of the second experiment as it demonstrates reproducibility [108]. The first experiment, with the positive result, was done only once and only with the tetrassaccharide. The second experiment was carried out with a veriety of oligosaccharide sizes, including the larger oligosaccharides which are known to bind with higher affinity to heparin binding proteins then the smaller oligosaccharides do. Taken together with the NMR titration results, there is now good evidence that fH-13 does not bind to heparin.

Furthermore, it is Dr Lyon's opinion that the first experiment may truely show an interaction between the protein sample and the tetrasaccharide, but that the balance of evidence points to the protein sample being affected by handling or by transport to Manchester [108]. Perhaps the protein sample was partially denatured or not fully folded and this exposed an artifactual non-specific binding site. Or alternatively, the protein may have aggregated and so the binding species was not a true monomer.

7.4 Conclusion

The general consensus in the literature has long been that the position of the second heparin binding site of fH is located on, or includes, module 13. Further evidence for this view was also found by our group and our collaborators. Clair Egan's method for the purification of the recombinant fH-13 depended on the binding of the protein to an immobilized heparin column, although this is not evidence of binding to native heparin and could be attributed to non-specific binding. Dr. Lyon's Initial studies with GMSA also provided evidence that the heparin derived tetrasaccharide also bound to fH-13. However, there was no evidence of binding of this tetrasaccharide in the chemical shift mapping experiments and in the subsequent, more sophisticated GMSA experiments.

While GMSA is a very useful tool to examine the interactions of proteins and ligands, our results demonstrate that the information should always be treated with caution. Clear evidence of binding between the protein and the ligand should be acquired, preferably using more then one method eg. ITC, GMSA and NMR to confirm binding. Binding to a heparin column is not in itself sufficient proof of a specific

interaction. The high theoretical pI value of fH-13 does not guarantee binding to negatively charged saccharides. Instead what is required is a complementary distribution of opposing charges on both molecules

Chapter 8

Conclusion

Factor H (fH) is a crucial regulator of the alternative pathway of the complement system, a part of the innate immune system in mammals. Structural information regarding factor H is of great value for investigating the various functions of fH, its interactions with the host's molecules and those of pathogens. As part of our ongoing efforts to solve a complete structure of factor H, the structure of module 13 (fH-13) has been solved by NMR spectroscopy.

The recombinant fH-13 protein produced in *Pichia pastoris* by Claire Egan was used successfully to produce unlabeled, ^{15}N-labeled and ^{15}N,^{13}C-labelled samples of this module. However, the method of dilution of the supernatant to avoid the time consuming process of concentration and buffer exchange was not optimal for the purification of the recombinant protein. After diluting the supernatant, only a very small fraction of the available protein would bind to SP-Sepharose, requiring the diluted supernatant to be passed through the column three times in order to extract enough protein for an NMR sample. Nevertheless, dilution was suitable to avoid buffer exchange prior to injection of the protein onto the immobilized heparin column, thus saving time.

A set of NMR experiments was performed yielding the near complete assignment of ^{1}H, ^{13}C and ^{15}N resonances in the recombinant protein. A set of 2D homonuclear, 3D ^{15}N-edited and ^{13}C-edited NOESY spectra were also obtained. From these spectra an extensive set of NOE restraints were extacted for structure calculation. Residual dipolar coupling restraints and hydrogen bond restraints were obtained from subsequent

experiments. All were used successfully in the structure caluation of fH-13.

As a part of the familiarization with the procedures for the assignment of homonuclear NMR spectra of proteins, the four short polypeptide sequences from the KefC ion channel were assigned. Our analysis of the NMR data did not support the existence of secondary or tertiary structural elements in these peptides.

The final ensemble of structures of fH-13 produced good RMSD statistics, but the Ramachandran statistics were less favourable. The structure of fH-13 reflects it's short primary sequence and is unusual amongst the complement control proteins (CCP modules). fH-13 possess the expected disulfide-bonding pattern and consensus tryptophan, but lacks many overall 3D-structural features that characterise a "typical" CCP-module. fH-13 possesses only two β-strands out of a maximum of eight. The most similar structure to fH-13 is fH-15, while the most dissimilar CCP module is CR1-16. One side of the fH-13 domain reveals a highly localised positively charged patch composed of eight residues.

To distinguish host from non-host cell membranes, factor H binds to polyanions such as sialic acid or heparan sulphate which are bound to the surface of host cells. There are three putative polyanion binding sites, located in modules 7, 13 and 20, whose involvement in this process is, to various extents, supported by experimental evidence. The one in module 13 is the most disputed of the three polyanion binding sites.

The heparin binding studies performed on fH-13 included gel mobility shift assay and NMR titrations with a fully sulphated heparin-derived tetrasaccharide. While initial GMSA results were promising, subsequent GMSA experiments involving the binding of fH-13 to a range of heparin derived oligosaccharides from disaccharide to dodecasaccharide produced negative results. Similarly, NMR titrations using the tetrasaccharide also yielded a negative result. This is despite the considerable accumulation of positive charge on one side of the fH-13 molecule. These results point to the importance of an appropriate spacial distribution of positively charged residues for the binding of polyanions. High charge density alone is insufficient to initiate binding in this case.

The recombinant protein of fH-13 produced in our lab does not include the N-terminal linker sequence. This linker contained the consensus sequence consisting of alternate arginine and lycine residues that has been identified as a binding sequence for heparin. Between domains 12 and 13 at residues 731-736 this has the sequence KLKKCK. This should be considered a candidate for the elusive second heparin binding site on fH.

Bibliography

[1] S.K.A. Law and K.B.M. Reid. *Complement.* Oxford University Press, 1995.

[2] T.E. Hugli. The structural basis for anaphylatoxin and chemotactic functions of C3a, C4a and C5a. *Critical Rev. in Immunol.*, 1:321–366, 1981.

[3] J.D. Lambris. The multifunctional role of C3, the third component of complement. *Immunology Today*, 9:387–393, 1988.

[4] S.K.A. Law and A.W. Dodds. The internal thioester and the covalent binding properties of the complement proteins C3 and C4. *Protein Sci.*, 6:263–274, 1997.

[5] A. Sahu and J.D. Lambris. Structure and biology of complement protein C3, a connecting link between innate and acquired immunity. *Immunological Reviews*, 180:35–48, 2001.

[6] B.P. Morgan and C.L. Harris. *Complement regulatory proteins.* Academic Press, 1999.

[7] K. Ikeda, T. Sannoh, N. Kawasaki, and I. Yamashina. Serum lectin with known structure activates complement through the classical pathway. *J. Biol. Chem.*, 262:7451–7455, 1987.

[8] G. Lindahl, U. Sjöbring, and E. Johnsson. Human complement regulators; a major target for pathogenic microorganisms. *Current opinion in Immunol.*, 12:44–51, 2000.

[9] J.H. Weis, C.C. Morton, G.A. Bruns, J.J. Weis, L.B. Klickstein, W.W. Wong, and D.T. Fearon. A complement receptor locus: genes encoding C3b/C4b receptor and C3d/Epstein-Barr virus receptor map to 1q32. *J. Immunol.*, 138:312–315, 1987.

[10] G. Black. *Studies of the structure and dynamics of the functional sites within complement receptor type 1.* PhD thesis, The University of Edinburgh, 2004.

[11] D.E. Hourcade, M.K. Liszewski, M. Krych-Goldberg, and J.P. Atkinson. Functional domains, structural variations and pathogen interactions in MCP, DAF and CR1. *Immunopharmacology*, 49:103–166, 2000.

[12] M.D. Kirkitadze and P.N. Barlow. Structure and flexibility of the multiple domain proteins that regulate complement activation. . *Immunol. Rev.*, 180:146–161, 2001.

[13] D. Soares. http://www.bru.ed.ac.uk/dinesh/ccp-db.html.

[14] P.N. Barlow and I.D. Campbell. Strategy for studying modular proteins: application to complement modules. *Method. Enzymol.*, 239:464–485, 1994.

[15] A. Herbert, J. O'Leary, M. Krych-Goldberg, J.P. Atkinson, and P.N. Barlow. Three-dimensional structure and flexibility of proteins of the RCA family - a progress report. *Biochem. Soc. Trans.*, 30:990–996, 2002.

[16] B.O. Smith, R.L. Mallin, M. Krych-Goldberg, X. Wang, R.E. Hauhart, K. Bromek, Uhrín D., J.P. Atkinson, and P.N. Barlow. Structure of the C3b binding site of CR1 (CD35), the immune adherence receptor. *Protein Sci.*, 13:1238–1250, 2004.

[17] A.P. Wiles, G. Shaw, J. Bright, A. Perczel, I.D. Campbell, and P.N. Barlow. NMR studies of a viral protein that mimics the regulators of complement activation. *J. Mol. Biol.*, 272:253–265, 1997.

[18] D.C. Soares, D.L. Gerloff, N.R. Syme, A.F. Coulson, J. Parkinson, and P.N. Barlow. Large-scale modelling as a route to multiple surface comparisons of the CCP module family. *Protein Eng. Des. Sel.*, 18:379–388, 2005.

[19] P. Lubacik, P. Roversi, J. White, D. Esser, G.P. Smith, J. Billington, P. A. Williams, P.M. Rudd, M.R. Wormald, D.J. Harvey, M.D.M. Crispin, C.M. Radcliffe, R.A. Dwek, D.J. Evans, B.P. Morgan, and S. Lea. Complement regulation at the molecular level: the structure of decay-accelerating factor. *Proc. Nat. Acad. Sci.*, 101:1279–1284, 2004.

[20] I.D. Campbell and A.K. Downing. NMR of modular proteins. *Nat. Struc. Biol.*, 5:496–499, 1998.

[21] D.G. Norman, P.N. Barlow, M. Baron, A.J. Day, R.B. Sim, and I.D. Campbell. Three-dimensional structure of a complement control protein module in solution. *J. Mol. Biol.*, 219:717–725, 1991.

[22] P.N. Barlow, A. Steinkasserer, D.G. Norman, B. Kieffer, A.P. Norman, R.B. Sim, and I.D. Campbell. Solution structure of a pair of complement modules by nuclear magnetic resonance. *J. Mol. Biol.*, 232:268–284, 1993.

[23] P.N. Barlow, D.G. Norman, A. Steinkasserer, T.J. Horne, J. Pearce, R.B. Sim, and I.D . Campbell. Solution structure of the fifth repeat of factor H: a second example of the complement control protein module. *Biochemistry*, 31:3626–3634, 1992.

[24] A.P. Herbert, D. Uhrin, M. Lyon, M.K. Pangburn, and P.N. Barlow. Disease-associated sequence variations congregate in a polyanion recognition patch on human factor H revealed in three-dimensional structure. *J. Mol. Biol.*, 232:268–284, 1993.

[25] A.P. Herbert, J.A. Deakin, C.Q. Schmidt, B.S. Blaum, C. Egan, V.P. Ferreira, M.K. Pangburn, M. Lyon, D. Uhrin, and P.N. Barlow. Structure shows

glycosaminoglycan- and protein-recognition site in factor H is perturbed by age-related macular degeneration-linked SNP (Epub ahead of print). *J Biol Chem*, 2007.

[26] P. Williams, Y. Chaudry, I.G. Goodfellow, J. Billington, R. Powell, O.B. Spiller, D.J. Evans, and S. Lea. Mapping function in CD55. The structure of two pathogen-binding domains at 1.7 Å. *J. Biol. Chem.*, 278:10691–10696, 2003.

[27] S. Uhrínová, F. Lin, G. Ball, K. Bromek, D. Uhrín, M.E. Medof, and P.N. Barlow. Solution structure of a functionally active fragment of decay-accelerating factor. *Proc. Nat. Acad. Sci.*, 100:4718–4723, 2003.

[28] J.M. Casasnovas, M. Larvie, and T. Stehle. Crystal structure of two CD46 domains reveals an extended measles virus-binding surface. *EMBO J.*, 18:2911–2922, 1999.

[29] J.M. O'Leary, K. Bromek, G.M. Black, S. Uhrinova, C. Schmitz, X. Wang, M. Krych, Uhrin D., J.P. Atkinson, and P.N. Barlow. Backbone dynamics of complement control protein (CCP) modules reveals mobility in binding studies. *Cell.*, 108:1–20, 2002.

[30] G. Szakonyi, J.M. Guthridge, D. Li, K. Young, V.M. Holers, and X.S. Chen. Structure of complement receptor 2 in complex with its C3d ligand. *Science*, 292:1725–1728, 2001.

[31] H.T. Jenkins, L. Mark, G. Ball, J. Persson, G. Lindahl, D Uhrin, A.M. Blom, and P.N. Barlow. Structural analysis of the complement control protein (CCP) modules of GABA(B) receptor 1a: only one of the two CCP modules is compactly folded. . *J. Mol. Biol.*, 281:3690–3697, 2006.

[32] B. Nilsson and H.J. Müller. Isolation of β_{iF} Globulin from Human Serum and it's Characterization as the Fifth Component of Complement . *J. Exp. Med.*, 122:277–298, 1965.

[33] P.F. Zipfil. Complement Factor H: Physiology and Pathophysiology . *Seminars in Thrombosis and Hemostasis*, 27:191–199, 2001.

[34] P.F. Zipfel, T. Sakari-Jokiranta, J Hellwage, V. Koistinen, and S. Meri. The Factor H Protein Family . *Immunopharmacology*, 42:53–60, 1999.

[35] S. Meri and M. Pangburn. Discrimination Between Activators and Nonactivators of the Alternative Pathway of Complement: Regulation Via Sialic Acid/Polyaninon Binding Site on Factor H . *Immunology*, 87:3982–3986, 1990.

[36] M.K. Pangburn, K.L.W. Pangburn, V. Koistinen, S. Meri, and A.K. Sharma. Molecular Mechanisms of Target Recognition in an Innate Immune System: Interaction Among Factor H, C3b, and Target in the Alternative Pathway of Human Complement . *J. Immunology*, 164:4742–4751, 2000.

[37] M.K. Pangburn, M.A.L. Atkinson, and S. Meri. Localization of the Heparin-Binding Site on Complement Factor H. *J. Biol. Chem.*, 266:16847–16853, 1991.

[38] T. Seya, K. Kakamura, T Masaki, C. Ichihara-Itoh, M Matsumoto, and S. Magasawa. Human Factor H and C4b-Binding Protein Serve as Factor I-Cofactors Both Encompassing Inactivation of C3b and C4b . *Mol. Immunology*, 32:355–360, 1995.

[39] M. Aslam and S.J. Perkins. Folded-back Solution Structure of Monomeric Factor H of Human Complement By Synchrotron X-ray and Neutron Scattering, Analytical Ultracentrifugation and Constrained Molecular Modelling . *J. Mol. Biol.*, 309:1117–1138, 2001.

[40] S. Kuhn and P.F. Zipfel. Mapping of the Domains Required for Decay Accelerating Activity of the Human Factor H-like Protein 1 and Factor H. *Eur. J. Immunology*, 26:2383–2387, 1996.

[41] D.L. Gordon, R.M. Kaufman, T.K. Blackmore, J. Kwong, and D.M. Lublin. Identification of the Complement Regulatory Domains in Human Factor H. *J. Immunology*, 155:348–356, 1995.

[42] S. Kuhn, C. Skerka, and P.F. Zipfel. Mapping of the Complement Regulatory Domains in the Human Factor H-like Protein 1 and in Factor H1. *J. Immunology*, 155:5663–5670, 1995.

[43] T.S. Jokiranta, J. Hellwage, V. Koistinen, P.F. Zipel, and S. Meri. Each of the Three Binding Sites on Complement Factor H Interacts With a Distinct Site on C3b. *J. Biol. Chem.*, 275:27657–27662, 2000.

[44] T.S. Jokiranta, V.P. Jaakola, M.J. Lehtinen, M. Parepalo, S. Meri, and A. Goldman. Structure of complement factor H carboxyl-terminus reveals molecular basis of atypical haemolytic uremic syndrome. *EMBO J.*, 25:1784–1794, 2006.

[45] S. Ram, A.K. Sharma, S.D. Simpson, S. Gulati, D.P. McQuillen, M.K. Pangburn, and P.A. Rice. A Novel Sialic Acid Binding Site on Factor H Mediates Serum Resistance of Sialylated *Neisseria gonorrhoeae*. *J. Exp. Med.*, 187:743–752, 1998.

[46] T.K. Blackmore, T.A. Sadlon, H.M. Ward, D.M. Lublin, and D.L. Gordon. Identification of a heparin binding domain in the seventh short consensus repeat of complement factor H. *J. Immunology*, 157:5422–5427, 1996.

[47] T.K. Blackmore, J. Hellwage, T.A. Sadlon, N. Higgs, P.F. Zipfel, H.M. Ward, and D.L. Gordon. Identification of a second heparin-binding domain in human complement factor H. *J. Immunology*, 160:3342–3348, 1998.

[48] R.J. Klein, C. Zeiss, E.Y. Chew, J. Tsai, R. Sackler, C. Haynes, A.K. Henning, J.P. SanGiovanni, S.M. Mane, S.T. Mayne, M.B. Bracken, F.L. Ferris, J. Ott, C. Barnstable, and J. Hoh. Complement Factor H Polymorphism in Age-Related Macular Degeneration. *Science*, 308:385–389, 2005.

[49] R.J. Ormsby, T.S. Jokiranta, T.G. Duthy, K.M. Griggs, T.A. Sadlon, E. Giannakis, and D.L. Gordon. Localization of the third heparin-binding site in the human complement regulator factor H. *Mol. Immunology*, 43:1624–1632, 2006.

[50] G. Remuzzi and P. Rugentti. The Hemolytic Uremic Syndrome. *Kidney Int.*, 53:S54–S57, 1998.

[51] P.F. Zipfel. Complement Factor H and the Hemolytic Uremic Syndrome. *The Lancet*, 358:1200–1202, 2001.

[52] P. Sanchez-Corral, D. Perez-Caballero, H Olatz, A.M. Simckes, E. Goicoechea, M. Lopez-Trascasa, and S. Rodriguez de Cordoba. Structural and Functional Characterization of Factor H Mutations Associated With Atypical Hemolytic Uremic Syndrome. *Am. J. Hum. Genet.*, 71:1285–1295, 2002.

[53] P.F. Zipfel, J. Hellwage, M.A. Friese, G. Hegasy, S.T. Jokiranta, and S. Meri. Factor H and Disease: a Complement Regulator Affects Vital Body Functions. *Mol. Immuno.*, 36:241–248, 1999.

[54] J. Caprioli, M. Noris, S. Brioschi, G. Pianetti, F. Castelletti, P. Bettinaglio, C. Mele, E. Bresin, L. Cassis, S. Gamba, F. Porrati, S. Bucchioni, G. Monteferrante, C.J. Fang, M.K. Liszewski, D. Kavanagh, J.P. Atkinson, and G. Remuzzi. Genetics of HUS: the impact of MCP, CFH, and IF mutations on clinical presentation, response to treatment, and outcome. *Blood*, 108:1267–1279, 2006.

[55] D. Perez-Caballero, C. Gonzalez-Rubio, M. E. Galardo, M. Vera, M. Lopez-Trascasa, S. Rodriguez de Cordoba, and P. Sanchez-Corral. Clustering of Missense Mutations in the C-Terminus of Factor H in Associated With Atypical Hemolytic Uremic Syndrome. *Am. J. Hum. Genet.*, 68:478–484, 2001.

[56] P.F. Zipel. Hemolytic Uremic Syndrome: how do Factor H Mutants Mediate Endothelial Damage. *Trends in Immuno.*, 22:345–348, 2001.

[57] D.S. Friedman, B.J. O'Colmain, and B. *et al* Munoz. Prevalence of Age-Related Macular Degeneration in the United States. *Arch. Ophthalmol.*, 122:564–572, 2004.

[58] J. Tuo, C.M. Bojanowski, and C.C. Chan. Genetic Factors of Age-Related Macular Degeneration. *Prog. Retin. Eye Res.*, 23:229–249, 2004.

[59] S. Haddad, C.A. Chen, S.L. Santangelo, and J.M. Seddon. The Genetics of Age-Related Macular Degeneration: A Review of Progress to Date. *Survey of Ophthalmology*, 51:316–363, 2006.

[60] E.H. Souied, N. Leveziel, F. Richard, M.A. Dragon-Durey, G. Coscas, G. Soubrane, P. Benlian, and V. Fremeaux-Bacchi. Y402H Complement Factor H Polymorphism Associated with Exudative Age-Related Macular Degeneration in the French Population. *Mol. Vis.*, 11:1135–1140, 2005.

[61] S. Zareparsi, K.E. Branham, M. Li, S. Shah, R.J. Klein, J. Ott, J. Hoh, G.R. Abecasis, and A. Swaroop. Strong Association of the Y402H Variant in Complement Factor H at 1q32 with suseptability to Age-Related Macular Degeneration. *Am. J. Hum. Genet.*, 77:149–153, 2005.

[62] H. Wang, M. Eberstadt, E.T. Olejniczak, R.P Meadows, and S.W. Fesik. A Liquid Crystalline medium for measuring residual dipolar couplings over a wide range of temperatures. *J. Biomol. NMR*, 12:443–446, 1998.

[63] G.M. Clore, M.R. Starich, and A.M. Gronenborn. Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod-shaped viruses. *J. Am. Chem. Soc.*, 120:10571–10572, 1998.

[64] A.M. Gronenborn. The importance of being ordered: improving NMR structures using residual dipolar couplings. *C. R. Biologies*, 325:957–966, 2002.

[65] P.J. Kraulis. A program for the assignment of protein ^{1}H 2D NMR spectra by interactive graphics. *J. Magn. Reson.*, 24:627–633, 1989.

[66] G.C.K. Roberts (Editor). *NMR of Macromolecules, a Practical Approach.* Oxford University Press, 1993.

[67] R.H. Fogh, W. Boucher, W.F. Vranken, A. Pajon, T.J. Stevens, J.M. Ionides, E.D. Laue, M. Llinas, E.L. Ulrich, J.L. Markley, and J. Ionides. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*, 59:687–696, 2005.

[68] R.H. Fogh, W. Boucher, W.F. Vranken, A. Pajon, T.J. Stevens, T.N. Bhat, J. Westbrook, J.M. Ionides, and E.D. Laue. A framework for scientific data modeling and automated software development. *Bioinformatics*, 21:1678–1684, 2005.

[69] W. Boucher. http://www.bio.cam.ac.uk/azara/.

[70] L.G. Barrientos, C. Dolan, and A.M. Gronenborn. Characterization of surfactant liquid crystal phases suitable for molecular alignment and measurement of dipolar couplings. *J. Biomol. NMR*, 16:329–337, 2000.

[71] M. Ottiger, F. Delaglio, and A. Box. Measurement of J and Dipolar Couplings from Simplified Two Dimensional Spectra. *J. Magn. Reson*, 131:373–378, 1998.

[72] A.T. Brünger, P.D. Adams, G.M. Clore, W.L. DeLano, P. Gros, W.R. Grosse-Kunstleve, J. Jiang, J. Kuszewski, M. Nilges, N.S. Pannu, R.J. Read, L.M. Rice, T. Simonson, and G.L. Warren. Crystallography and NMR system. *Acta Cryst.*, 54:905–921, 1998.

[73] J. P. Linge and Nilges M. Influence of non-bonded parameters on the quality of NMR structures: a new forcefield for NMR structure calculation. *J. Biomol. NMR*, 13:51–59, 1999.

[74] M. Nilges. Structure calculation from NMR data. *Current opinion in Struc. Biol.*, 6:617–623, 1996.

[75] S. Maltke and S. Gizesiek. Structure contraints from residual tensorial couplings in high resolution NMR, without an explicit term for the alignment tensor. *J. Biomol. NMR*, 245:77–82, 1999.

[76] C.A.E.M. Spronk, S.B. Nabours, E. Krieger, G. Vriend, and G.W. Vuister. Validation of Protein Structures Derived By NMR Spectroscopy. *J. Progress in Nuclear Magnetic Resonance Spectroscopy*, 45:315–337, 2004.

[77] A. J. Nederveen et al. RECOORD: A recalculated coordinated database of 500+ proteins from the PDB using restraints from the BioMagResBank. *Proteins: Structure Function and Bioinformatics*, 59:662–672, 2005.

[78] S. Blein, R. Ginham, D. Uhrin, B.O. Smith, D.C. Soares, S. Veltel, R.A. McIlhinney, J.H. White, , and P.N. Barlow. Structural analysis of the complement control protein (CCP) modules of GABA(B) receptor 1a: only one of the two CCP modules is compactly folded. . *J. Biol. Chem.*, 279:48292–48306, 2004.

[79] I.N. Shindyalov and P.E Bourne. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.*, 11:739–747, 1998.

[80] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne. The Protein Data Bank. *Nucleic Acids Res.*, 28:235–242, 2000.

[81] M. Budayova-Spano, M. Lacroix, N.M. Thielens, G.J. Arlaud, J.C. Fontecilla-Camps, and C. Gaboriaud. The crystal structure of the zymogen catalytic domain of complement protease C1r reveals that a disruptive mechanical stress is required to trigger activation of the C1 complex. . *Embo. J.*, 21:231–239, 2002.

[82] M. Rickert, X. Wang, M.J. Boulanger, N. Goriatcheva, , and K.C. Garcia. The structure of interleukin−2 complexed with its alpha receptor. *Science*, 308:1477–1480, 2005.

[83] D.J. Stauber, E.W. Debler, P.A. Horton, K.A. Smith, and I.A. Wilson. Crystal structure of the IL−2 signaling complex: Paradigm for a heterotrimeric cytokine receptor. . *Proc. Natl. Acad. Sci. USA.*, 2006.

[84] X. Wang, M. Rickert, and K.C. Garcia. Structure of the quaternary complex of interleukin−2 with its alpha, beta, and gamma receptors. . *Science*, 310:1159–1163, 2005.

[85] D. Lupyan, A. Leo-Macias, and A.R. Ortiz. A new progressive-iterative algorithm for multiple structure alignment. *Bioinformatics*, 21:3255–3263, 2005.

[86] http://www.scsb.utmb.edu/cgi-bin/get_a_form.tcl.

[87] R. Fraczkiewicz and W. Braun. Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J. Comput. Chem.*, 19:319–333, 1998.

[88] K.A. Nicholls, A. Sharp and B. Honig. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, 11:281–296, 1991.

[89] W. Heiden, G. Moeckel, and J. Brickmann. A new approach to analysis and display of local lipophilicity/hydrophilicity mapped on molecular surfaces. *J. Comput. Aided Mol. Des.*, 7:503–514, 1993.

[90] S. Miller, R.M. Douglas, P. Carter, and I.R. Booth. Mutations in the Glutathione-gated KefC K^+ Efflux System of *Escherichia coli* That Cause Constitutive Activation. *J. Bio. Chem.*, 272:24942–24947, 1997.

[91] T.P Rooslid, S. Miller, and I.R. Booth. A Mechanism of Regulating Transmembrane Potassium Flux through a Ligand-Mediated Conformational Switch. *Cell*, 109:781–791, 2002.

[92] K. Wütrich. *NMR of Proteins and Nucleic Acids.* Wiley Interscience, 1986.

[93] Prof Ian Booth. personal communication.

[94] R. Koradi and K. Billeter, M.and Wütrich. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics*, 14:51–55, 1996.

[95] K. Wüthrich. http://www.mol.biol.ethz.ch/wuthrich/software/molmol/.

[96] R. Laskowski, M. MacArthur, D. Moss, and J. Thornton. PROCHECK: a program to check stereochemistry quality of protein structures. *J. Appl. Crystallogr.*, 26:283–290, 1993.

[97] G. Vriend. WHAT IF: a molecular modelling and drug design program. *J. Mol. Graph.*, 8:52–56, 1990.

[98] http://www.pymol.org .

[99] P.J. Kraulis. Molscript - a Program to Produce Both Detailed and Schematic Plots of Protein Structures. . *J. Appl. Crystallogr.*, 24:946–950, 1991.

[100] D.C. Soares and P.N. Barlow. *Structural Biology of the Complement System (D. Morikis, and J.D. Lambris (Editors)).* CRC Press, Taylor, Francis Group, Boca Raton, 2005.

[101] J.D. Thompson, T.J. Gibson, F. Plewniak, F. Jeanmougin, , and D.G Higgins. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, 25:4876–4882, 1997.

[102] http://www.ch.embnet.org/software/BOX_form.html.

[103] http://weblogo.berkeley.edu/.

[104] G.E. Crooks, G. Hon, J.M. Chandonia, , and S.E Brenner. WebLogo: a sequence logo generator. . *Genome Res.*, 14:1188–1190, 2004.

[105] R.A. Sayle and E.J. Milner-White. RASMOL: biomolecular graphics for all. . *Trends Biochem. Sci.*, 20:374, 1998.

[106] C.E. Henderson, K. Bromek, N.P. Mullin, B.O. Smith, D. Uhrin, and P.N. Barlow. Solution structure and dynamics of the central CCP module pair of a poxvirus complement control protein. *J. Mol. Biol.*, 307:323–339, 2001.

[107] T. Meri, T.S. Jokiranta, J. Hellwage, A. Bialonski, P.F. Zipfel, and S. Meri. Onchocerca volvulus microfilariae avoid complement attack by direct binding of factor H. . *J Infect Dis*, 185:1786–1793, 2002.

[108] Dr Malcolm Lyon. personal communication.

Appendix A

Chemical Shift Tables of KefC polypeptides

Chemical shift tables for the KefC polypeptides are listed below:

Hal20

Residue Number	Residue Type	Atom	Shift
1	LEU	HA	4.007
1	LEU	HG	1.674
1	LEU	HB3	1.735
1	LEU	HB2	1.697
1	LEU	HD2	0.931
1	LEU	HD1	0.931
2	ALA	H	8.725
2	ALA	HA	4.414
2	ALA	HB	1.415
3	SER	H	8.431
3	SER	HA	4.399
3	SER	HB3	3.947
3	SER	HB2	3.863
4	SER	H	8.377
4	SER	HA	4.393
4	SER	HB3	3.904
4	SER	HB2	3.832
5	GLU	H	8.241
5	GLU	HG2	2.263
5	GLU	HG3	2.311
5	GLU	HA	4.253
5	GLU	HB3	1.964
5	GLU	HB2	1.903
6	TYR	H	8.02
6	TYR	HA	4.512
6	TYR	HB3	3.033
6	TYR	HB2	2.914
6	TYR	HE	6.774
6	TYR	HD	7.07
7	ARG	H	8.082
7	ARG	HE	7.158
7	ARG	HA	4.179
7	ARG	HD3	3.141
7	ARG	HD2	3.141
7	ARG	HG2	1.516
7	ARG	HG3	1.516
7	ARG	HB3	1.752
7	ARG	HB2	1.682
8	HIS	H	8.347
8	HIS	HE1	8.592
8	HIS	HD2	7.278
8	HIS	HA	4.586
8	HIS	HB3	3.247
8	HIS	HB2	3.159
9	ALA	H	8.303
9	ALA	HA	4.261
9	ALA	HB	1.377
10	LEU	H	8.212
10	LEU	HA	4.298
10	LEU	HG	1.585
10	LEU	HB3	1.605
10	LEU	HB2	1.605
10	LEU	HD2	0.903
10	LEU	HD1	0.858

Residue Number	Residue Type	Atom	Shift
11	GLU	H	8.266
11	GLU	HG2	2.448
11	GLU	HG3	2.448
11	GLU	HA	4.341
11	GLU	HB3	2.113
11	GLU	HB2	1.99
12	SER	H	8.183
12	SER	HA	4.401
12	SER	HB3	3.869
12	SER	HB2	3.812
13	ASP	H	8.351
13	ASP	HA	4.713
13	ASP	HB3	2.888
13	ASP	HB2	2.824
14	ILE	H	7.934
14	ILE	HA	4.166
14	ILE	HB	1.848
14	ILE	HG12	1.148
14	ILE	HG13	1.413
14	ILE	HG2	0.856
14	ILE	HD1	0.825
15	GLU	H	8.266
15	GLU	HG2	2.437
15	GLU	HG3	2.437
15	GLU	HA	4.646
15	GLU	HB3	2.03
15	GLU	HB2	1.882
16	PRO	HD3	3.731
16	PRO	HD2	3.63
16	PRO	HG2	1.922
16	PRO	HG3	1.922
16	PRO	HA	4.354
16	PRO	HB3	2.183
16	PRO	HB2	1.743

Residue Number	Residue Type	Atom	Shift
17	PHE	HZ	7.287
17	PHE	H	8.178
17	PHE	HA	4.574
17	PHE	HB3	3.083
17	PHE	HB2	3.083
17	PHE	HE	7.334
17	PHE	HD	7.247
18	LYS	H	8.148
18	LYS	HA	4.238
18	LYS	HD3	1.635
18	LYS	HD2	1.635
18	LYS	HE2	2.959
18	LYS	HE3	2.959
18	LYS	HG2	1.344
18	LYS	HG3	1.344
18	LYS	HB3	1.784
18	LYS	HB2	1.666
18	LYS	HZ	7.512
19	GLY	H	7.712
19	GLY	HA2	3.856
19	GLY	HA3	3.856
20	LEU	H	7.97
20	LEU	HA	4.312
20	LEU	HG	1.627
20	LEU	HB3	1.629
20	LEU	HB2	1.629
20	LEU	HD2	0.913
20	LEU	HD1	0.876

Hal22

Residue Number	Residue Type	Atom	Shift
1	LEU	HA	3.986
1	LEU	HG	1.614
1	LEU	HB3	1.701
1	LEU	HB2	1.668
1	LEU	HD1	0.907
1	LEU	HD2	0.907
2	ALA	H	8.614
2	ALA	HA	4.394
2	ALA	HB	1.585
3	ALA	H	8.499
3	ALA	HA	4.298
3	ALA	HB	1.384
4	SER	H	8.286
4	SER	HA	4.365
4	SER	HB3	3.933
4	SER	HB2	3.837
5	SER	H	8.309
5	SER	HA	4.387
5	SER	HB3	3.895
5	SER	HB2	3.829
6	GLU	H	8.192
6	GLU	HG2	2.24
6	GLU	HG3	2.289
6	GLU	HA	4.234
6	GLU	HB3	1.944
6	GLU	HB2	1.892
7	TYR	H	8.008
7	TYR	HA	4.51
7	TYR	HB3	3.02
7	TYR	HB2	2.905
7	TYR	HD	7.066
7	TYR	HE	6.764
8	ARG	H	8.065
8	ARG	HE	7.136
8	ARG	HA	4.181
8	ARG	HD3	3.131
8	ARG	HD2	3.131
8	ARG	HG2	1.496
8	ARG	HG3	1.496
8	ARG	HB3	1.735
8	ARG	HB2	1.663
9	HIS	H	8.356
9	HIS	HE1	8.584
9	HIS	HD2	7.266
9	HIS	HA	4.578
9	HIS	HB3	3.228
9	HIS	HB2	3.139
10	ALA	H	8.32
10	ALA	HA	4.255
10	ALA	HB	1.358
11	LEU	H	8.239
11	LEU	HA	4.291
11	LEU	HB3	1.598
11	LEU	HB2	1.579
11	LEU	HD1	0.856
11	LEU	HD2	0.856
12	GLU	H	8.294
12	GLU	HG2	2.432
12	GLU	HG3	2.432
12	GLU	HA	4.339
12	GLU	HB3	2.098
12	GLU	HB2	1.973
13	SER	H	8.212
13	SER	HA	4.391
13	SER	HB3	3.852
13	SER	HB2	3.795
14	ASP	H	8.368
14	ASP	HA	4.699
14	ASP	HB3	2.868
14	ASP	HB2	2.8
15	ILE	H	7.954
15	ILE	HA	4.145
15	ILE	HB	1.833
15	ILE	HG12	1.135
15	ILE	HG13	1.402
15	ILE	HD1	0.815
15	ILE	HG2	0.842
16	GLU	H	8.293
16	GLU	HG2	2.432
16	GLU	HG3	2.432
16	GLU	HA	4.644
16	GLU	HB3	2.024
16	GLU	HB2	1.869
17	PRO	HD3	3.733
17	PRO	HD2	3.629
17	PRO	HG2	1.924
17	PRO	HG3	1.924
17	PRO	HA	4.343
17	PRO	HB3	2.183
17	PRO	HB2	1.743
18	PHE	HZ	7.283
18	PHE	H	8.217
18	PHE	HA	4.558
18	PHE	HB3	3.099
18	PHE	HB2	3.056
18	PHE	HD	7.243
18	PHE	HE	7.328
19	LYS	H	8.159
19	LYS	HA	4.215
19	LYS	HD3	1.622
19	LYS	HD2	1.622
19	LYS	HE2	2.94
19	LYS	HE3	2.94
19	LYS	HG2	1.32
19	LYS	HG3	1.32
19	LYS	HB3	1.769
19	LYS	HB2	1.638
19	LYS	HZ	7.497
20	GLY	H	7.617
20	GLY	HA2	3.822
20	GLY	HA3	3.822
21	LEU	H	8.043
21	LEU	HA	4.359
21	LEU	HB3	1.604
21	LEU	HB2	1.604
21	LEU	HD1	0.88
21	LEU	HD2	0.88
22	LEU	H	8.154
22	LEU	HA	4.313
22	LEU	HB3	1.62
22	LEU	HB2	1.62
22	LEU	HD1	0.831
22	LEU	HD2	0.831

Hel

Residue Number	Residue Type	Atom	Shift
1	LEU	HA	3.996
1	LEU	HG	1.667
1	LEU	HB3	1.728
1	LEU	HB2	1.667
1	LEU	HD1	0.921
1	LEU	HD2	0.921
2	ALA	H	8.72
2	ALA	HA	4.405
2	ALA	HB	1.408
3	SER	H	8.445
3	SER	HA	4.398
3	SER	HB3	3.853
3	SER	HB2	3.939
4	SER	H	8.393
4	SER	HA	4.391
4	SER	HB3	3.824
4	SER	HB2	3.896
5	GLU	H	8.222
5	GLU	HG2	2.265
5	GLU	HG3	2.313
5	GLU	HA	4.26
5	GLU	HB3	1.963
5	GLU	HB2	1.897
6	TYR	H	8.04
6	TYR	HA	4.504
6	TYR	HB3	2.92
6	TYR	HB2	3.02
6	TYR	HD	7.072
6	TYR	HE	6.772
7	ARG	H	8.087
7	ARG	HE	7.139
7	ARG	HA	4.165
7	ARG	HD3	3.138
7	ARG	HD2	3.138
7	ARG	HG2	1.495
7	ARG	HG3	1.495
7	ARG	HB3	1.751
7	ARG	HB2	1.684
8	HIS	H	8.405
8	HIS	HE1	8.59
8	HIS	HD2	7.269
8	HIS	HA	4.585
8	HIS	HB3	3.159
8	HIS	HB2	3.251
9	GLU	H	8.354
9	GLU	HG2	2.415
9	GLU	HG3	2.415
9	GLU	HA	4.286
9	GLU	HB3	2.067
9	GLU	HB2	1.969
10	LEU	H	8.277
10	LEU	HA	4.291
10	LEU	HG	1.567
10	LEU	HB3	1.618
10	LEU	HB2	1.571
10	LEU	HD1	0.837
10	LEU	HD2	0.884
11	LEU	HA	3.996
11	GLU	H	8.317
11	GLU	HG2	2.453
11	GLU	HG3	2.453
11	GLU	HA	4.385
11	GLU	HB3	2.112
11	GLU	HB2	1.991

Residue Number	Residue Type	Atom	Shift
12	THR	H	8.013
12	THR	HB	4.193
12	THR	HA	4.258
12	THR	HG2	1.169
13	ALA	H	8.172
13	ALA	HA	4.315
13	ALA	HB	1.356
14	ILE	H	8.013
14	ILE	HA	4.079
14	ILE	HB	1.795
14	ILE	HG12	1.148
14	ILE	HG13	1.425
14	ILE	HD1	0.82
14	ILE	HG2	0.837
15	ASP	H	8.462
15	ASP	HA	4.943
15	ASP	HB3	2.728
15	ASP	HB2	2.912
16	PRO	HD3	3.697
16	PRO	HD2	3.734
16	PRO	HG2	1.649
16	PRO	HG3	1.872
16	PRO	HA	4.322
16	PRO	HB3	2.119
16	PRO	HB2	2.119
17	PHE	HZ	7.289
17	PHE	H	8.014
17	PHE	HA	4.575
17	PHE	HB3	3.036
17	PHE	HB2	3.16
17	PHE	HD	7.27
17	PHE	HE	7.343
18	LYS	H	7.932
18	LYS	HA	4.23
18	LYS	HD3	1.653
18	LYS	HD2	1.653
18	LYS	HE2	2.97
18	LYS	HE3	2.97
18	LYS	HG2	1.357
18	LYS	HG3	1.376
18	LYS	HB3	1.808
18	LYS	HB2	1.696
18	LYS	HZ	7.51
19	GLY	H	7.89
19	GLY	HA2	3.871
19	GLY	HA3	3.871
20	LEU	H	7.998
20	LEU	HA	4.328
20	LEU	HG	1.628
20	LEU	HB3	1.64
20	LEU	HB2	1.64
20	LEU	HD1	0.863
20	LEU	HD2	0.912

Dik

Residue Number	Residue Type	Atom	Shift
1	LEU	HA	4
1	LEU	HG	1.661
1	LEU	HB3	1.718
1	LEU	HB2	1.656
1	LEU	HD1	0.913
1	LEU	HD2	0.913
2	ALA	H	8.719
2	ALA	HA	4.413
2	ALA	HB	1.399
3	SER	H	8.426
3	SER	HA	4.399
3	SER	HB3	3.935
3	SER	HB2	3.853
4	SER	H	8.37
4	SER	HA	4.396
4	SER	HB3	3.893
4	SER	HB2	3.823
5	GLU	H	8.245
5	GLU	HG2	2.242
5	GLU	HG3	2.294
5	GLU	HA	4.253
5	GLU	HB3	1.954
5	GLU	HB2	1.882
6	TYR	H	8.037
6	TYR	HA	4.518
6	TYR	HB3	3.024
6	TYR	HB2	2.912
7	ARG	H	8.088
7	ARG	HE	7.155
7	ARG	HA	4.188
7	ARG	HD3	3.134
7	ARG	HD2	3.134
7	ARG	HG2	1.486
7	ARG	HG3	1.486
7	ARG	HB3	1.728
7	ARG	HB2	1.651
8	HIS	H	8.359
8	HIS	HA	4.592
8	HIS	HB3	3.238
8	HIS	HB2	3.148
9	ALA	H	8.322
9	ALA	HA	4.268
9	ALA	HB	1.356
10	LEU	H	8.249
10	LEU	HA	4.307
10	LEU	HB3	1.615
10	LEU	HB2	1.576
11	GLU	H	8.307
11	GLU	HG2	2.431
11	GLU	HG3	2.431
11	GLU	HA	4.343
11	GLU	HB3	2.097
11	GLU	HB2	1.978
12	SER	H	8.225
12	SER	HA	4.399
12	SER	HB3	3.866
12	SER	HB2	3.804
13	ASP	H	8.378
13	ASP	HA	4.705
13	ASP	HB3	2.858
13	ASP	HB2	2.786
14	ILE	H	7.942
14	ILE	HA	4.124
14	ILE	HB	1.82
14	ILE	HG12	1.139
14	ILE	HG13	1.414
14	ILE	HG2	0.829
14	ILE	HD1	0.804
15	LYS	H	8.32
15	LYS	HA	4.59
15	LYS	HD3	1.661
15	LYS	HD2	1.661
15	LYS	HE2	2.985
15	LYS	HE3	2.985
15	LYS	HG2	1.421
15	LYS	HG3	1.421
15	LYS	HB3	1.752
15	LYS	HB2	1.68
15	LYS	HZ	7.513
16	PRO	HD3	3.769
16	PRO	HD2	3.588
16	PRO	HG2	1.944
16	PRO	HG3	1.944
16	PRO	HA	4.364
16	PRO	HB3	2.199
16	PRO	HB2	2.121
17	PHE	HZ	7.339
17	PHE	H	8.263
17	PHE	HA	4.57
17	PHE	HB3	3.106
17	PHE	HB2	3.041
17	PHE	HE	7.363
17	PHE	HD	7.274
18	LYS	H	8.187
18	LYS	HA	4.224
18	LYS	HD3	1.619
18	LYS	HD2	1.619
18	LYS	HE2	2.944
18	LYS	HE3	2.944
18	LYS	HG2	1.306
18	LYS	HG3	1.306
18	LYS	HB3	1.751
18	LYS	HB2	1.636
18	LYS	HZ	7.5
19	GLY	H	7.729
19	GLY	HA2	3.854
19	GLY	HA3	3.854
20	LEU	H	7.979
20	LEU	HA	4.296
20	LEU	HG	1.581
20	LEU	HB3	1.622
20	LEU	HB2	1.584
20	LEU	HD1	0.869
20	LEU	HD2	0.869

Appendix B

Chemical Shift Tables for the Unlabelled and ^{15}N-labeled fH~13

Chemical shift tables for the unlabelled and ^{15}N-labeled fH~13 are listed below:

Residue Number	Residue Type	Atom	Shift
747	Ala	HA	4.311
747	Ala	HB1	1.371
748	Glu	HA	4.154
748	Glu	HB1	1.946
748	Glu	HB2	1.946
748	Glu	HG1	2.2432
748	Glu	HG2	2.2432
748	Glu	HN	8.5864
748	Glu	N	121.5348
749	Ala	HA	4.164
749	Ala	HB1	1.2535
749	Ala	HN	8.4079
749	Ala	N	126.3222
750	Gly	HA1	3.4988
750	Gly	HA2	3.767
750	Gly	HN	8.5932
750	Gly	N	109.6893
751	Ala	HA	4.026
751	Ala	HB1	1.1098
751	Ala	HN	8.2441
751	Ala	N	123.868
752	Lys	HA	4.657
752	Lys	HB1	1.5535
752	Lys	HB2	1.5535
752	Lys	HD1	1.482
752	Lys	HD2	1.482
752	Lys	HE1	2.936
752	Lys	HE2	2.936
752	Lys	HG1	1.418
752	Lys	HG2	1.418
752	Lys	HN	7.818
752	Lys	N	117.6582
753	Cys	HA	4.4745
753	Cys	HB1	1.638
753	Cys	HB2	1.7075
753	Cys	HN	8.9801
753	Cys	N	118.0891
754	Lys	HA	4.9897
754	Lys	HB1	1.839
754	Lys	HB2	2.109
754	Lys	HD1	1.737
754	Lys	HD2	1.737
754	Lys	HE1	3.014
754	Lys	HE2	3.014
754	Lys	HG1	1.617
754	Lys	HG2	1.617
754	Lys	HN	8.6677
754	Lys	N	121.945
755	Ser	HA	4.4757
755	Ser	HB1	3.9845
755	Ser	HB2	3.9845
755	Ser	HN	8.4305
755	Ser	N	117.0453
756	Ser	HA	4.7106
756	Ser	HB1	3.3989
756	Ser	HB2	3.9635
756	Ser	HN	7.7295
756	Ser	N	114.0442
757	Asn	HA	4.741
757	Asn	HB1	2.878
757	Asn	HB2	2.928
757	Asn	HD21	6.927
757	Asn	HD22	7.639
757	Asn	HN	8.9194
757	Asn	ND2	112.062
757	Asn	N	122.479
758	Leu	HA	4.222
758	Leu	HB1	1.461
758	Leu	HB2	1.544
758	Leu	HD11	0.752
758	Leu	HD21	0.828
758	Leu	HG	1.556
758	Leu	HN	8.2049
758	Leu	N	118.4221
759	Ile	HA	4.931
759	Ile	HB	1.7352
759	Ile	HD11	0.1044
759	Ile	HG11	1.05
759	Ile	HG12	1.05
759	Ile	HG21	0.331
759	Ile	HN	6.8093
759	Ile	N	108.931

Residue Number	Residue Type	Atom	Shift
760	Ile	HA	4.2557
760	Ile	HB	1.563
760	Ile	HD11	0.73
760	Ile	HG11	1.01
760	Ile	HG12	1.345
760	Ile	HG21	0.9237
760	Ile	HN	9.1927
760	Ile	N	123.6314
761	Leu	HA	4.929
761	Leu	HB1	1.588
761	Leu	HB2	1.586
761	Leu	HD11	0.911
761	Leu	HD21	0.911
761	Leu	HG	1.78
761	Leu	HN	8.2832
761	Leu	N	125.9
762	Glu	HA	4.22
762	Glu	HB1	1.7935
762	Glu	HB2	2.356
762	Glu	HG1	1.974
762	Glu	HG2	2.226
762	Glu	HN	8.9973
762	Glu	N	119.2869
763	Glu	HA	3.9423
763	Glu	HB1	2.116
763	Glu	HB2	2.214
763	Glu	HG1	2.357
763	Glu	HG2	2.4247
763	Glu	HN	8.8371
763	Glu	N	120.7214
764	His	HA	4.607
764	His	HB1	3.0583
764	His	HB2	3.3337
764	His	HD2	7.019
764	His	HN	8.5312
764	His	N	113.83
765	Leu	HA	4.54
765	Leu	HB1	1.687
765	Leu	HB2	1.841
765	Leu	HD11	0.9485
765	Leu	HD21	1.1037
765	Leu	HG	0.6317
765	Leu	HN	7.6129
765	Leu	N	120.5956
766	Lys	HA	4.003
766	Lys	HB1	1.791
766	Lys	HB2	1.958
766	Lys	HD1	1.726
766	Lys	HD2	1.726
766	Lys	HE1	3.02
766	Lys	HE2	3.02
766	Lys	HG1	1.3617
766	Lys	HG2	1.504
766	Lys	HN	7.2649
766	Lys	N	117.8032
767	Asn	HA	4.7662
767	Asn	HB1	2.7465
767	Asn	HB2	2.935
767	Asn	HD21	7.041
767	Asn	HD22	7.6747
767	Asn	HN	8.5919
767	Asn	ND2	113.813
767	Asn	N	115.838
768	Lys	HA	4.087
768	Lys	HB1	1.412
768	Lys	HB2	1.6188
768	Lys	HD1	1.456
768	Lys	HD2	1.5215
768	Lys	HE1	2.762
768	Lys	HE2	2.8578
768	Lys	HG1	0.9405
768	Lys	HG2	1.2363
768	Lys	HN	7.7303
768	Lys	N	122.4816
769	Lys	HA	4.4377
769	Lys	HB1	1.747
769	Lys	HB2	1.917
769	Lys	HG1	1.409
769	Lys	HG2	1.479
769	Lys	HN	8.6457
769	Lys	N	122.1

Residue Number	Residue Type	Atom	Shift
770	Glu	HA	5.1246
770	Glu	HB1	1.678
770	Glu	HB2	1.902
770	Glu	HG1	1.991
770	Glu	HG2	1.991
770	Glu	HN	7.3195
770	Glu	N	116.4347
771	Phe	HA	4.629
771	Phe	HB1	2.6193
771	Phe	HB2	3.141
771	Phe	HD1	6.0758
771	Phe	HD2	6.0758
771	Phe	HE1	6.5599
771	Phe	HE2	6.5599
771	Phe	HZ	7.2751
771	Phe	HN	9.0348
771	Phe	N	119.9547
772	Asp	HA	4.5247
772	Asp	HB1	2.4623
772	Asp	HB2	2.738
772	Asp	HN	8.6405
772	Asp	N	119.7458
773	His	HA	4.1167
773	His	HB1	2.745
773	His	HB2	3.1201
773	His	HD2	6.8799
773	His	HE1	7.6895
773	His	HN	8.9925
773	His	N	123.3252
774	Asn	HA	4.0929
774	Asn	HB1	2.7568
774	Asn	HB2	2.9317
774	Asn	HD21	6.6058
774	Asn	HD22	7.4787
774	Asn	HN	9.3791
774	Asn	ND2	110.303
774	Asn	N	122.7903
775	Ser	HA	4.5129
775	Ser	HB1	4.013
775	Ser	HB2	4.013
775	Ser	HN	8.0281
775	Ser	N	115.667
776	Asn	HA	5.462
776	Asn	HB1	2.7582
776	Asn	HB2	2.8267
776	Asn	HD21	7.1375
776	Asn	HD22	7.5292
776	Asn	HN	8.9339
776	Asn	ND2	113.9193
776	Asn	N	122.8619
777	Ile	HA	4.7655
777	Ile	HB	1.525
777	Ile	HD11	-0.5086
777	Ile	HG11	0.4019
777	Ile	HG12	0.7204
777	Ile	HN	8.7551
777	Ile	N	116.4374
778	Arg	HA	5.4351
778	Arg	HB1	1.2858
778	Arg	HB2	1.2858
778	Arg	HD1	2.4165
778	Arg	HD2	2.728
778	Arg	HG1	1.242
778	Arg	HG2	1.242
778	Arg	HE	6.5637
778	Arg	HN	7.6685
778	Arg	NE	83.458
778	Arg	N	118.3299
779	Tyr	HA	5.6984
779	Tyr	HB1	2.3704
779	Tyr	HB2	2.8436
779	Tyr	HD1	6.3836
779	Tyr	HD2	6.3836
779	Tyr	HE1	6.3079
779	Tyr	HE2	6.3079
779	Tyr	HN	9.3129
779	Tyr	N	119.8876

Residue Number	Residue Type	Atom	Shift
780	Arg	HA	4.6209
780	Arg	HB1	0.8933
780	Arg	HB2	1.0404
780	Arg	HD1	1.7537
780	Arg	HD2	2.381
780	Arg	HG1	0.3282
780	Arg	HG2	0.6192
780	Arg	HE	6.0986
780	Arg	HN	8.7601
780	Arg	NE	84.0125
780	Arg	N	116.9741
781	Cys	HA	4.953
781	Cys	HB1	2.4462
781	Cys	HB2	3.3394
781	Cys	HN	9.057
781	Cys	N	120.8743
782	Arg	HA	3.815
782	Arg	HB1	1.696
782	Arg	HB2	1.696
782	Arg	HD1	3.169
782	Arg	HD2	3.169
782	Arg	HG1	1.5805
782	Arg	HG2	1.5805
782	Arg	HN	8.6252
782	Arg	N	122.3165
783	Gly	HA1	3.714
783	Gly	HA2	4.046
783	Gly	HN	8.7197
783	Gly	N	111.9922
784	Lys	HA	4.6693
784	Lys	HB1	1.67
784	Lys	HB2	1.755
784	Lys	HD1	1.548
784	Lys	HD2	1.548
784	Lys	HE1	2.945
784	Lys	HE2	2.945
784	Lys	HG1	1.276
784	Lys	HG2	1.276
784	Lys	HN	7.7567
784	Lys	N	119.1926
785	Glu	HA	4.0386
785	Glu	HB1	1.7995
785	Glu	HB2	1.9485
785	Glu	HG1	2.174
785	Glu	HG2	2.2035
785	Glu	HN	8.6288
785	Glu	N	122.3367
786	Gly	HA1	3.639
786	Gly	HA2	3.99
786	Gly	HN	8.0892
786	Gly	N	110.3069
787	Trp	HA	4.6084
787	Trp	HB1	2.7059
787	Trp	HB2	2.9473
787	Trp	HD1	7.2371
787	Trp	HH2	6.9683
787	Trp	HZ2	7.4638
787	Trp	HZ3	7.024
787	Trp	HE1	10.1605
787	Trp	HN	8.3308
787	Trp	NE1	129.1445
787	Trp	N	122.8658
788	Ile	HA	4.0558
788	Ile	HB	0.3757
788	Ile	HD11	0.638
788	Ile	HG11	1.155
788	Ile	HG12	1.155
788	Ile	HG21	0.745
788	Ile	HN	9.5116
788	Ile	N	126.9983
789	His	HA	5.4233
789	His	HB1	3.0333
789	His	HB2	3.1067
789	His	HD2	7.0801
789	His	HE1	7.57
789	His	HN	8.3262
789	His	N	123.173
790	Thr	HA	4.7481
790	Thr	HB	4.2747
790	Thr	HG21	0.561
790	Thr	HN	8.5576
790	Thr	N	119.0811
791	Val	HA	5.2548
791	Val	HB	2.08
791	Val	HG11	0.891
791	Val	HG21	0.9084
791	Val	HN	8.2045
791	Val	N	125.3641

Residue Number	Residue Type	Atom	Shift
792	Cys	HA	4.1662
792	Cys	HB1	2.068
792	Cys	HB2	2.6621
792	Cys	HN	8.5963
792	Cys	N	125.5213
793	Ile	HA	4.7082
793	Ile	HB	2.009
793	Ile	HD11	1.5495
793	Ile	HG11	0.823
793	Ile	HG12	1.2544
793	Ile	HG21	0.957
793	Ile	HN	9.1856
793	Ile	N	132.5426
794	Asn	HA	4.2829
794	Asn	HB1	2.4604
794	Asn	HB2	2.9805
794	Asn	HD21	6.7886
794	Asn	HD22	7.6036
794	Asn	HN	12.1449
794	Asn	ND2	112.2371
794	Asn	N	131.127
795	Gly	HA1	3.3156
795	Gly	HA2	4.0383
795	Gly	HN	6.597
795	Gly	N	102.5028
796	Arg	HA	4.5688
796	Arg	HB1	1.5467
796	Arg	HB2	1.792
796	Arg	HD1	3.2118
796	Arg	HD2	3.2118
796	Arg	HG1	1.665
796	Arg	HG2	1.665
796	Arg	HE	7.251
796	Arg	HN	7.7076
796	Arg	NE	85.042
796	Arg	N	122.81
797	Trp	HA	4.109
797	Trp	HB1	2.816
797	Trp	HB2	3.491
797	Trp	HD1	7.1806
797	Trp	HE1	9.5774
797	Trp	HN	8.1532
797	Trp	NE1	125.0468
797	Trp	N	126.1265
798	Asp	HA	-99
798	Asp	HB1	2.3032
798	Asp	HB2	3.0097
798	Asp	HN	9.4237
798	Asp	N	124.6167
799	Pro	HA	4.985
799	Pro	HB1	2.14
799	Pro	HB2	2.5461
799	Pro	HD1	3.723
799	Pro	HD2	3.9123
799	Pro	HG1	1.976
799	Pro	HG2	1.976
800	Glu	HA	4.1053
800	Glu	HB1	2.021
800	Glu	HB2	2.021
800	Glu	HG1	2.417
800	Glu	HG2	2.417
800	Glu	HN	8.6472
800	Glu	N	120.56
801	Val	HA	3.6551
801	Val	HB	1.649
801	Val	HG11	0.2022
801	Val	HG21	0.3272
801	Val	HN	8.5442
801	Val	N	124.8533
802	Asn	HA	4.853
802	Asn	HB1	2.6151
802	Asn	HB2	2.813
802	Asn	HD21	6.9285
802	Asn	HD22	7.611
802	Asn	HN	8.2451
802	Asn	ND2	112.533
802	Asn	N	125.8455
803	Cys	HA	5.061
803	Cys	HB1	2.4245
803	Cys	HB2	3.4763
803	Cys	HN	8.4651
803	Cys	N	121.6331
804	Ser	HA	4.2625
804	Ser	HB1	3.8623
804	Ser	HB2	3.8623
804	Ser	HN	8.1257
804	Ser	N	123.1729

Appendix C

Chemical Shift Tables for the ^{13}C,-^{15}N-labeled fH~13

Chemical shift tables for the ^{13}C-labeled fH~13 are listed below:

Residue Number	Residue Type	Atom	Shift
747	ALA	CA	52.604
747	ALA	CB	19.021
747	ALA	HA	4.298
747	ALA	HB	1.36
748	GLU	H	8.581
748	GLU	CA	56.664
748	GLU	HG2	2.24
748	GLU	HG3	2.24
748	GLU	N	121.505
748	GLU	CB	30.159
748	GLU	CG	36.147
748	GLU	HA	4.15
748	GLU	HB3	1.944
748	GLU	HB2	1.927
749	ALA	H	8.403
749	ALA	CA	52.432
749	ALA	N	126.273
749	ALA	CB	18.978
749	ALA	HA	4.164
749	ALA	HB	1.25
750	ALA	H	8.242
750	ALA	CA	51.69
750	ALA	N	123.865
750	ALA	CB	18.84
750	ALA	HA	4.025
750	ALA	HB	1.098
751	GLY	H	8.59
751	GLY	HA2	3.491
751	GLY	HA3	3.765
751	GLY	CA	45.846
751	GLY	N	109.661

Residue Number	Residue Type	Atom	Shift
752	LYS	H	7.818
752	LYS	CA	56.054
752	LYS	CG	25.925
752	LYS	CD	29.446
752	LYS	N	117.666
752	LYS	HA	4.647
752	LYS	HD3	1.476
752	LYS	HD2	1.476
752	LYS	HE2	2.93
752	LYS	HE3	2.93
752	LYS	HG2	1.418
752	LYS	HG3	1.418
752	LYS	HB3	1.56
752	LYS	CB	34.189
752	LYS	HB2	1.56
753	CYS	H	8.977
753	CYS	CA	57.026
753	CYS	N	118.062
753	CYS	CB	42.381
753	CYS	HA	4.462
753	CYS	HB3	1.706
753	CYS	HB2	1.614
754	LYS	H	8.665
754	LYS	CA	55.863
754	LYS	CG	25.186
754	LYS	CE	41.952
754	LYS	CD	29.107
754	LYS	N	121.929
754	LYS	HA	4.999
754	LYS	HD3	1.743
754	LYS	HD2	1.743
754	LYS	HE2	3.023
754	LYS	HE3	3.023
754	LYS	HG2	1.606
754	LYS	HG3	1.613
754	LYS	HB3	2.104
754	LYS	CB	34.506
754	LYS	HB2	1.844

Residue Number	Residue Type	Atom	Shift
755	SER	H	8.43
755	SER	CA	57.36
755	SER	N	117.033
755	SER	CB	63.87
755	SER	HA	4.472
755	SER	HB3	3.979
755	SER	HB2	3.979
756	SER	H	7.729
756	SER	CA	56.756
756	SER	N	114.033
756	SER	CB	64.688
756	SER	HA	4.708
756	SER	HB3	3.955
756	SER	HB2	3.401
757	ASN	ND2	112.16
757	ASN	HD22	7.641
757	ASN	HD21	6.925
757	ASN	H	8.918
757	ASN	CA	54.221
757	ASN	N	122.395
757	ASN	CB	37.678
757	ASN	HA	4.733
757	ASN	HB3	2.935
757	ASN	HB2	2.873
758	LEU	H	8.203
758	LEU	CA	55.688
758	LEU	N	118.39
758	LEU	CD1	22.511
758	LEU	CD2	25.098
758	LEU	HA	4.206
758	LEU	HG	1.559
758	LEU	CB	42.901
758	LEU	HB3	1.492
758	LEU	HB2	1.471
758	LEU	HD2	0.818
758	LEU	HD1	0.759
759	ILE	H	6.812
759	ILE	CA	59.493
759	ILE	N	108.939
759	ILE	CD1	12.386
759	ILE	CG1	24.209
759	ILE	CG2	18.84
759	ILE	HA	4.935
759	ILE	HB	1.729
759	ILE	HG12	1.047
759	ILE	HG13	1.047
759	ILE	CB	40.498
759	ILE	HG2	0.327
759	ILE	HD1	0.096
760	ILE	H	9.192
760	ILE	CA	60.471
760	ILE	N	123.62
760	ILE	CD1	13.196
760	ILE	CG1	27.356
760	ILE	CG2	17.273
760	ILE	HA	4.252
760	ILE	HB	1.557
760	ILE	HG12	1.01
760	ILE	HG13	1.341
760	ILE	CB	39.354
760	ILE	HG2	0.92
760	ILE	HD1	0.726
761	LEU	H	8.283
761	LEU	CA	53.404
761	LEU	CG	26.524
761	LEU	N	125.882
761	LEU	CD1	24.151
761	LEU	HA	4.918
761	LEU	CB	44.337
761	LEU	HB3	1.778
761	LEU	HB2	1.589
761	LEU	HD2	1.047
761	LEU	HD1	0.818
762	GLU	H	8.989
762	GLU	CA	57.966
762	GLU	HG2	1.972
762	GLU	HG3	2.343
762	GLU	N	119.277
762	GLU	CB	31.495
762	GLU	CG	38.521
762	GLU	HA	4.217
762	GLU	HB3	2.216
762	GLU	HB2	1.79
763	GLU	H	8.832
763	GLU	CA	60.127
763	GLU	HG2	2.366
763	GLU	HG3	2.417
763	GLU	N	120.753
763	GLU	CB	30.045
763	GLU	CG	35.742
763	GLU	HA	3.941
763	GLU	HB3	2.211
763	GLU	HB2	2.111
764	HIS	CE1	139.018
764	HIS	H	8.533
764	HIS	CA	58.183
764	HIS	HE1	7.785
764	HIS	N	113.838
764	HIS	HD2	7.032
764	HIS	CD2	118.768
764	HIS	CB	29.107
764	HIS	HA	4.6
764	HIS	HB3	3.33
764	HIS	HB2	3.055
765	LEU	H	7.616
765	LEU	CA	54.226
765	LEU	N	120.57
765	LEU	CD1	23.364
765	LEU	CD2	26.325
765	LEU	HA	4.536
765	LEU	HG	0.633
765	LEU	CB	43.414
765	LEU	HB3	1.839
765	LEU	HB2	1.69
765	LEU	HD2	1.109
765	LEU	HD1	0.95
766	LYS	H	7.259
766	LYS	CA	59.231
766	LYS	CG	24.167
766	LYS	CE	42.374
766	LYS	CD	29.669
766	LYS	N	117.829
766	LYS	HA	4.001
766	LYS	HD3	1.717
766	LYS	HD2	1.717
766	LYS	HE2	3.005
766	LYS	HE3	3.005
766	LYS	HG2	1.356
766	LYS	HG3	1.504
766	LYS	HB3	1.949
766	LYS	CB	32.519
766	LYS	HB2	1.784
767	ASN	ND2	113.81
767	ASN	HD22	7.674
767	ASN	HD21	7.04
767	ASN	H	8.591
767	ASN	CA	53.226
767	ASN	N	115.854
767	ASN	CB	38.439
767	ASN	HA	4.758
767	ASN	HB3	2.925
767	ASN	HB2	2.745
768	LYS	H	7.731
768	LYS	CA	57.084
768	LYS	CG	24.62
768	LYS	CE	42.018
768	LYS	CD	29.076
768	LYS	N	122.468
768	LYS	HA	4.076
768	LYS	HD3	1.5
768	LYS	HD2	1.466
768	LYS	HE2	2.755
768	LYS	HE3	2.854
768	LYS	HG2	0.93
768	LYS	HG3	1.232
768	LYS	HB3	1.616
768	LYS	CB	32.782
768	LYS	HB2	1.412
769	LYS	H	8.639
769	LYS	CA	56.66
769	LYS	CG	24.874
769	LYS	CE	41.99
769	LYS	CD	28.748
769	LYS	N	122.078
769	LYS	HA	4.432
769	LYS	HD3	1.669
769	LYS	HD2	1.65
769	LYS	HE2	2.962
769	LYS	HE3	3.108
769	LYS	HG2	1.433
769	LYS	HG3	1.445
769	LYS	HB3	1.919
769	LYS	CB	33.76
769	LYS	HB2	1.746

Residue Number	Residue Type	Atom	Shift
770	GLU	H	7.325
770	GLU	CA	54.567
770	GLU	HG2	1.998
770	GLU	HG3	1.998
770	GLU	N	116.472
770	GLU	CB	33.118
770	GLU	CG	35.88
770	GLU	HA	5.125
770	GLU	HB3	1.905
770	GLU	HB2	1.688
771	PHE	HZ	7.31
771	PHE	H	9.032
771	PHE	CA	56.927
771	PHE	N	119.952
771	PHE	HA	4.627
771	PHE	CZ	131.478
771	PHE	CB	42.755
771	PHE	HB3	3.141
771	PHE	HB2	2.615
771	PHE	CD	131.478
771	PHE	HE	7.31
771	PHE	HD	7.305
771	PHE	CE	131.478
772	ASP	H	8.642
772	ASP	CA	54.521
772	ASP	N	119.768
772	ASP	CB	41.766
772	ASP	HA	4.526
772	ASP	HB3	2.739
772	ASP	HB2	2.46
773	HIS	CE1	139.052
773	HIS	H	8.989
773	HIS	CA	59.195
773	HIS	HE1	7.69
773	HIS	N	123.303
773	HIS	HD2	6.877
773	HIS	CD2	118.394
773	HIS	CB	31.347
773	HIS	HA	4.109
773	HIS	HB3	3.112
773	HIS	HB2	2.744
774	ASN	ND2	110.304
774	ASN	HD22	7.479
774	ASN	HD21	6.604
774	ASN	H	9.378
774	ASN	CA	54.445
774	ASN	N	122.769
774	ASN	CB	36.685
774	ASN	HA	4.09
774	ASN	HB3	2.928
774	ASN	HB2	2.755
775	SER	H	8.028
775	SER	CA	60.806
775	SER	N	115.678
775	SER	CB	64.042
775	SER	HA	4.509
775	SER	HB3	4.007
775	SER	HB2	4.007
776	ASN	ND2	113.925
776	ASN	HD22	7.529
776	ASN	HD21	7.137
776	ASN	H	8.934
776	ASN	CA	52.57
776	ASN	N	122.834
776	ASN	CB	41.084
776	ASN	HA	5.462
776	ASN	HB3	2.818
776	ASN	HB2	2.764

Residue Number	Residue Type	Atom	Shift
777	ILE	H	8.754
777	ILE	CA	59.724
777	ILE	N	116.431
777	ILE	CD1	12.448
777	ILE	CG1	25.966
777	ILE	CG2	17.704
777	ILE	HA	4.763
777	ILE	HB	1.518
777	ILE	HG12	0.352
777	ILE	HG13	0.713
777	ILE	CB	42.706
777	ILE	HG2	0.395
777	ILE	HD1	-0.514
778	ARG	NE	83.464
778	ARG	H	7.665
778	ARG	CA	53.835
778	ARG	CG	26.308
778	ARG	CD	43.683
778	ARG	N	118.33
778	ARG	HE	6.561
778	ARG	HA	5.435
778	ARG	HD3	2.726
778	ARG	HD2	2.416
778	ARG	HG2	1.214
778	ARG	HG3	1.214
778	ARG	CB	33.135
778	ARG	HB3	1.28
778	ARG	HB2	1.28
779	TYR	H	9.311
779	TYR	CA	55.669
779	TYR	N	119.891
779	TYR	HH	9.919
779	TYR	HA	5.699
779	TYR	CB	43.13
779	TYR	HB3	2.84
779	TYR	HB2	2.366
779	TYR	HD	6.384
779	TYR	CE	118.007
779	TYR	HE	6.307
779	TYR	CD	133.255
780	ARG	NE	84.014
780	ARG	H	8.761
780	ARG	CA	53.62
780	ARG	CG	26.187
780	ARG	CD	43.457
780	ARG	N	116.951
780	ARG	HE	6.097
780	ARG	HA	4.62
780	ARG	HD3	2.385
780	ARG	HD2	1.748
780	ARG	HG2	0.316
780	ARG	HG3	0.609
780	ARG	CB	33.482
780	ARG	HB3	1.032
780	ARG	HB2	0.875
781	CYS	H	9.055
781	CYS	CA	54.312
781	CYS	N	120.918
781	CYS	CB	38.652
781	CYS	HA	4.959
781	CYS	HB3	3.331
781	CYS	HB2	2.448

Residue Number	Residue Type	Atom	Shift
782	ARG	H	8.617
782	ARG	CA	58.582
782	ARG	CG	27.026
782	ARG	CD	43.009
782	ARG	N	122.314
782	ARG	HA	3.814
782	ARG	HD3	3.149
782	ARG	HD2	3.149
782	ARG	HG2	1.573
782	ARG	HG3	1.573
782	ARG	CB	29.471
782	ARG	HB3	1.692
782	ARG	HB2	1.692
783	GLY	H	8.724
783	GLY	HA2	3.708
783	GLY	HA3	4.042
783	GLY	CA	45.377
783	GLY	N	112.048
784	LYS	H	7.755
784	LYS	CA	54.371
784	LYS	CG	24.449
784	LYS	CE	42.124
784	LYS	CD	28.967
784	LYS	N	119.184
784	LYS	HA	4.664
784	LYS	HD3	1.54
784	LYS	HD2	1.54
784	LYS	HE2	2.94
784	LYS	HE3	2.94
784	LYS	HG2	1.257
784	LYS	HG3	1.294
784	LYS	HB3	1.755
784	LYS	CB	35.068
784	LYS	HB2	1.67
785	GLU	H	8.624
785	GLU	CA	56.652
785	GLU	HG2	2.166
785	GLU	HG3	2.205
785	GLU	N	122.299
785	GLU	CB	30.808
785	GLU	CG	36.114
785	GLU	HA	4.037
785	GLU	HB3	1.944
785	GLU	HB2	1.805
786	GLY	H	8.09
786	GLY	HA2	3.635
786	GLY	HA3	3.99
786	GLY	CA	44.867
786	GLY	N	110.311
787	TRP	HH2	6.969
787	TRP	CZ2	114.266
787	TRP	CZ3	122.334
787	TRP	CA	56.584
787	TRP	HZ2	7.462
787	TRP	N	122.89
787	TRP	CE3	119.954
787	TRP	HA	4.604
787	TRP	CD1	127.791
787	TRP	HE1	10.16
787	TRP	HD1	7.236
787	TRP	HE3	7.016
787	TRP	CH2	124.189
787	TRP	HB3	2.944
787	TRP	HZ3	7.017
787	TRP	H	8.329
787	TRP	NE1	129.143
787	TRP	CB	31.012
787	TRP	HB2	2.701

Residue Number	Residue Type	Atom	Shift
788	ILE	H	9.512
788	ILE	CA	60.634
788	ILE	N	126.994
788	ILE	CD1	13.459
788	ILE	CG1	27.332
788	ILE	CG2	18.632
788	ILE	HA	4.052
788	ILE	HB	0.369
788	ILE	HG12	0.724
788	ILE	HG13	1.147
788	ILE	CB	39.519
788	ILE	HG2	0.744
788	ILE	HD1	0.633
789	HIS	HE2	7.57
789	HIS	CE1	136.814
789	HIS	H	8.326
789	HIS	CA	55.161
789	HIS	HE1	8.076
789	HIS	N	123.144
789	HIS	HD2	7.083
789	HIS	CD2	120.697
789	HIS	CB	31.106
789	HIS	HA	5.421
789	HIS	HB3	3.086
789	HIS	HB2	3.042
790	THR	CG2	21.376
790	THR	H	8.557
790	THR	CA	61.186
790	THR	HG1	5.768
790	THR	HB	4.274
790	THR	N	119.059
790	THR	CB	67.919
790	THR	HA	4.74
790	THR	HG2	0.559
791	VAL	CG2	22.065
791	VAL	H	8.203
791	VAL	CA	59.849
791	VAL	N	125.346
791	VAL	CG1	21.114
791	VAL	CB	35.891
791	VAL	HB	2.067
791	VAL	HA	5.256
791	VAL	HG2	0.911
791	VAL	HG1	0.885
792	CYS	H	8.596
792	CYS	CA	54.322
792	CYS	N	125.509
792	CYS	CB	38.12
792	CYS	HA	4.159
792	CYS	HB3	2.657
792	CYS	HB2	2.064
793	ILE	H	9.182
793	ILE	CA	58.844
793	ILE	N	132.538
793	ILE	CD1	11.717
793	ILE	CG1	25.987
793	ILE	CG2	17.47
793	ILE	HA	4.704
793	ILE	HB	2.005
793	ILE	HG12	1.244
793	ILE	HG13	1.544
793	ILE	CB	37.543
793	ILE	HG2	0.949
793	ILE	HD1	0.817

Residue Number	Residue Type	Atom	Shift
794	ASN	ND2	112.249
794	ASN	HD22	7.603
794	ASN	HD21	6.788
794	ASN	H	12.141
794	ASN	CA	53.785
794	ASN	N	131.132
794	ASN	CB	37.885
794	ASN	HA	4.28
794	ASN	HB3	2.975
794	ASN	HB2	2.456
795	GLY	H	6.597
795	GLY	HA2	3.31
795	GLY	HA3	4.022
795	GLY	CA	45.75
795	GLY	N	102.492
796	ARG	NE	85.046
796	ARG	H	7.705
796	ARG	CA	53.794
796	ARG	CG	27.018
796	ARG	CD	43.459
796	ARG	N	122.775
796	ARG	HE	7.245
796	ARG	HA	4.562
796	ARG	HD3	3.205
796	ARG	HD2	3.202
796	ARG	HG2	1.645
796	ARG	HG3	1.645
796	ARG	CB	33.1
796	ARG	HB3	1.788
796	ARG	HB2	1.545
797	TRP	HH2	6.073
797	TRP	CZ2	112.594
797	TRP	CZ3	121.539
797	TRP	CA	56.841
797	TRP	HZ2	6.556
797	TRP	N	126.122
797	TRP	CE3	122.812
797	TRP	HA	4.106
797	TRP	CD1	126.638
797	TRP	HE1	9.577
797	TRP	HD1	7.18
797	TRP	HE3	7.273
797	TRP	CH2	123.636
797	TRP	HB3	3.486
797	TRP	HZ3	6.558
797	TRP	H	8.151
797	TRP	NE1	125.042
797	TRP	CB	29.628
797	TRP	HB2	2.808
798	ASP	H	9.423
798	ASP	CA	50.921
798	ASP	N	124.586
798	ASP	CB	44.151
798	ASP	HB3	3.01
798	ASP	HB2	2.298

Residue Number	Residue Type	Atom	Shift
799	PRO	HD3	3.91
799	PRO	HD2	3.723
799	PRO	CB	34.866
799	PRO	CA	63.086
799	PRO	HG2	1.969
799	PRO	HG3	2.098
799	PRO	CD	51.095
799	PRO	CG	25.987
799	PRO	HA	4.986
799	PRO	HB3	2.545
799	PRO	HB2	2.135
800	GLU	H	8.646
800	GLU	CA	57.103
800	GLU	HG2	2.42
800	GLU	HG3	2.42
800	GLU	N	120.531
800	GLU	CB	30.317
800	GLU	CG	36.422
800	GLU	HA	4.097
800	GLU	HB3	2.009
800	GLU	HB2	2.006
801	VAL	CG2	21.36
801	VAL	H	8.54
801	VAL	CA	63.377
801	VAL	N	124.836
801	VAL	CG1	19.96
801	VAL	CB	31.486
801	VAL	HB	1.644
801	VAL	HA	3.652
801	VAL	HG2	0.325
801	VAL	HG1	0.2
802	ASN	ND2	112.53
802	ASN	HD22	7.612
802	ASN	HD21	6.928
802	ASN	H	8.241
802	ASN	CA	52.843
802	ASN	N	125.8
802	ASN	CB	39.906
802	ASN	HA	4.848
802	ASN	HB3	2.806
802	ASN	HB2	2.611
803	CYS	H	8.467
803	CYS	CA	53.208
803	CYS	N	121.643
803	CYS	CB	39.004
803	CYS	HA	5.063
803	CYS	HB3	3.476
803	CYS	HB2	2.423
804	SER	H	8.123
804	SER	CA	60.589
804	SER	N	123.157
804	SER	CB	64.828
804	SER	HA	4.26
804	SER	HB3	3.86
804	SER	HB2	3.86