Human C4b-binding Protein, Structural Basis for Interaction with Streptococcal M Protein and DNA

Huw Thomas Jenkins

PhD

The University of Edinburgh

2006

Abstract

Human C4b-binding protein (C4BP) protects host tissue, and those pathogens able to hijack this plasma glycoprotein, from complement-mediated destruction. For example, C4BP localises to apoptotic and necrotic cells, via its affinity for DNA, whereupon this regulator helps to prevent complement activation and subsequent inflammation and tissue damage. The first two CCP modules of the C4BP α-chain, plus the four residues connecting them, are necessary and sufficient for binding a bacterial virulence factor, the *Streptococcus pyogenes* M4 (Arp4) protein. Structure determination of this region of C4BP by NMR reveals two tightly coupled CCP modules in an elongated arrangement. Chemical shift perturbation studies demonstrate that the N-terminal, hyper-variable, region of M4 binds to a site including strand 1 of CCP module 2. This interaction is accompanied by an intermodular reorientation within C4BP. The DNA-binding site of C4BP also involves CCP1 and CCP2 of the α-chain. Chemical shift changes locate the binding site for DNA to a groove at the CCP1/2 interface enabling the use of data-driven docking to produce a model of the C4BP12:DNA complex. The work described in this thesis thus provides detailed pictures of interactions whereby a pathogen evades complement and that help protect host tissue during programmed cell death.

Acknowledgments

I would like to thank the following people for their invaluable help:

Paul, Anna, Linda, Jenny, Gunnar, Dušan, Juraj, Andy, Graeme, Brian, Janice, Alan, Margaret, Christoph, Dinesh, Elin, Eve, Henry, Lena, the MRC, my parents, and most of all Nicky.

Unless stated in the text, the work described in this thesis is my own work and has not been submitted for in whole or in part for a degree or other qualification at this or any other university. Work described in this thesis has been published in Jenkins, H. T., Mark, L., Ball, G., Persson, J., Lindahl, G., Uhrin, D., Blom, A. M, and Barlow, P. N. (2006) *J. Biol. Chem.* 281(6):3690-7. This paper is included in Appendix B of this thesis and reproduced with the permission of the American Society for Biochemistry and Molecular Biology, Inc.

Huw Jenkins

Contents

Table of Figures

List of Tables

Abbreviations

2D	two-dimensional
3D	three-dimensional
ADR	ambiguous distance restraint
C4BP	C4b-binding protein
CCP	complement control protein
CNS	Crystallography and NMR system
CR1	complement receptor type 1
CR2	complement receptor type 2
CSI	chemical shift indexing
DAF	decay accelerating factor
DGSA	distance geometry simulated annealing
fH	factor H
fI	factor I
FID	free induction decay
FT	Fourier transform
HSQC	heteronuclear single quantum coherence
ISPA	isolated spin pair approximation
ITC	isothermal titration calorimetry
K_d	equilibrium dissociation constant
M_r	relative molecular mass
MAC	membrane attack complex
MASP	mannan-binding lectin associated serine protease
MBL	mannan-binding lectin
MCP	membrane cofactor protein
MDSA	molecular dynamics-based simulated annealing
MEM	maximum entropy method
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
PDB	protein data bank
PS	protein S
RCA	regulator of complement activation
RF	radio frequency
rmsd	root mean square deviation
SCR	short consensus repeat
SPR	surface plasmon resonance
TOCSY	total correlation spectroscopy
TROSY	transverse relaxation optimised spectroscopy
VCP	*Vaccinia* virus complement-binding protein

Chapter 1

INTRODUCTION: THE COMPLEMENT SYSTEM AND THE REGULATORS OF COMPLEMENT ACTIVATION

1.1 Overview

The complement system is a key molecular component of innate immunity. It consists of more than thirty proteins that participate in a well co-ordinated series of specific protein-protein recognition and catalytic events accompanied by the formation and destruction of multi-protein complexes. The complement system is potentially destructive to the host organism and is thus tightly regulated. Several inhibitors, known as the regulators of complement activation (RCA) family, are particularly important in the early stages of the complement cascade. The majority of RCA proteins act on the bimolecular C3 convertases that are the crucial enzymatic complexes of both the classical and alternative pathways of complement activation. The C3 convertase of the classical and lectin pathways is C4bC2a and this is regulated by C4b-binding protein (C4BP) which competes with C2 for binding to C4b and acts a co-factor for the factor I mediated cleavage of C4b (Fujita *et al.* 1978; Fujita *et al.* 1979); it also accelerates the decay of the C4bC2a complex (Gigli *et al.* 1979).

Like many members of the RCA family, C4BP is a target of several major pathogenic bacteria. This was first shown for *Streptococcus pyogenes* which expresses surface M proteins that in many strains have the ability to bind C4BP (Thern *et al.* 1995). More recently, C4BP has been shown to bind to all strains of *Bordetella pertussis* (Berggård *et al.* 2001b), some strains of *Neisseria gonorrhoeae* (Ram *et al.* 2001) and *Neisseria meningitidis* (Jarva *et al.* 2005), *Escherichia coli* (Prasadarao *et al.* 2002; Wooster *et al.* 2006), *Candida albicans* (Meri *et al.* 2004) and *Moraxella catarrhalis* (Nordström *et al.* 2004). Binding of this potent inhibitor of complement activation has been shown to contribute to bacterial virulence (Carlsson *et al.* 2003).

Unlike in the case of all other RCA proteins, until now there has been no detailed experimental 3D structural information for C4BP. Hence the atomic resolution basis of its

interactions with pathogen-borne and natural ligands has remained unknown. Opportunities for rational design of therapeutic interventions – e.g. to counter streptococcal infections, or to inhibit complement-mediated inflammation - have consequently been limited.

The work described in this thesis focuses on the structural details of the interaction between C4BP and *Streptococcus pyogenes* M proteins. In this chapter an overview of the complement system and the role of C4BP in complement regulation is described. A discussion of the pathogens that 'hijack' C4BP to evade the complement system follows, with emphasis on the *S. pyogenes* M proteins.

1.2 The complement system

Complement was first described in the 1890s as a heat labile serum protein that 'complemented' antibodies in the destruction of bacteria. The complement system is now known to consist of over thirty proteins that undergo a series of specific protein-protein recognition and enzymatic cleavage events (Figure 1.1). The first major outcome of attack by the complement system is the deposition of C3b on the surface of unprotected particles such as an invading microorganism. This process, known as opsonisation, marks the particle as a target for phagocytosis by cells such as macrophages and neutrophils. These cells are attracted to the site of infection by the chemoattractant properties of the anaphylotoxins C3a and C5a, which are the products of C3 and C5 cleavage respectively. Further fragmentation of C3b generates C3d (and C3dg) which stimulate antibody production by B-cells. C3b also nucleates assembly of further C3 convertase complexes, driving the progression of the cascade towards the assembly of the membrane attack complex (MAC) - a multi-component, membrane-spanning pore that punches holes in the outer membranes of invading microorganisms leading to lysis.

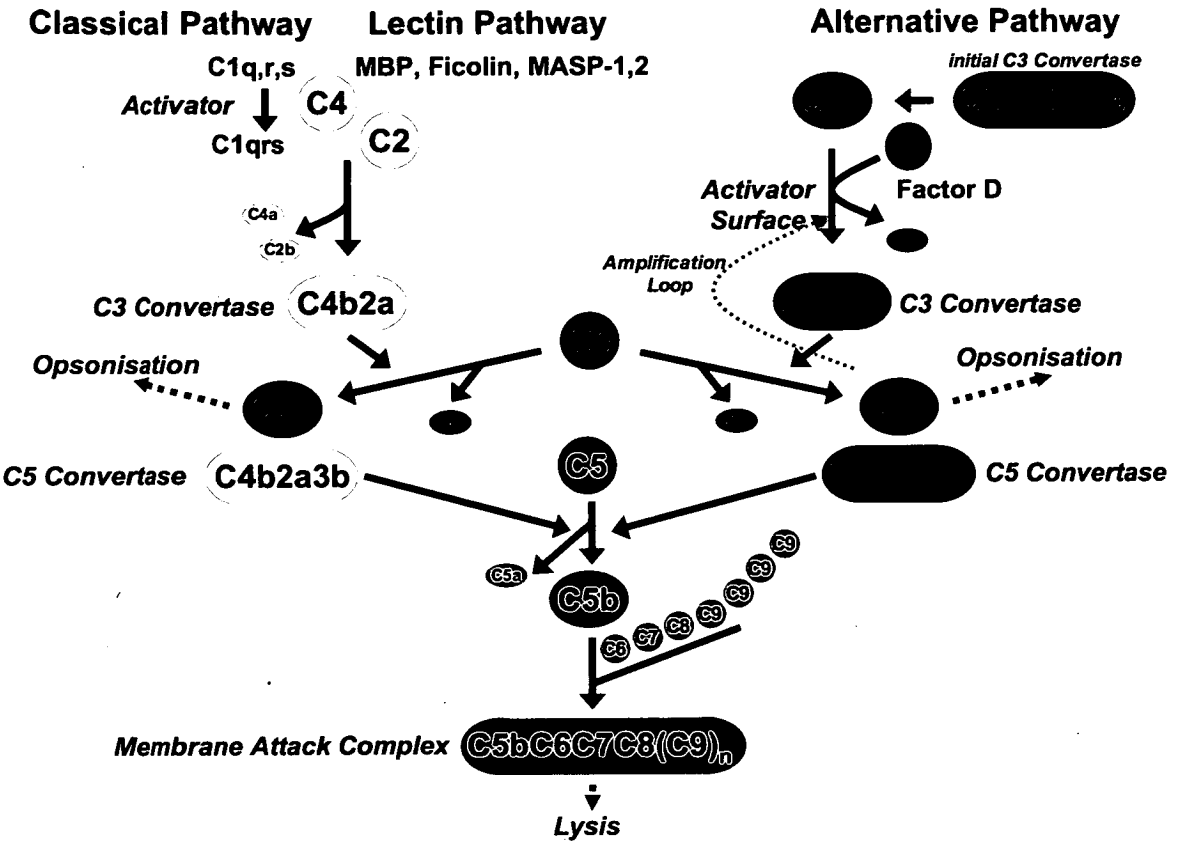

Figure 1.1 The complement system

1.2.1 The role of complement in disease

The complement system plays an important role in the pathogenesis of many diseases. Much of the understanding of the role of complement in disease has come from study of patients with inherited deficiencies of individual complement proteins (Colten *et al.* 1992). There are three types of disease linked with complement deficiencies: infectious diseases, systemic lupus erythematosus (SLE)-like diseases and C1 inhibitor deficiency-linked diseases.

The role of complement in infection is illustrated by the increased susceptibility to pyogenic infections caused by complement deficiency. Three types of infection are observed. Firstly, increased susceptibility to pyogenic bacteria such as *Streptococcus pyogenes* occurs in

patients with defects in either the C3 gene or in the control proteins factor I and factor H, deficiency of which leads to unregulated C3 consumption. This demonstrates that the normal pathway of defence against pyogenic bacteria is opsonisation with antibody, followed by complement activation, opsonisation with the covalently-bound cleavage products of C3 - C3b and iC3b, phagocytosis and intracellular killing. Secondly, deficiencies of the components of the MAC lead to infections with *Neisseria* demonstrating that complement-mediated lysis is critical for host defence against neisserial infection (Figueroa *et al.* 1993). Thirdly, recurrent pyogenic infections associated with mannose binding lectin (MBL) deficiency in young children suggest that the MBL pathway is important during the interval between the loss of passively acquired maternal antibody and the development of a mature immunologic repertoire (Summerfield *et al.* 1997).

SLE-like diseases are linked to deficiencies of the classical pathway components C1q, C1r, C1s, C4 and C2. These diseases are thought to illustrate the role of complement in the clearance of immune complexes and apoptotic cells (Botto *et al.* 1998).

In the autosomal dominant disease of C1 inhibitor deficiency, the single normal allele of the C1 inhibitor gene cannot ensure sufficient production of C1 inhibitor (Agostoni *et al.* 1992). This leads to loss of control of the activation of the complement serine esterases C1r and C1s, kallikrein of the kinin system and activated factors XI and XII of the coagulation system. The main clinical feature of C1 inhibitor deficiency is recurrent angioedema (swelling of the deep layers of the skin or mucous membranes) which may cause severe illness if it affects the intestinal submucosa or death by suffocation through obstruction of the upper airways.

1.2.2 The classical pathway

The complement system is activated by one of three routes: the classical, lectin or alternative pathways (Figure 1.1). All three pathways converge on the central step of C3 activation through their respective C3 convertases, C4b2a (classical and lectin) and C3bBb (alternative), but differ in the method of recognition of the invading particle. The recognition event in the classical pathway is the binding of C1q to either IgG or IgM in immune complexes resulting from recognition of the invading particle by the adaptive immune system or natural IgM antibodies. C1q circulates in blood bound to the two serine proteases C1r and C1s in the C1 complex ($C1qr_2s_2$) (Arlaud *et al.* 2002). Upon binding of the recognition component of C1q to the Fc subunit of an immunoglobulin the serine protease C1r is auto-activated and cleaves and activates C1s (Arlaud *et al.* 1993). Activated C1s translates the activation of the C1 complex into activation of the classical pathway by cleaving C2 and C4 to C2a and C4b respectively (releasing C2b and C4a) which together form the C3 convertase of the classical pathway - C4b2a (Arlaud *et al.* 1998).

1.2.3 The lectin pathway

The recognition event in the lectin pathway is mediated by the binding of either mannose binding lectin (MBL) (Presanis *et al.* 2003) or ficolin (Matsushita *et al.* 2000), associated with the mannose associated serine proteases (MASPs) MASP1 and MASP2, to specific patterns of carbohydrate groups on the surface of a bacterial cell. This leads to the cleavage of C2 and C4 by MASP2 leading to formation of the same C3 convertase (C4b2a) as in the classical pathway. MASP1 is also able to cleave C3 directly (Matsushita *et al.* 1995).

1.2.4 The alternative pathway

In contrast to the specific recognition events leading to activation of the classical and lectin pathways, the alternative pathway is constantly activated at a low 'tick over' level by the hydrolysis of C3 to C3(H_2O) - also known as C3i - which binds to factor B. Once factor B is cleaved and activated by factor D the C3iBb compex cleaves C3 to form C3b which then binds randomly via a thioester bond to any nearby host or bacterial surface. Factor B then binds to surface bound C3b and is cleaved to the active form Bb by factor D resulting in the formation of the C3 convertase of the alternative pathway - C3bBb. This activates an amplification loop resulting in deposition of many molecules of C3b on the surface surrounding the region of complement activation.

1.3 The regulators of complement activation

The C3 convertases represent, therefore, a crucial step in the complement system and it is at this point that many of the regulators of complement activation act. The serine proteases of the classical and alternative pathway C3 convertases - C2 and factor B, respectively - are only active when cleaved by proteolysis and transiently complexed to C4b or C3b. There is no natural inhibitor of the C3 convertases but the activity of the C4b2a and C3bBb complexes is regulated in other ways. Firstly, the complexes are unstable and dissociate with a half-life of a few minutes. Secondly, the human complement regulatory proteins C4BP, factor H (fH), decay accelerating factor (DAF, CD55), membrane cofactor protein (MCP, CD46) and complement regulator type 1 (CR1, CD35) interact with the complexes and promote dissociation - a process termed decay acceleration. Once C2a or Bb dissociates they do not reassociate with C4b or C3b. Thirdly, C4BP, fH, MCP or CR1 bind to the C4b or C3b subunits of the convertase and promote their cleavage by factor I to C4c plus C4d or iC3b respectively - a process termed (factor I) cofactor activity.

The properties of the human RCA proteins C4BP, factor H, DAF, MCP, CR1 and CR2 are summarised in Table 1.1. CR2 has neither decay accelerating nor cofactor activity but acts as the receptor for C3d and iC3b and thus enhances the humoral immune response of B lymphocytes (Molina *et al.* 1996) and follicular dendritic cells (Fang *et al.* 1998). Factor H-like protein 1 (FHL-1), a splice variant of fH containing the N-terminal seven CCP modules and a C-terminal extension of four residues (Zipfel *et al.* 1999) also acts as a complement regulator. The role of the five factor H related proteins (FHR 1-5) is less well understood.

RCA	Mw (kDa)	Number of CCP modules	Primary Ligand	Decay accelerating activity	Factor I cofactor activity	Key reference
C4BP	570	α-chain: 7* β-chain: 3*	C4b†	X	X	(Scharfstein *et al.* 1978)
FH	~155	20	C3b	X	X	(Sim *et al.* 1982)
FHL-1	43	7	C3b	X	X	(Misasi *et al.* 1989)
DAF	70	4	C3b/C4b/ C3 Convertases	X‡		(Nicholson-Weller *et al.* 1982)
MCP	~65/55§	4	C3b/C4b		X	(Seya *et al.* 1986)
CR1	~250	30*	C3b/C4b	X	X	(Fearon 1979)
CR2	145	15*	iC3/C3d/C3dg			(Weis *et al.* 1988)

Table 1.1 Members of the RCA family. Notes: * Most common isoform in human plasma; † Also binds C3b *in vitro*; ‡ Also accelerates decay of C5 convertase; § Two forms found in plasma.

1.3.1 The complement control protein module

The proteins involved in the complement system are often 'mosaic' in nature, built up from a mixture of module (domain) types (Bork *et al.* 1996). Each module type is defined by a specific consensus sequence. The most common protein motif found in the RCA family is the complement control protein (CCP) module - also referred to as a short consensus repeat (SCR) or sushi domain. Of the RCA proteins two – factor H and the viral homologues produced by pox virii such as vaccinia virus complement control protein (VCP) – are composed entirely of multiple CCP modules (Reid *et al.* 1989), the remaining five – C4BP, MCP, DAF, CR1 and CR2 – are composed almost entirely of CCP modules (Table 1.1).

In 1990 a single CCP module – the 16th module of factor H (fH~16) – was expressed in *Saccharomyces cerevisiae* and studied by two-dimensional homonuclear NMR (Barlow *et al.* 1991) This confirmed the presence of β-strands as predicted by circular dichroism (Sim *et al.* 1982) and Fourier transform-infrared spectroscopy (FT-IR) (Perkins *et al.* 1988). Subsequently the three-dimensional (3D) solution structure of fH~16 was solved (Norman *et al.* 1991) – the first of any CCP module. This structure (Figure 1.2) showed that fH~16 is based on a β-sandwich arrangement with one β-sheet composed of three strands and the other of two. As the last strand is composed of two segments separated by a β-bulge this may be regarded as two separate strands (as shown in Figure 1.2). A hydrophobic core is formed from side chains from both sheets and the whole arrangement is stabilised by two disulphide bonds between the four consensus cysteines in a 1-3, 2-4 arrangement. Overall the module has an elongated shape with the N- and C-termini at opposite ends of the long axis. The two disulphides lie far apart, forming the boundaries of the hydrophobic core.

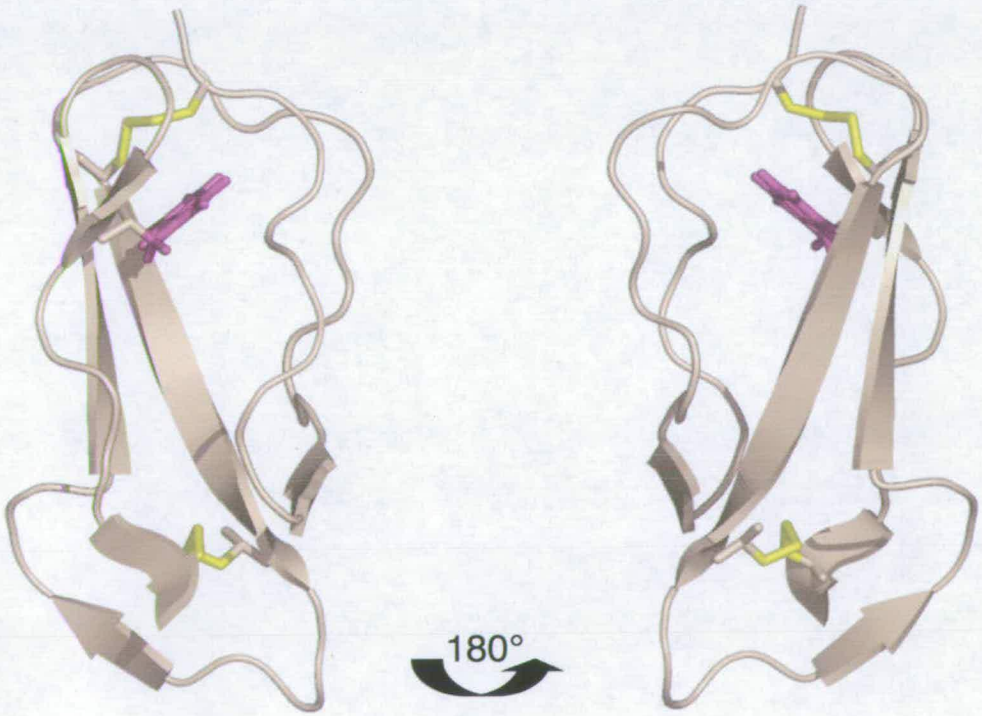

Figure 1.2 Cartoon showing the solution structure of fH~16 solved by NMR (pdb = 1HCC). The 2 conserved disulphide bonds are shown in yellow and the conserved tryptophan in purple. Figure produced using PyMol (www.pymol.org).

Examination of sequence alignments (Figure 1.3) of CCP modules with reference to the structure of fH~16 showed that many residues that are structurally important – for example, the buried tryptophan prior to the last consensus cysteine – are conserved or conservatively replaced. This suggested that the structure of other CCP modules would be similar to that of fH~16. This has been confirmed by the solution NMR structures of CCP modules from other RCA proteins such as VCP (Wiles *et al.* 1997; Henderson *et al.* 2001), CR1 (Smith *et al.* 2002), fH (Herbert *et al.* 2006) and DAF (Uhrinova *et al.* 2003) as well as the structures of modules from MCP (Casasnovas *et al.* 1999), VCP (Murthy *et al.* 2001; Ganesh *et al.* 2004), DAF (Williams *et al.* 2003; Lukacik *et al.* 2004), fH (Jokiranta *et al.* 2006) and CR2 (Szakonyi *et al.* 2001; Prota *et al.* 2002) solved by X-ray crystallography. The structures show a common scaffold-like structure with up to eight β-strands with variable loops, bulges and turns.

```
Name        SS β1        β2  HV-loop  β3    β4       β5        β6          β7     β8
C4BPA_01      NCGPPPTLSFAAPM-DITLTET-RFKTGTTLKYTCLPGYVRSH-STQTLTCNSDG---EWV-YNTFCIYK
DAF_02        SCEVPTRLNSASLK-QPYITQN-YFPVGTVVEYECRPGYRREPSLSPKLTCLQNL---KWSTAVEFCKKK
CR1_15        HCQAPDHFLFAKL--KTQTNAS-DFPIGTSLKYECRPEYYGRP---FSITCLDNL---VWSSPKDVCKRK
CR1_01        QCNAPEWLPFARP--TNLTDEF-EFPIGTYLNYECRPGYSGRP---FSIICLKNS---VWTGAKDRCRRK
C4BPA_02      RCRHPGELRNG-QVE---IKT-DLSFGSQIEFSCSEGFFLIG--STTSRCEVQDRGVGWSHPLPQCEIV
DAF_03        SCPNPGEIRNG-QID---VPG-GILFGATISFSCNTGYKLFG--STSSFCLISGSSVQWSDPLPECREI
CR1_16        SCKTPPDPVNG-MVH---VIT-DIQVGSRINYSCTTGHRLIG--HSSAECILSGNTAHWSTKPPICQRI
CR1_02        SCRNPPDPVNG-MVH---VIK-GIQFGSQIKYSCTKGYRLIG--SSSATCIISGDTVIWDNETPICDRI
```

Figure 1.3 Sequence alignment of CCP1 and CCP2 of the C4BP α-chain with CCP2 and CCP3 of DAF, and CCP1, CCP2, CCP15 and CCP16 of CR1. The location of the eight β-strands and the hypervariable (HV-) loop (indicated with *italics*) commonly observed in the structures of CCP modules is indicated in the line designated ss (secondary structure)

In cases where the structures of the same CCP module were solved by both NMR and x-ray crystallography, the agreement varies from good (e.g C^{α} rmsd of 1.14 Å for VCP~3) to poor (C^{α} rmsd of 3.31 Å for VCP~2). In the case of VCP~3 this difference could be explained by the presence or absence of the neighbouring modules as the NMR structure was of the VCP~2,3 module pair whereas the crystal structure was of the intact protein. However the NMR structures of CR1~16 as a single module or in the context of the CR1~15,16 and

CR1~16,17 pairs are essentially identical. The variation in structures may, therefore, represent differences in conformation under the different conditions (i.e pH, salt concentration, presence/absence of detergent) under which the structures were solved.

1.4 C4b-binding protein

C4BP is a 570 kDa plasma glycoprotein with a plasma concentration of ~200 mg/l (Dahlbäck *et al.* 1983). C4BP is the only circulating complement inhibitor with a polymeric structure. In humans the most common isoform (75-80% of C4BP in plasma) consists of seven identical α-chains containing eight CCP modules and a C-terminal region that forms two disulphide bridges with the C-terminal region of a single β-chain containing three CCP modules (Scharfstein *et al.* 1978; Hillarp *et al.* 1990). The less abundant forms consist of seven α-chains or six α-chains and one β-chain (Hillarp *et al.* 1989; Sánchez-Corral *et al.* 1995).

C4BP was first visualised by electron microscopy (EM) (Dahlbäck *et al.* 1983) and reported to have a spider like structure with the α-chains forming 330 Å-long tentacular structures extending from a small central body together with the short arm of the β-chain. This chain length corresponds to 41 Å per module – suggesting that the average tilt angle between successive modules is shallow. Some arms are bent in one or two places. X-ray scattering and hydrodynamics (Perkins *et al.* 1986) suggest a less sprawling arrangement with an angle of approximately 10° between the α-chains. A schematic representation of C4BP (with the spider-like arrangement shown in the EM studies) is shown in Figure 1.4.

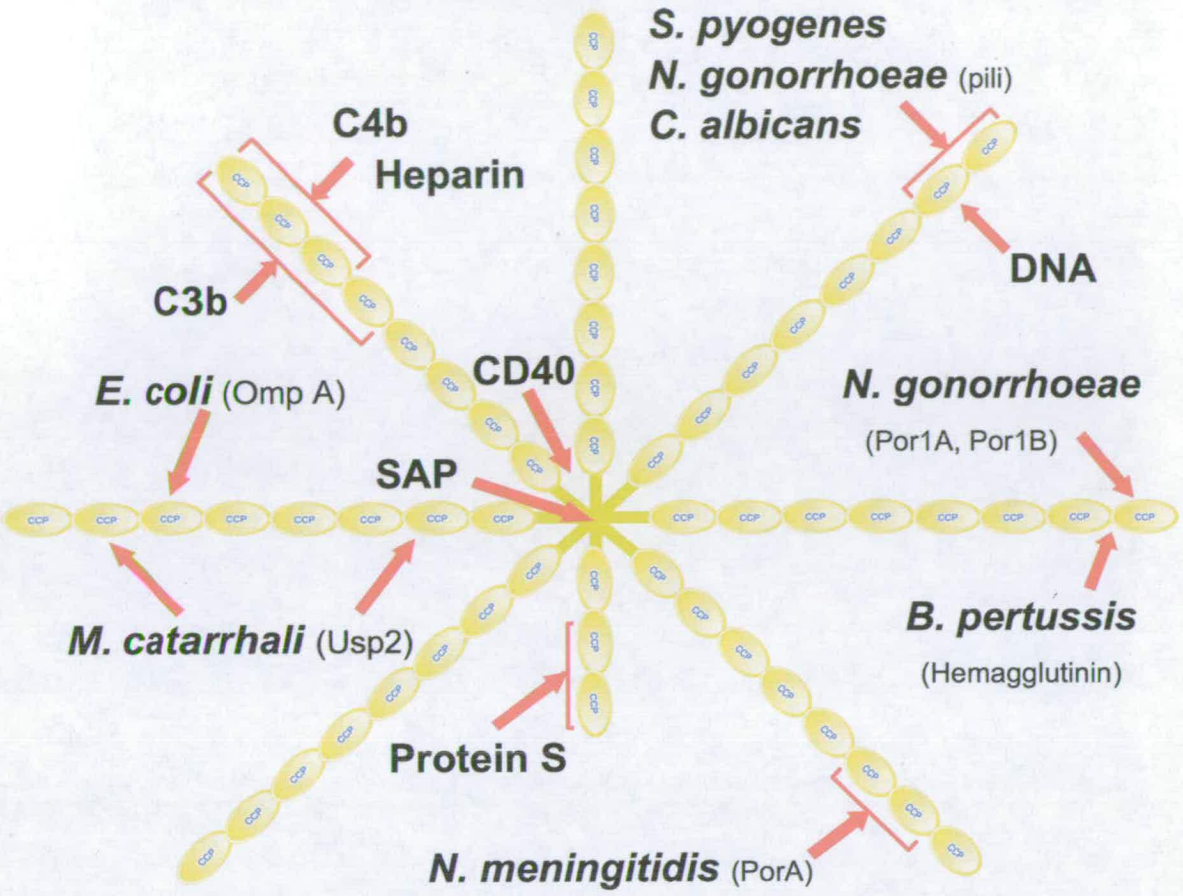

Figure 1.4 Schematic representation of C4BP with indicated binding sites for ligands

1.5 C4BP interactions

To date, there has been no high-resolution structural work on C4BP, however, homology modelling has been used in combination with site-directed mutagenesis to investigate structure-function relationships (Blom *et al.* 1999; Villoutreix *et al.* 1999). C4BP controls C4b-mediated reactions and, therefore, inhibits the classical pathway of complement activation in at least three ways. First C4BP acts as a cofactor in the proteolytic inactivation of C4b by the serine protease factor I (Fujita *et al.* 1978; Fujita *et al.* 1979). The mechanism by which C4BP acts as a cofactor to factor I is not fully understood. Second, C4BP prevents assembly of C4bC2a – the classical C3-convertase. Third, C4BP accelerates decay of the C4bC2a complex (Gigli *et al.* 1979). C4BP has been shown to exist in complex with the

vitamin K-dependent anti-coagulant protein S (PS) (Dahlbäck *et al.* 1981), and to bind heparin (Hessing *et al.* 1990), serum amyloid P component (SAP) (Schwalbe *et al.* 1990), DNA (Trouw *et al.* 2005) and C-reactive protein (CRP) (Sjoberg *et al.* 2006).

1.5.1 C4b binding

The EM studies of Dahlbäck *et al.* showed that up to six molecules of C4b bind to C4BP and that the binding site for C4b is located at the end of each α-chain. Härdig *et al.* showed, using chimeric proteins composed of C4BP α-chains with one, two or three of the amino-terminal CCPs replaced by corresponding CCPs from the C4BP β-chain, that the N-terminal CCP module of the C4BP α-chain (CCP1) is crucial for C4b binding and factor I-cofactor activity (Härdig *et al.* 1997). Previously Ogata *et al.* had used cell surface-bound forms of murine C4b-binding protein (mC4BP), which enabled the binding of mC4BP to C4b to be monitored with relatively simple erythrocyte rosette assays, to demonstrate that CCP 1-3 of mC4BP are necessary and sufficient for binding to C4b (Ogata *et al.* 1993). To further investigate C4BP binding to C4b, Blom *et al.* constructed a panel of eight mutants in which individual CCP modules were removed one at a time. Truncated monomeric forms of the α-chain were also created by inserting stop codons after each CCP (Figure 1.5). The recombinant proteins were expressed in a eukaryotic cell line and purified by affinity chromatography using monoclonal antibiodies. These mutants were assayed for their ability to bind C4b using a competition assay in which the variants were allowed to compete with ^{125}I-labeled wild-type C4BP for binding to immobilised C4b. The ability to bind C4b was severely impaired in the case of C4BP mutants lacking CCP1, CCP2 (in particular) or CCP3 indicating that CCPs 1-3 are crucial for binding (Blom *et al.* 2001a). The spacing between the CCPs was also shown to be important – alanine insertions between CCP1 and CCP2, and between CCP2 and CCP3, both disrupted C4b binding (Blom *et al.* 2001a).

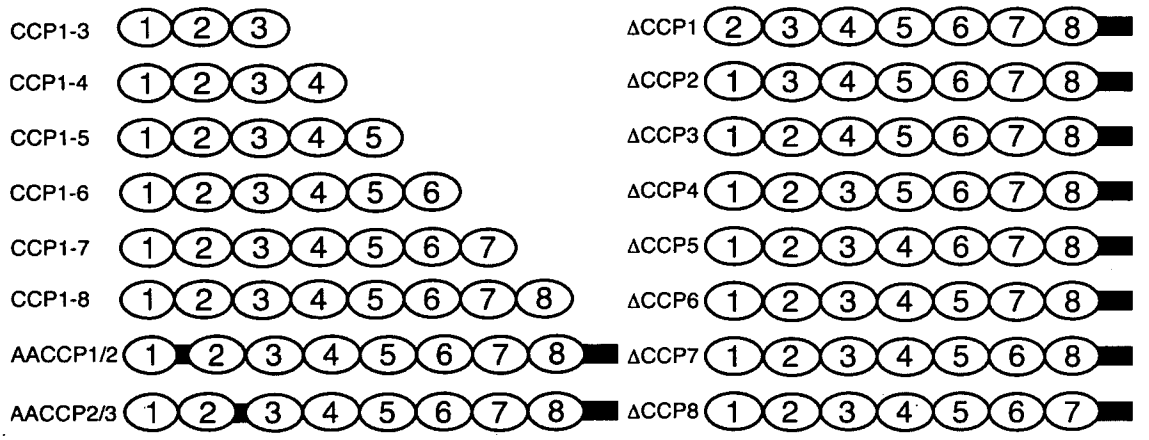

Figure 1.5 Recombinant forms of C4BP α-chain constructed by Blom *et al.* (left) C-terminal truncated forms and alanine insertion forms, (right) individual CCP module deletions.

The interaction between C4b and C4BP has been shown to be highly sensitive to salt concentration (Blom *et al.* 2001a) suggesting that ionic interactions between charged residues are involved. Using information from a homology-based model of α-chain CCP1-2 to guide site-directed mutagenesis, Blom *et al.* found that Arg^{39}, Lys^{63}, Arg^{64} and His^{67}, which are all located in a positively charged cluster on the CCP1/CCP2 interface of the model structure, are crucial for C4b binding (Blom *et al.* 1999; Blom *et al.* 2000). Using these mutants, and truncated α-chains of C4BP, this group also showed that CCPs 1-3 of the α-chain are sufficient for cofactor activity in the cleavage of C4b molecules (Blom *et al.* 2001a). Recently the same group have demonstrated that C4BP can serve as a cofactor to factor I in the cleavage of fluid phase C3b (a role normally played by fH) and thus influence the alternative pathway of complement (Blom *et al.* 2003). All four N-terminal CCP modules of the α-chain are required for this activity, with CCP2 and 3 being the most important. Since C4BP has also been shown to regulate the lectin pathway of complement (due at least in part to mannan-binding lectin enhancement of C4BP binding to C4b) (Suankratay *et al.* 1999), C4BP is now known to be involved in regulating all three pathways of complement.

C4BP has very recently been shown to bind to both the C4c and C4dg subfragments of C4b. (Leung *et al.* 2006). CCPs 1-3 were required for binding, and analysis of the panel of mutants used in previous studies by Blom *et al.* suggested the C4c and C4dg binding sites were different but cooperative. The kinetics observed in the surface plasmon resonance (SPR) study led the authors to conclude that there were two affinity states of the C4c binding subsite in each C4BP molecule and that the equilibrium between these two states is shifted towards the high affinity state upon C4dg binding to the C4dg subsite.

1.5.2 Heparin binding

C4BP has been shown to bind heparin, a model for glycosaminoglycans, tightly (Hessing *et al.* 1990) although the affinity constant for this interaction has not been determined. High concentrations of heparin inhibit the interaction between C4BP and C4b suggesting that the C4b- and heparin-binding sites overlap (Garcia de Frutos *et al.* 1994). Blom *et al.* used heparin affinity chromatography and surface plasmon resonance (SPR) with biotinylated heparin immobilised on a streptavadin chip, to characterise the mutants of C4BP described previously. These experiments showed that deletion of CCP1 and CCP3 had minor effects on heparin binding. Deletion of CCP2 or insertion of two alanine residues between CCP1 and CCP2 compromised binding suggesting that CCP2 is the most important module for heparin binding. Additionally mutation of the positively charged residues on the putative CCP1/2 interface almost abolished binding (Blom *et al.* 1999). This suggests that the interaction between C4BP and heparin is similar in nature to the interaction between C4BP and C4b. The C4BP-heparin interaction is physiologically interesting as the multivalency of C4BP would permit occupancy of some heparin-binding sites by host cell-surface glycosaminoglycans, leaving others to mediate decay acceleration and cofactor activity, thus protecting host cells from the destructive and inflammatory results of complement activation.

1.5.3 Protein S

The β-chain of C4BP contains a high affinity binding site for the vitamin K-dependent protein S (PS) and protein S is bound to all C4BP circulating in plasma (Dahlbäck *et al.* 1983; Hillarp *et al.* 1988; Hardig *et al.* 1996). The Gla domain of protein S binds to negatively charged phospholipids which aids in the localisation of C4BP to apoptotic cells (Webb *et al.* 2002). Studies have suggested that early complement components are important for the rapid clearance of apoptotic cells (Korb *et al.* 1997; Taylor *et al.* 2000). However, the cell must be protected from assembly of later components in order not to provoke an inflammatory response triggered by the complement system through the release of the anaphylatoxins C5a and C3a. Thus the localisation of C4BP to apoptotic cells may be vital to prevent inflammation in the surrounding area.

1.5.4 DNA binding

Recently C4BP was shown to bind to 25-bp double-stranded DNA with an affinity of 190 nM (Trouw *et al.* 2005). Dot blot analysis showed binding to both single and double stranded DNA and also to RNA. Module deletion studies localised the DNA binding site to CCP2 with a small contribution from CCP1. Mutagenesis studies showed that the positively charged residues Arg^{39}, Lys^{63} and Arg^{66} were essential for DNA binding with some contribution from Arg^{64}. Mutation of His^{67} (which is required for C4b-binding) had no effect on DNA binding. C4b and heparin both compete with DNA for C4BP binding. The C4BP-PS complex was shown to localise to necrotic cells both via the interactions between PS and phosphatidylserine, and that between C4BP and DNA. C4BP then acts to inhibit complement activation on these cells preventing C5a release, inflammation and cell lysis.

1.6 Interaction with pathogens

Many pathogens have developed strategies that enable them to avoid clearance and destruction by complement, either by hijacking host RCA proteins to locally down-regulate complement activation, or by expressing their own regulators that are remarkably similar to the host proteins [reviewed in (Lindahl *et al.* 2000)]. The list of pathogens that bind C4BP in order to evade the immune system is growing rapidly and to date a direct link between C4BP binding and resistance to serum-mediated killing has been shown for three pathogens: *Streptococcus pyogenes* (Carlsson *et al.* 2003), *Neisseria gonorrhoeae* (Ngampasutadol *et al.* 2005) and *Escherichia coli* (Wooster *et al.* 2006). These interactions are summarised in Table 1.2, together with the binding sites for the 'natural' ligands: C4b, protein S and heparin, and also DNA and C3b.

Ligand/organism	Interaction site on C4BP	Interaction	References
C4b	CCP1-**2-3** of α-chain: +ve charged aa at CCP1/2 interface	Ionic	(Blom *et al.* 1999)
Heparin	CCP1-**2**-3 of α-chain: +ve charged aa at CCP1/2 interface	Ionic	(Blom *et al.* 1999)
C3b	CCP1-**2-3**-4 of α-chain	Ionic	(Blom *et al.* 2003)
Protein S	CCP **1**-2 of β-chain	Hydrophobic	(Hardig *et al.* 1996)
DNA	CCP1-**2** of α-chain: +ve charged aa at CCP1/2 interface	Ionic	(Trouw *et al.* 2005)
S. pyogenes M proteins	CCP**1**-2 of α-chain	Hydrophobic	(Blom *et al.* 2000)
B. pertussis	+ve charged aa at CCP1/2 interface	Ionic	(Berggård *et al.* 2001b)
N. gonorrhoeae Por1A/Por1B/pili	Por1A: CCP1 of α-chain Por1B: CCP1 of α-chain Pili: CCP1-2 of α-chain	Hydrophobic Ionic Ionic	(Ram *et al.* 2001) (Ram *et al.* 2001) (Blom *et al.* 2001b)
N. meningitidis PorA	CCP2-3 of α-chain	Ionic	(Jarva *et al.* 2005)
C. albicans	CCP1-2 of α-chain	Ionic	(Meri *et al.* 2004)
M.catarrhalis UspA1, UspA2	CCP**2**, 5 & 7 of α-chain	Hydrophobic	(Nordström *et al.* 2004)
E. coli OmpA	CCP3 of α-chain	Hydrophobic	(Prasadarao *et al.* 2002)

Table 1.2 Interaction sites on C4BP for various ligands and pathogens. For some cases the major interaction site is indicated with bold text.

1.6.1 Streptococcus pyogenes

The most extensively studied *S. pyogenes* virulence factor, identified over 75 years ago (Lancefield 1928), is M protein – a dimeric coiled coil protein that has antiphagocytic function and is a target for protective antibodies. This protective immunity is directed against an amino terminal hypervariable region (HVR) that it a source of antigenic variation (Robinson *et al.* 1992). The HVR, which has a length of ~50-100 residues, is stable within a strain, but ~100 different HVRs with highly divergent sequences have been identified through analysis of different clinical isolates. The function of the HVR was unclear until it emerged that many HVRs, with very different sequences, share the ability to bind to C4BP (Thern *et al.* 1995; Johnsson *et al.* 1996). More recently it has been shown that C4BP-binding HVRs bind their ligand with high specificity; moreover these HVRs can be emulated by synthetic peptides that bind only C4BP among all human serum proteins (Morfeldt *et al.* 2001). The five different C4BP-binding HVRs (shown in Figure 1.6) exhibit no or very little sequence identity. The conserved L-X(2)-E-X(8)-D motif (highlighted in Figure 1.6) occurs in many streptococcal proteins that do not bind C4BP and do not cross-react immunologically; these include M5 as shown in Figure 1.6.

```
M2              NSKNPVPVKKEAKLSEAELHDKIKNLEEEKAELFEKLDKVEEEHKKVE
M4            AEIKKPQADSAWNWPKEYNALLKENEELKVEKEKYLSYADD-KEKD
M22           ESSNNAESSNISQESKLINTLTDENEKLREELQQYYALSDA-KEEEPRYKALR
M60   ESSTVKAESSTVKAESSTISKERELINTLVDENNKLMEERARHLDLIDNIREKDPQYRALRGENQD
PrtH   EGAKIDWQEEYKKLDEDNAKLVEVVETTSLENEKLKSENEENKKNLDK-LSKD

M5    YELENHDLKTKNEGLKTENEGLKTENEGLKTENEGLKTEKKEHEAENDKLKQQRDTLST
```

Figure 1.6 Multiple sequence alignment of the hypervariable regions of five M proteins that bind C4BP with the sequence of the non-C4BP-binding M5 protein shown for comparison. The sequences of the M2, M4 and M22 proteins are those of the synthetic peptides used by Morfeldt *et al.* (2001). The conserved L-X(2)-E-X(8)-D motif is highlighted.

The M protein can be divided up into a the HVR, a semi-variable and a conserved region as shown for M4 in Figure 1.7. Two well defined ligand binding regions have been identified in many M proteins – a C4BP-binding region as previously described, and an IgA-Fc binding region (Johnsson *et al.* 1994; Johnsson *et al.* 1996) in the semivariable region. These sites are non-overlapping – binding of C4BP does not affect binding of IgA and vice-versa (Thern *et al.* 1995). Also synthetic peptides corresponding to each region specifically bind either C4BP or IgA (Johnsson *et al.* 1999; Morfeldt *et al.* 2001), further evidence that each corresponds to a distinct protein domain.

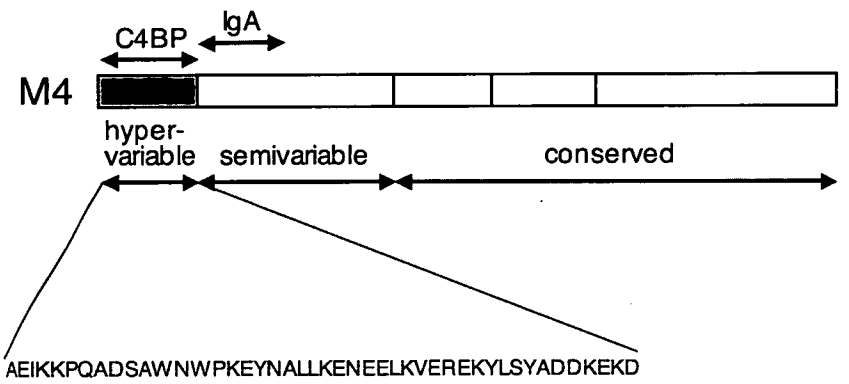

Figure 1.7 Schematic representation of the M4 protein showing the hypervariable, semi-variable and conserved regions. The amino acid sequence of the 45 residue hypervariable C4BP-binding region is shown.

Until recently no high resolution structural work had been published on the M proteins but they had been proposed to have a dimeric coiled-coil structure (Nilson *et al.* 1995; Cedervall *et al.* 1997). It was not known whether this coiled-coil structure extended into the HVR. Originally circular dichroism (CD) spectroscopy indicated that the peptides corresponding to M2, M4 and M22 are 30-50% α-helical but they do not appear to have a coiled-coil structure when dimerised via a C-terminal disulphide bond between a non-native cysteine residue in each peptide (Morfeldt *et al.* 2001). Computational modelling of secondary structure supported the CD data, with three different neural network-based prediction servers

[PsiPred2 (McGuffin *et al.* 2000), Jnet (Cuff *et al.* 2000) and and PhD (Rost 1996)] all predicting helix (60% in each HVR) and coil only. For all the HVRs except M2 the consensus prediction amongst the three servers includes two α-helices with the connecting loops well-aligned when the sequences are positioned as in the multiple sequence alignment (Figure 1.6). Analysis with fold recognition methods such as 3D-PSSM (Kelley *et al.* 2000) indicated that all five C4BP-binding HVRs adopt a helix-turn-helix conformation (Morfeldt *et al.* 2001).

Recent NMR studies on a peptide corresponding to the HVR of the M4 protein dimerised via a non-native C-terminal cysteine - termed M4-N - led the authors to conclude this peptide does in fact form a coiled-coil structure in solution, but only the central 27 residues are structured with residues 1-12 and 40-46 unstructured (André *et al.* 2006). No evidence to support a helix-turn-helix structure was found. Upon binding to C4BP12 the disappearance of certain peaks in the ^{15}N-HSQC spectra showed that all of the 27 structured residues are affected by C4BP binding and that the remaining residues remain flexible when bound to C4BP.

The important role of C4BP binding in phagocytosis resistance of *S. pyogenes*, potentially through cooperativity with IgA binding, has been well documented (Berggård *et al.* 2001a; Carlsson *et al.* 2003).

1.6.2 Bordetella pertussis

All clinical isolates of *Bordetella pertussis*, the Gram-negative bacterium that causes the childhood disease whooping cough, bind C4BP through at least two different surface-exposed components. One of these is thought to be the virulence factor filamentous hemagglutinin (FHA) on the basis that *fha* mutants show reduced C4BP binding. Purified

FHA does not bind C4BP, however, so FHA may act indirectly. The mode of binding of *B. pertussis* to C4BP resembles that of C4b – i.e. ionic in nature and involving a cluster of positively charged residues at the CCP1/2 module interface, with mutagenesis data suggesting that Arg^{64} and Arg^{66} are involved (Berggård *et al.* 2001b).

1.6.3 Neisseria gonorrhoeae

Neisseria gonorrhoeae outer membrane porin molecules (Por1A and Por1B) have been shown to bind to C4BP (Ram *et al.* 2001). This study showed that C4BP-Por1B interactions are ionic in nature and can be inhibited by high salt or by heparin – whereas C4BP-Por1A binding is predominantly hydrophobic. Only C4BP molecules that contain CCP1 bound to Por1A and Por1B gonococci, suggesting that CCP1 contributes to both porin-binding sites. Inhibiting C4BP binding with fAb fragments against C4BP CCP1 resulted in complete killing of otherwise fully serum-resistant strains in serum that had been diluted 10-fold; this demonstrates the importance of C4BP in mediating gonococcal serum resistance. Isolated type IV pili from *N. gonorrhoeae* have also been demonstrated to bind human C4BP (Blom *et al.* 2001b). An inhibition assay with C4b, and a competition assay in which Blom *et al.* tested mutants of C4BP lacking individual CCPs, led to the conclusion that the binding area for pili is localized to CCP1 and CCP2 of the α-chain. The binding between pili and C4BP was abolished at 0.25 M NaCl, implying that like C4b-C4BP binding it is mainly mediated by ionic interactions.

More recently, Ngampasutadol *et al.* showed that *N. gonorrhoeae* strains expressing Por1A bound only to human C4BP and not to chimpanzee or rhesus monkey C4BP (Ngampasutadol *et al.* 2005). Strains expressing Por1B could bind human and chimpanzee C4BP but not rhesus C4BP. Comparing the sequences of C4BP from these three species suggests that up to four residues that are different in CCP1 between human and chimpanzee C4BP - Ala^{12},

Met^{14}, Arg^{22} and Leu^{34} are critical in Por1A binding to C4BP. Three of these substitutions - M14V, R22H and L34R - are also present in rhesus C4BP, however there are an additional ten amino acid differences between rhesus and human C4BP, suggesting that other residues are involved in Por1B binding to C4BP.

1.6.4 Neisseria meningitidis

In 2005 Jarva *et al.* showed that serogroup B *Neisseria meningitidis*, one of the most common serotypes isolated from patients with meningitis, binds C4BP (Jarva *et al.* 2005). Analysis of wild-type Group B meningococcus and 11 isogenic mutant strains showed that all strains expressing the surface porin PorA were able to bind C4BP, whereas strains lacking in PorA expression showed significantly reduced C4BP binding. Through a C4b-cofactor assay for the factor I-mediated degradation of C4b, the bound C4BP was shown to be active and this suggested that the resistance of PorA-expressing strains to complement-mediated killing was due to these strains binding functionally active C4BP, leading to down-regulation of the classical pathway of complement deposition on the surface of the bacterium.

Module-deletion studies suggested that the binding site for PorA on C4BP involved CCP2 and CCP3. Peptide scanning assays with radiolabelled C4BP and peptide sequences corresponding to putative surface-exposed loops in PorA identified two similar sequences from loop 1 that bound C4BP most strongly.

1.6.5 Candida albicans

Both the fungal and hyphal forms of *Candida albicans*, an important pathogenic yeast, have been shown to bind C4BP (Meri *et al.* 2004). *C. albicans* also binds fH and FHL-1 and

simultaneous binding of C4BP and fH was observed by confocal microscopy. Surface-attached C4BP was shown to retain activity and thus down-regulate complement activation, leading to immune evasion by *C. albicans*.

1.6.6 Moraxella catarrhalis

Ubiquitous surface proteins A1 (UspA1) and A2 (UspA2) of the mucosal pathogen *Moraxella catarrhalis* have been shown to bind C4BP (Nordström *et al.* 2004) with the K_d of the UspA2:C4BP interaction being 10-fold smaller than that of the UspA1:interaction. Module-deletion studies suggested that CCP2, and surprisingly CCP5 and CCP7 - which have not previously been implicated in interactions between C4BP and either its 'natural' ligands or pathogens - were involved. Again C4b-degradation assays showed that bound C4BP maintained its cofactor activity suggesting that C4BP-binding enables *M. catarrhalis* to evade the classical pathway of complement activation.

1.6.7 Escherichia coli

The pathogenic K1 strain of the *Escherichia coli,* which is responsible for meningitis in neonates, binds C4BP (Prasadarao *et al.* 2002). The interaction site was localised to the N-terminus of outer protein A (OmpA). Module deletion studies showed that the interaction site on C4BP was CCP3 of the α-chain. The interaction was not significantly inhibited by C4b or heparin and was not sensitive to salt concentration – suggesting that it is mainly hydrophobic in nature. This study also showed that synthetic peptides corresponding to CCP3 sequences block the binding of C4BP to OmpA and significantly enhance serum bactericidal activity.

1.6.8 Borrelia recurrentis and Borrelia duttonii

The spirochetes *Borrelia recurrentis* and *Borrelia duttonii* are, respectively, the causative agents of louse-borne relapsing fever (transmitted by the human body louse) and endemic or tick-borne relapsing fever (transmitted by soft-bodied ticks). Both these pathogens have recently been shown to evade the complement system by specifically binding both factor H and C4BP (Meri *et al.* 2006). Both complement regulators were shown to retain functional activity when bound to the surface of the spirochetes. This is the fourth example of a pathogen that has been demonstrated to bind factor H and C4BP - the others are: *N. gonorrhoeae*, *S. pyogenes* (although the strains that bind C4BP do not appear to bind fH and vice versa) and *C. albicans*.

1.7 Conclusions

C4b-binding protein, the key fluid phase regulator of the classical pathway is targeted by many pathogenic micro-organisms, which 'hi-jack' C4BP in order to evade the complement system. The most widely studied interaction of a pathogen with C4BP is that of *S. pyogenes* M proteins. This interaction is becoming increasingly relevant as the prevalence of antibiotic resistant strains of *S. pyogenes* increases (Lamagni *et al.* 2005) and therefore, the need to develop alternative therapeutics to treat *S. pyogenes* infections becomes more pressing.

The C4BP-binding region of the M protein has been localised to the N-terminal hypervariable region (HVR) and the high sequence variation in this region between M proteins has to date prevented the identification of a C4BP binding motif in the M proteins. The interaction site on C4BP has been more closely defined through a series of mutagenesis studies, however, without the 3D structure of the M protein binding region of C4BP - CCP modules 1 and 2 of the α-chain - the role of these residues in binding M protein versus their role in maintaining the structure of the CCP1/2 interface has remained unclear. Therefore, to understand the structural basis for the interaction of M proteins with C4BP the structure of

this region of C4BP is required. This will allow the residues involved in the interaction of M protein with C4BP to be identified by chemical shift mapping experiments and plotted onto the 3D structure of C4BP12. The mutagenesis data can then be interpreted in the light of the structure and the method of interaction of M protein with C4BP can be elucidated.

It has been shown that CCPs 1-3 of the C4BP α-chain are required for binding of the 'natural' ligand C4b (Blom *et al.* 2001a). Therefore, whilst the involvement of residues shown by mutagenesis to be critical for C4b binding in CCPs 1 and 2 can be investigated in relation to the 3D structure of this region, to fully interpret the structural basis of C4b-binding the structure of modules 1-3 is required. This work is in progress but is beyond the scope of this thesis.

1.8 Aims of this thesis

The aims of this thesis are as follows:

1. Use NMR spectroscopy to solve the 3D structure of C4BP12.
2. Interpret the mutagenesis data gathered to date in light of this 3D structure.
3. Investigate the binding of the M4 HVR from *S. pyogenes* to C4BP12 using NMR spectroscopy and other biophysical techniques.
4. Map the binding site for DNA onto the C4BP12 structure.

Chapter 2

RESONANCE ASSIGNMENT OF C4BP12

2.1 Introduction

The first aim of this project was to solve the 3D solution structure of C4BP12. In order to achieve this, as many as possible of the ^{13}C, ^{15}N and ^{1}H resonances of C4BP12 must first be assigned to specific nuclei. This procedure, known as resonance assignment, is now routine for proteins of less than ~25 kDa that are uniformly isotopically labelled with ^{13}C and ^{15}N. After assignment has been completed, the conversion of NOE cross-peaks into inter-proton distance restraints (Chapter 3) becomes possible. Resonance assignment requires the recording of a series of 3D NMR spectra (as detailed in Section 2.3) followed by assignment of firstly the backbone resonances (Section 2.4.3), then the side-chain resonances (Sections 2.4.4 and 2.4.5) and finally assignment of the NOESY spectra (Section 2.4.7).

An in-depth discussion of the theory of nuclear magnetic resonance spectroscopy is beyond the scope of this thesis and for more information the reader is referred to (Cavanagh *et al.* 1996).

2.2 Sample preparation

The C4BP12 samples were provided by Prof. Anna Blom (Lund University, Sweden). The details of the expression and purification are described in full in (Jenkins *et al.* 2006). Briefly, human C4BP CCP1–2 cDNA was amplified by PCR yielding the protein sequence: MNCGPPPTLSFAAPMDITLTETRFKTGTTLKYTCLPGYVRSHSTQTLTCNSDGEWVY NTFCIYKRCRHPGELRNGQVEIKTDLSFGSQIEFSCSEGFFLIGSTTSRCEVQDRGVWS HPLPQCEILEHHHHHH. This was cloned in the Bluescript vector and transferred into the pET16 vector. The DNA was transfected into BL21 (DE3) CodonPlus-RP *E. coli*, which were cultured in Luria-Bertani broth containing kanamycin and chloramphenicol at 37 °C. After cooling to 30 °C, expression was induced with IPTG and bacteria grown for 5 h. The bacteria were then centrifuged and resuspended in cold phosphate-buffered saline, lysed, and

sonicated. The sonicate was centrifuged and the pellet resuspended in 6 M guanidine HCl, 20 mM Tris-HCl, pH 8.0, 10 mM reduced glutathione. The crude material was applied to a nickel-nitrilotriacetic acid Superflow column (~100 ml), and the protein eluted with the same guanidinium buffer plus 0.7 M imidazole. Refolding conditions [from (Heiring *et al.* 2001)] were screened using a conformation-dependent monoclonal antibody and buffer 11 (55 mM Tris-HCl, pH 8.5, 10.6 mM NaCl, 2.2 mM $CaCl_2$, 2.2 mM $MgCl_2$, 0.055% polyethyleneglycol-4000, 0.55 M arginine, 0.1 mM oxidized glutathione, 1 mM reduced glutathione) selected. Pooled fractions containing C4BP12 were diluted (20 mg/litre) in buffer 11 and incubated overnight. Iodoacetamide was added to 5 mM, and the solution incubated and dialyzed against 50 mM Tris-HCl, pH 8.5. The protein solution was then applied to an ~10-ml SourceQ column and C4BP12 eluted with 50 mM sodium phosphate buffer, pH 8.5. ^{15}N and $^{13}C,^{15}N$ labelling were achieved using M9 medium supplemented with $^{15}NH_4Cl$ and [^{13}C]-glucose.

NMR samples were prepared by adding a protease inhibitor mixture (Sigma-Aldrich) and 0.05% (v/v) NaN_3 and concentrating to 500 µl, followed by buffer-exchange through three rounds of 10-fold dilution in NMR buffer (20 mM deuterated NaOAc, pH 4.5, 0.05% NaN_3) and concentration. The final sample volume was 600 µl with protein concentrations of 1.3 mM (^{15}N sample) and 2.5 mM ($^{13}C,^{15}N$ sample), containing 10% (v/v) D_2O. The SDS-PAGE of the final $^{13}C,^{15}N$-labelled NMR sample of C4BP12 is shown in Figure 2.1.

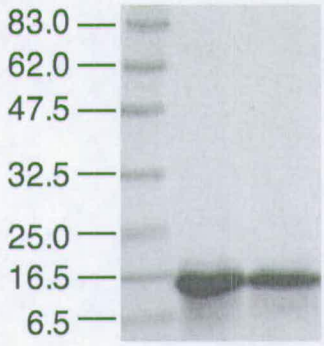

Figure 2.1 SDS-PAGE of C4BP12. Lanes as follows (left to right): Molecular weight markers labelled in kDa; purified C4BP12 as received from Lund University; C4BP12, in NMR buffer with protease inhibitor cocktail and NaN_3 added, after 10 days incubation at 37° C - i.e. to demonstrate the sample is stable over this time scale.

2.3 Data collection

NMR spectra were acquired at 37 °C on Bruker AVANCE 600 and 800 MHz spectrometers. The majority of the experiments were acquired at 600 MHz (^{1}H frequency) with the exception of the ^{13}C- and ^{15}N- edited NOESY spectra, the ^{13}C- and ^{15}N-HSQC spectra and the HBCBCGCDHD and HBCBCGCDCEHE spectra which were acquired at 800 MHz (^{1}H frequency). All spectra were acquired by the author with the assistance of Dr Dušan Uhrín and Mr Juraj Bella (University of Edinburgh). The spectra recorded are summarised in Table 2.1. All the spectra were processed using AZARA (W. Boucher, Department of Biochemistry, University of Cambridge, UK).

The 1D ^{1}H spectrum of u[^{13}C,^{15}N]-labelled C4BP12 is shown in Figure 2.2. There is no decoupling of carbon in this pulse sequence so the peaks appear as doublets due to the ^{13}C-^{1}H coupling. The wide chemical shift dispersion in the amide region and the clear upfield shifted methyl resonances are indicative of a well-folded protein. The sharp lines at 2 and 3.5 ppm arise from compounds in the protease inhibitor cocktail used in the sample preparation.

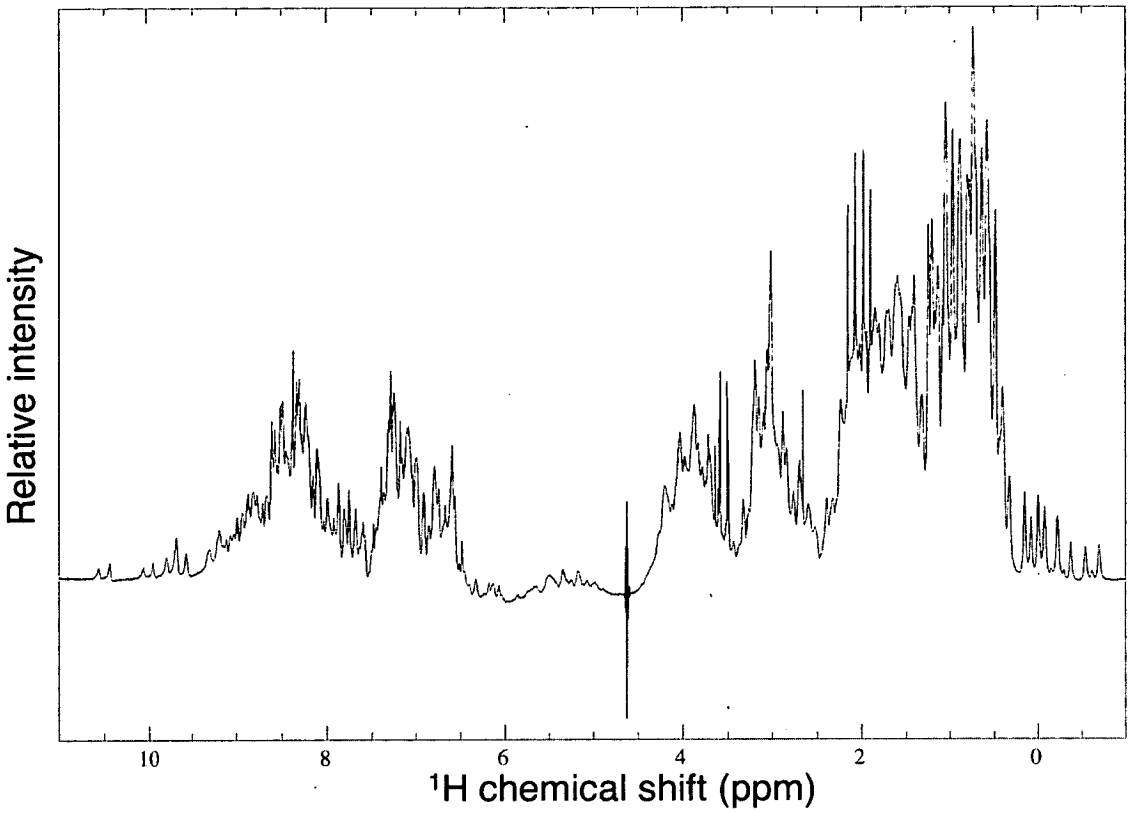

Figure 2.2 1D ^{1}H NMR spectrum of 2.5 mM u[^{13}C,^{15}N]-labelled C4BP12 (20 mM NaOAc pH 4.5, 37 °C). The Shaka method of water suppression was used (Hwang *et al.* 1995).

2.4 Resonance assignment

2.4.1 General strategy

The assignment of all the resonances in a protein is a stepwise procedure. Firstly the backbone shifts (C^{α},H^{α}, N^{H}, H^{N} and CO) together with the C^{β}, H^{β} shifts of sequential residues are identified. Then experiments that utilise TOCSY transfer along the sidechain atoms are used to identify the ^{13}C and ^{1}H shifts of all assignable resonances in the residue. This information is used to assign the cross-peaks in the HCCH-TOCSY spectrum which contains signals for all the covalently linked atoms in the sidechain of each residue in the protein. These assignments can then be transferred to the ^{13}C- and ^{15}N-edited NOESY spectra and thus cross-peaks arising from NOE transfer between non-covalently linked protons that are close in space can be identified and assigned.

The backbone assignment makes use of triple-resonance spectra, so called as the pulse sequence involves magnetisation transfer between all three of the NMR active nuclei (^{1}H,^{15}N,^{13}C) in the isotopically labelled protein. These experiments make use of the large J-couplings between nuclei in the protein backbone (Figure 2.3) which enable efficient transfer of magnetisation along the protein backbone. These 3D experiments finish with acquisition on amide H^{N}, thus the ^{15}N and ^{1}H projections of these spectra overlay on the ^{15}N-HSQC (Figure 2.8) which provides a convenient frame of reference during the assignment procedure.

The names of the triple-resonance experiments reflects the order of the nuclei that the magnetisation is transferred through during the pulse sequence. Nuclei that are not frequency-labelled in the experiment and hence whose chemical shifts cannot be determined from the spectrum acquired, are included in parentheses in the name of the pulse sequence, for example ^{13}CO in the CBCA(CO)NH experiment illustrated in Figure 2.5Figure 2.6.

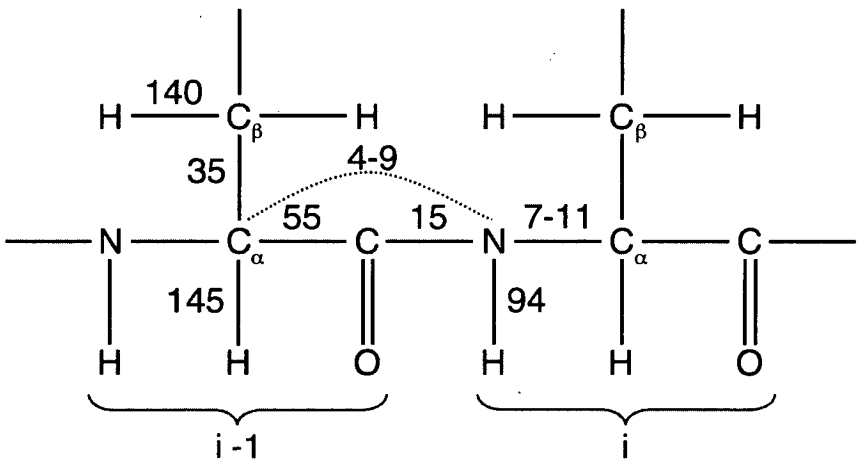

Figure 2.3 J-coupling constants in protein NMR (figures in Hz).

Experiment	Dimension 1			Dimension 2			Dimension 3			Sample	Mixing time (ms)
	Nucleus	No. of complex points	Sweep width (Hz)	Nucleus	No. of complex points	Sweep width (Hz)	Nucleus	No. of complex points	Sweep width (Hz)		
^{15}N-HSQC[1]	^{1}H	1024	11160.7	^{15}N	256	1800.5				u[^{13}C,^{15}N]-C4BP12	
^{13}C-HSQC[2]	^{1}H	1024	11160.7	^{13}C	256	7648.2				u[^{13}C,^{15}N]-C4BP12	
CG(CB)H[3]	^{1}H	1024	7183.9	$^{13}C_{arom}$	160	6035.0				u[^{13}C,^{15}N]-C4BP12	
CG(CD)H[3]	^{1}H	1024	7183.9	$^{13}C_{arom}$	120	6035.0				u[^{13}C,^{15}N]-C4BP12	
CG(C^{arom})H-TOCSY[3]	^{1}H	1024	7183.9	$^{13}C_{arom}$	120	6035.0				u[^{13}C,^{15}N]-C4BP12	7.6
(HB)CB(CGCD)HD[4]	^{1}H	2048	9615.4	$^{13}C_{arom}$	64	4025.0				u[^{13}C,^{15}N]-C4BP12	
(HB)CB(CGCDCE)HE[4]	^{1}H	1024	9615.4	$^{13}C_{arom}$	64	4025.0				u[^{13}C,^{15}N]-C4BP12	
CBCA(CO)NH[5]	^{1}H	1024	8389.3	$^{13}C_{\alpha/\beta}$	128	12070.0	^{15}N	64	1349.7	u[^{13}C,^{15}N]-C4BP12	
CBCANH[6]	^{1}H	1024	8389.3	$^{13}C_{\alpha/\beta}$	128	11312.2	^{15}N	64	1349.7	u[^{13}C,^{15}N]-C4BP12	
HBHA(CO)NH[7]	^{1}H	1024	8389.3	$^{1}H_{\alpha/\beta}$	128	6598.5	^{15}N	64	1349.4	u[^{13}C,^{15}N]-C4BP12	
HBHANH[8]	^{1}H	1024	8389.3	$^{1}H_{\alpha/\beta}$	128	5997.9	^{15}N	64	1349.4	u[^{13}C,^{15}N]-C4BP12	
HNCO[9]	^{1}H	1024	8389.3	^{13}CO	64	1659.6	^{15}N	64	1349.4	u[^{13}C,^{15}N]-C4BP12	
HN(CA)CO[10]	^{1}H	1024	8389.3	^{13}CO	64	1659.3	^{15}N	64	1349.4	u[^{13}C,^{15}N]-C4BP12	
H(C)(CO)NH-TOCSY[11]	^{1}H	1024	8389.3	$^{1}H_{sidechain}$	128	5399.6	^{15}N	64	1349.7	u[^{13}C,^{15}N]-C4BP12	20.1
(H)C(CO)NH-TOCSY[11]	^{1}H	1024	8389.3	$^{13}C_{sidechain}$	128	11312.2	^{15}N	64	1349.4	u[^{13}C,^{15}N]-C4BP12	20.1
HCCH-TOCSY[12]	^{1}H	1024	8389.3	^{1}H	128	5398.1	^{13}C	64	5731.5	u[^{13}C,^{15}N]-C4BP12	15.2
13C-edited NOESY[13]	^{1}H	2048	11160.7	^{1}H	192	9615.4	^{13}C	64	7646.7	u[^{13}C,^{15}N]-C4BP12	100
15N-edited NOESY[14]	^{1}H	2048	11160.7	^{1}H	256	9615.4	^{15}N	96	1800.5	u[^{13}C,^{15}N]-C4BP12	100
15N-edited TOCSY[14]	^{1}H	1024	8389.3	^{1}H	196	7500.5	^{15}N	64	1349.7	u[^{15}N]-C4BP12	60

Table 2.1 Acquisition parameters of the NMR spectra recorded for C4BP12.

[1](Bodenhausen *et al.* 1980), [2](Vuister *et al.* 1992) [3](Prompers *et al.* 1998), [4](Yamazaki *et al.* 1993), [5](Grzesiek *et al.* 1992a), [6](Grzesiek *et al.* 1992b), [7](Grzesiek *et al.* 1993a), [8] (Wang *et al.* 1994), [9](Grzesiek *et al.* 1992c), [10](Clubb *et al.* 1992) [11](Montelione.*et al.* 1992), [12](Kay *et al.* 1993), [13](Pascal *et al.* 1994), [14](Sklenar *et al.* 1993)

An example of the magnetisation transfer in a triple-resonance experiment is shown for the HBHANH pulse sequence (Wang *et al.* 1994) in Figure 2.4. This pulse sequence gives, for a certain H^N, cross-peaks for the H^β and H^α resonances of both the residue containing that H^N (labelled i in Figure 2.4) and also the preceding residue (labelled i-1 in Figure 2.4) in the protein backbone. By overlaying this spectrum with the HBHA(CO)NH spectrum (Grzesiek *et al.* 1993c), which has cross-peaks corresponding to the H^β and H^α resonances of only the preceding residue (labelled i-1 in Figure 2.4) at the shift of the H^N of residue i, pairs of sequential amino acids can be identified.

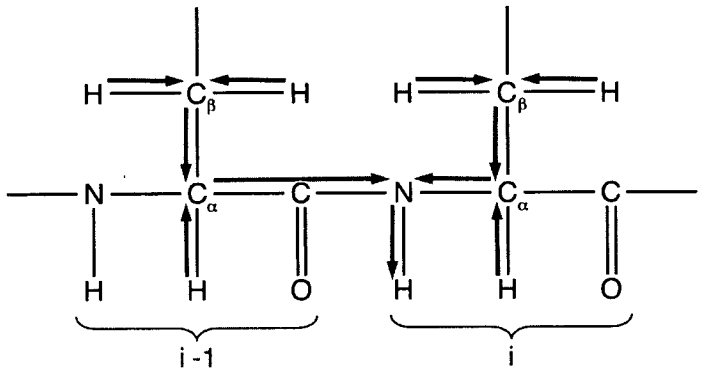

Figure 2.4 Magnetisation transfer pathway in the HBHANH experiment (Wang *et al.* 1994)

The magnetisation transfer pathway in the triple-resonance 3D CBCA(CO)NH experiment (Grzesiek *et al.* 1992a) is illustrated in Figure 2.5.This experiments connects the C^α and C^β shifts of residue i-1 with the N^H and H^N shifts of resonance i. The pulse sequence for this experiment is reproduced from (Grzesiek *et al.* 1992a) in Figure 2.6.The localisation of the magnetisation during the two constant-time chemical shift evolution periods t_1 and t_2 and the acquisition time t_3 is shown in Figure 2.5.

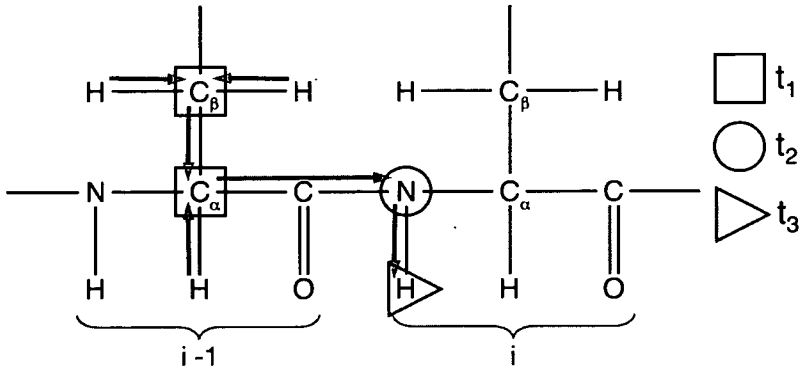

Figure 2.5 Magnetisation transfer pathway in the CBCA(CO)NH experiment.

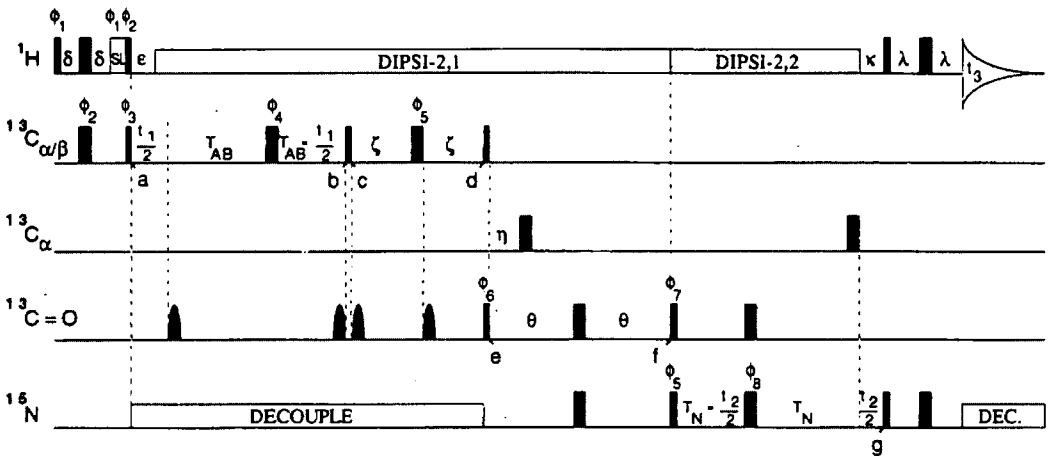

Figure 2.6 The pulse scheme of the CBCA(CO)NH experiment reproduced from (Grzesiek *et al.* 1992a). Narrow pulses correspond to 90° flip angles and wider pulses to 180°. Pulses for which no phase is indicated are applied along the *x* axis.

A brief description of the mechanism of the CBCA(CO)NH experiment is as follows. The pulse sequence starts with INEPT (insensitive nuclei enhanced by polarisation transfer, (Morris *et al.* 1979)) transfer from protons to $^{13}C^{\alpha/\beta}$. From points a to b in Figure 2.6, the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ shifts are labelled in a constant-time evolution period (t_1) of duration $1/(4J_{CC})$. During t_1 the J_{CCO} coupling is removed by the 180° shaped pulse applied $t_1/2$ after point a. The application of this ^{13}CO pulse changes the resonance frequency of the $^{13}C^{\alpha}$ due to the Bloch-Siegert effect (Bloch *et al.* 1940). The phase error due to this Bloch-Siegert effect is

removed by the identical 180° shaped ^{13}CO pulse applied at the end of the constant time evolution period. At time b the 90° pulse transfers magnetisation from $^{13}C^{\beta}$ to $^{13}C^{\alpha}$. Between c and d INEPT transfer of magnetisation between $^{13}C^{\alpha}$ and ^{13}CO occurs. The Bloch-Siegert phase error caused by the fourth 180° shaped ^{13}CO pulse that is part of this INEPT transfer step is compensated for by the third shaped ^{13}CO pulse. Another INEPT transfer step between e and f transfers magnetisation to ^{15}N. During this transfer the $J_{C\alpha CO}$ coupling is removed by the 180° $^{13}C^{\alpha}$ pulse (with an identical 180° $^{13}C^{\alpha}$ pulse to compensate for Bloch-Siegert phase error applied during t_2). The $^{15}N^{H}$ chemical shift is labelled during the second constant time evolution period (t_2) during which the J_{NCO} coupling is destroyed by the 180° ^{13}CO pulse. Finally, at point g, ^{15}N magnetisation is transferred to its directly attached $^{1}H^{N}$ by a reverse INEPT transfer before detection during the acquisition time t_3.

2.4.2 Visualisation of 3D spectra

The programme ANSIG (Kraulis 1989) was used to display the 3D experiments as a series of stacked 2D planes (Figure 2.7). A scrollbar corresponding to the 3rd dimension enabled the selection of the plane corresponding to the desired chemical shift in this dimension.

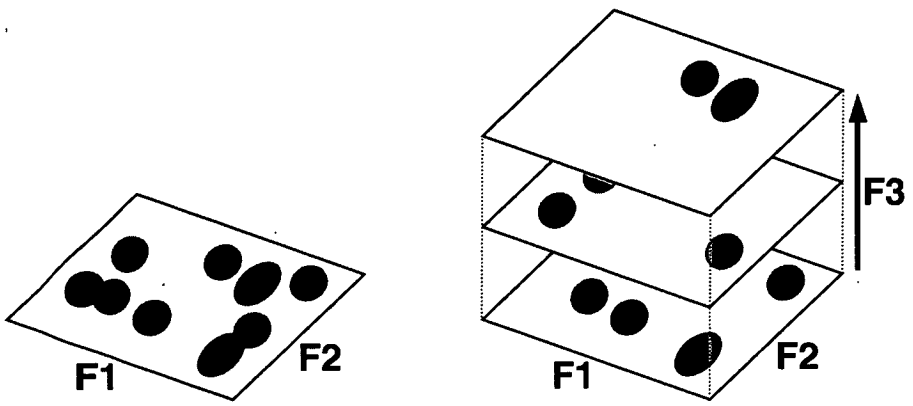

Figure 2.7 Representation of a 3D spectrum as a stack of 2D planes. Overlap in the 2D spectrum on the left is resolved by the addition of a third dimension on the right, separating the peaks by a third chemical shift (F3).

2.4.3 Backbone assignment

The C^{α}, C^{β}, CO, N^{H} and H^{N} resonances of C4BP12 were assigned using the 2D ^{15}N-HSQC (Figure 2.8) and the 3D CBCA(CO)NH, CBCANH, HNCO and HN(CA)CO experiments (Bodenhausen *et al.* 1980; Clubb *et al.* 1992; Grzesiek *et al.* 1992b; Grzesiek *et al.* 1992c; Grzesiek *et al.* 1992a). For a certain H^{N} peak (residue i) in the ^{15}N-HSQC the CBCA(CO)NH and CBCANH spectra corresponding to this H^{N} were overlaid in order to identify the C^{α} and C^{β} from this residue and from the preceding (i-1) residue. A macro called `close_in_C` was used to find cross-peaks in the CBCANH spectrum that matched cross-peaks in the CBCA(CO)NH spectrum. By matching peaks in this way a series of connected strips (each corresponding to a single H^{N} cross-peak in the ^{15}N-HSQC) was produced. The strips were located in the primary sequence of C4BP12 through identification of residues with characteristic chemical shift patterns - primarily glycine, serine and threonine. An example series of strips is shown in Figure 2.9. A similar procedure using the 3D HBHA(CO)NH and HBHANH (Grzesiek *et al.* 1993b; Wang *et al.* 1994) spectra and the macro `close_in_H_F1` was used to resolve ambiguity in possible matches in the ^{13}C spectra, and to assign the H^{α}, and H^{β} resonances. The strips for the same residues as Figure 2.9 are shown in Figure 2.10. For cases in which both the pairs of ^{13}C and ^{1}H triple-resonance spectra were unable to provide a single match, the 3D HNCO and HN(CA)CO spectra (Clubb *et al.* 1992; Grzesiek *et al.* 1992c) were used to resolve this ambiguity. These spectra were also used to assign the carbonyl resonances of C4BP12. Example strips from these ^{13}CO spectra are shown in Figure 2.11.

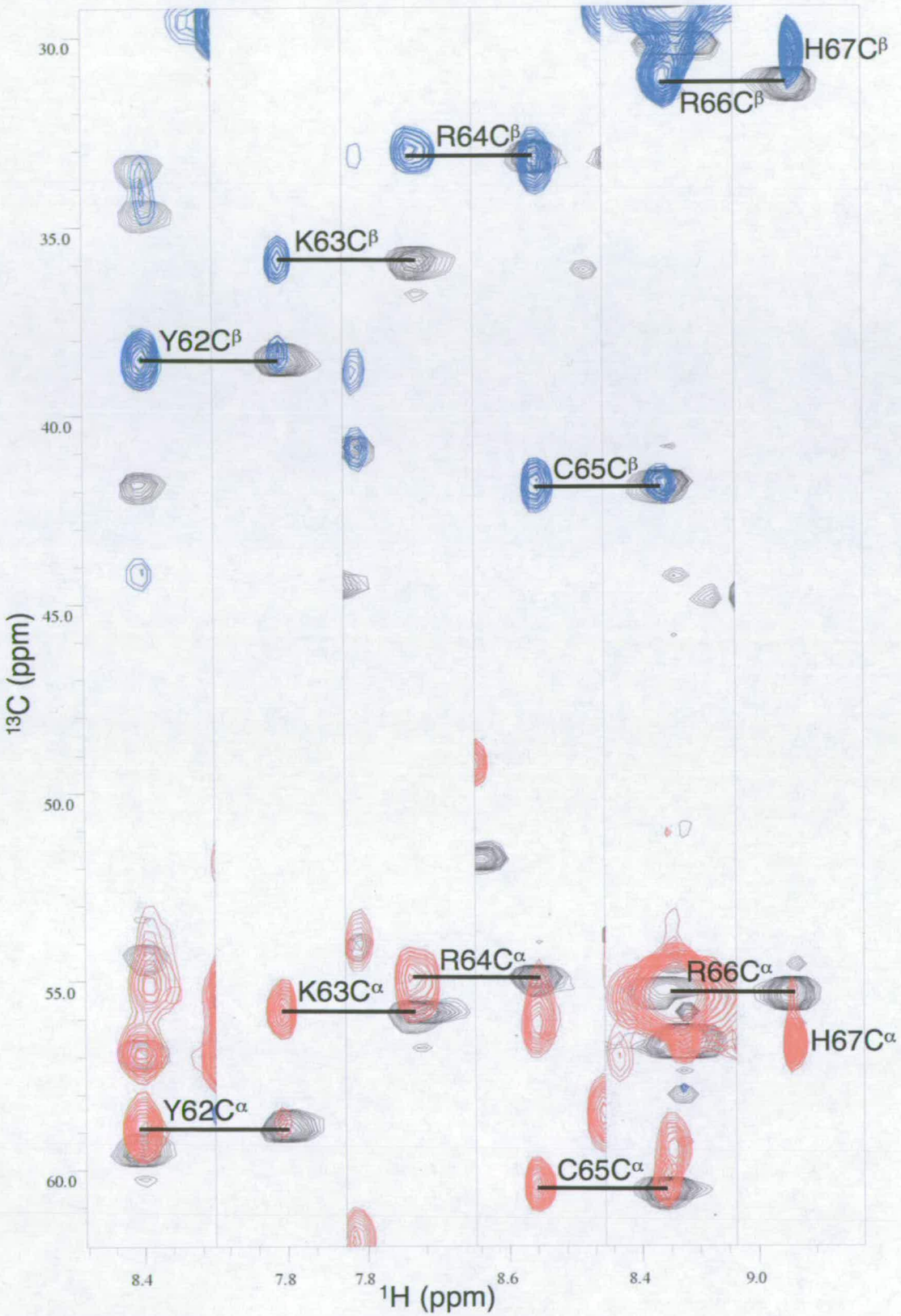

Figure 2.9 Series of strips from the CBCA(CO)NH (black) and CBCANH (blue and red) spectra showing sequential assignment of residues from Tyr[62] to His[67]

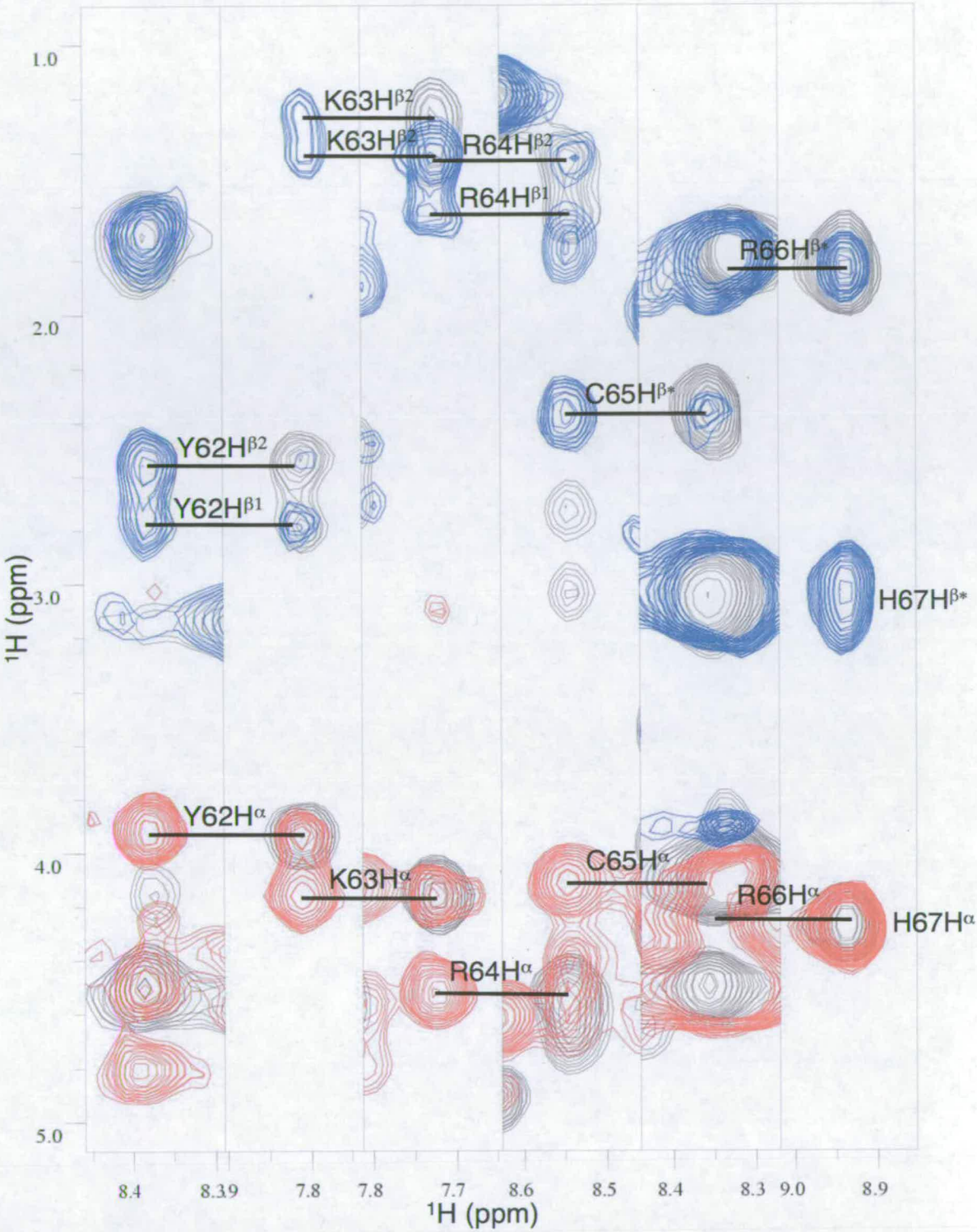

Figure 2.10 Series of strips from the HBHA(CO)NH (black) and HBHANH (red and blue) spectra showing sequential assignment of residues from Tyr[62] to His[67]

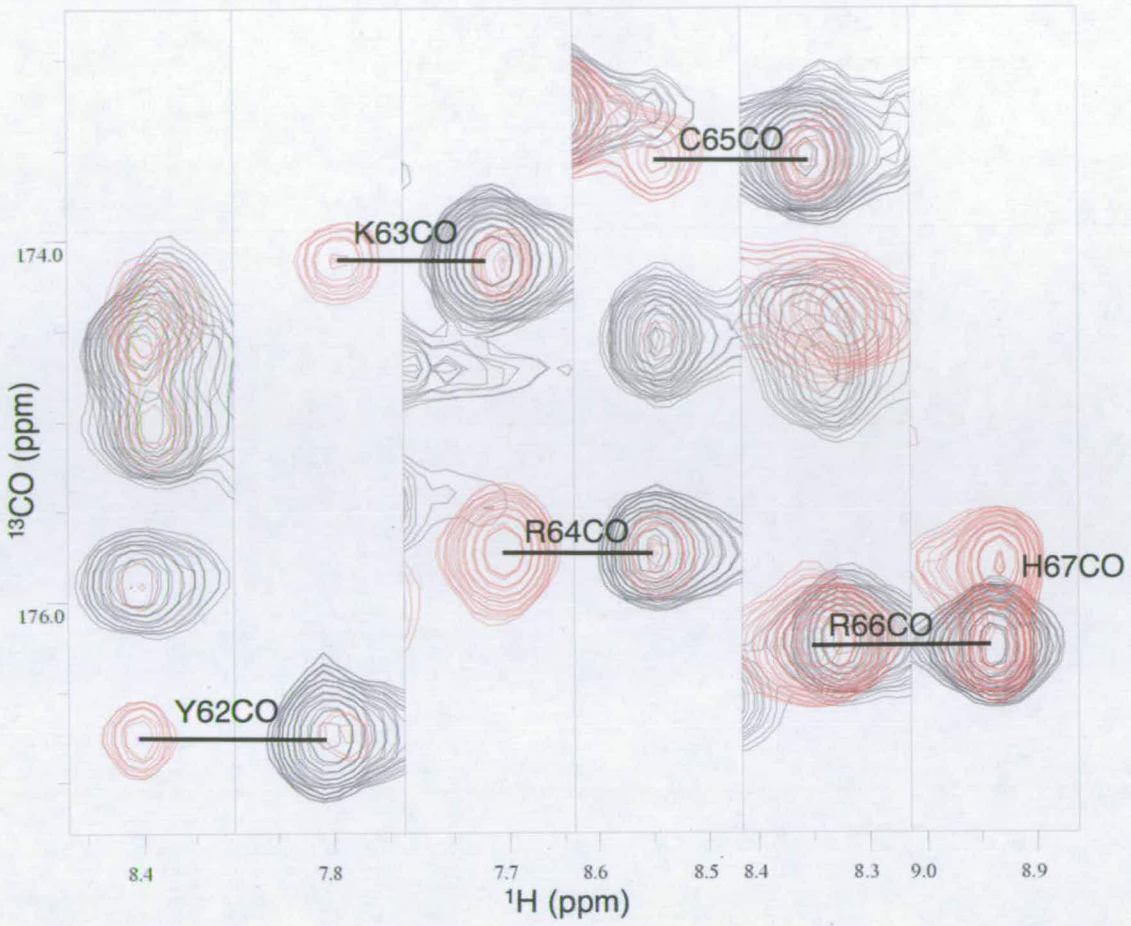

Figure 2.11 Series of strips from the HNCO (black) and HN(CA)CO (red) spectra showing sequential assignment of residues from Tyr^{62} to His^{67}

2.4.4 Aliphatic side-chain assignment

The ^{13}C and ^{1}H shifts of resonances from aliphatic side-chains were assigned using the 3D H(C)(CO)NH-TOCSY and (H)C(CO)NH-TOCSY spectra (Montelione *et al.* 1992) which give the complete ^{1}H and ^{13}C sidechain shifts, respectively, for the residue preceding a certain H^N in the ^{15}N-HSQC. These assignments were transferred to the ^{13}C-HSQC and this acted as a reference point for the assignment of the HCCH-TOCSY spectrum (Kay *et al.* 1993) which enabled the resolution of ambiguities such as arise from methyl resonances in isoleucine and leucine residues.

2.4.5 Aromatic side-chain assignment

The HBCBCGCDHD and HBCBCGCDCEHE spectra (Yamazaki *et al.* 1993) that aid aromatic side-chain assignment were dominated by strong un-phaseable peaks arising from the H^{δ}s in the flexible histidine residues in the His_6-tag at the C-terminus of C4BP12. Unfortunately this meant that it was impossible to use these spectra for assignment. Although the ^{13}C-edited NOESY spectrum can be used to assign the aromatic side-chain resonances this does rely on the assumption that the most intense cross-peaks arise from intra-residue NOEs between protons in the aromatic ring. It is, therefore, beneficial to have spectra which make use of magnetisation transfer via through-bond couplings between atoms in the aromatic side-chain.

In an alternative series of experiments described by Prompers *et al.* the $^{13}C^{\gamma}$ resonances are first correlated with the H^{β}s using the CG(CB)H pulse sequence and then correlated with the aromatic protons in 4 other experiments: with H^{δ} [CG(CD)H], with H^{ε} [CG(CDCE)H], with H^{ζ} [CG(CDCECZ)H] or with all aromatic proton resonances [CG(C^{aro})H-TOCSY] (Prompers *et al.* 1998). Juraj Bella (University of Edinburgh) implemented the CG(CB)H, CG(CD)H and the CG(C^{aro})H-TOCSY pulse sequences and these spectra were recorded by the author with his assistance (using two different mixing times - 7.6 and 15.2 ms - for the CG(C^{aro})H-TOCSY) on C4BP12. The CG(C^{aro})H-TOCSY with a mixing time of 7.6 ms produced a better spectrum and so this was used for assignment.

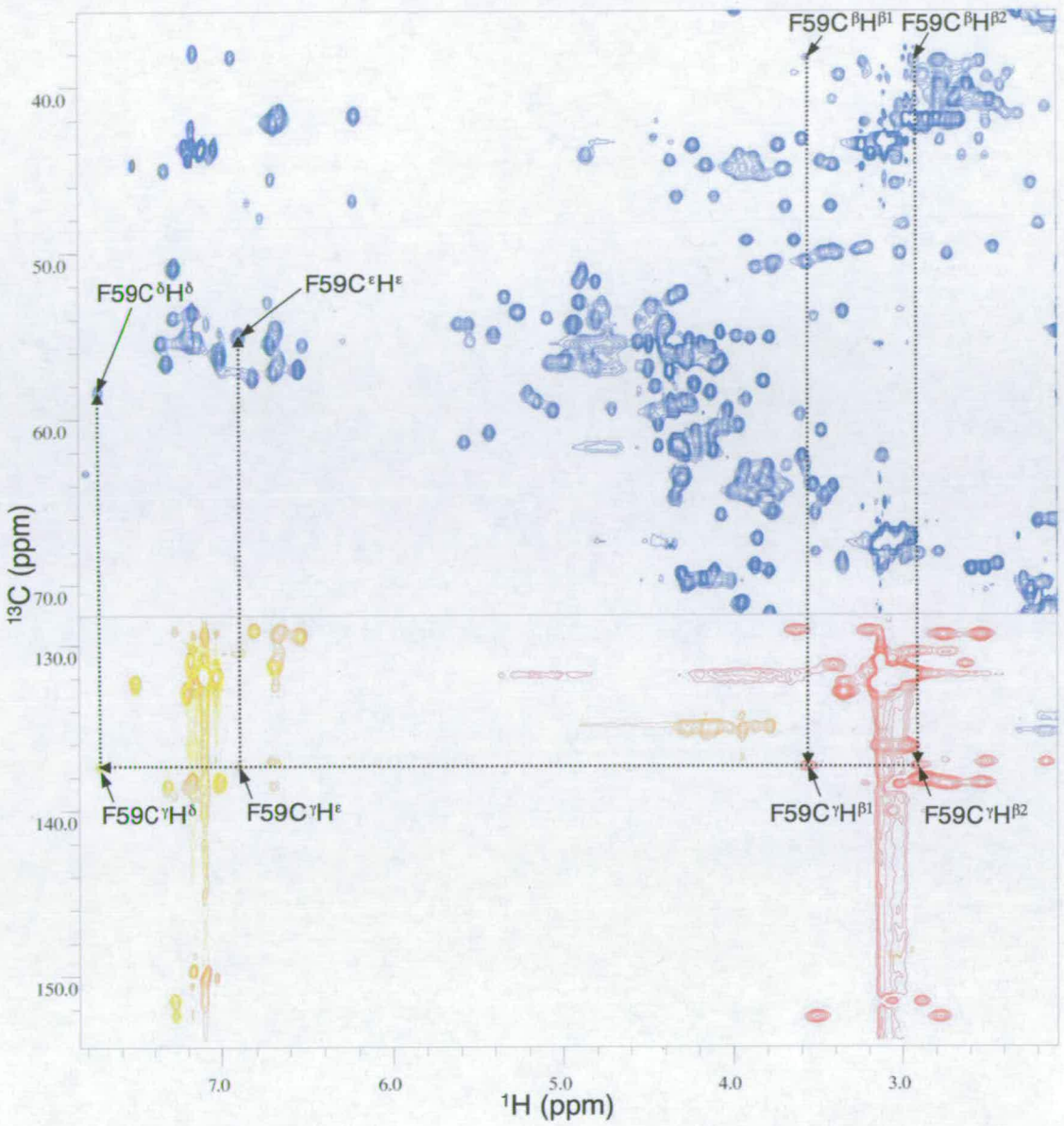

Figure 2.12 Overlay of ^{13}C-HSQC (blue), CG(CB)H (red), CG(CD)H (yellow) and CG(C^{aro})H-TOCSY spectra (orange). The assignment pathway for Phe^{59} is illustrated with arrows and labelled peaks.

These C^{γ} correlation spectra still contain intense un-phaseable peaks from the His_6-tag, however, the wider dispersion of C^{γ} shifts compared to that of C^{β} means that these signals do not overlap with the cross-peaks from the other aromatic side-chains in these spectra, unlike in the HBCBCGCDHD and HBCBCGCDCEHE spectra. Figure 2.12 illustrates the assignment procedure for aromatic side-chain protons using the CG(CB)H, CG(CD)H and

CG(C^{aro})H-TOCSY spectra together with the ^{13}C-HSQC. First a pair of $C^{\beta}H^{\beta}$ cross-peaks from an aromatic residue are located in the ^{13}C-HSQC. By matching the H^{β} shifts the pair of cross-peaks corresponding to the $C^{\gamma}H^{\beta}$ correlation can be found in the CG(CB)H spectrum - bottom right of Figure 2.12. This enables the assignment of the C^{γ} of this residue. The C^{γ} shift can then be matched in the CG(CD)H spectrum to find the cross-peak corresponding to the $C^{\gamma}H^{\delta}$ correlation and the CG(C^{aro})H-TOCSY spectrum to also find the cross-peak corresponding to the $C^{\gamma}H^{\varepsilon}$ correlation (bottom left of Figure 2.12). This gives the shifts for the H^{δ} and H^{ε} of the aromatic side-chain and these assignments can be transferred to the aromatic region of the ^{13}C-HSQC (top left of Figure 2.12).

Using these spectra it was possible to assign all the aromatic H^{δ}s and H^{ε}s and also the tryptophan $H^{\delta 1}$s. The remaining assignments - phenylalanine H^{ζ}s and the rest of the side-chain protons in the two tryptophan residues in C4BP12 were made using the ^{13}C-NOESY spectrum. The aromatic side-chain protons of the native histidines could not be assigned as this region of the CG(CB)H, CG(CD)H and CG(C^{aro})H-TOCSY spectra is dominated by strong signals from the His_6-tag. The side-chain assignments of His^{41}, His^{67} (except for $H^{\delta 2}$) and His^{117} are, therefore, missing.

2.4.6 Proline isomers

The configuration of the eight proline residues in C4BP12 was determined by calculating the difference in chemical shift δC^{β} - δC^{γ} for each residue and by inspecting the proline C^{α} and C^{δ} strips in the ^{13}C-edited NOESY spectrum. The difference in the chemical shifts, δC^{β} - δC^{γ} for Pro^4, Pro^5, Pro^6, Pro^{13}, Pro^{35}, Pro^{68} and Pro^{118} (Pro^{120} C^{γ} could not be assigned) is shown in Table 2.2. Based on a statistical analysis of the ^{13}C chemical shifts of 1,033 prolines from 304 proteins deposited in the BioMagRes database (Schubert *et al.* 2002) it is known that

δC^{β} - δC^{γ} is 4.51 ± 1.17 ppm for *trans* and 9.64 ± 1.27 for *cis* proline residues. The shift differences in Table 2.2 indicate that these seven proline residues are in the *trans* conformation.

Residue	δC^{β} - δC^{γ}
4	2.64
5	3.61
6	5.37
13	4.88
35	3.63
68	5.23
118	6.16

Table 2.2 Chemical shift difference δC^{β} - δC^{γ} for seven of the eight proline residues in C4BP12

The pattern of sequential NOE cross-peaks observed in the ^{13}C-edited NOESY spectrum is also indicative of the conformation of proline residues. As shown in Figure 2.13, strong H^{δ} H^{α} NOEs and weak H^{α} H^{α} NOEs between the proline and the previous residue indicate a *trans* conformation whilst weak H^{δ} H^{α} NOEs and strong H^{α} H^{α} NOEs indicate a *cis* conformation. The NOE patterns observed confirmed that the seven prolines listed in Table 2.2 were in the *trans* conformer and that Pro^{120} was also in the *trans* conformer. This information was used in building the template structure used in the calculation of the structure of C4BP12 as described in Chapter 3.

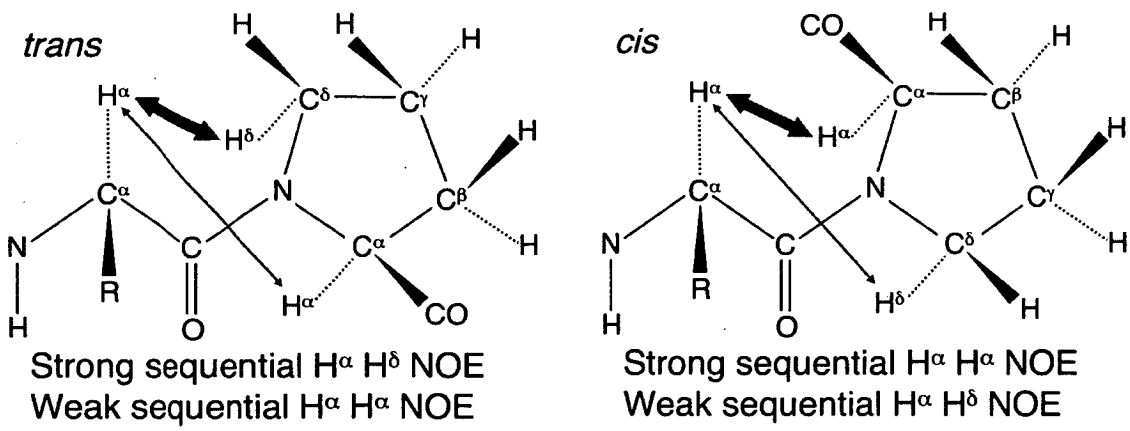

Figure 2.13 Proline isomers. The thickness of the arrows indicates the intensity of the sequential NOEs observed.

2.4.7 NOE assignment

Overlaying the ^{1}H-^{1}H planes of the HCCH-TOCSY and ^{13}C-NOESY allowed identification and assignment of *intra*residue NOEs. *Inter*residue NOEs were assigned by inspecting pairs of symmetry-related cross-peaks and with reference to the assigned nuclei. When no unique match to a specific nucleus existed the cross-peak was ambiguously assigned. A total of 3266 NOEs were assigned in the ^{13}C-edited NOESY spectrum together with an additional 1170 NOEs in the ^{15}N-edited NOESY spectrum. There were a total of 895 ambiguous NOEs (included in the above figures) - 581 and 314 in the ^{13}C- and ^{15}N-edited NOESY spectra respectively.

2.4.8 Extent of assignment

With the exception of Phe^{59} the backbone H^N and N^H resonances for all of the native residues (Asn^{1}-Ile^{124}), excluding the eight proline residues, could be assigned. The C^{α}, H^{α}, C^{β}, H^{β} and CO resonances for all native residues as well as the cloning residues Met^{0}, Leu^{125} and Glu^{126} were completely assigned. For the native residues Asn^{1}-Ile^{124}, 96.4% of the total resonances were assigned. The aromatic side-chain resonances of the native histidines were not assigned due to overlap with strong signals from the His_6 tag as explained in section 2.4.5. Cross-peaks for the majority of the side-chain resonances of Gln^{44} were also missing probably due to this residue being in a very mobile loop in CCP1 - as judged from relaxation studies on C4BP12 (Section 4.5). Resonance assignments for C4BP12 are listed in Appendix A. The assignments have been deposited in the BioMagResBank under accession number 6712.

2.4.9 Chemical shift index

Chemical shift indexing (CSI) (Wishart *et al.* 1992; Wishart *et al.* 1994) is a fast and relatively accurate method for predicting secondary structure based on the deviation of H^{α}, C^{α}, C^{β}, and CO shifts of each amino acid from those observed in "random coil" regions of proteins with known 3D structure. For H^{α}, C^{α} and CO shifts, a shift greater than the range given is assigned the CSI value of 1, a shift smaller than the range is assigned -1, and a shift within the range is assigned 0. A series of four or more values of -1, not interrupted by a 1, is taken to indicate an α-helix and a series of three or more values of 1, not interrupted by a -1, indicates a β-strand. For C^{β} shifts the deviation from "random coil" values are in the opposite direction and only β-strands can be identified (Wishart *et al.* 1994). A series of four or more values of -1, not interrupted by a 1, is taken to indicate a β-strand. The CSI program (Wishart *et al.* 1994) was used to calculate the CSI for H^{α}, C^{α}, C^{β}, and CO shifts of C4BP12, and also to output the consensus chemical shift index from the results for the four nuclei used in the calculation. The results are shown in Figure 2.14. As expected for a protein known to contain two CCP modules (Section 1.4), the consensus CSI predicted only β-strands - five in CCP1, four in CCP2 and one in the linker.

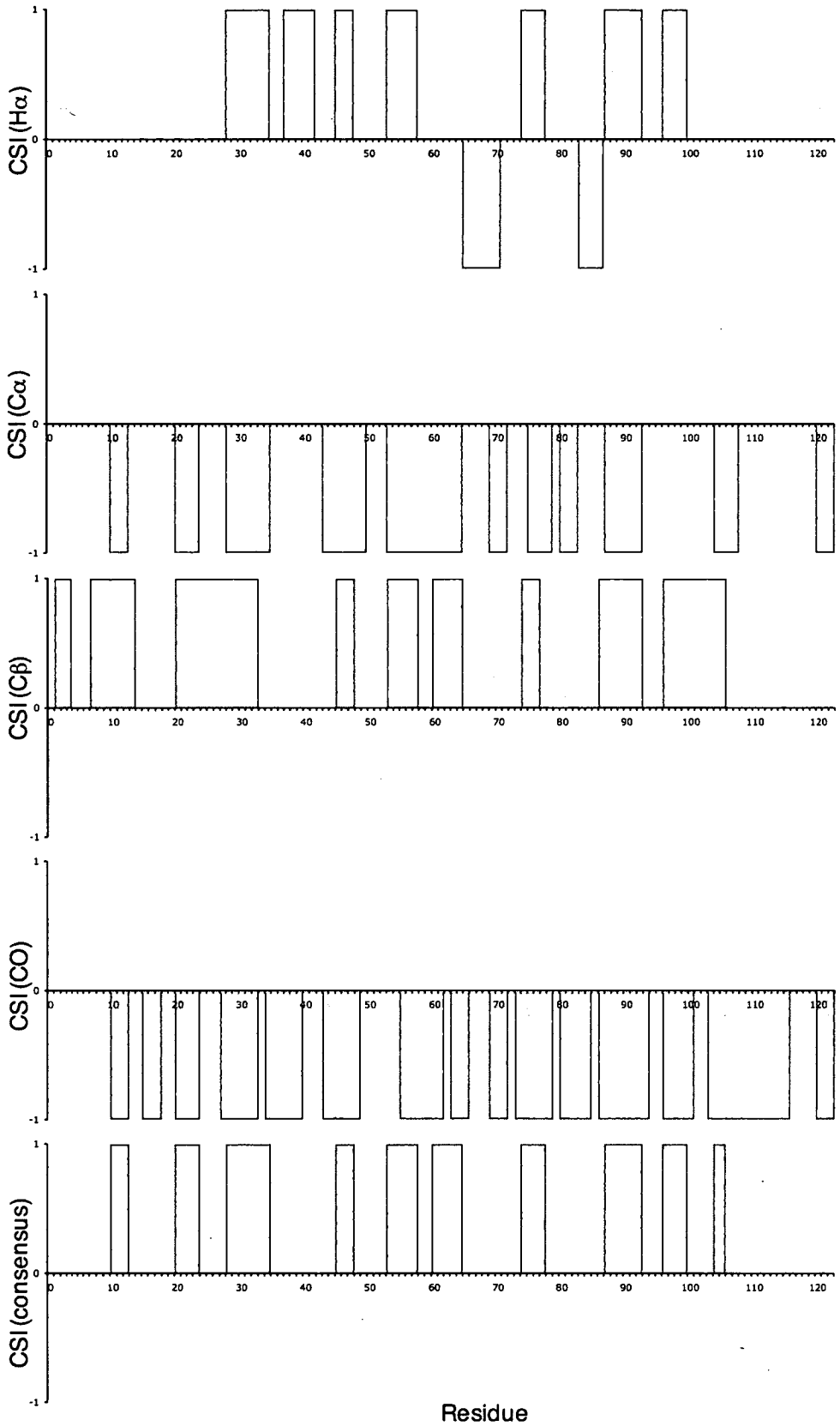

Figure 2.14 Chemical shift index for H^{α}, C^{α}, C^{β}, and CO and the consensus. Values calculated using the CSI programme (Wishart *et al.* 1994).

2.5 Conclusions

Almost complete resonance assignment of C4BP12 was achieved. These assignments were transferred to the NOESY spectra from which interproton distances were derived and used as input in structure calculations to solve the 3D structure of C4BP12 using the method of restrained molecular dynamics as described in Chapter 3.

Chapter 3

CALCULATION OF THE STRUCTURE OF C4BP12

3.1 Introduction

The process of converting a set of inter-proton distance restraints calculated from the intensities of cross-peaks in a NOESY spectrum into a 3D structure of the protein of interest is non-trivial. This is predominantly because the limited amount of distance restraints obtained from the observed NOEs (much fewer than the theoretical maximum of two symmetry-related cross-peaks for each pair of protons less than ~5Å apart in the structure) means that in order to solve the structure, considerable non-experimentally determined information must be used in the structure calculation. Therefore, weighting of this information relative to the experimentally-derived information is of critical importance (Habeck *et al.* 2006).

The strategy used to calculate the structure of C4BP12 is based on that described by Prof. Michael Nilges (Nilges 1995). In addition to NOE-derived distance restraints, dihedral angle restraints based on the chemical shift index (Wishart *et al.* 1994) and residual dipolar couplings (RDCs) measured on a sample of C4BP12 aligned in filamentous phage were incorporated. As experiments to directly measure hydrogen bonds (Cordier *et al.* 1999) have never been successfully applied to CCP module-pairs, and H^N deuterium-exchange data requires inference of the carbonyl oxygen acceptor, no hydrogen bond restraints were included in the structure calculation. Restrained molecular dynamics from random starting structures was used with a simple repulsive non-bonded potential. Simulated annealing was used during the calculation procedure to enable the calculation to escape local minima and to improve the sampling of conformational space.

3.2 NOE based distance restraints

Despite the increasing use of RDC restraints in structure calculation and the demonstration that 3D structures can be calculated solely using RDC based restraints (Hus *et al.* 2001), distance restraints obtained from the volume of NOESY cross-peaks remain the predominant source of experimental data used in the calculation of structures from NMR data. Although many cross-peaks are present in NOESY spectra, the distance restraints obtained by calculating the volume of many of these cross-peaks are redundant. For example, each pair of symmetry related cross-peaks describes only a single distance between a pair of protons. Moreover, there are numerous *intra*residue NOEs that give distance restraints for covalently bonded protons (for example the distance between a pair of H^{β}s) that are already defined by the covalent geometry for each amino acid present in the topology files used as input for the structure calculation. Thus the number of 'long-range' NOEs, i.e. from protons separated by more than four residues in the primary sequence, is often only 20% of the total number of NOEs. Furthermore, while the integration of a NOE cross-peak to obtain the volume is accurate, the conversion of this to a distance restraint is problematic. This is because spin diffusion and local dynamics can alter the volume of the peak so that it no longer reflects the true distance between the protons involved. This is discussed further in Sections 3.2.2 and 3.2.3 (and in Chapter 7).

3.2.1 The Nuclear Overhauser effect

In a two dimensional NOESY spectrum obtained using the pulse sequence in Figure 3.1, cross-peaks arise due to magnetisation transfer between the two spins arising from cross-relaxation, commonly referred to as the nuclear Overhauser effect (NOE). The NOESY pulse sequence starts with a 90° pulse that creates transverse magnetisation and the spins precess during t_1 at their characteristic Larmor frequencies. When the second 90° pulse is applied along the $+x$ axis, the magnetisation vectors are rotated so a component exists along the $-z$

axis. These *z*-magnetisation components can exchange due to cross relaxation during the mixing period t_m. The transverse components are removed by phase cycling. The final 90° *x* pulse regenerates transverse magnetisation which is detected during the acquisition time t_2. By varying the length of t_1 the frequency of the first spin can be labelled, and after two-dimensional Fourier transformation the incrementation of t_1 gives rise to the second dimension in the 2D NOESY. The position of the cross-peak is determined by the chemical shift of the second proton (acquired in t_2) in F2 - the directly detected dimension and that of the first spin in F1 - the indirectly detected dimension.

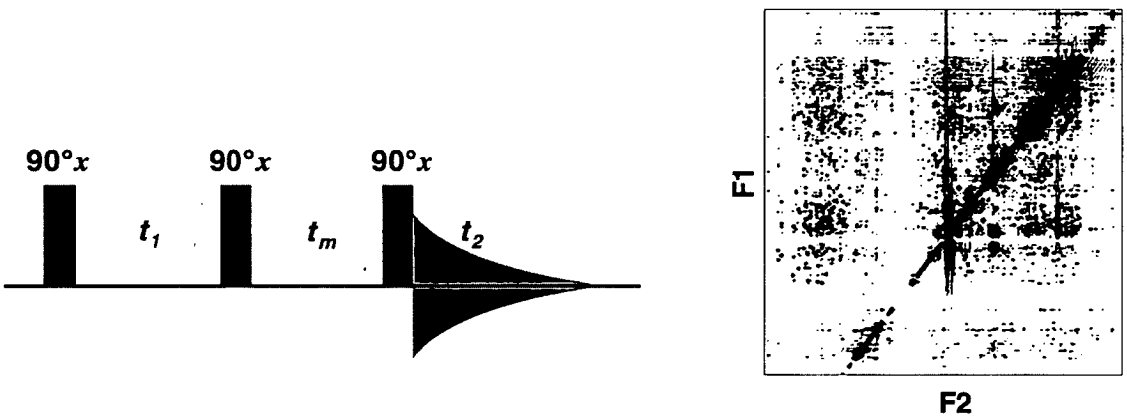

Figure 3.1 The 2D NOESY pulse sequence (Jeener *et al.* 1979) and the resulting 2D NOESY spectrum of C4BP12

In a two-spin system the NOE is proportional to the inverse sixth power of the distance between the two nuclei involved (Solomon 1955). For a cross-peak in the NOESY spectrum between spins I and S, (with frequencies of ω_I and ω_S, respectively) which are part of a multi-spin system, the initial rate of build up of the cross-peak (with intensity a_{IZ}) is given by (Macura *et al.* 1980)

$$d(a_{IZ})/d\tau_m\big|_{\tau_m=0} = -\sigma_{IS}M_0 \qquad (3.1)$$

Where τ_m is the mixing time and M_0 is the intensity of the diagonal peak at $\tau_m = 0$. This initial rate approximation gives the important result that only the cross-relaxation rate between spins I and S, σ_{IS}, controls the size of the NOE cross-peak. For the protons I and S σ_{IS} is given by

$$\sigma_{IS} = \left(\frac{\mu_0}{4\pi}\right)^2 \frac{\hbar^2\gamma^4}{10}\left[\frac{6\tau_c}{1+(\omega_I+\omega_S)^2\tau_C^2} - \frac{\tau_C}{1+(\omega_I-\omega_S)^2\tau_C^2}\right] r_{IS}^{-6} \qquad (3.2)$$

where $\hbar$ is $h/2\pi$ (i.e Plank's constant divided by 2π), μ_0 is the permeability constant in a vacuum (these constants arise from the quantum mechanics used to calculate the transition probabilities from the dipolar interaction Hamilitonian) and γ is the proton gyromagnetic ratio. τ_c is the correlation time (the time taken by the molecule to rotate by roughly one radian about any axis). Therefore, r_{IS} - the internuclear distance between the two protons - is the only distance in the multi-spin network that controls the size of the observed NOE. The r^{-6} dependence arises because dipolar interactions have a r^{-3} dependence and cross-relaxation is a second order effect, hence $(r^{-3})^2$.

This is the basis of the isolated spin pair approximation (ISPA) which enables, for NOESY spectra with short mixing times, the reduction of the magnetisation transfer in a multi-spin system (e.g. a protein) to the transfer between pairs of spins such that:

$$D_{ij}^{-6} \propto NOE_{ij} \qquad (3.3)$$

where NOE_{ij} is the intensity of the observed NOE cross-peak and D_{ij} is the internuclear distance between the protons involved. Experimentally the NOE can deviate significantly from this approximation as a result of spin diffusion and internal dynamics.

3.2.2 Spin diffusion

In a protein, the dense network of protons means that magnetisation can be transferred indirectly between a pair of protons as illustrated in Figure 3.2. This magnetisation transfer via indirect pathways usually leads to increased NOE intensities. Thus distances derived from NOEs using the ISPA are often underestimated (Linge *et al.* 2004). It is possible to include spin diffusion corrections in the structure calculation. This can be achieved by direct refinement against the NOE intensities, although this is computationally expensive as the NOESY spectra must be simulated from the coordinates at every minimization step. Another method uses complete relaxation matrix analysis to derive distances corrected for spin diffusion that can be used in conventional structure calculation protocols (Linge *et al.* 2004). In practice, however, the least complicated and, therefore, most common method of accounting for spin diffusion is the use of a minimal mixing time and appropriately wide bounds on the distance restraints used in the calculation (this is discussed further in Chapter 7). This simple method was the approach used in the calculation of the structure of C4BP12.

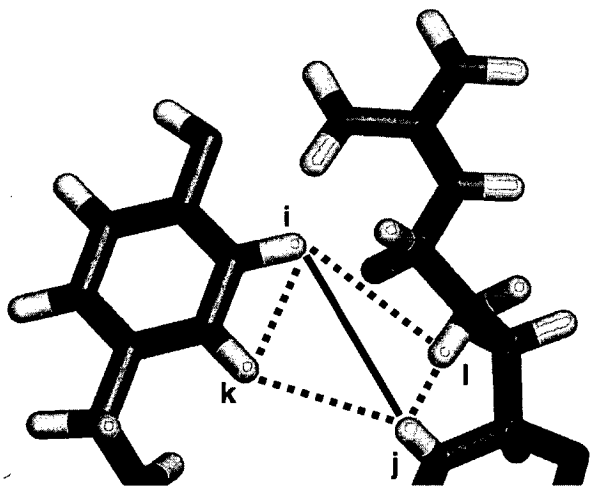

Figure 3.2 Spin diffusion in a protein. Indirect magnetisation transfer via spins k and l (dashed lines) influences the size of the NOE between i and j (solid line).

3.2.3 Local dynamics

Accounting for the effect of dynamics on the intensity of the NOE is much more difficult than correcting for spin diffusion as, depending on the type of fluctuation, the NOE intensity can be either increased or decreased. The common approach to this problem is again to use large error bounds for the distances and, subsequently, to estimate the dynamics from the spread of structures in the final ensemble. This is far from ideal as the diversity of the ensemble reflects solely the distribution of the experimental data and while dynamics does influence this distribution, other factors such as the presence of experimental artefacts in the spectra and overlap will also contribute. Furthermore the diversity depends on the protocol used to calculate the distance bounds from the NOE data. It is also important to bear in mind that the measured NOE is a time- and ensemble-averaged value whereas the bounds of a distance restraint refer to instantaneous distances in a single structure. There is currently much work on improving the way NOE-derived distance restraints are used in structure calculation such as using a potential for these restraints in the structure calculation that reflects the error in calculating these distances rather than using wide upper and lower bounds to account for the error (Nilges *et al.* 2006), this is discussed in Chapter 7. For reasons of simplicity the only correction for spin diffusion and local dynamics used in the calculation of the structure of C4BP12 was the use of large error bounds on the distances derived from the volume of NOE cross-peaks.

3.2.4 Ambiguous distance restraints

In the above section an underlying assumption was made that each experimental distance restraint can be unambiguously assigned to a pair of protons. Clearly there are special cases such as methyl protons, certain pairs of methylene protons and equivalent protons on aromatic rings, which have degenerate shifts. In the calculation of the structure of C4BP12

these cases were treated in exactly the same way as the ambiguous NOE restraints described below. Moreover, in proteins, several unconnected protons may have the same chemical shift. This leads to NOE cross-peaks that can not be immediately converted to a distance restraint between two atoms and these are termed ambiguous NOEs. They cannot be assigned even after the complete chemical shift assignment of the protein has been completed. Traditionally, ambiguous NOEs were assigned manually in an iterative procedure based on inspection of preliminary structures calculated using the distance restraints derived from unambiguously assigned NOEs, followed by calculation of a new set of structures calculated with these additional restraints. These new structures were used to derive further discrimination between possible matches for an ambiguous NOE and so on, until no further assignments could be made. This method relied heavily on the quality of the preliminary structures and there is a danger that errors made in the initial assignment may be further propagated throughout the calculation, leading to an incorrect final structure.

In an alternative method proposed by Nilges (1995), the distance restraints derived from the NOESY spectra are incorporated into a target function that allows all possibilities to be specified for each restraint. Such restraints are termed ambiguous distance restraints (ADRs). The target function is minimised using simulated annealing (Section 3.5.1). This method proposes that an ambiguous NOE cross-peak at coordinates $F1,F2$ contains contributions from all proton pairs with those chemical shift assignments:

$$NOE_{F1F2} = \sum_{i\in\{F1,\Delta1\},\, j\in\{F2,\Delta2\}} NOE_{ij} \qquad (3.4)$$

where $\{F1,\Delta1\}$ is defined as the set of all protons with chemical shifts within a tolerance Δ from shift F. Under the ISPA, contributions to the volume of this cross-peak depend on the distances D_{ij} for each spin pair i,j. The rate of build up of this NOE is given by:

$$\frac{d}{d\tau_m} NOE_{ij} = cD_{ij}^{-6} \tag{3.5}$$

where τ_m is the mixing time and c is a constant. If the assumption is made that the order parameters and correlation times for all the spin pairs are identical, combining equations 3.4 and 3.5 results in

$$\frac{d}{d\tau_m} NOE_{ij} = c\left(\sum_{i\in\{F1,\Delta 1\}, j\in\{F2,\Delta 2\}} NOE_{ij} \right) \tag{3.6}$$

As the ambiguous NOE depends on the sum of the inverse sixth powers of the individual distances a "r^{-6}-summed" distance $\overline{D}$ can be defined:

$$\overline{D}_{F1,F2} = \left(\sum_{i\in\{F1,\Delta 1\}, j\in\{F2,\Delta 2\}} D_{ij}^{-6} \right)^{-1/6} \tag{3.7}$$

enabling equation 3.6 to be written as:

$$\frac{d}{d\tau_m} NOE_{F1,F2} = c\overline{D}_{F1,F2}^{-6} \tag{3.8}$$

This is similar to the $\langle r^{-6}\rangle^{-1/6}$ average distance (Brünger *et al.* 1986) originally defined for methyl and methylene groups. However, since Equation 3.7 is a sum, a scale factor $n^{-1/6}$ is included where n is the number of spin pairs in the sum. As this approach can be applied to highly ambiguous data sets (i.e n is large), this factor is not negligible. Another difference is

that the value of the $\langle r^{-6}\rangle^{-1/6}$ average distance is always between that of the smallest and largest distances, whereas $\overline{D}$ is always shorter than any of the contributing distances.

3.2.5 Filtering of ADRs

The number of possible contributions to an ADR can be incrementally reduced as the structure calculation progresses. This process, which increases the precision of the calculated structures, is termed 'filtering'. In the filtering procedure, possible assignments are compared with the corresponding distances in the ensemble of low-energy structures from the previous round of structure calculation. Using an approach developed for the programme ARIA (Ambiguous Restraints for Iterative Assignment) (Nilges *et al.* 1997), the relative size of the contribution of the N different assignment possibilities to the volume of a cross-peak at frequencies $F1$ and $F2$ can be estimated as follows. For each contribution k to the ADR the minimum distance $D^k_{\min}$ in the ensemble of converged structures is determined. The contribution C^k of assignment k to the cross-peak is then estimated as

$$C^k = \frac{{D^k_{\min}}^{-6}}{\sum_i^{N(F1,F2)} {D^{ij}_{\min}}^{-6}} \tag{3.9}$$

The C^k are then sorted by size and the N_p largest contributions chosen such that:

$$\sum_i^{N_p} C^i > p \tag{3.10}$$

where p is a parameter set by the user. The parameter p is often expressed as a percentage - e.g - "filtering at 99%" - i.e $p = 0.99$ which removes assignment possibilities contributing less than 1% of the intensity of the NOE cross-peak that gives rise to the restraint.

3.2.6 Calculation of NOE-derived distance restraints

NOE-derived distance restraints were generated by integration of the NOESY cross-peaks in ANSIG. The chemical shift ranges over which the cross-peak intensity was integrated were set to the halfwidth at half height of the peaks (Figure 3.3) and were as follows: $^{13}C = 0.3$ ppm, $^{15}N = 0.5$ ppm, ^{1}H (indirect dimension) = 0.1 ppm, ^{1}H (acquisition dimension) 0.05 ppm. The intensities were scaled relative to the average intensity which was set to a value 1.0. After scaling, the restraints were divided into four classes or 'bins'. The classes used were: strong: intensity ≥ 3.0; medium: $1.5 \leq$ intensity < 3.0; weak: $0.3 \leq$ intensity < 1.5 and very weak: $0 \leq$ intensity < 0.3. The distance bounds chosen for these classes are shown in Table 3.1.

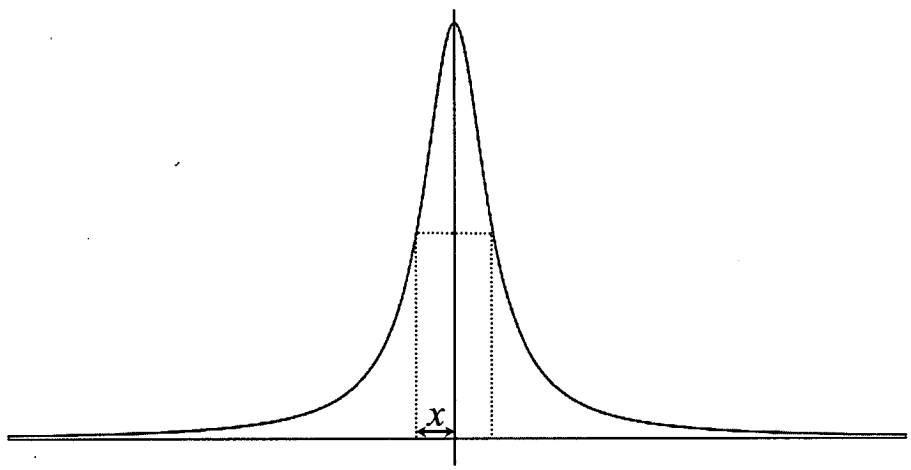

Figure 3.3 The halfwidth at half-height of a peak is distance x.

Distance Class	Lower Bound (Å)	Upper Bound (Å)
Strong (intensity ≥ 3.0)	0	2.7
Medium (1.5 ≤ intensity < 3.0)	0	3.3
Weak (0.3 ≤ intensity < 1.5)	0	5.0
Very Weak (0 ≤ intensity < 0.3)	0	6.0

Table 3.1 Distance restraint classes used in the structure calculation of C4BP12

The lower limit of the distance classes was set to 0 Å to account for the effects of spin diffusion and local dynamics on the size of the observed NOEs.

3.3 CSI-derived distance restraints

As described in the previous chapter, secondary distance restraints can be used to predict secondary structure in proteins via the programme CSI (Wishart *et al.* 1994). An in house script (`CSItoDIHE.awk`) was used to generate dihedral angle restraints based on the consensus chemical shift index. The weak restraints (angles are allowed to vary by ± 30°) used are shown in Figure 3.4. No special treatment was applied to residues predicted to be at the beginning or end of a strand. The predicted β-strand in the linker was omitted from the input data as restraining dihedral angles in the linker could exert a large effect on the intermodular orientation.

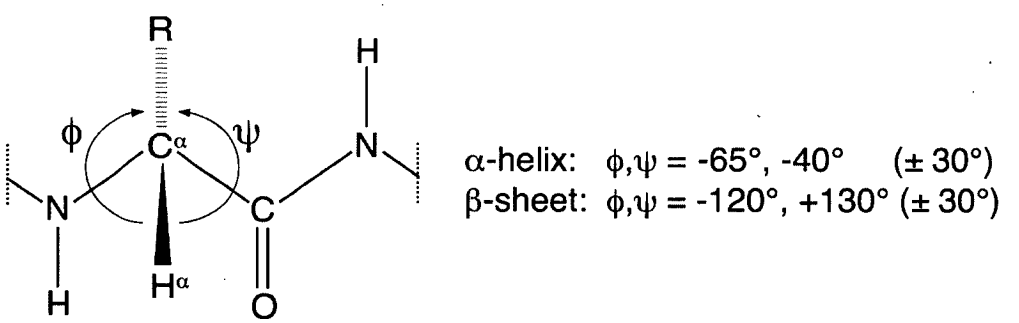

Figure 3.4 Dihedral angle restraints used in the structure calculation of C4BP12

3.4 Residual dipolar coupling restraints

For experiments intended to measure residual dipolar couplings (RDCs) the sample was aligned in *Pseudomonas* filamentous (Pf1) phage with a final ^2H splitting of 0.9 Hz. The sample conditions were as follows: 0.70 mM C4BP12, 7.3 mg/ml Pf1 phage, 20 mM NaOAc, pH 5.5, 90 mM NaCl. $^1D_{NH}$ (Ottiger *et al.* 1998), $^1D_{C^\alpha C'}$ (Permi *et al.* 1999) and $^1D_{C^\alpha H^\alpha}$ (Ball *et al.* 2006) couplings were measured for both aligned and non-aligned samples. The aligned sample was prepared by Dr Graeme Ball who also recorded the experiments and measured the values of the RDCs from the spectra. RDCs for residues 1, 15-23, 43, 79, 110-112 and 125+ were excluded on the basis that heteronuclear NOE data indicate significant local motion occurs at these residues (Section 4.5) Further details of the use of RDC restraints in the calculation of the structure of C4BP12 can be found in (Ball 2005).

3.5 Structure calculation from NMR data

There are two optimisation methods in general use for calculating 3D structures from NMR data. The first is minimisation in torsion-angle space with residue-by-residue build up of distance and non-bonded information (Braun 1987; Guntert *et al.* 1997). The second is based on global optimisation of a hybrid energy function by molecular dynamics-based simulated annealing (MDSA) in Cartesian space with a simplified soft non-bonding potential. The calculation strategy for C4BP12 was based on that described in (Nilges 1995) and is detailed below.

3.5.1 Molecular dynamics based simulated annealing

The basis of simulated annealing (Kirkpatrick *et al.* 1983) is the raising of the temperature of the system followed by slow cooling steps in order to allow local minima to be overcome on the path towards the global minimum of the target function. Originally the Metropolis algorithm (Metropolis *et al.* 1953) was used for simulation of the system, but in MDSA integration of Newton's equations of motion is used in a similar manner to that used in

molecular dynamics (Brooks *et al.* 1983). The target function F_{tot} for which the global minimum is searched contains both experimental and empirically determined terms:

$$F_{tot} = F_{covalent} + F_{repel} + F_{experimental} \tag{3.11}$$

where $F_{covalent}$ maintains the correct bond lengths, angles, planes and chirality, F_{repel} is a simple repulsive term to prevent unduly close contacts and $F_{experimental}$ contains the experimentally derived information. For example, the target function used in the final round of structure calculation of C4BP12 was as in equation 3.12

$$E_{tot} = E_{bonds} + E_{angles} + E_{improper} + E_{vdw} + E_{NOE} + E_{dihe} + E_{rdc} \tag{3.12}$$

where $E_{covalent}$ consists of E_{bonds}, E_{angles} and $E_{improper}$, E_{repel} is E_{vdw} and $E_{experimental}$ consists of E_{NOE}, E_{dihe} and E_{rdc}. The terms in equation 3.12 are defined in the following section. E_{tot} represents the total effective potential energy in the calculation, and as the force on an atom can be expressed as the gradient of the potential energy this can be calculated by integrating Newton's laws of motion for all n atoms in the system:

$$\frac{\partial^2 x_i}{\partial t^2} = -\frac{1}{m_i}\frac{\partial}{\partial x_i} E_{tot}(x_1, x_2, \ldots, x_n) \tag{3.13}$$

where x_i is the coordinate of an atom (of mass m_i) in the system. The temperature of the system can be calculated from the sum of the kinetic energies of all n atoms in the system at time t:

$$T_t = \frac{2}{k_B(3n-6)}\left(\sum_{i=1}^{n}\frac{m_i v_i^2}{2}\right)_t \tag{3.14}$$

where k_B is the Boltzmann constant and v_i is the velocity of an atom with mass m_i.

$E_{covalent}$, the first of the two empirical energy terms, maintains the correct bond angles, lengths, planes and chirality, and is defined as:

$$E_{covalent} = \sum_{bonds} k_b (r - r_0)^2 + \sum_{angles} k_\theta (\theta - \theta_0)^2 + \sum_{impropers} k_\phi (\phi - \phi_0)^2 + \sum_{dihe} k_\omega (1 + \cos\omega)^2 \quad (3.15)$$

where k_b, k_θ, k_ϕ, k_ω, are the force constants for bonds, angles, improper angles (angle terms that are used to maintain the planarity and chirality of the peptide bond which is assumed to be planar and trans and also the planarity of aromatic rings), and peptide bond dihedral angles respectively. The dihedral angle force constants for other rotatable bonds are set to 0 as the motion of these groups is covered by the non-bonded interactions. The values of the force constants are are varied throughout the calculation.

Originally, NOE-derived distance restraints were represented by a square well potential with a force constant k_{NOE}:

$$E_{NOE} = k_{NOE} \begin{cases} (r_{ij}^l - r_{ij})^2 & \text{, if } r_{ij} < r_{ij}^l \\ 0 & \text{, if } r_{ij}^l \le r_{ij} \le r_{ij}^u \\ (r_{ij} - r_{ij}^u)^2 & \text{, if } r_{ij} > r_{ij}^u \end{cases} \quad (3.16)$$

where r_{ij}^u and r_{ij}^l are the upper and lower bounds of the distance restraint and r_{ij} the calculated values. This potential can cause distorted structures due to the large energy penalty for distances that slightly exceed the upper bounds and so for the calculation of the

structure of C4BP12 a modified harmonic flat-bottom potential with linear behaviour for large deviations termed a 'soft potential' was used:

$$E_{NOE} = k_{NOE}\begin{cases}\left(r_{ij}^{l} - r_{ij}\right)^2 & ,\text{if } r_{ij} < r_{ij}^{l} \\ 0 & ,\text{if } r_{ij}^{l} \le r_{ij} \le r_{ij}^{u} \\ \left(r_{ij} - r_{ij}^{u}\right)^2 & ,\text{if } r_{ij}^{u} < r_{ij} \le r^{u} + r_{sw} \\ \left(\alpha + \beta\left(r_{ij} - r_{ij}^{u}\right)^{-1} + \gamma\left(r_{ij} - r_{ij}^{u}\right)\right), & \text{if } r_{ij} > r_{ij}^{u} + r_{sw}\end{cases} \tag{3.17}$$

where γ is the gradient of the asymptote, r_{sw} determines the distance where the potential switches from the square well function to the 'soft' asymptote and α and β are determined such that the potential is continuous and differentiable at $r_{ij}^{u} + r_{sw}$. These potentials are illustrated in Figure 3.5.

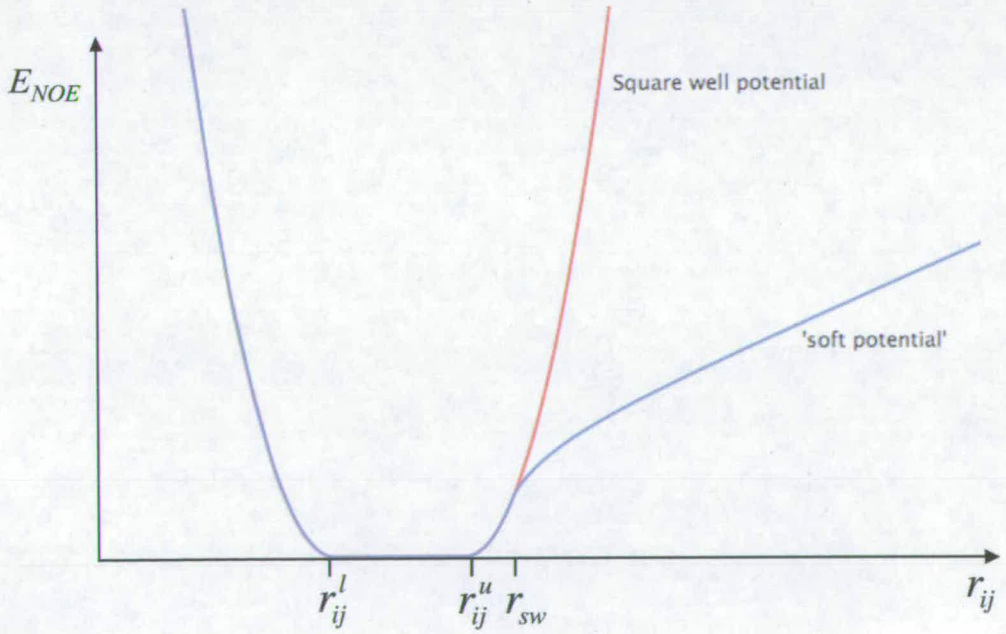

Figure 3.5 Potentials for E_{NOE} used in structure calculation by MDSA.

The covalent bond energy is defined by the equation of a simple harmonic spring:

$$E_{bond} = \sum_{bonds} k_b (r - r_0)^2 \tag{3.18}$$

where r is the actual bond length, k_b is the bond energy constant and r_0 is the equilibrium bond length as defined in the force field.

The angle bond energy is defined by the term:

$$E_{angle} = \sum_{angles} \left(k_\theta (\theta - \theta_0)^2 + k_{ub} (r_{13} - r_{ub})^2 \right) \tag{3.19}$$

where k_θ and k_{ub} are energy constants and θ_0 and r_{ub} are equilibrium values. θ is the actual value of the angle. The second term in Equation 3.19 is the Urey-Bradley term which is used only in certain forcefields and hence the default value of k_{ub} is 0.

The torsion angles between planes are defined by dihedral and improper statements with energy terms of E_{dihe} and $E_{improper}$ respectively and is of the general form:

$$E = \sum \sum_{i=1,m} \begin{cases} k_{\phi_i} \left(1 + \cos(n\phi_i + \partial_i)\right) & \text{if } n_i > 0 \\ k_{\phi_i} (\phi_i - \partial_i)^2 & \text{if } n_i = 0 \end{cases} \tag{3.20}$$

ϕ_i is the actual torsion angle, k_ϕ is the energy constant and n_i, m_i and δ_i are periodicities, multiplicities and phase shifts respectively. The definition of dihedral and improper angles is identical but two separate topology and parameter lists are maintained in CNS for these angles. As described above, improper angles are used with $n_i = 0$ to maintain planarity or

chirality, whereas dihedral angles are used with $n_i > 0$ to describe multi-minimum torsion potentials. Therefore, the actual energy terms for $E_{improper}$ and E_{dihe} are:

$$E_{improper} = \sum_{impropers} k_{\phi}(\phi - \phi_0)^2 \tag{3.21}$$

$$E_{dihe} = \sum_{dihedrals} k_{\omega}(1 + \cos\omega) \tag{3.22}$$

where k_{ϕ} and k_{ω} are the improper and dihedral energy constants and ω is the actual dihedral angle.

The non-bonded interactions are described by a simple repulsive term - E_{repel} with a force constant of k_{rep} which prevents close contacts:

$$E_{repel} = \begin{cases} 0 & , \text{if } r \geq s.r_{min} \\ k_{rep}\left(s^2 r_{min}^2 - r^2\right)^2 & , \text{if } r \leq s.r_{min} \end{cases} \tag{3.23}$$

The values of r_{min} are taken from the CHARMM empirical energy function (Brooks *et al.* 1983) and are standard values for the van der Waals radii as calculated from the Lennard-Jones potential. The factor s, which allows for scaling of the value of r_{min}, is varied throughout the calculation.

3.5.2 Calculation strategy

The calculation strategy is divided into three consecutive protocols: randomise, regularise and refine. This is because when starting from random Cartesian coordinates a modified

simulated annealing protocol is required. In this protocol the weights on the covalent geometry terms are very low at the start and are increased throughout the calculation. The first stage 'randomise' is a high temperature search phase. 'Regularise' involves enantiomer selection and then 'refine' contains two cooling stages.

Randomise

The random starting conformation is generated by assigning random values to the *x*, *y* and *z* coordinates of the atoms. Qualitatively this represents a high-temperature conformation of the system. To remove very close non-bonded contacts, 50 cycles of Powell (Powell 1977) minimisation are performed with all the force constants set to very low values. In the early stages of the 'random' phase, the residues in the protein are represented by a reduced non-bonded form (Nilges 1993; Nilges *et al.* 1997) to increase the efficiency of computation. One non-bonded sphere with a van der Waals radius of 2.25 Å is centred on the C^{α} atom, and a second sphere with the same radius is placed at the position of a single carbon atom in each side chain as defined in Table 3.2.

Amino Acid	**Non-bonded atoms**
Gly	C^{α}
Ala, Ser	C^{α}, C^{β}
Pro, Cys, Thr, Asn, Asp	C^{α}, C^{γ}
Val, Ile	C^{α}, $C^{\gamma 1}$
Met, Gln, Glu, Lys, Arg	C^{α}, C^{δ}
Leu, Phe, Tyr, His	C^{α}, $C^{\delta 1}$
Trp	C^{α}, $C^{\varepsilon 2}$

Table 3.2 Atoms used in the reduced non-bonded representation

This minimisation is followed by three steps of molecular dynamics (MD); the first step uses the reduced non-bonded representation and the subsequent two use all atoms. The first step is carried out at 2000 K and the second two at 1500 K. The weighting of the experimental

restraints is kept low (k_{NOE} = 0.5 kcal.mol^{-1}.Å^{-2}) in the first stage of MD and increased (k_{NOE} = 5.0 kcal.mol^{-1}.Å^{-2}) in the second two stages. The force constants for bond, angle, and non-bonded interactions (k_b, k_θ and k_{rep}) are increased throughout the three stages of MD. Throughout the 'random' stage all the torsion angle terms are removed from E_{total} (i.e k_θ and k_ω = 0) and these are re-introduced together with the complete non-bonded interactions in the regularisation stage.

Regularise

This step is based on the standard X-PLOR DGSA protocol, which was derived from the protocol published in (Nilges *et al.* 1988b). The input structures are those produced by the previous 'random' step. Amino acids in an extended conformation (ϕ = -120, ψ = 120) are fitted to the structure residue-by-residue and a mirror image of the structure is generated. This is because distance geometry methods cannot discriminate between enantiomers so both enantiomers are refined and the one with the lowest energy is chosen.

Two rounds of energy minimisation, the first with bond angle restraints turned off and the second with angles included, are followed by MD at 2000 K during which the weightings of the improper and dihedral angles are increased. In this round of MD the value of k_{rep} is decreased to a level which is sufficient to maintain the global structure but small enough to allow atoms to get very close to each other (and even move through each other) in order to improve the structure locally. Next the correct handedness of the structure is determined by evaluating the energies of the two enantiomers. A round of MD at 2000 K involving prochiral swapping (see section 3.5.4) is carried out with increasing weighting of the experimental terms (k_{NOE} = 5.0 → 50.0 kcal.mol^{-1}.Å^{-2}) and the non-bonded term as before. After energy minimisation the first cooling stage begins. MD is performed as the system is

cooled from 2000 to 100 K and the weighting of the non-bonded term is increased. Prochiral swapping is included in this step and the weighting of the experimental energy terms is kept high. After further prochiral swapping there is a final energy minimisation step and the structures are output for the final round.

Refine

Starting with the structures produced at the end of the previous rounds, three rounds of MD are performed. The first is at high temperature (2000 K) and the system is cooled from 2000-1000 K in the second step, and from 1000K-100 K in the third step. Prochiral swapping is included in the second and third stages. In the high-temperature stage the weighting of the experimental terms is set very high compared to the non-bonded and covalent terms. During the first cooling stage the weighting of the non-bonded and covalent terms is increased so that in the second cooling stage the weighting of the experimental terms is comparable to that of the empirically defined terms. After a final round of energy minimisation the final structures are output.

3.5.3 RDC restraints

RDC restraints are included in a modified 'refine' stage starting from structures calculated from random starting coordinates with NOE- and CSI-derived dihedral angle restraints as the only experimentally determined restraints. The structures are taken after the 'regularise' round of simulated annealing. RDC restraints were introduced by including the TENSO energy term (Sass *et al.* 2001) within CNS using a harmonic potential. The weighting is kept low (k_{TENSO} = 0.001) in the high temperature stage and increased (to k_{TENSO} = 1.0) in the first

cooling stage. Further details of the use of RDC restraints in the calculation of the structure of C4BP12 can be found in (Ball 2005).

3.5.4 Prochiral swapping

Amino acid side chains include numerous prochiral centres - carbon atoms with two identical and two different substituents, such as methylene protons and isopropyl groups. The stereospecific assignment of these groups has been shown to increase the accuracy and precision of the ensemble of calculated structures (Driscoll *et al.* 1989). However, stereospecific assignment by careful analysis of intra-residue and sequential NOE patterns is time consuming. One method for dealing with these missing assignments is to introduce pseudoatoms to replace the methylene or methyl protons. This requires widening of the bounds of the experimental distance restraints to correct for the position of the pseudoatom relative to those of the protons for which the NOE has been measured and results in a loss of information from the distance restraint. Prochiral floating or swapping (Fólmer *et al.* 1997) can be used to allow for missing stereospecific assignments, without the use of pseudoatoms. In this procedure the NOEs are measured for both individual resonances of a methylene or isopropyl group which are arbitrarily assigned (e.g $H^{\beta 2}$ and $H^{\beta 3}$ for a β-methylene group). Then, during the calculation, the stereo-related atoms or methyls are allowed to float between the *pro-R* and *pro-S* configurations. In this way the groups can move to find the most energetically favoured configuration. This requires removal of all the energy terms that define chirality at the prochiral centre and also a reduction in the weighting of the bond angle energy terms that define the bonds to the methyls or methylene protons so that these atoms or methyls can swap positions during the calculation. The prochiral swapping protocol was modified from the original X-PLOR protocol (Folmer *et al.* 1997) by Dr Brian Smith, Dr Andrew Raine and Dr Peter Domaille (University of Cambridge) to use the Metropolis

criterion thus allowing swaps that temporarily increase the energy of the system. Prochiral swapping was included at several points in the regularise and refine steps as detailed above.

3.5.5 Disulphide restraints

In initial rounds of structure calculation the two conserved disulphide bonds in each CCP module were not defined in the template structure. In the structure calculations the sulphur atoms of all the pairs of cysteine residues expected to form disulphide bonds based on homology with other CCP modules (Cys^{2} and Cys^{48}, Cys^{33} and Cys^{60}, Cys^{65} and Cys^{106}, Cys^{92} and Cys^{122}) converged to within 4 Å. The template was, therefore, modified to remove the sulphydryl protons and replace them with a bond between the pairs of sulphur atoms 2 Å long.

3.6 Calculation of the structure of C4BP12

3.6.1 Initial structures

The distance restraints were generated from the 3266 NOE cross-peaks assigned in the ^{13}C-edited NOESY spectrum, and the 1170 NOE cross-peaks assigned in the ^{15}N-edited NOESY spectrum, using the `connect` tool in AZARA. Ambiguous distance restraints were generated in the same manner using the chemical shift assignments from the ^{15}N- and ^{13}C-HSQC spectra, the ^{13}C- and ^{15}N-edited NOESY spectra and the HCCH-TOCSY spectrum. Structure calculations were performed in CNS (Brünger *et al.* 1998) using the protocols described in section 3.5.2.

3.6.2 Refinement of structures

Initial calculations of 100 structures used only distance restraints derived from unambiguously assigned NOEs. Restraints that were violated in all 100 structures (a violation is reported where a distance in a calculated structure exceeds the upper bound of the distance restraint by more than a preset threshold) were subsequently analysed in ANSIG (Kraulis 1989) by inspecting the cross-peak in the NOESY spectrum from which the restraint was generated. Where violations were judged to have arisen from mistakes in assignment, or noise peaks that were incorrectly picked, the assignment was corrected or the peak removed. Restraints were then regenerated from the cross-peaks in the NOESY spectra and the calculation repeated using the new restraints.

The second stage of the refinement strategy involved 'filtering' (Section 3.2.5) the ADRs to reduce the number of contributions to each restraint. After each round of filtering the new restraint list was used to calculate another set of structures, which was then used to filter the original restraint list again. At each filtering stage duplicate restraints generated from symmetry-related cross-peaks in the NOESY spectra were removed by selecting the restraint that more closely matched the distance in the ensemble of low energy structures. This avoids bias due to the duplication of a subset of the total restraints, and also reduces the number of restraints, which decreases the computational time required for the structure calculation. As before, restraints that were violated in structures that were within the ensemble of low energy structures were analysed in ANSIG by inspecting the NOESY cross-peak from which the restraint was generated, and assignments were corrected as required.

The final structures of C4BP12 were calculated using NOE and dihedral angle restraints (from CSI) in three rounds, as shown in Table 3.3. The structures after regularisation (Section 3.5.2) were used as input for a final refinement stage in which the RDC restraints

were incorporated as described in Section 3.5.3. In each round 100 structures were calculated.

Round	Level of filtering	Restraints used				No. of converged structures†
		UnAm	Dihe	ADR	RDC	
1	n/a	X	X			24
2	$p = 0.99$‡	X	X	X		27
3	$p = 0.99$	X	X	X		28
RDC	n/a	X	X	X	X	40

Table 3.3 Calculation strategy. Notes: Unamb = unambiguous restraints; Dihe = dihedral angle restraints derived from CSI; † these structures were used to filter the restraints for the next round of structure calculation; ‡ ADRs filtered before use in structure calculation to remove incorrect intra-modular restraints which prevented convergence of the structures.

The final restraints used in the calculation of C4BP12 are summarised in Table 3.4.

Unambiguous NOEs:	
Intraresidue	1639
Sequential	899
Short-range ($2 \leq \|i-j\| \leq 4$)	240
Long-range ($\|i-j\| > 4$)	763
Intermodular	7
Intralinker	111
From Module 1 to linker	35
From Module 2 to linker	21
Total unambiguous NOEs	3541
Total ambiguous NOEs	895
Final unique distance restraints	3032
Unique *inter*residue distance restraints	1250
Residual dipolar couplings	
$^1D_{NH}$	62
$^1D_{NC'}$	65
$^1D_{C^\alpha C'}$	41
$^1D_{C^\alpha H^\alpha}$	24
Total RDC restraints	192

Table 3.4 Final restraints used in the calculation of C4BP12

The unambiguous restraints summarised in Table 3.4 are plotted per residue in Figure 3.6. The nature of ADRs mean that it is not meaningful to plot these restraints in this way. This

figure over estimates the total number of NOEs as each restraint is included twice - once for each residue involved.

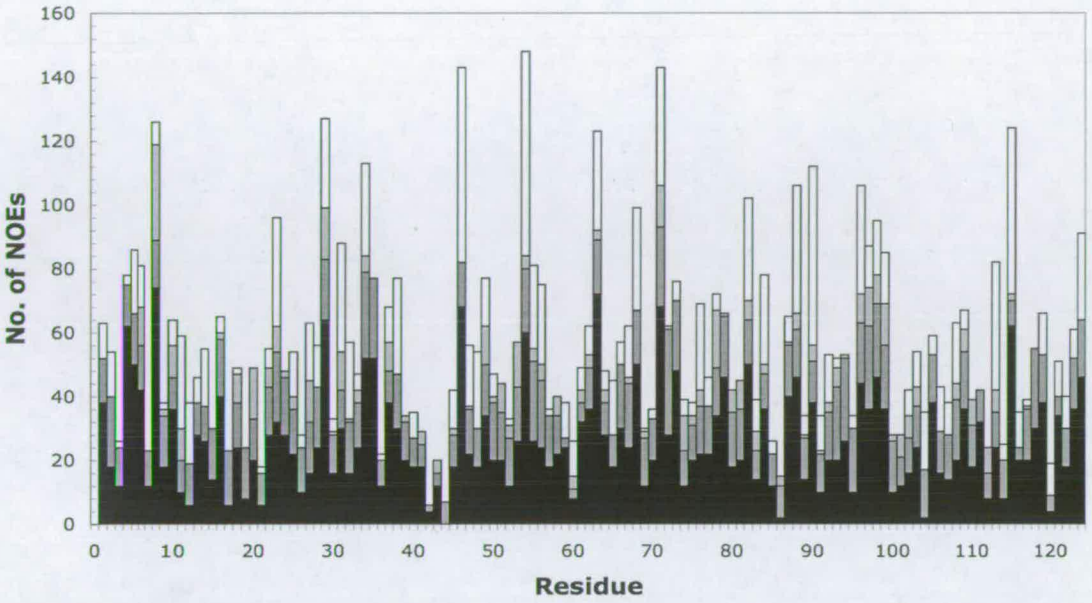

Figure 3.6 The number of unambiguous distance restraints used in the structure calculation plotted by sequence number: intra-residue in black, sequential (residue i – residue i+1) in dark grey, short-range ($2 \leq |i-j| \leq 4$) in light grey and long-range ($|i-j| > 4$) in white.

A contact plot of the final set of distance restraints, after filtering and removal of duplicate restraints, is shown in Figure 3.7. All assignment possibilities contributing to > 30% were included.

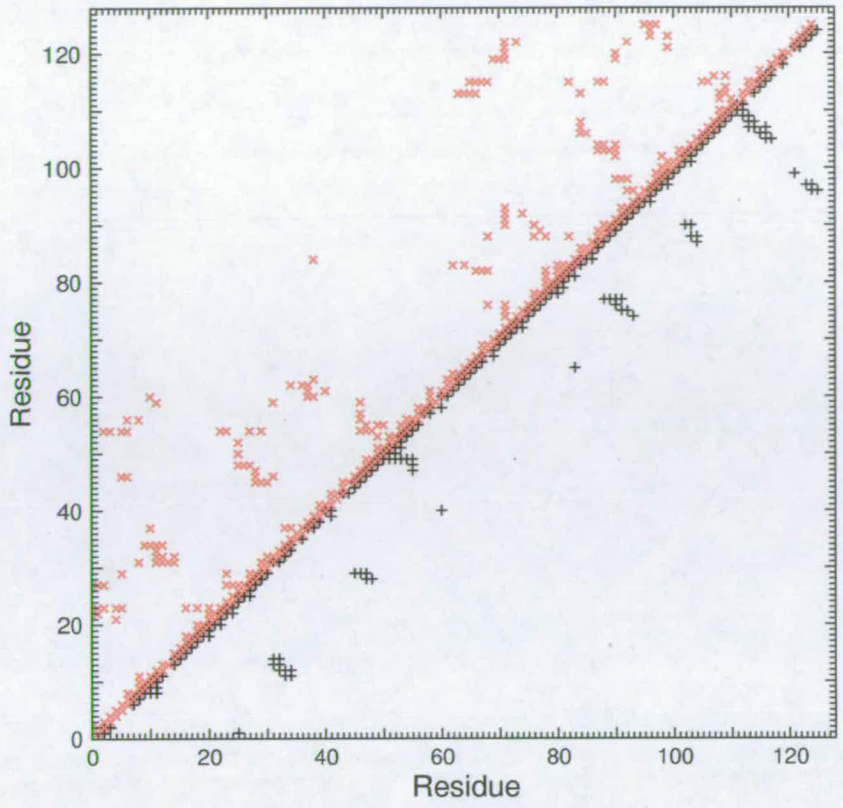

Figure 3.7 Contact map of final NOE distance restraints used for calculation of the structure of C4BP12. (x) sidechain-sidechain and sidechain-backbone restraints; (+) backbone-backbone restraints.

3.7 Conclusions

The large number of distance and RDC restraints led to excellent convergence (40 of 100) structures in the final round of structure calculation. The NOE contact plot (Figure 3.7) shows the characteristic pattern of a protein composed of β-sheets. The structure of C4BP12 is discussed further in Chapters 4 and 7.

Chapter 4

THE STRUCTURE OF C4BP12

4.1 Introduction

The methods described in Chapter 3 were employed to calculate the 3D structure of C4BP12 using NOE-derived distance restraints, RDC-derived orientation restraints and dihedral angle restraints calculated from CSI information (Table 3.4). The structures show the features expected of a pair of CCP modules. The interface between modules 1 and 2 is well defined due to interactions between the bulky side-chains of residues, both within the linker and within the individual modules close to the interface. Regions of the structure that show low convergence in the ensemble of structures are inferred to be flexible from the T_1/T_2 ratio and heteronuclear NOE measurements (Section 4.5).

4.2 Convergence of the ensemble of structures

A plot of the overall and NOE energy terms, per structure, is shown in Figure 4.1. The level of convergence appears to be excellent - based on the low energies of 40 (out of the 100 calculated) structures. The similarity in NOE energy for this ensemble of 40 structures indicates that these structures are all equally good solutions to the experimentally derived restraints. Outside this ensemble, the values of the NOE energy term diverge and the overall energy rises rapidly which is indicative of structures that do not fit the experimental data. The lowest energy 40 structures were selected as the final ensemble and these structures were submitted to the protein databank (PDB - http://www.rcsb.org/pdb) with accession code 2A55. A plot of the C^{α} rmsd (compared to the closest to mean structure) is overlaid on the plot of the number of NOE-derived distance restraints per residue in Figure 4.2. As expected, a clear correlation is observed between the number of distance restraints and the precision of the structure - i.e. regions that have few NOE-derived distance restraints show disorder in the ensemble.

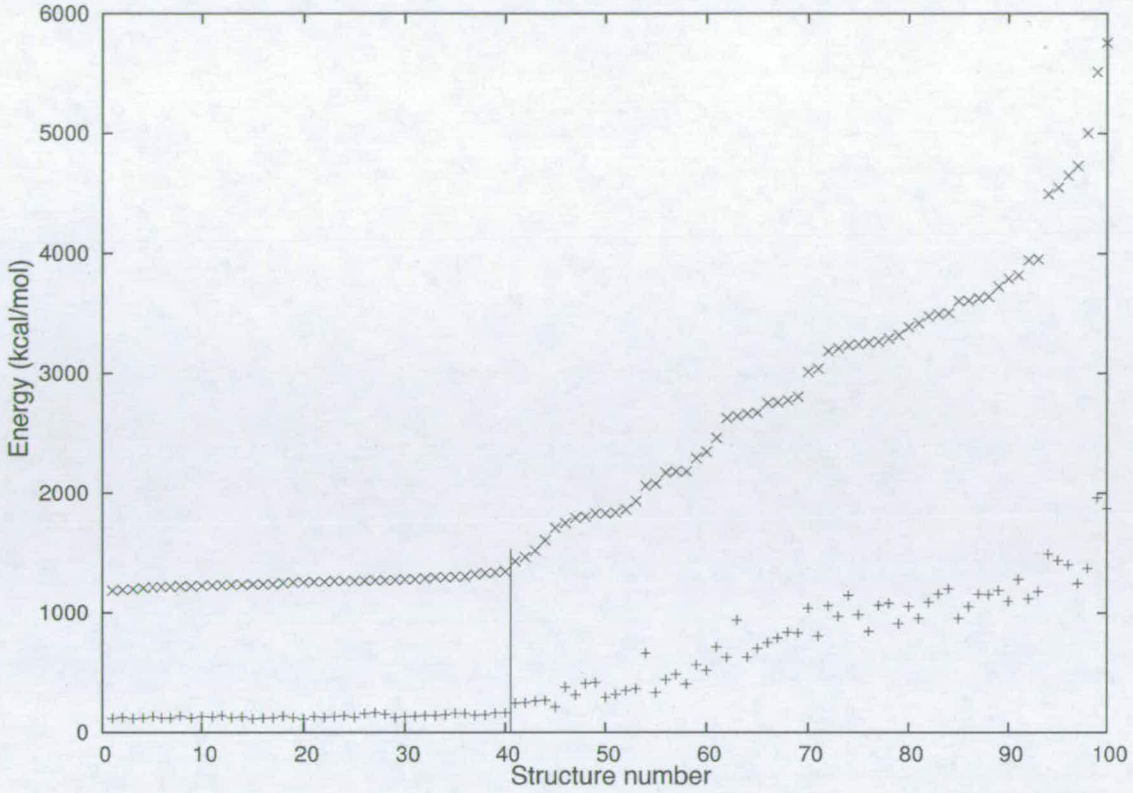

Figure 4.1 Plot of overall (x) and NOE (+) energy per structure. The low energy ensemble of 40 structures is indicated by the vertical line. Structures are ranked by overall energy.

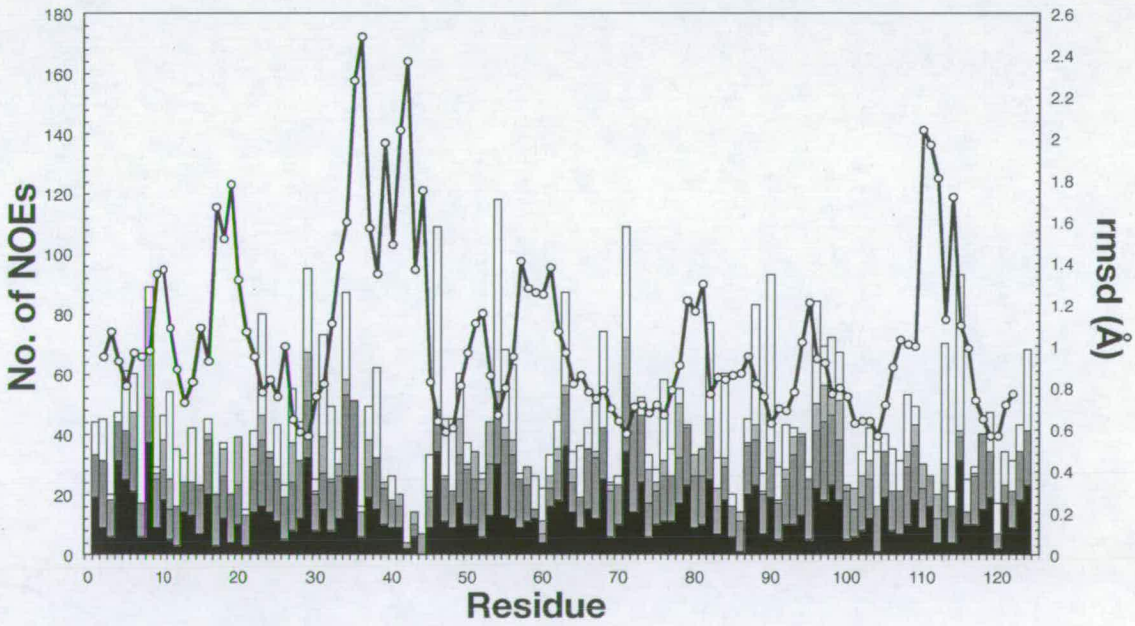

Figure 4.2 The number of unambiguous distance restraints per residue (bars, left-hand scale): intra-residue in black, sequential (residue i – residue i+1) in dark grey, short-range ($2 \leq |i-j| \leq 4$) in light grey and long-range ($|i-j| > 4$) in white. The rmsd of the C^{α} coordinates of residues 2-122 from the mean structure overlaid on both modules is also shown (right-hand scale).

4.3 Structural statistics

The structural statistics for the 40 structures in the low-energy ensemble are presented in Table 4.1. The precision of the structures is high and the PROCHECK (Laskowski *et al.* 1993; Laskowski *et al.* 1996) quality scores are acceptable for a small protein domain with disordered loops ; 90% of residues within the ensemble occur in the most favoured and additionally allowed regions of the Ramachandran (Ramachandran *et al.* 1963) plot (Figure 4.3).

rmsd for ensemble of 40 structures ± SD		
NOE (Å)	0.0423 ± 0.0028	
Bond Lengths (Å)	0.0025 ± 0.0001	
Bond Angles (°)	0.5123 ± 0.0177	
Coordinate rmsd (Å)		
Backbone atoms (C^{α}, N, CO): Excluding loops[a] (all residues[b])		
Module 1	0.528	(0.895)
Module 2	0.488	(0.597)
Both Modules	0.856	(1.042)
All heavy atoms : Excluding loops[a] (all residues[b])		
Module 1	0.851	(1.487)
Module 2	0.922	(1.108)
Both Modules	1.147	(1.507)
Number of violations > 0.5 Å	46	
Ramachandran assessment (%)		
Most favoured	52.4	
Additionally allowed	37.5	
Generously allowed	7.1	
Disallowed	3.1	

[a] excluding residues 15-21, 35-45 (module 1) 78-82, 107-111 (module 2)
[b] residues From CysI to CysIV of each module

Table 4.1 Structural statistics for the 40 lowest energy structures.

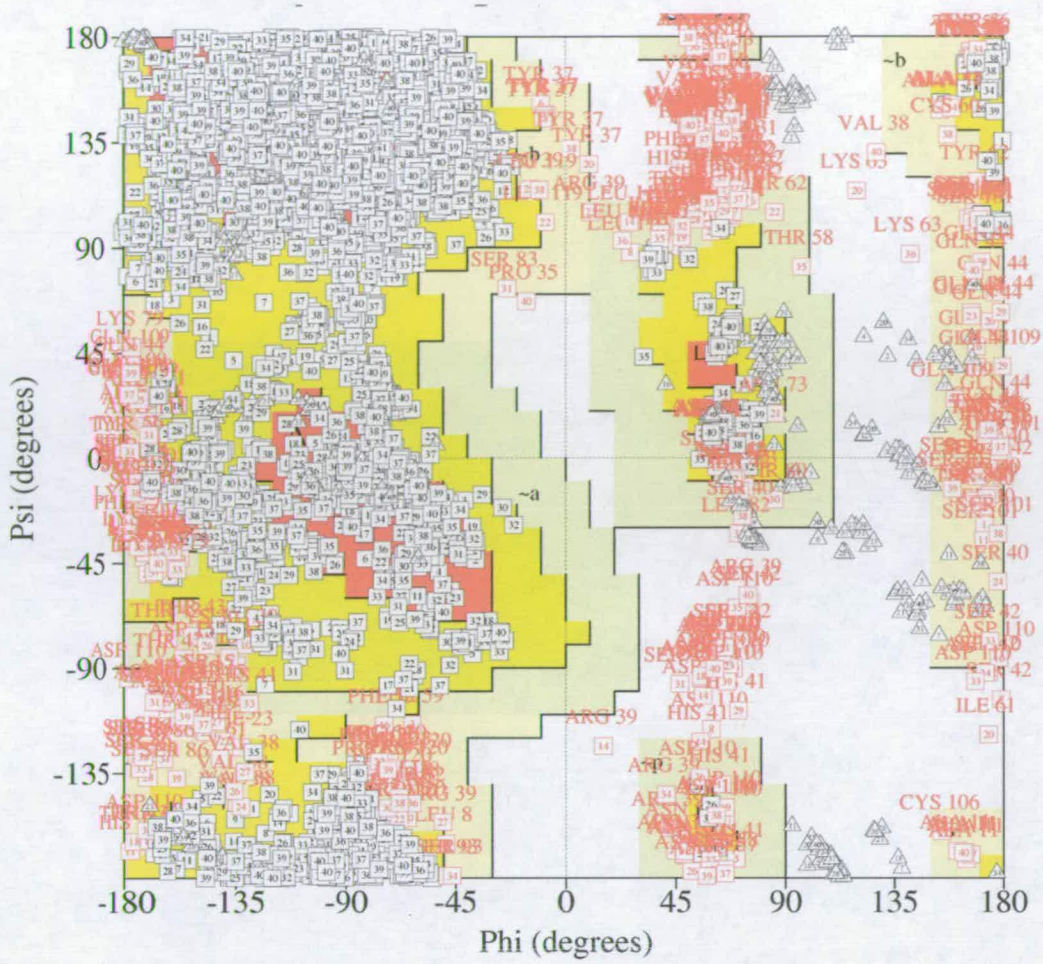

Figure 4.3 Ramachandran plot for the 40 lowest energy structures. Boxed numbers indicate the structure number; residue numbers indicate residues that lie outside the favoured regions.

4.4 The ensemble of structures

Figure 4.4 - Figure 4.6 show the backbone overlays of the low-energy ensemble of 40 structures (from 100 calculated) of C4BP12. These structures are overlaid on CCP1 (Figure 4.4), CCP2 (Figure 4.5) and both CCP modules (Figure 4.6). The residues used to fit the backbone overlays are the same as were used for the calculation of rmsd excluding loops in Table 4.1 and were obtained by analysis of the relaxation data (Section 4.5).

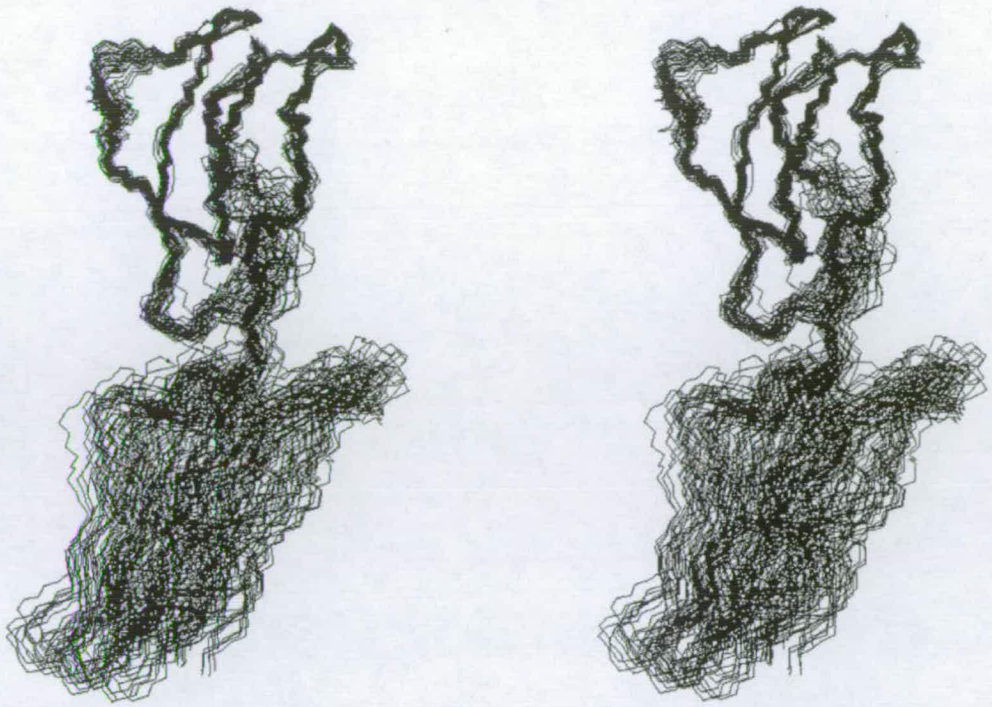

Figure 4.4 Stereo-view (cross-eyed) showing backbone overlay of the 40 lowest energy structures. Residues 2-122 are shown. Residues 2-60 (CysI to CysIV) excluding the loops consisting of residues 15-21 and 35-45 were used for the overlay.

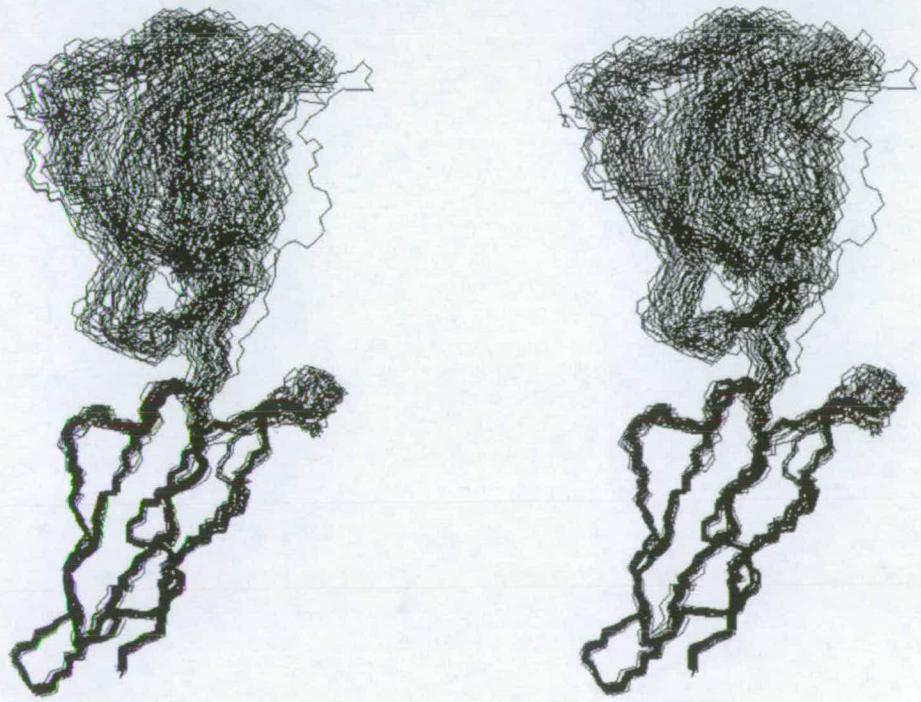

Figure 4.5 Stereo-view (cross-eyed) showing backbone overlay of 40 lowest energy structures. Residues 2-122 are shown. Residues 65-122 (CysI to CysIV) excluding the loops consisting of residues 78-82 and 107-111 used for the overlay

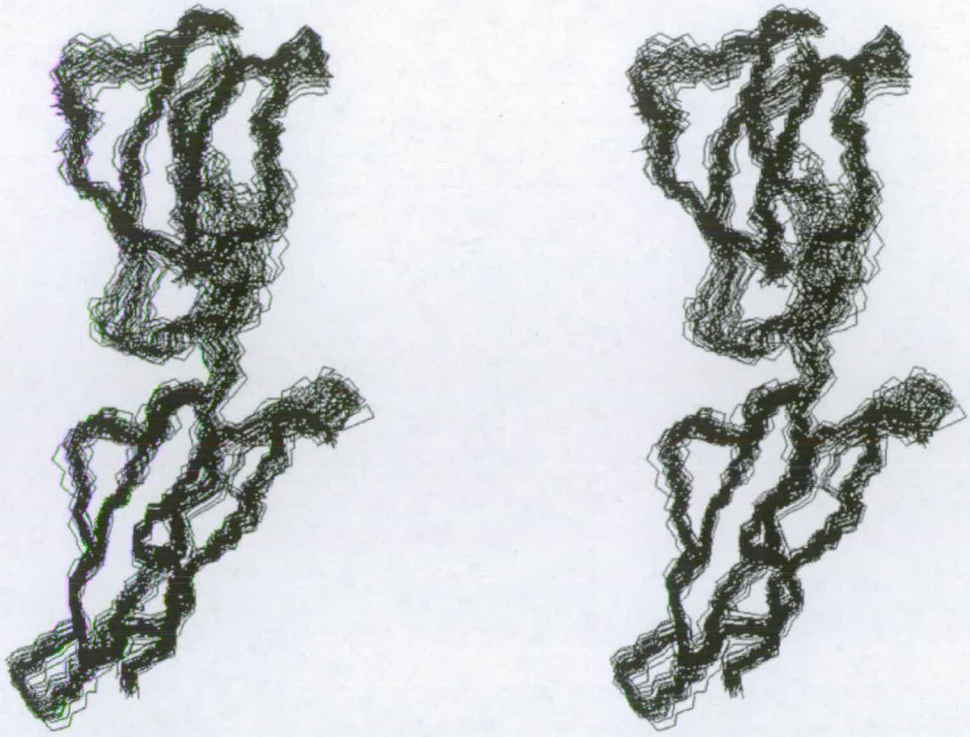

Figure 4.6 Stereo-view (cross-eyed) showing backbone overlay of 40 lowest energy structures. Residues 2-122 are shown. Residues 2-60 and 65-122, excluding loops as described above, were used for the overlay.

4.5 Relaxation data for C4BP12

Steady-state $\{^1H\}$–^{15}N NOEs and ^{15}N T_1 and T_2 values were measured at 600 and 800 MHz These experiments were recorded by the author with the assistance of Dr Dusan Uhrin. The data was processed and analysed with the help of Dr Graeme Ball. The T_1 and T_2 relaxation times were calculated from data at 600 MHz and 800 MHz by nonlinear least-squares fitting. To allow estimation of the experimental error of the measured peak intensities the spectrum corresponding to one of the relaxation delay values was re-acquired. Heteronuclear $\{^1H\}$–^{15}N NOEs were calculated from the ratio of the intensities of the cross-peaks in the reference spectra to those recorded with saturation of the ^1H signal. The plots of T_1, T_2 and $\{^1H\}$–^{15}N NOEs are shown in Figure 4.7.

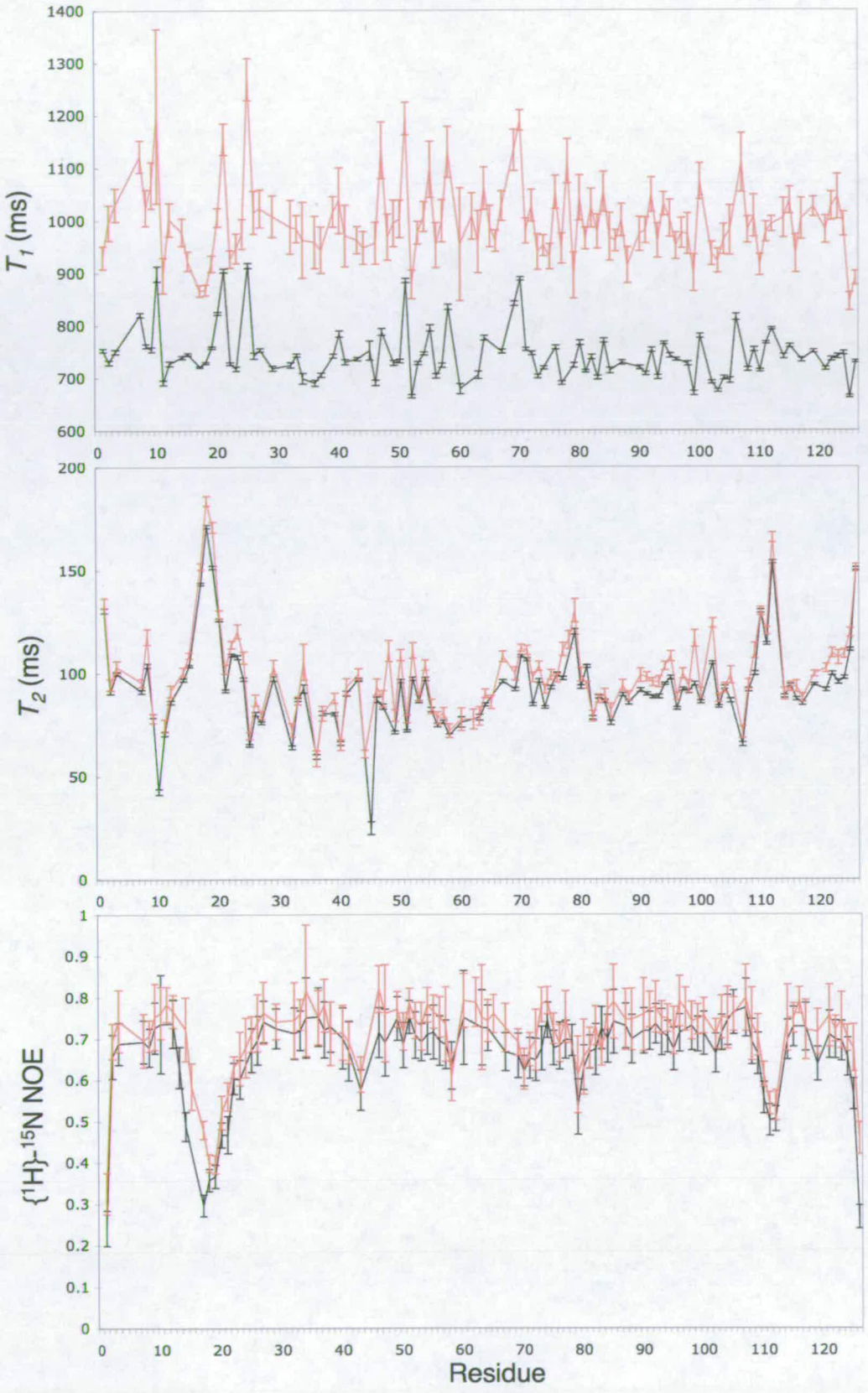

Figure 4.7 Relaxation data versus residue number. (top) T_1; (middle) T_2; (bottom) {^{1}H}-^{15}N NOE . Data collected at 600 MHz (black) and 800 MHz (red).

4.6 Discussion of the structure of C4BP12

The programme uwmn (M. Hartshorn and L. Caves, University of York) was used to identify the structure from the ensemble that was closest to the mean structure of the ensemble. The structures were fitted over the backbone of residues Cys^{2} to Cys^{122} - the first cysteine of CCP1 to the last cysteine of CCP2. The structure refine_rdcs_21.pdb was closest to the mean structure with a rmsd of 0.655 Å. Secondary structure was calculated for all the members of the low-energy ensemble using STRIDE (Frishman *et al.* 1995) and the secondary structure present in > 50% of the ensemble is shown in Figure 4.8.

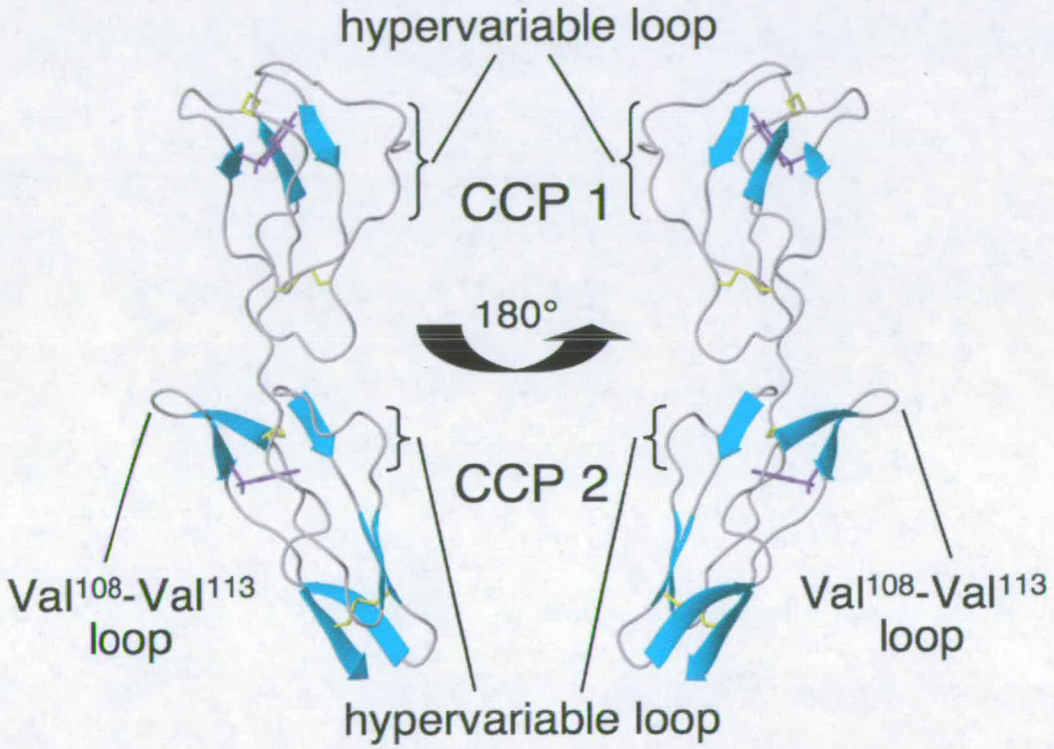

Figure 4.8 Two views of the closest to mean structure, with secondary structure as predicted by STRIDE (Frishman *et al.* 1995) for > 50% of structures in the ensemble, shown. Cysteine side chains, yellow; Tryptophan side chains, purple. Selected loops are labelled.

4.6.1 The individual modules

As can been seen from Figure 4.4 and Figure 4.5, the structures of the individual modules converge well. Each CCP module has a similar, elongated, shape (Figure 4.6) in which short

β-strands and other extended segments [that do not satisfy the criteria used to define β-strands (Frishman *et al.* 1995)] are aligned with the long axis. Five strands/extended segments wrap around a hydrophobic core that is bounded by the two invariant disulphide bridges. The strands/extended segments run up-down-up-down-up such that the N- and C-termini are at opposite ends of the module. The connecting turns and loops also generally lie towards the ends of the module, but the "hypervariable loop" - a site of high sequence variation and of insertions or deletions (Figure 1.3) - projects laterally (labelled in Figure 4.6). The longer hypervariable loop of CCP1 (residues Asp^{15}-Thr^{21}) is poorly defined by the data and corresponds to a prominent dip in the plot of $\{^1H\}$-^{15}N NOEs (Figure 4.7) that is consistent with motion on the nanosecond timescale. The $\{^1H\}$-^{15}N NOE data and the high value of the T_1/T_2 ratio for Thr^{45} indicate a second, poorly defined, flexible region in CCP1 that includes residues 35-45 (Figure 4.7) and is close to, but does not participate in, the interface. A prominent feature of CCP2 is the loop (residues Val^{108}-Val^{113}), between the fourth and fifth extended segments, that carries a three-residue insertion relative to CCP1 (see Figure 1.3). This loop is flexible, as judged by $\{^1H\}$-^{15}N NOEs, (see Figure 4.7), and it lies close to the intermodular interface. In general, the values of T_1, T_2 and heteronuclear NOEs are more variable in CCP1 than in CCP2, reflecting a higher level of flexibility in the first module.

4.6.2 Intermodular flexibility

The ensemble of structures converge well when superimposed over both modules (Figure 4.6) although not as well as when superimposed over the individual modules. The intermodular interface is mainly hydrophobic with contributions from all four linker residues (Ile^{61}, Tyr^{62}, Lys^{63} and Arg^{64}) and from Tyr^{37} and Val^{38} of CCP1, along with Phe^{84}, Val^{108} and Val^{113} of CCP2. The Tyr^{37}, Tyr^{62} and Arg^{64} side-chains line up on one side, while on the other, Val^{38} and Val^{108} bracket Phe^{84} and all three contact Lys^{63}, which in turn lies alongside

Ile[61]. The observed arrangement comprises a well-structured hydrophobic pocket exhibiting only limited flexibility (as indicated by the $\{^1H\}$-^{15}N NOE data for this region). This region is shown in detail in Figure 4.9.

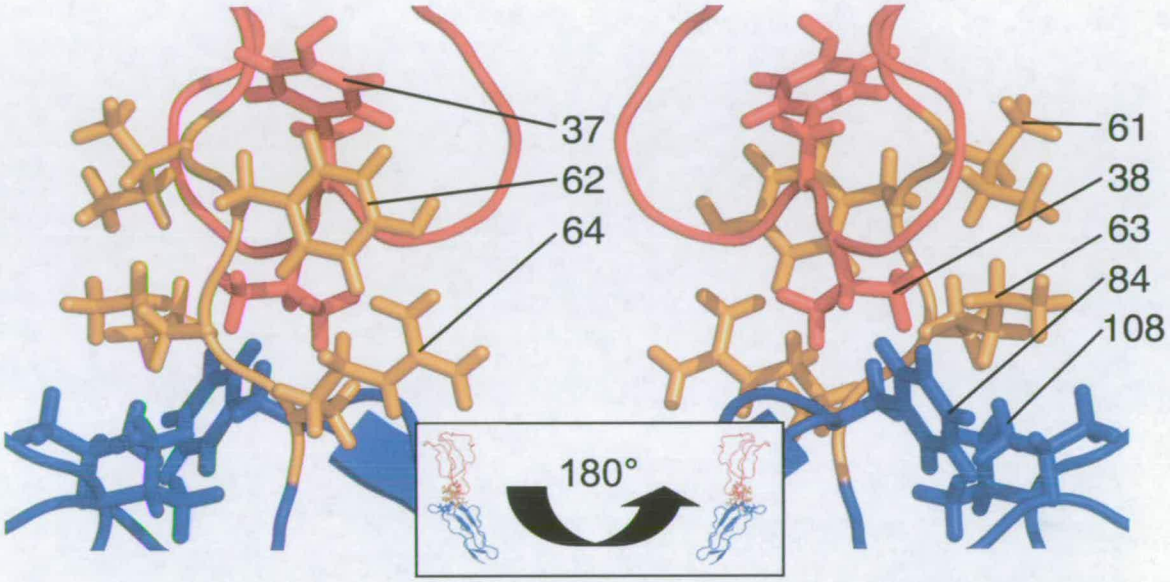

Figure 4.9 The CCP1/2 interface in the closest to mean structure. Colours: CCP1- red; CCP2 - blue; linker - orange. The side-chains of residues discussed in the text are shown in stick representation and labelled. The orientation of C4BP12 is indicated in the boxed insert. Figure produced with PyMOL (www.pymol.org)

There is no evidence from the relaxation measurements for fast- or slow-timescale motion in the backbone of the linker (Figure 4.7). However, a superposition based on individual modules is better than a superposition of the module-pair (Figure 4.4 - Figure 4.6), reflecting a small range of intermodular angles amongst the ensemble. These angles (illustrated in Figure 4.10) are calculated using for reference a vector connecting the principal inertia tensor of each module with the C^{α} of its consensus tryptophan. The values of the intermodular angles are listed in Table 4.1 and shown graphically in Figure 4.10. This could be interpreted in terms of a restricted degree of intermodular movement, but could also reflect a limitation of the experimental data. Both the mean tilt and twist angles are relatively

small so that the two modules form an elongated structure in which equivalent features – such as the hypervariable loops – lie on the same face of the molecule (Figure 4.8).

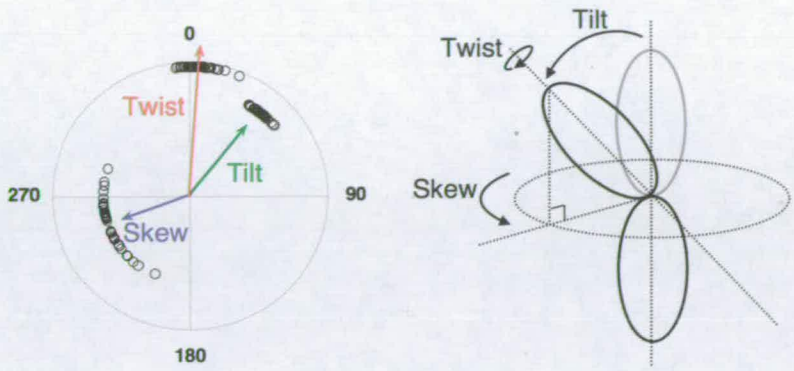

Figure 4.10 Intermodular angles for the ensemble of structures. Arrows show mean values

4.6.3 Comparison with other CCP module structures

The structures of CCP1 and CCP2 of C4BP12 were compared to each other and to all the experimentally solved CCP module structures using the programme Combinatorial Extension (Shindyalov *et al.* 1998). Comparison of CCP1 with CCP2 yielded a C^{α} rmsd of 3.4 Å over 57 residues. The second module of the C4BP α-chain is highly similar in structure to the 16th CCP module of CR1 (pdb = 1GKN) and the 3rd CCP module of decay accelerating factor (DAF, pdb = 1H03) (C^{α} rmsd of ~1.6 Å over 62 residues in both cases). It is also similar to the 2nd module of MCP (pdb = 1CKL) (2.0 Å over 62 residues). All of these modules belong to the same sequence cluster (cluster C) as defined by Soares *et al.* (Soares *et al.* 2005) and occupy the second-module positions within the C3b/C4b-binding sites of their respective parent proteins (Kirkitadze *et al.* 2001). On the other hand, CCP 1 of C4BP is a rather more structurally divergent module. Its closest known structural relatives include the 15th module of CR1 (2.2 Å over 60 residues) and the 2nd module of DAF (2.5 Å over 63 residues), both of which are first modules within C3b/C4b-binding sites. Therefore, both at

the level of individual module structures, and from a functional perspective, there are parallels between module pairs C4BP12, DAF23 and CR1-1516. These structures are illustrated in Figure 4.11.

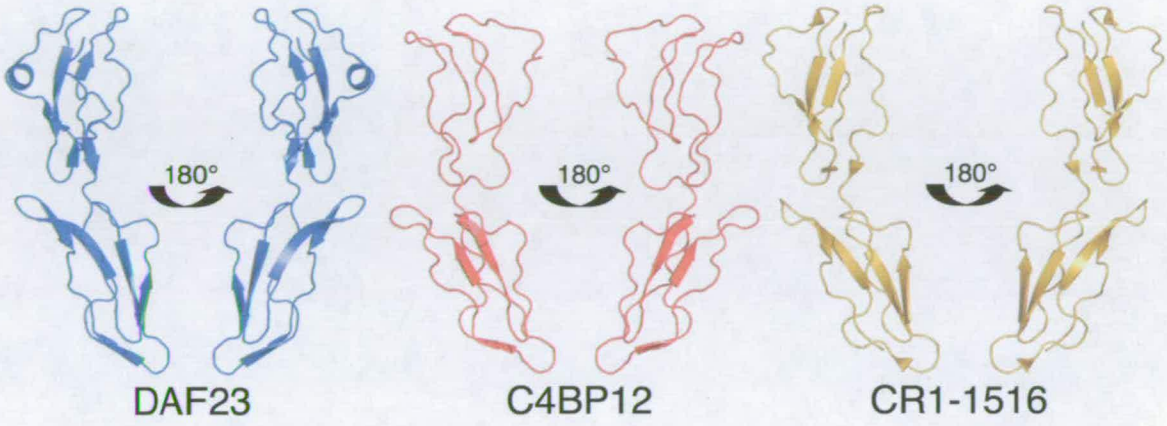

Figure 4.11 Comparison of the structures of C4BP12 (closest to mean structure) with modules 2 and 3 of DAF (crystal structure, pdb = 1OK3) and modules 15 and 16 of CR1 (closest to mean structure of ensemble of 24 NMR structures, pdb = 1GKN).

From inspection of Figure 4.11 it can be seen that whilst the tilt and skew angles of C4BP12, DAF23 and CR1-1516 are similar the twist angle for C4BP12 is almost 180° to that of DAF23 and CR1-1516. As these module pairs represent part of the C4b-binding site in CR1, DAF and C4BP it appears that the ability of a protein to bind C3b/C4b does not appear to require a specific intermodular arrangement in the free protein. This is discussed further in Chapter 7.

A prominent structural feature of C4BP12 is the loop in module 2 formed by the residues between Val^{108} and Val^{113} (sequence – QDRG). This loop is marked in Figure 4.8. The presence of such an exposed loop - four mostly polar residues (which are not involved in the intermodular interface) bounded by two hydrophobic residues that contribute to the interface with the preceding module - is unique to cluster-C CCP modules. As mentioned above cluster-C members form the second modules of C3b/C4b recognition sites in C4BP, MCP,

DAF, and both sites 1 and 2 of CR1. Some conservation amongst these modules in the use of specific features for C4b-binding might be expected and this loop would be an obvious candidate. It is indeed critical to the C3b/C4b-binding activities of CR1 sites 1 and 2 (Krych *et al.* 1994a; Krych *et al.* 1998a). This means that substitution of residues between Val^{108} and Val^{113} of C4BP represents a worthwhile exercise; this is currently in progress.

The electrostatic surface of the closest to mean structure was calculated using APBS (Baker *et al.* 2001) and is shown in Figure 4.12.

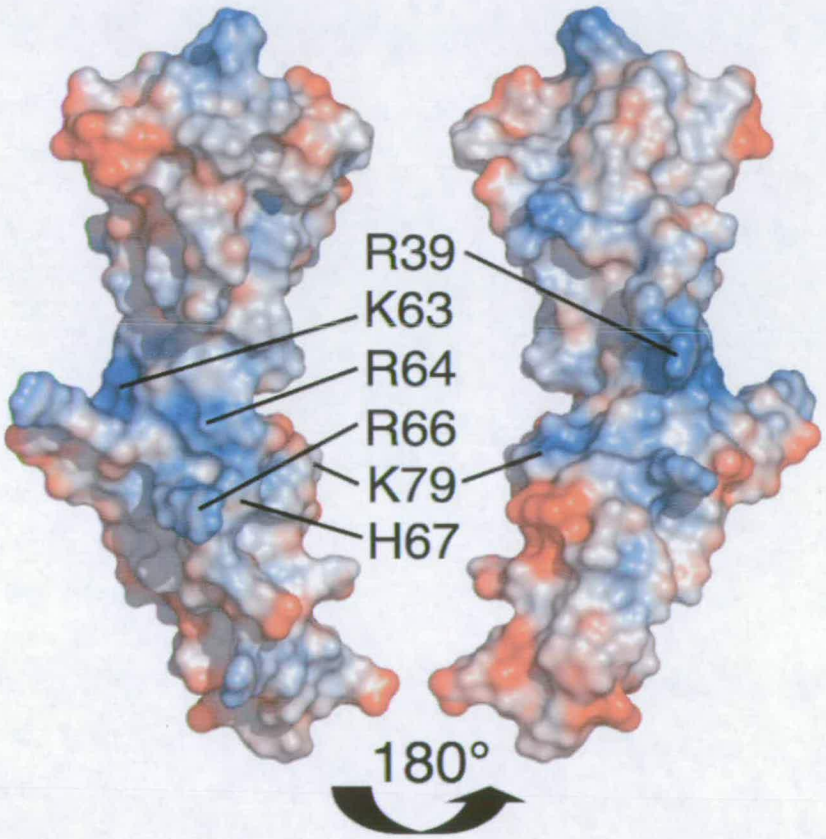

Figure 4.12 Electrostatic surface potential of closest to mean structure. Calculated using the Adaptive Poisson-Boltzmann Solver (Baker *et al.* 2001) plug-in within PyMOL: red is negative charge and blue is positive charge. A range of -5/+5 kT was used. Those amino acids substituted in mutagenesis studies by Blom *et. al.* are marked.

4.6.4 Interpretation of mutagenesis data for C4b binding

The design and interpretation of site-directed mutagenesis data for C4BP (mutated residues marked on Figure 4.12) previously relied upon models built on the basis of homology with other RCA protein structures. The experimentally determined structure of C4BP12 allows the mutagenesis data pertaining to C4BP12 to be considered in a more accurate structural context. The proposed contribution of Arg^{39}, Lys^{63}, Arg^{64} and His^{67} – all of which are functionally critical for C4b binding, (Blom *et al.* 1999; Blom *et al.* 2000) - to a contiguous positively charged surface patch is not consistent with the observed solution structure (Figure 4.13). The intermodular twist rotates Arg^{39} in module 1 away from His^{67} in module 2. Furthermore, within the linker, the positive charges of Arg^{64} and Lys^{63} are exposed on opposite faces. When an α-chain of C4BP binds to C4b either it may lie within a valley or groove on the surface of the bigger protein or, alternatively, an inter-modular motion occurs that brings these four critical residues onto the same face of the molecule. This is discussed further in Chapter 7.

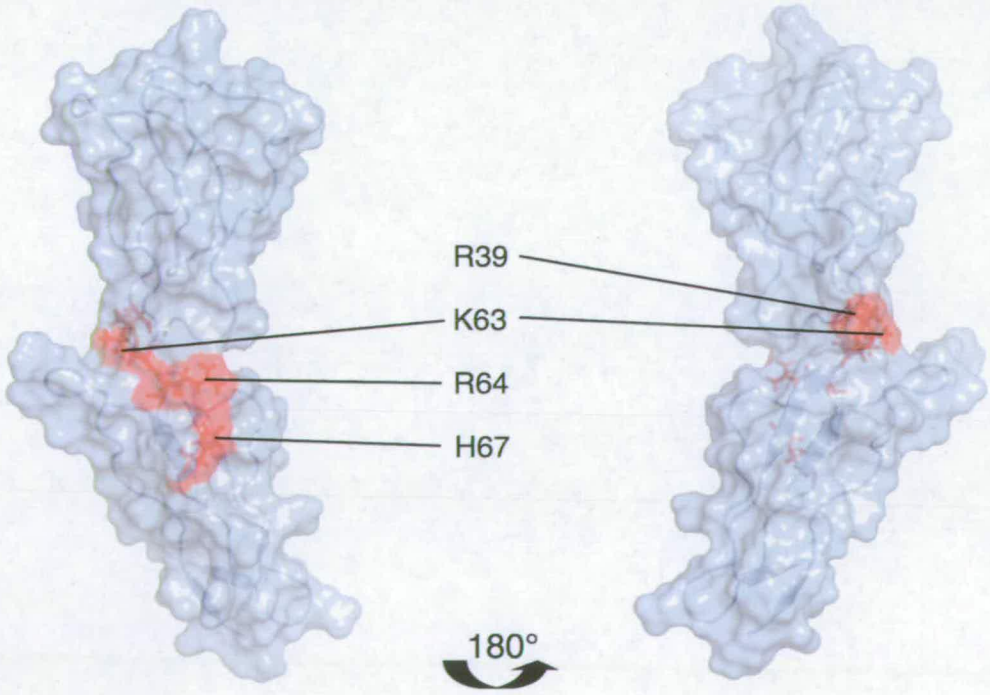

Figure 4.13 Residues shown to be critical for C4b binding mapped on the closest to mean structure of C4BP12.

4.7 Refinement of the structures in explicit solvent.

Refinement of NMR structures in a full molecular dynamics force field including electrostatic and Lennard-Jones non-bonded potentials and interaction with solvent has been shown to improve the quality of the structures (Linge *et al.* 2003). This technique was successfully applied to over 500 NMR structures from the PDB to generate the RECOORD (recalculated coordinate) database (Nederveen *et al.* 2005). In order to quickly investigate the effect of water refinement on the structure of C4BP12, the RECOORD scripts (available from http://www.ebi.ac.uk/msd/recoord) were used to refine the 40 low-energy structures in explicit solvent. The final NOE-derived distance restraints were used in the refinement but the RDC and dihedral angle restraints were omitted as these are not incorporated into the RECOORD protocols. As found in the RECOORD study, the RMSD of the ensemble increased with water refinement but the overall fold remained identical (Figure 4.14). The backbone rmsd for Cys^{2}-Cys^{122} between the two closest to mean structures was 1.39 Å. The quality scores were significantly improved by water refinement (Table 4.2, Figure 4.15).

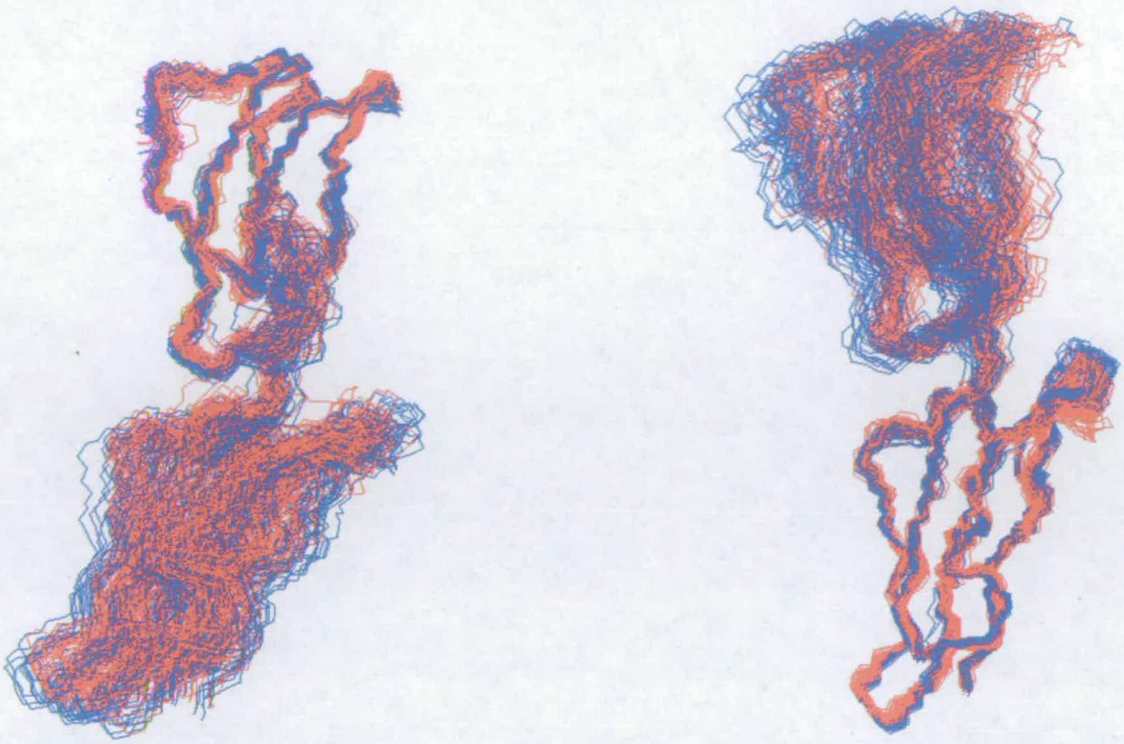

Figure 4.14 Backbone overlays of 40 lowest energy structures before (blue) and after (red) water refinement. Residues 2-122 are shown. (left) overlaid on CCP1 (right) overlaid on CCP2, the residues used for the overlay are as in Figure 4.4 and Figure 4.5.

	Original structures	After water refinement
Coordinate rmsd (Å)		
Backbone atoms (C^{α}, N, CO): Excluding loops[a] (all residues[b])		
Module 1	0.528 (0.895)	0.616 (0.890)
Module 2	0.488 (0.597)	0.588 (0.723)
Both Modules	0.856 (1.042)	0.898 (1.073)
Ramachandran assessment (%)		
Most favoured	52.4	70.2
Additionally allowed	37.5	22.9
Generously allowed	7.1	4.1
Disallowed	3.1	2.8
WHATCHECK structure Z-scores, positive is better than average		
2nd generation packing quality	-3.002	-2.424
Ramachandran plot appearance	-6.405	-4.489
$\chi 1/\chi 2$ rotamer normality	-3.623	-2.915
Backbone conformation	-8.294	-9.481

[a] excluding residues 15-21, 35-45 (module 1) 78-82, 107-111 (module 2)
[b] residues From CysI to CysIV of each module

Table 4.2 Structural quality scores before and after water refinement

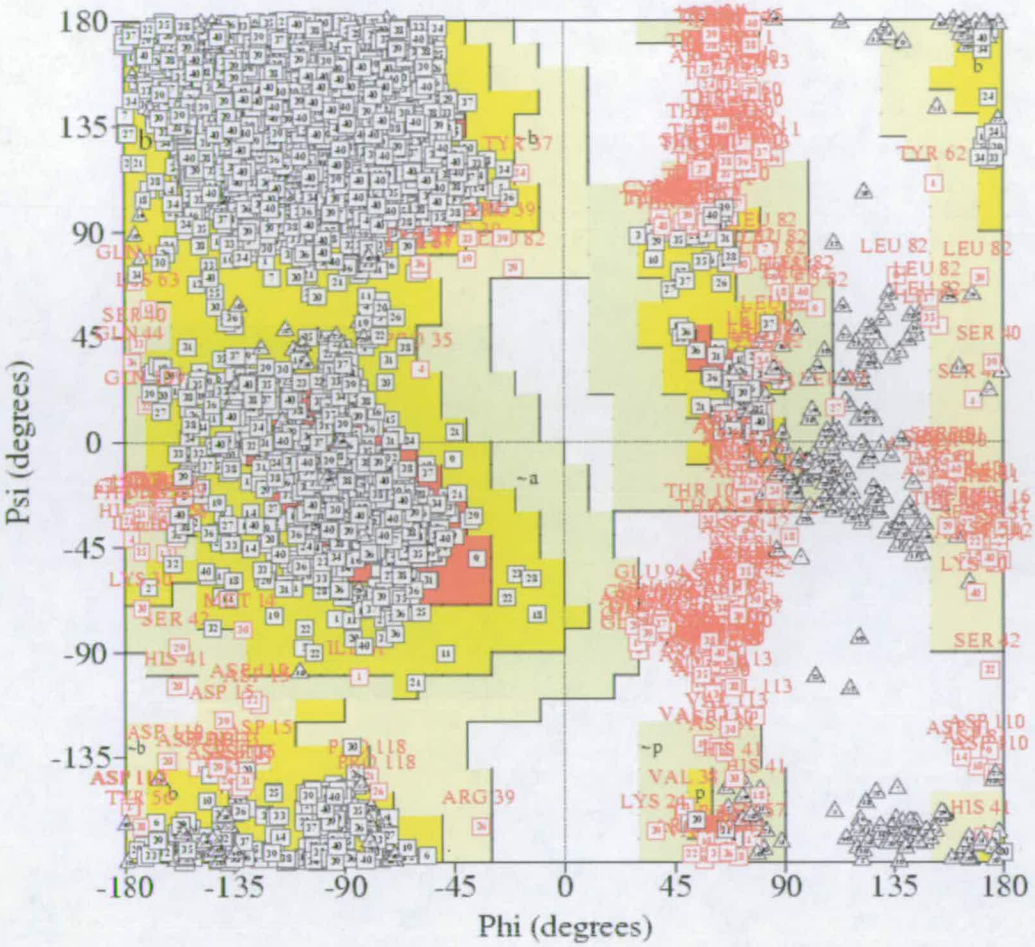

Figure 4.15 Ramachandran plot for the 40 lowest energy structures after water refinement. Boxed numbers indicate the structure number in the ensemble; red boxed numbers annotated with residue numbers indicate residues that lie outside the favoured regions of the plot

4.8 Conclusions

The structure of C4BP12 was solved using the methods described in Chapter 3. The structures show the features expected of a pair of CCP modules and compare well to previously solved CCP-module structures. The cluster of positively charged residues present on the model structure of C4BP, and implicated in function by mutagenesis studies, is not present in the experimental structure and this suggests that intermodular re-orientation may occur on C4b binding. This is discussed further in Chapter 7. Whilst the quality statistics for the C4BP12 structure are acceptable, refinement in explicit solvent can improve the quality

of the structures without changing the overall fold. As the inclusion of RDC derived restraints into the water refinement protocol is non-trivial these restraints were not included in the water-refined structures. For this reason the refined structures were not judged to be an improvement on the original structures that were submitted to the PDB and so the PDB deposition has not been updated.

Chapter 5

M PROTEIN BINDING

5.1 Introduction

The interaction between C4BP12 and M proteins from *Streptoccus pyogenes* is the most well characterised host-pathogen interaction involving C4BP. As C4BP is active on the surface of *S. pyogenes*, this interaction allows the bacterium to avoid opsonisation and subsequent phagocytosis (Carlsson *et al.* 2003). Previous experiments had located the binding site for M protein to CCP1-3 of the α-chain (Accardo *et al.* 1996) and shown that the site overlapped with, but was not identical to, the C4b-binding site (Blom *et al.* 2000). Prior to the structure of C4BP12 presented in this thesis, there was no high-resolution information for this region of C4BP. Therefore, a primary motivation for solving the structure of C4BP12 was to map the binding site for M proteins at an atomic level. This was achieved and the unexpected result that binding of M protein involves intermodular re-orientation of C4BP12 was obtained.

5.2 CCP1 and CCP2 are necessary for M protein binding

Prof. Anna Blom performed M4-binding assays with full length *S. pyogenes* M4 protein and the module-deletion and alanine-insertion C4BP mutants described in Section 1.5.1. More details of these experiments are given in (Jenkins *et al.* 2006). These studies showed that none of CCP modules 3-8 were required for M4 binding but that deletion of CCP2 led to significantly (~10-fold) reduced affinity. Deletion of CCP1 destroyed binding almost completely. The addition of alanine residues between CCP2 and CCP3, or between CCP3 and CCP4, had no effect on binding, whereas insertions between CCP1 and CCP2 caused a loss of M4 binding (Figure 5.1). These results show that CCP1 and CCP2 are necessary for M4 binding to C4BP.

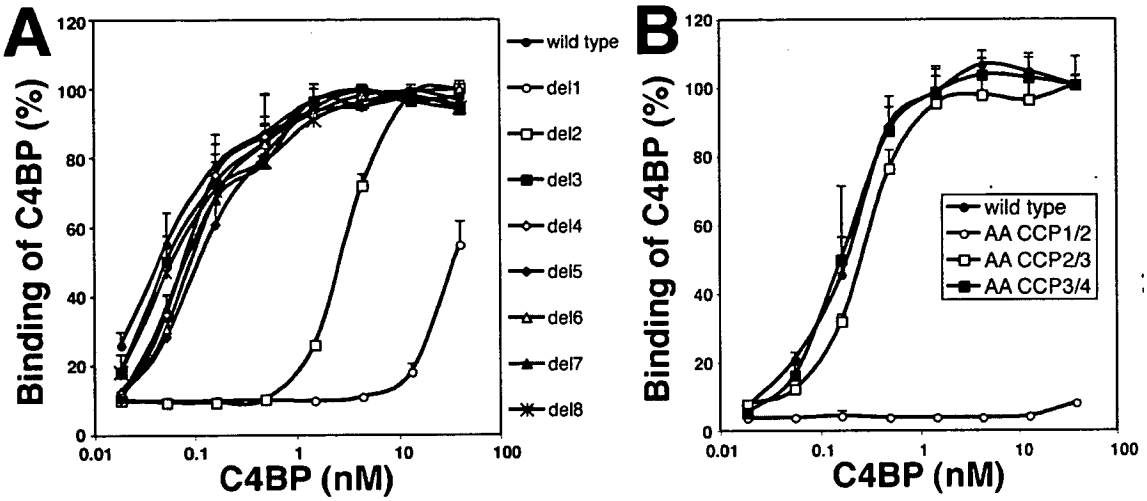

Figure 5.1 Binding of CCP module deletion mutants (A), and double-alanine insertion mutants (B) of C4BP to M4. Microtiter plates coated with M4 were exposed to recombinant C4BP molecules in which one of the eight CCP modules of the α-chain was deleted (A), or in which two Ala residues were inserted into the linker between CCPs 1 and 2, 2 and 3 or 3 and 4 (B). Binding was expressed as a percentage of the maximum binding of wild type observed in each experiment. Figure reproduced from Jenkins *et al.* (2006).

5.3 CCP1 and CCP2 are sufficient for M4 binding

In order to demonstrate that modules CCP1 and CCP2 were sufficient for M4 binding, isothermal titration calorimetry (ITC) was used to measure the binding affinity between C4BP12 and a dimerised peptide corresponding to the hypervariable region of M4. This peptide consisted of residues 1-45 of the M4 (also known as Arp4) protein with a C-terminal non-native cysteine added and was termed M4-N. A dimer was used as previous studies had shown that dimerisation of M4 strongly enhanced C4BP binding (Morfeldt *et al.* 2001). The peptide was expressed in *E. coli* by Jenny Persson (Lund University, Sweden) as described in (André *et al.* 2006). The ITC was performed at the BBSRC/EPSRC Biological Microcalorimetry Facility, University of Glasgow, by the author with the assistance of Margaret Nutley and Prof. Alan Cooper. The results were consistent with a 1:1 stoichiometry and a K_d of 0.5 μM. The data was corrected for heat of dilution effects by subtracting the results of a blank titration (i.e. injecting M4-N into buffer) and fitted using a one-site model

resulting in the following: stoichiometry, N = 0.787 ± 0.015; K_d = 0.5 ± 0.07 μM; ΔH = -9450± 243 cal mol^{-1}; ΔS = -2.96 cal $mol^{-1}K^{-1}$. These results are shown in Figure 5.2.

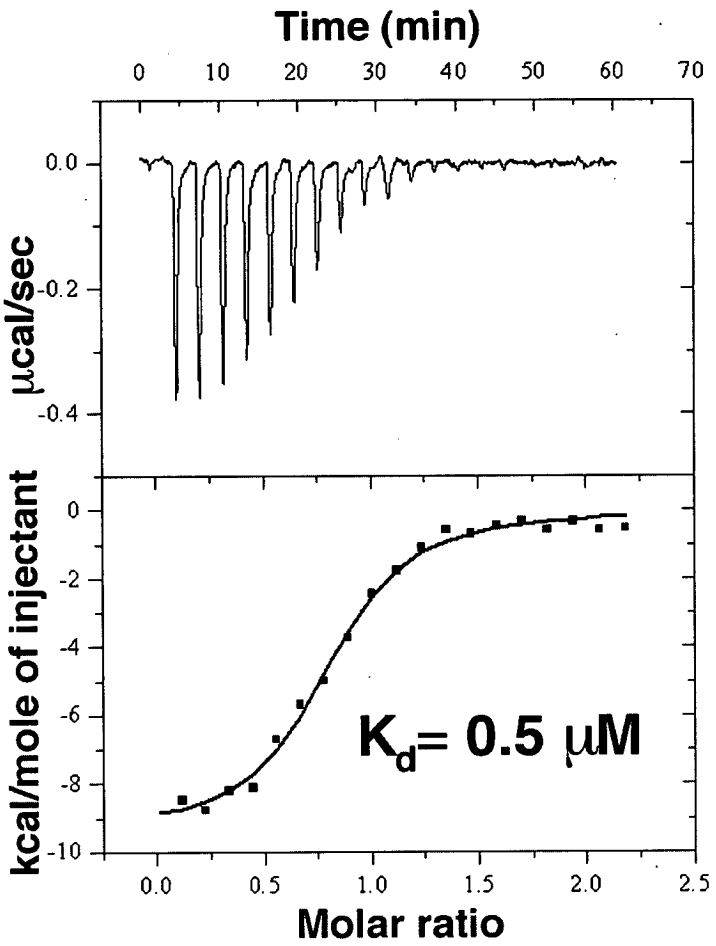

Figure 5.2 Isothermal titration calorimetry profile for the interaction of M4-N with C4BP12. Top, heat differences obtained for 20 injections (corrected for heat of dilution effects). Bottom, integrated curve with experimental points (■) and the best fit (—).

5.4 C4BP12 and M4-N model system

The results summarised in Figure 5.1 and Figure 5.2 prove that the N-terminal two CCP modules of the C4BP α-chain, along with the four residues that link them, are both necessary and sufficient for M protein recognition. The data confirm that M4-N represents a useful model of the C4BP-binding region of the bacterial protein (Morfeldt *et al.* 2001). Therefore,

the C4BP12 M4-N interaction could be used to map the binding site of *S. pyogenes* M proteins on C4BP (Section 5.6).

5.5 Initial studies with M22

Prior to the studies with M4, Prof Gunnar Lindahl sent two peptides corresponding to the HVR (53 residues, termed M22-N) and the HVR and IgA binding domain (83 residues, termed M22-rec) of the C4BP-binding M22 (also called Sir22) protein. Both peptides were dimerised via a non-native C-terminal cysteine. The M22-N was produced by peptide synthesis and M22-rec was produced in *E. coli.* Due to problems exchanging the M22-N peptide into NMR buffer it was abandoned and M22-rec was used for a preliminary titration with 50 μM u[^{15}N]-C4BP12. This was performed using a 500 MHz spectrometer equipped with a cryoprobe at the University of Dundee with the assistance of Dr David Norman.

The results of this titration are shown in Figure 5.3. Addition of the peptide led to the broadening and disappearance of many peaks in the ^{15}N-HSQC spectrum. This could either be due to the formation of a large, slowly tumbling complex, non-specific aggregation of the C4BP12:M22-rec complex, or an exchange rate between free and bound C4BP12 that is intermediate on the NMR time-scale. Adding more peptide pushed the equilibrium towards formation of the complex and led to the reappearance of many peaks. Overlaying the first and last spectra in Figure 5.3 showed that the shifts of the cross-peaks induced by the peptide were small. A ^{15}N-TROSY spectrum of the complex (at 500 MHz) yielded no improvement over the ^{15}N-HSQC, suggesting that it was not the size of the complex (~35 kDa) that led to the poor quality of the spectrum but rather the intermediate exchange regime. Adjusting the experimental conditions (i.e buffer, pH, salt concentration and temperature) might improve the quality of the spectra. As the initial titration studies with the related peptide M4-N were much more successful, no further work on M22 was performed.

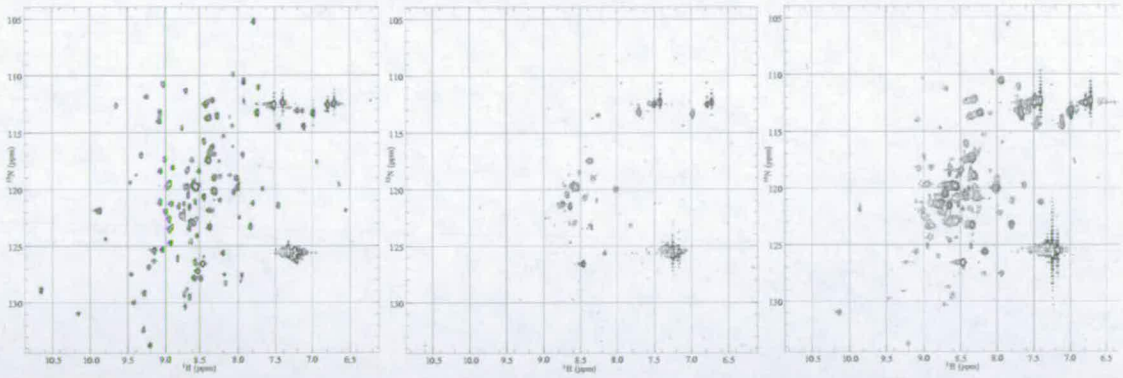

Figure 5.3 ^{15}N-HSQC spectra for 50 μM C4BP12 before addition of M22-rec (left) and at M22-rec:C4BP ratios of 4:1 (centre) and 8:1 (right). Samples were in 20 mM deuterated NaOAc buffer, pH 4.6. Spectra acquired at 500 MHz (^{1}H frequency).

5.6 Titration with M4-N

Significant changes in the ^{15}N-HSQC spectrum of C4BP12 occurred upon addition of M4-N (Figure 5.4). At a low ratio (0.5:1) of M4-N to C4BP12, some cross-peaks disappear (peaks labelled in Figure 5.4) while others move a small distance within the spectrum. As this ratio is increased, those peaks that had previously vanished reappear at a new frequency, while those that had moved shift further from their original positions. The disappearance and subsequent reappearance of some peaks reflects exchange between free and bound forms of C4BP12 with a frequency comparable to the difference between the frequencies of the original peak (free) and the new one (bound). On the other hand, when the difference between the frequencies of the original peak and the new one is smaller, then the exchange rate is faster relative to the frequency difference, giving rise to the peaks that move incrementally during the course of the titration. The line-widths of the cross-peaks for the complex are not significantly broader than those of free C4BP12, despite the increased molecular weight of the complex (15 kDa for C4BP12 + 11.3 kDa for M4-N). One explanation is that the complex is less anisotropic than free C4BP12 and tumbles accordingly – this could happen if the bound form of M4-N is relatively compact and binds towards the centre of C4BP12, as opposed to at either end.

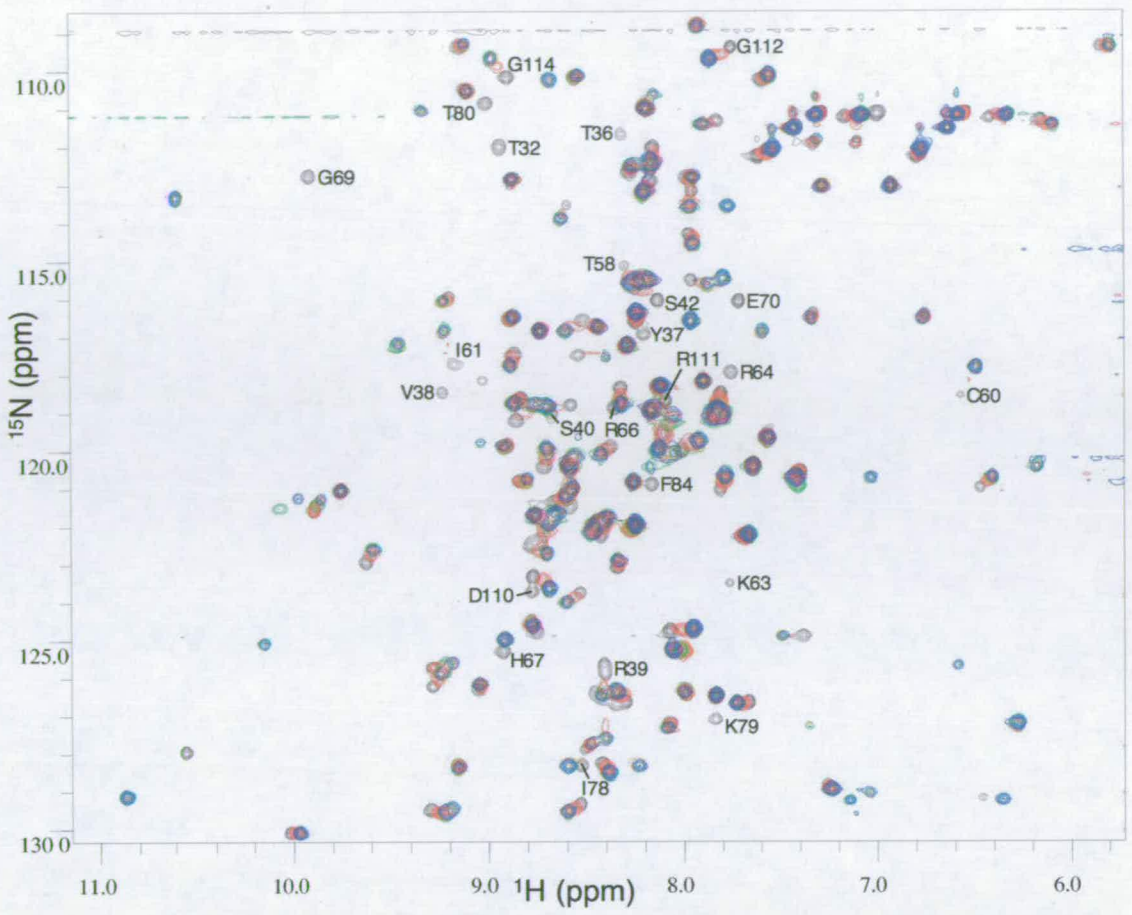

Figure 5.4 Overlay of five ^{15}N-HSQC spectra of 20 μM C4BP12 (20 mM NaOAc, pH 6.0, 37 °C) with increasing concentrations of M4-N. Colours as follows: black - 0 μM M4-N, red - 10 μM M4-N, orange - 25 μM M4-N, green - 50 μM M4-N, blue - 100 μM M4-N (all concentrations are for the dimer). Cross-peaks that disappear and then reappear are labelled at their positions in the free spectrum (black).

5.7 Location of the residues on the structure of C4BP12

Most of the significantly perturbed residues (Figure 5.5) are in the intermodular linking sequence, or in loops or turns near the intermodular interface. From the surface representation (Figure 5.5) it is obvious that these residues do not all reside on one face of the molecule. It is therefore improbable that they are all simultaneously involved in contacting the peptide. More likely the chemical shifts of some of these residues are perturbed by re-orientation of the two modules upon M4-N binding. Within CCP2, however, two stretches of residues distant from the interface (Gly^{69}-Glu^{70} and Ile^{78}-Thr^{80}) have

significantly perturbed chemical shifts. These do form a contiguous surface, and together comprise a feasible binding patch. The protein engineering experiments described above establish that residues from the interface-proximal region of module 1 and/or in the linker additionally contribute directly to binding.

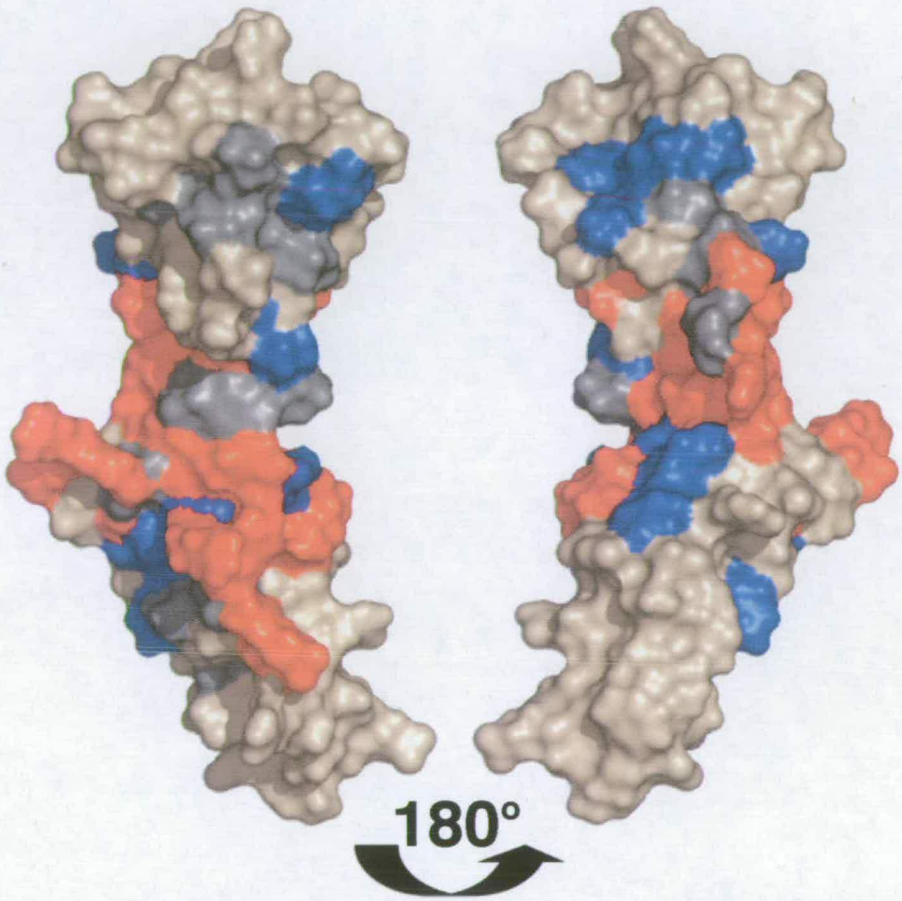

Figure 5.5 (Top) surface representation (N-terminus at top), indicating residues that show chemical shift changes upon M4-N binding. Blue – small combined chemical shift difference (see below); red – large chemical shift difference, unable to track cross-peak. Residues for which no NH shift information could be obtained (either proline residues or due to overlap in the HSQC spectra) are coloured grey. (Bottom) sequence of C4BP12 illustrating the NH chemical shift changes on addition of M4-N. Residues coloured as follows: Black – no apparent change [combined chemical shift difference $\sqrt{((\Delta\delta^{N}H)^2+(\Delta\delta^{15}N/5)^2)} < 0.08$ ppm]; grey – data missing; blue - small shift (combined chemical shift difference of > 0.08ppm, < 0.2 ppm but able to trace peak movement); red – large shift (unable to track peak).

5.8 Intermodular re-orientation upon M4-N binding

The changes in C4BP12 chemical shifts that accompany association with M4-N implicate residues towards the middle of the module pair in binding, while the N-terminus of CCP1 and the C-terminus of CCP2 are not involved. This is consistent with a complex that is globular, rather than linear, as also inferred from the relatively narrow line-widths of its NMR signals. The extent and distribution of affected residues are most convincingly explained by a ligand-induced intermodular conformational adjustment. There are three aromatic side-chains that can exert ring-current shifts in the vicinity of the interface (Figure 5.6) - one from each of CCPs 1 and 2, and one in the linker (Tyr^{37}, Tyr^{62} and Phe^{84}, respectively). Therefore an intermodular re-orientation upon binding would result in large chemical shift changes to residues around the CCP1/2 interface as observed. The residues away from the interface that are likely to be in direct contact with M4-N are also labelled in Figure 5.6.

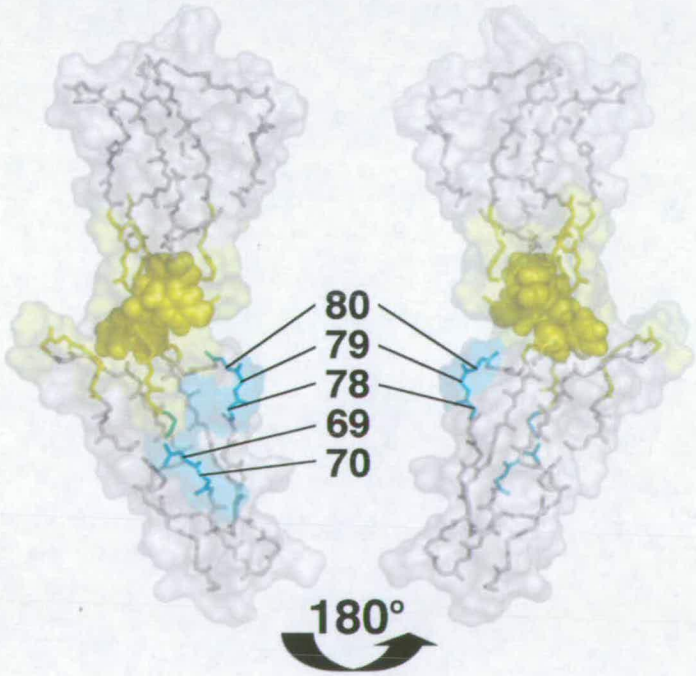

Figure 5.6 Closest to mean structure with residues that show chemical shift changes upon M4-N binding coloured according to position: Yellow – close to the interface (the chemical shift of these could be influenced by intermodular reorientation or direct contact); cyan – residues away from the interface likely to be in direct contact with M4-N. The aromatic side chains of Tyr37, Tyr62 and Phe84 are shown in space-fill representation.

5.9 Interpretation of the mutagenesis data for M4-N binding

In earlier studies, K63Q – a substitution that could not be accommodated without perturbing the intermodular interface observed in the structure – had little effect on M4 binding (Blom *et al.* 2000). When considered in conjunction with the Ala-insertion experiments, this implies that while the four-residue length of the linker is critical for M4 binding, the interface composition, and hence the relative arrangement and flexibility of the two modules in the free protein, is not. This conclusion is consistent with a change of modular orientation upon binding, i.e. the bound conformation being different from the free one.

The mutants R64Q and H67Q, each displayed reduced M4-binding – on the other hand R66Q and K79Q have increased affinity. All four of these mutated residues show significant chemical shift changes upon addition of M4-N. From the structure it is apparent that none of these mutations involve a residue that participates in the interface, rather Arg^{64} and His^{67} form a potential M4-binding patch on the surface of CCP2. This finding is supported by the observation that bovine, murine and rat C4BP lack one or both of Arg^{64} and His^{67} and are indeed unable to bind M protein (Accardo *et al.* 1996). Arg^{66} and Lys^{79} flank His^{67}, and the *gain* of affinity resulting from the R66Q and K79Q substitutions could be explained if the native residues impede access to the M4 binding site either because of their positive charge or by virtue of their steric bulk. The lack of dependence on salt and pH suggest that binding of M4 requires other forces apart from electrostatics (Blom *et al.* 2000). In general, the pattern of effects on M4 binding that result from mutagenesis of charged residues is consistent with an interaction in which electrostatics steer the two components towards a productive interaction, but other forces are critical to stabilise the complex eventually formed. Thus while some mutations inhibit binding, others increase it, and some mutations compensate for one another. It is unsurprising that selected substitutions in C4BP improve binding - the interaction is unlikely to be optimal in terms of affinity since the two partners

are under opposing evolutionary influences in this respect. Furthermore, the interaction is not specific to M4 - other M proteins with different hypervariable domains also bind to C4BP.

5.10 Conclusions

The module deletion and ITC studies showed that CCP1 and CCP2 of the C4BP α-chain and the linker residues between them are necessary and sufficient for M4 binding. The chemical shift changes observed upon M4-N binding to C4BP12 suggest that the binding site is close to the CCP1/2 interface and that intermodular re-orientation occurs upon binding. This is discussed with respect to C4b- and glycosaminoglycan-binding in Chapter 7.

Chapter 6

DNA BINDING

6.1 Introduction

C4BP binds to double stranded 25-bp DNA with a K_d of 190 nm (Trouw *et al.* 2005). This interaction is thought to localise C4BP to areas of necrosis and apoptosis where is can act to down-regulate the classical pathway of complement activation. This ensures that the late stages of complement - i.e. C5a release and MAC assembly - do not occur, thus preventing inflammation and tissue damage. Studies using deletion mutants have localised the binding site to CCP1 and CCP2 of the α-chain. Having solved the structure of this region of C4BP, the binding site for DNA could be mapped onto the structure. This was achieved using dsDNA in a NMR titration with C4BP12 to map NH chemical shift changes on addition of DNA. As the structure of both interacting components is known, data from the titration could be used to produce a model of the C4BP12:DNA complex by data-driven docking using the HADDOCK programme (Dominguez *et al.* 2003).

6.2 DNA used

The SPR studies performed by Trouw *et al.* used double stranded 25-bp DNA. For the initial NMR titration two complementary 20mer oligonucleotides with a 'randomly' selected sequence (chosen to have an equal number of A, T, G and C bases, and no self-complementarily to avoid secondary structure formation) were purchased from Sigma-Genosys and annealed as described in Section 8.3. In an initial titration with u[^{15}N]-C4BP12 considerable precipitation of the C4BP12-DNA complex occurred (as judged by the lack of ^{15}N-labelled material present in the sample after the sample was spun in a microfuge to remove the precipitate). The precipitation was presumed to result from multiple molecules of C4BP12 binding to one dsDNA molecule. Two shorter 10-bp oligos were purchased, and annealed and used for subsequent titrations. The 10-bp dsDNA caused much less precipitation; and any precipitate formed on addition of 10-bp DNA to C4BP12 could be redissolved by warming (to 37 °C) and stirring. The sequences used are shown in Table 6.1.

20mer 5'	5'- AAT ATA TGG CGC GCT TAC CG -3'
20mer 3'	3'- TTA TAT ACC GCG CGA ATG GC -5'
10mer 5'	5'- AAT CGC GCT T -3'
10mer 3'	3'- TTA GCG CGA A -3'

Table 6.1 DNA sequences used for titration with C4BP12

6.3 Titration of 10-bp dsDNA with C4BP12

An overlay of five ^{15}N-HSQC spectra recorded at increasing ratios of DNA:C4BP12 is shown in Figure 6.1. It is clear that the system is in fast exchange and the cross-peaks track with increasing concentration of DNA. Unlike the situation with M4-N, all the perturbed cross-peaks could be tracked and the chemical shift changes for each residue in C4BP12 (excluding the eight proline residues and Asn^{1}, His^{41} and Phe^{59} that could not be assigned at pH 6.0) could be measured. The combined chemical shift changes are calculated as $\sqrt{((\Delta\delta^{N}H)^2+(\Delta\delta^{15}N/5)^2)}$ and are plotted per residue in Figure 6.2.

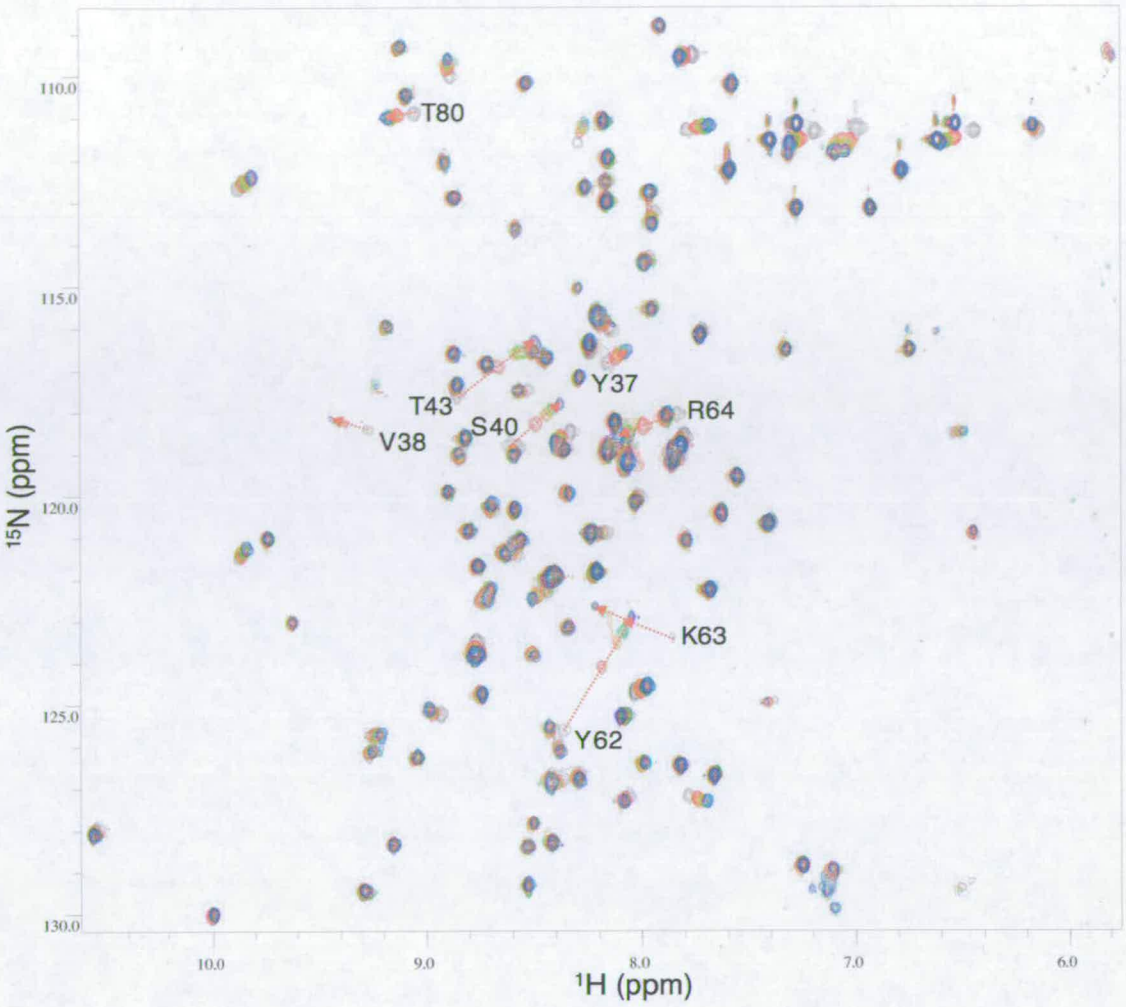

Figure 6.1 Overlay of five ^{15}N-HSQC spectra of 50 μM C4BP12 with increasing concentrations of DNA. Colours: black - 0 μM, red - 25 μM, orange - 50 μM, green - 100 μM, blue - 200 μM 10-bp dsDNA. Residues with combined chemical shift difference of > 0.1 ppm are labelled.

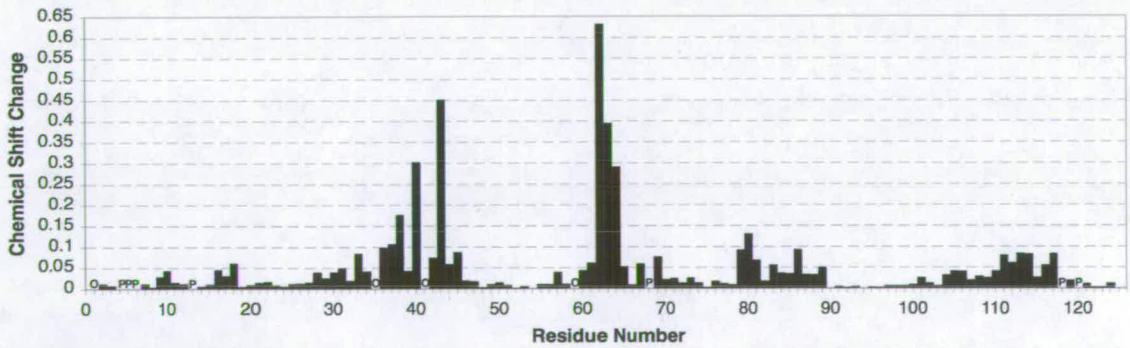

Figure 6.2 Combined chemical shift changes [$\sqrt{((\Delta\delta^{15}N/5)^2+(\Delta\delta^{1}H)^2)}$] per residue at 4:1 [DNA]:[C4BP12]. P indicates proline residues and O illustrates missing data.

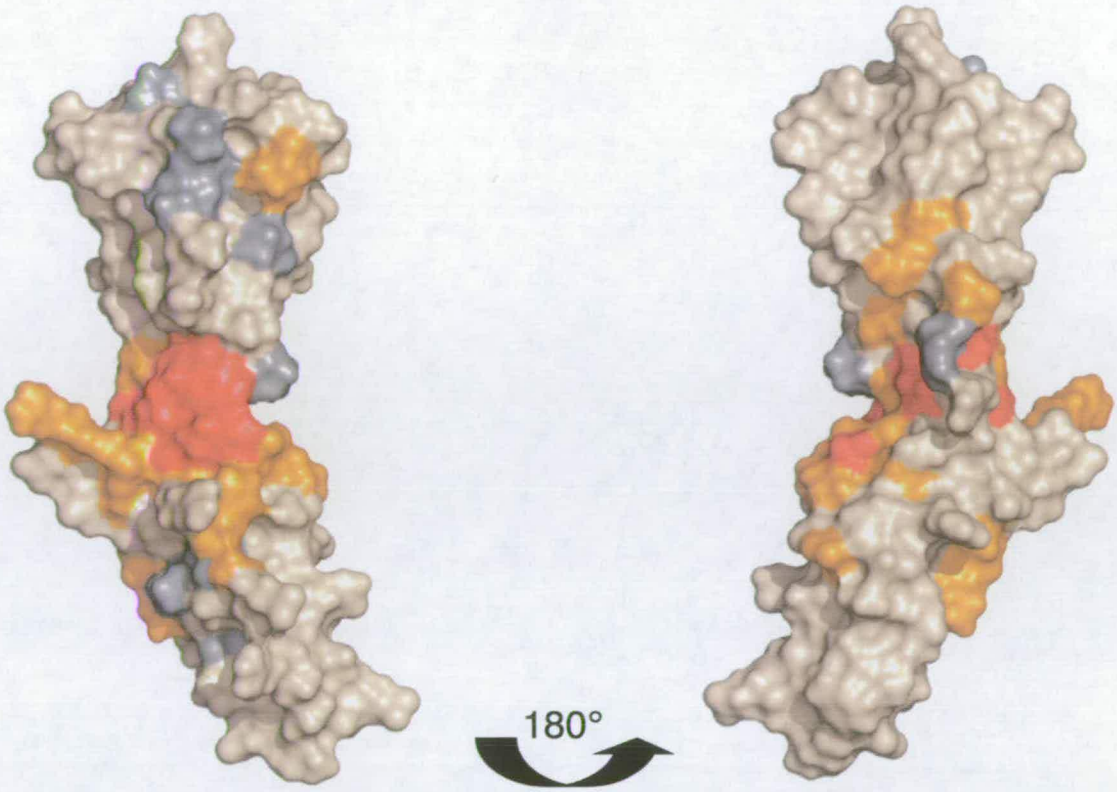

Figure 6.3 Surface of the closest to mean C4BP12 structure with residues showing chemical shift changes upon DNA binding coloured: red - combined chemical shift difference > 0.1 ppm; orange - chemical shift difference > 0.05, < 0.1 ppm. Residues for which no information could be obtained are coloured grey

6.4 Mapping shift changes onto the C4BP12 structure

The chemical shift changes at 4:1 [DNA]:[C4BP12] are shown on the structure of C4BP12 in Figure 6.3. The largest chemical shift changes occur at the CCP1/2 interface and all the shift changes are located proximal to the interface. The number and extent of shift changes is not as great as for the M4-N interaction. This suggests that the intermodular reorientation proposed to occur upon M4-N binding does not feature in DNA binding. Therefore, the localisation of shift changes at the CCP1/2 interface indicates that DNA binds at the interface between CCP1 and CCP2.

6.5 Using HADDOCK to generate the C4BP12:DNA complex

The programme HADDOCK (High Ambiguity Driven protein-protein Docking) was developed to produce the structure of a complex by docking the proteins (or protein and DNA) involved by making use of biochemical and/or biophysical information. Such data includes chemical shift perturbations obtained from NMR titration experiments or the results of mutagenesis experiments and is used to guide the two interacting partners together in the correct orientation. In HADDOCK the information regarding residues that interact is introduced as ambiguous interaction restraints (AIRs) to drive the docking. The AIRs are divided into two classes- 'active' and 'passive'. For NMR titration data, 'active' residues are defined as those residues that show a significant chemical shift perturbation upon complex formation as well as a high solvent accessibility in the free form. The 'passive' residues are defined as residues that show smaller chemical shift perturbations and/or that are surface neighbours of the active residues and have a high solvent accessibility. In the case of mutagenesis data, the 'active' residues are those that when mutated prevent complex formation and are also solvent-exposed.

6.5.1 Ambiguous interaction restraints

For the C4BP12:DNA docking the 'active' residues were defined as the eight residues that showed combined chemical shift changes on DNA binding of > 0.1 ppm and, in addition, residues Arg^{39} and Arg^{66} that were shown to be critical for DNA binding by mutagenesis but that did not show large shift changes on DNA binding. The 'passive' residues were solvent-exposed residues surrounding the active residues that showed smaller shift changes on DNA binding and also two residues that were missing from the shift mapping data - Pro^{35} and His^{41}. For the DNA, all the bases were defined as 'active' as there was no experimental information to define the location of C4BP12 on the DNA duplex. These residues are shown on the structure in Figure 6.4.

'Active' residues	Tyr[37], Val[38], **Arg[39]**, Ser[40], Thr[43], Tyr[62], Lys[63], Arg[64], **Arg[66]**, Thr[80]
'Passive' residues	Pro[35], Gly[36], His[41], Ser[42], Ile[61], Ser[83], Ser[86]

Table 6.2 The active and passive residues defined for the docking with HADDOCK. Residues highlighted in bold were implicated by mutagenesis but did not show shift changes upon DNA binding.

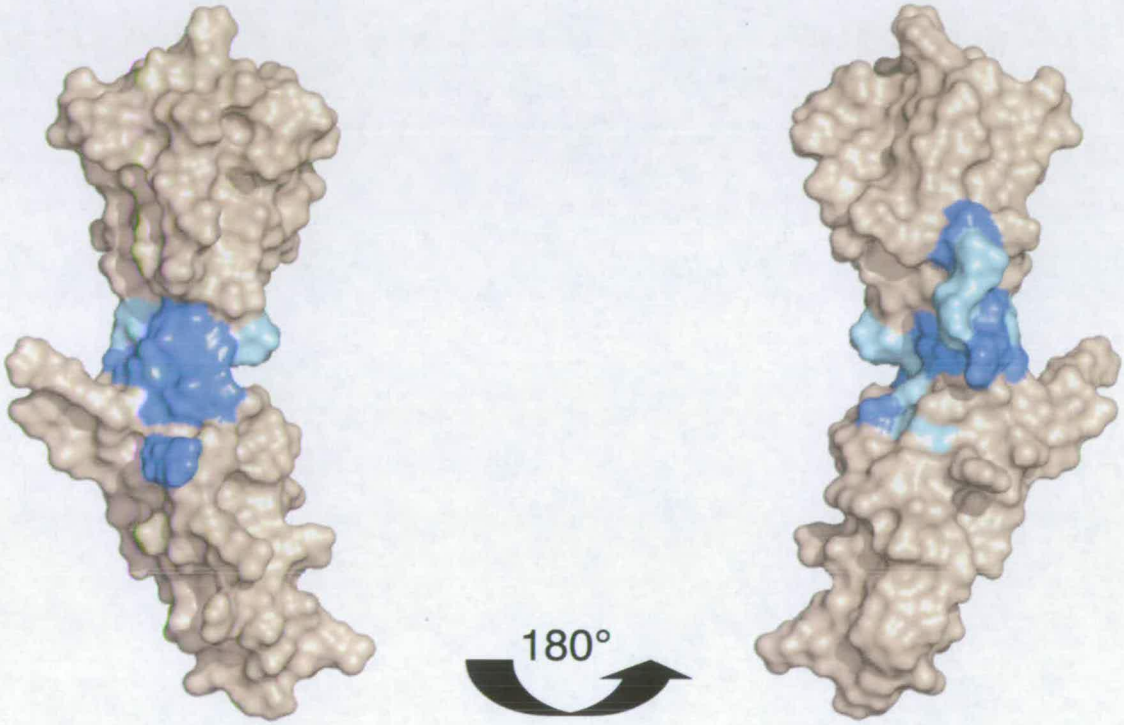

Figure 6.4 The active (blue) and passive (cyan) residues used in the C4BP12-DNA docking

6.5.2 Structures used

For C4BP12 all forty structures in the ensemble were used as input for HADDOCK. For DNA the NMR structure of the 12mer of double-stranded DNA with sequence 5'-GGCAAAAAACGG-3' (PDB id = 1FZX) was truncated to 5'-GGCAAAAAAC-3' prior to use in the HADDOCK docking. This structure serves only as a template as the DNA conformation is restrained to that of canonical B-form DNA in the starting structures for the rigid-body docking There is no reported sequence specificity for DNA binding by C4BP12

so the differences in sequence between the DNA used in the titration and the DNA used for docking is not considered likely to cause problems.

6.5.3 Flexible regions

The HADDOCK docking protocol consists of four stages: (i) a high temperature rigid body search, (ii) rigid body simulated annealing (SA), (iii) semi-flexible SA with flexible side-chains at the interface, (iv) semi-flexible SA with fully flexible interface (both backbone and side-chains). The residues at the intermolecular interface that were defined as semi-flexible and hence allowed to move in steps (iii) and (iv) of the C4BP12-DNA docking protocol were defined as those stretches of residues that showed chemical shift changes upon DNA binding (Table 6.3 and see Figure 6.2).

Segment 1	Thr^{32}-Thr^{45}
Segment 2	Cys^{60}-Cys^{65}
Segment 3	Ile^{78}-Gln^{87}
Segment 4	Asp^{110}-Trp^{115}

Table 6.3 Semi-flexible residues defined in the C4BP12:DNA docking.

In addition residues Gly^{36}-His^{41} and Lys^{79}-Asp^{81} were defined as fully flexible and allowed to move in all stages of the docking protocol because these are the two loops close to the intermodular interface containing active residues which are not part of the linker between CCP1 and CCP2.

6.5.4 The docking

1000 complex structures were calculated in the rigid docking stage. The 100 lowest energy structures from this stage were used in the semi-flexible refinement stages and then refined in a water shell to produce the final ensemble of complex structures. The energy term used to

rank the structures is the intermolecular energy which is the sum of the electrostatic, van der Waals, and AIR energy terms. The structures were analysed using `cluster_struc` (distributed with HADDOCK). This programme uses the pairwise-rmsd matrix of all the structures calculated to produce a cluster of structures whose rmsd to each other is less than a specified cut-off value. A rmsd cut-off of 1.2 Å resulted in a single cluster containing five structures (Figure 6.5). These structures are shown in (Figure 6.6).

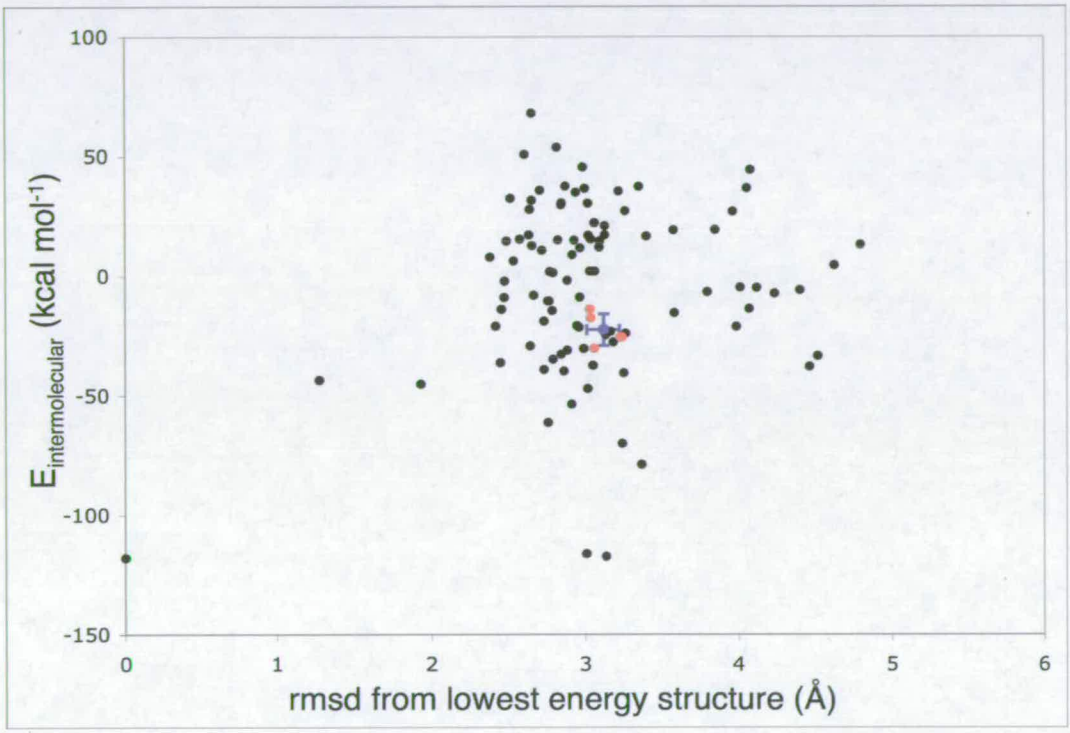

Figure 6.5 Plot of intermolecular energy vs backbone rmsd from the lowest-energy structure for the 100 water refined structures of the C4BP12-DNA complex produced by HADDOCK. The members of the cluster described in the text are coloured red. The mean intermolecular energy and rmsd of this cluster from the lowest-energy structure is shown by the blue point with the standard deviation shown by the error bars.

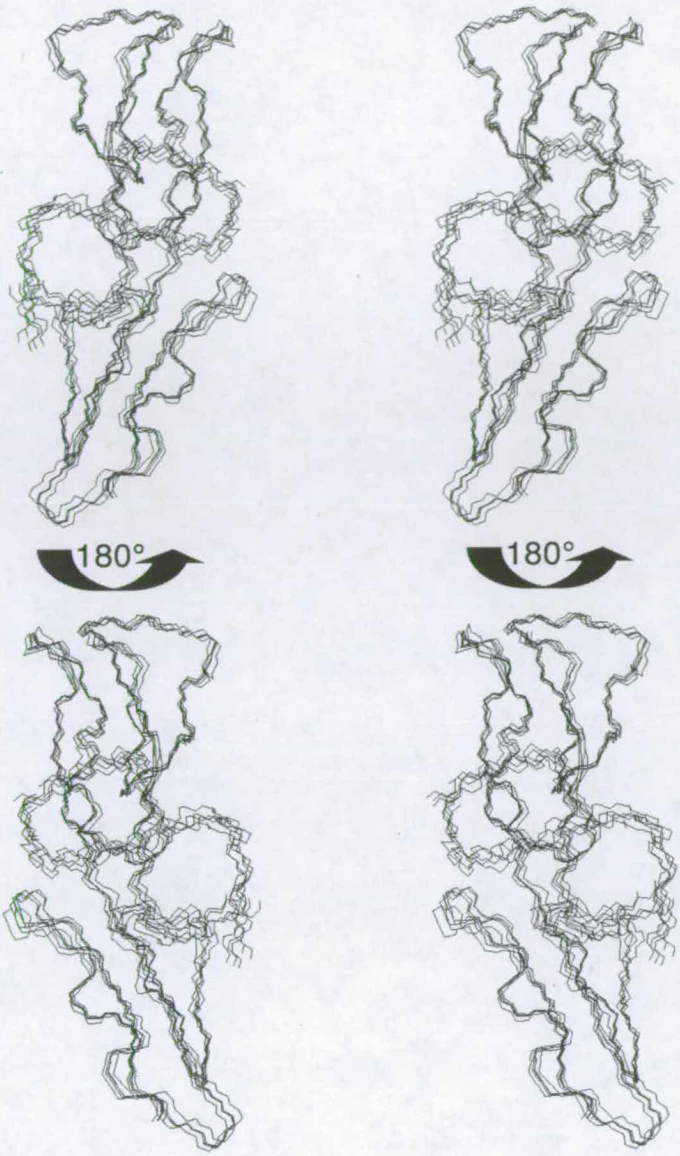

Figure 6.6 Two stereo views (cross-eyed) of the cluster of five structures of the C4BP12-DNA complex produced by HADDOCK. The backbone of residues 2-122 of C4BP12 and the 10bp DNA duplex is shown. Structures were overlaid on the backbone of residues 2-122.

6.5.5 The complex

The lowest-energy structure of the cluster is assumed to be the 'best' structure of the complex and this is shown in (Figure 6.7). This complex has a buried surface area of 1730 $Å^2$ (calculated by taking the difference between the sum of the solvent accessible surface area (calculated using a 1.4 Å water probe radius) for each molecule separately and

the solvent accessible area of the complex. The DNA is bound at the CCP1/2 interface with residues from both CCP1 and CCP2 contacting the DNA. The contacts are predominantly electrostatic interactions between positively charged side-chains and the negatively charged phosphates in the backbone of the DNA duplex.

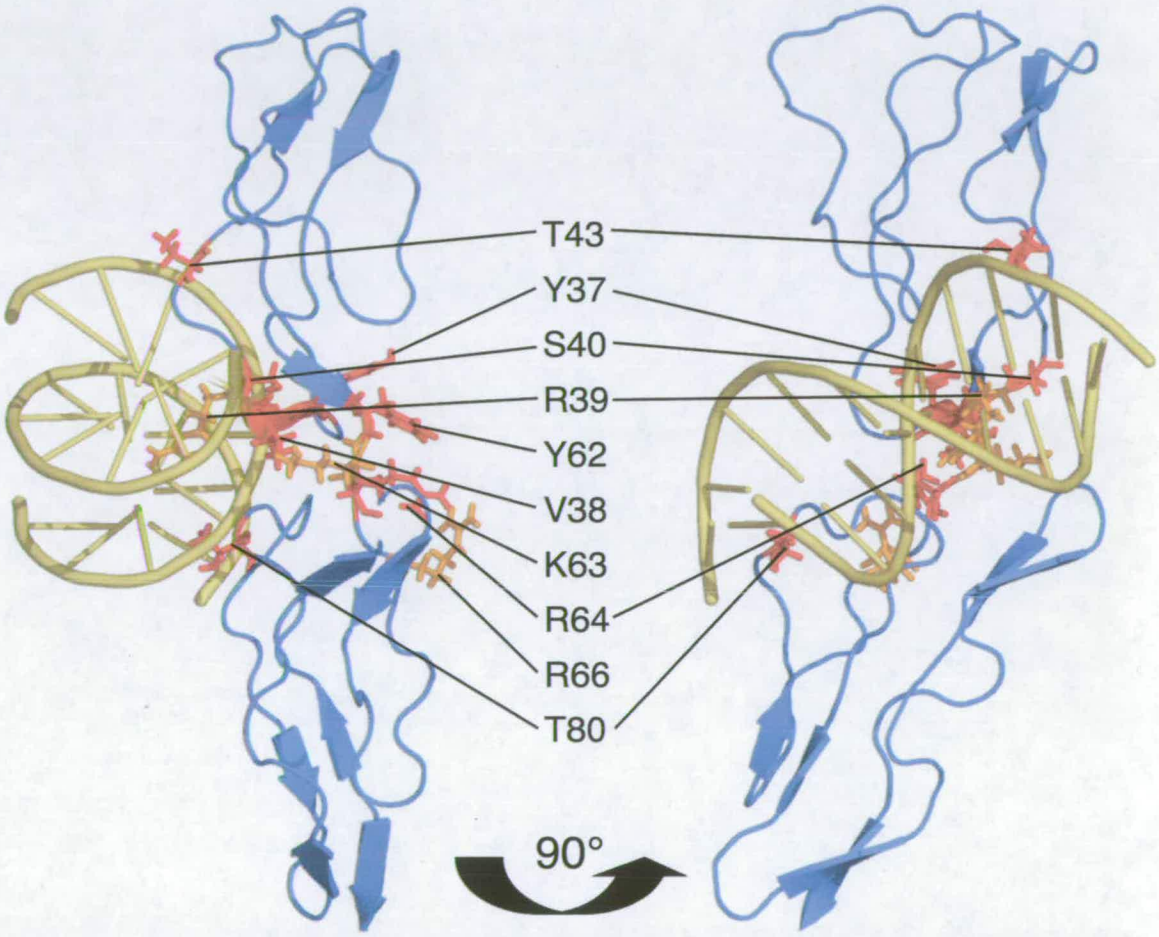

Figure 6.7 Two views of the lowest energy structure in the cluster. The active residues used in the the docking are shown in stick representation and coloured red. Residues that were shown to be critical by mutagenesis (Arg^{39}, Lys^{63} and Arg^{66}) are coloured orange.

6.6 Interpretation of the mutagenesis data for DNA binding

In the study by Trouw *et al.* (2003), DNA gel shift analysis with C4BP mutants showed that the positively charged residue Arg^{39} in CCP1, Lys^{63} in the linker and Arg^{66} in CCP2 were critical for DNA binding whereas Arg^{64} was of intermediate importance and His^{67} did not

seem to be required. On the other hand, in the chemical shift mapping study, the NH cross-peaks of Arg^{39} and Arg^{66} did not show significant shift changes upon DNA binding. Lys^{63} and Arg^{64} showed significant shift changes (combined chemical shift change > 0.1 ppm) and were included as 'active' residues in the data-driven docking. His^{67} shifted an intermediate amount (0.06ppm). These results are summarised in Table 6.4.

Residue	Effect of mutagenesis	Chemical shift change	Position in structure
Arg^{39}	destroys binding	small (<0.05 ppm)	side-chain contacts DNA
Lys^{63}	destroys binding	large (0.39 ppm)	side-chain contacts DNA
Arg^{64}	intermediate binding	large (0.29 ppm)	side-chain away from DNA
Arg^{66}	destroys binding	small (<0.05 ppm)	far from DNA
His^{67}	no effect	intermediate (0.06 ppm)	far from DNA

Table 6.4 Residues changed in mutagenesis study: effect on DNA binding, (combined) chemical shift changes upon DNA binding, position in HADDOCK complex.

In the HADDOCK C4BP12:DNA complex the side-chains of both Arg^{39} and Lys^{63} contact DNA, and it is possible that both are directly involved in DNA binding. The side-chain of Arg^{64}, however, points away from the DNA and in this case the amide shift changes observed upon DNA binding likely reflect a slight change in the structure of the intermodular interface rather than resulting directly from the interaction of this residue with DNA. Indeed, the structure of C4BP12 suggests that Arg^{64} is critical to the stability of the CCP1/2 interface (Section 4.6.2), so its mutation would disrupt the interface and thus likely reduce, but not eliminate, the ability of C4BP to bind DNA. This hypothesis is supported by the mutagenesis data indicating an intermediate role for Arg^{64}. The mutagenesis results for Arg^{66} and His^{67} are interesting since mutation of these two consecutive, charged, residues has such different effects. As Arg^{66} does not shift upon DNA binding, and it is located far from the DNA in the HADDOCK complex, its role is likely to be in steering the two components towards a productive interaction rather than in contacting the DNA in the final complex. The small chemical shift change of the NH cross-peak of His^{67} may arise from a

slight rearrangement of the backbone of CCP2 upon DNA binding at the interface. The absence of shift changes for Arg^{39} upon DNA binding is, however, not consistent with this residue contacting DNA in the complex. There are no DNA-induced shift changes seen for the H^{β}, $H^{\gamma 1}$, $H^{\gamma 2}$ or $H^{\delta *}$ cross-peaks of Arg^{39} in the ^{13}C-HSQC spectra (data not shown) yet the shift changes of residues around Arg^{39} and the mutagenesis data implicate this residue in binding. It is, therefore, possible that the role of Arg^{39} is again in electrostatic steering of the DNA and that the contacts with DNA seen in the HADDOCK model are an artefact of erroneously defining this residue as 'active' - based on the mutagenesis data - in the docking procedure.

6.7 Conclusions

NH chemical shift-mapping was used to locate the DNA binding site on C4BP12 to the CCP1/2 interface. This is consistent with module-deletion studies and mutagenesis data. The shift-mapping data was used in conjunction with the results of the mutagenesis study in data-driven docking with HADDOCK to generate a model of the C4BP12:DNA complex. The existing mutagenesis data was then interpreted in the light of this model. The role of Lys^{63} in directly binding DNA seems clear. Arg^{39} and Arg^{66}, which are both critical for DNA binding, are hypothesised to be involved in steering the DNA, via electrostatic interactions, towards the correct binding site, rather than being involved in direct contact with DNA in the final complex.

Chapter 7

DISCUSSION

7.1 Overview

The aims of this thesis were to:

1. Use NMR spectroscopy to solve the 3D structure of C4BP12.
2. Interpret the mutagenesis data gathered to date in light of this 3D structure.
3. Investigate the binding of the M4 HVR from *S. pyogenes* to C4BP12 using NMR spectroscopy and other biophysical techniques.
4. Map the binding site for DNA onto the C4BP12 structure.

Previous chapters of this thesis describe the determination of the 3D structure of C4BP12. This has allowed delineation of the structural basis for the interactions of C4BP with the M proteins of the pathogenic bacterium *S. pyogenes*, and with host DNA.

In this chapter the limitations of the methods used will be critically assessed. The implications of the structural studies for the function of C4BP will be considered, including its affinity for binding to C4b and cell-surface glycosaminoglycans. Future directions for study of the structure and function of C4BP will be discussed.

7.2 Methods used for calculation of the structure of C4BP12

The structure of C4BP12 was solved on the basis of NMR-derived structural restraints using MDSA protocols based on the work of Prof. Michael Nilges and others (Nilges *et al.* 1988a; Nilges *et al.* 1988b; Nilges 1995). The calculated structures showed a good level of convergence, and high precision (i.e. a low rmsd for the ensemble of 40 low-energy structures). The quality scores were deemed acceptable for a small protein domain with relatively little secondary structure and some disordered loops. However, the improvement in quality scores obtained following water refinement suggests that the protocols used to calculate the structures could be improved. In particular, the strategy for dealing with the effects of spin diffusion and local dynamics on the intensity of NOE cross-peaks, which entails the use of large error bounds on the distance restraints, discards much of the information content of the NOE. It was shown recently that the use in structure calculation of a potential for NOE-derived distance restraints that reflects the error in calculating these distances improves the quality of the calculated structures (Nilges *et al.* 2006). This potential has the form of a log normal distribution, rather than the previously used flat-bottomed harmonic well (FBHW) potential with wide upper and lower bounds. A further advantage of this potential is that the precision of the structures automatically decreases with the quality of the data, making the rmsd a more meaningful number. With a FBHW potential the rmsd is heavily influenced by the width of the bounds (Chalaoux *et al.* 1999) and is, therefore, not an unbiased measure of structural quality.

Recently, a new method of structure calculation has been described (Rieping *et al.* 2005). This method is based on a probability distribution that represents an unknown structure and its precision, derived using Bayesian inference, and simulated using Markov chain[1] Monte Carlo techniques and is known as inferential structure determination (ISD). ISD circumvents

[1] A Markov chain is a collection of random variables (describing the state of a system) with the property that, given the present, the future is conditionally independent of the past.

problems in conventional structure calculation protocols such as empirical choice of the parameters because these are determined directly from the probability distribution. However, at present this technique is too computationally expensive for routine use.

7.3 Intermodular orientation of C4BP12

The extent to which the intermodular orientation in the calculated structures reflect reality is a key issue. The ensemble of structures presented in Chapter 3 were calculated using a single alignment tensor. This would be expected to produce structures with a well-defined intermodular orientation. It is important to question whether this could be an artefact. There was good agreement between the alignment tensors produced either when separate alignment tensors for each module were employed, or when a single alignment tensor for the whole molecule was used. This indicates that the approximation that the whole molecule experiences the same alignment tensor is valid (Ball 2005).

RDCs (by definition) can only be measured from aligned molecules in the sample. If only a subset of the conformations of the protein in solution can align, the conformational space sampled in the RDC measurements, and hence the apparent flexibility of the structures refined using these RDCs will be reduced. Interactions between the protein and the alignment media could also influence artefactually the conformation of the protein. Therefore, it is possible that the intermodular orientations in the RDC-refined structures presented in Chapter 3 are overly tightly defined and do not reflect the full range of conformations sampled by C4BP12 in solution.

To investigate this possibility, structures were recalculated using identical protocols and restraints to those employed to produce the RDC-refined structures but with the TENSO term removed from the energy function. The intermodular angles amongst these recalculated

structures are compared with the RDC-structures in Figure 7.1. This comparison suggests that the spread of intermodular angles is indeed wider in the NOE-only structures than when RDC restraints are used in refinement. However, it is important to remember that the spread of structures reflects only the number and strength of intermodular (and module-linker) restraints. Although less NOEs might be expected if the interface were flexible due to averaging effects and local dynamics, the spread of orientations in the structure is not, of course, a reliable indication of intermodular flexibility.

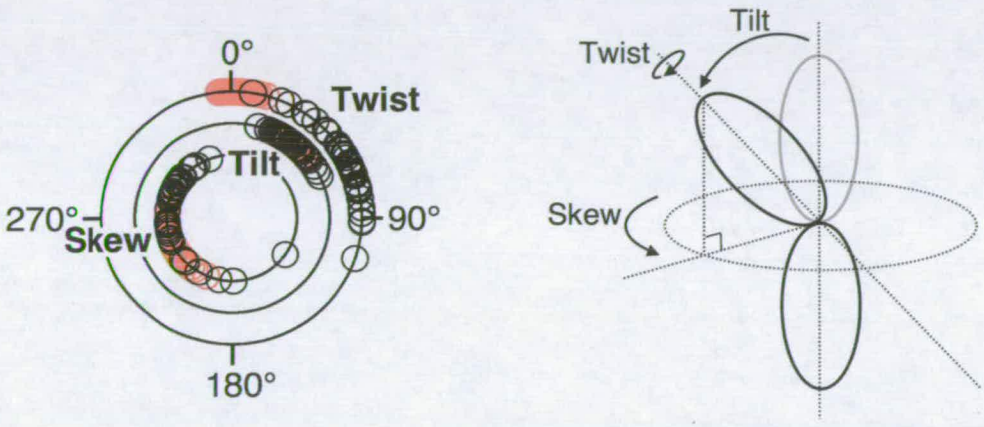

Figure 7.1 Comparison of intermodular twist, tilt and skew angles from the ensemble of 40 RDC-refined structures (red circles) and the ensemble of 42 structures calculated using only NOE-derived restraints. The NOE and dihedral angle restraints used were identical.

It is slightly concerning, however, that mean twist angle in the RDC-refined structures lies just outside the spread of values for the structures calculated using only NOE restraints. This difference in inter-modular orientations may reflect genuine differences in conformation between aligned and non-aligned media; it could account for some of the distance-restraint violations in the ensemble of RDC-refined structures that are not present in the NOE-only structures. If, as proposed in Section 7.4, the intermodular orientation of C4BP12 is able to change on ligand binding, it is conceivable that the twist angle could be different under different sets of conditions. The negatively charged surface of the phage used to align C4BP12 could well have a weak interaction with the protein that may alter the intermodular orientation. RDC data acquired from C4BP12 aligned in a different medium would provide

an independent check on the intermodular orientation presented in Chapter 3. Also comparison of experimental RDC values with values calculated from an alignment tensor based on the structure of C4BP12 using PALES (Zweckstetter *et al.* 2000; Zweckstetter *et al.* 2004) would indicate if all the conformations present in non-aligned media can align. For a protein interconverting rapidly between multiple conformations, the measured dipolar couplings must agree with the ensemble-weighted average of couplings predicted from these conformations.

The use of data from small-angle X-ray scattering (SAXS) in refinement of multi-domain protein structures has been recently reported (Grishaev *et al.* 2005). It is possible that this technique could be used to refine the C4BP12 structures presented in Chapter3. The most striking difference in the intermodular orientation of C4BP12 compared to DAF23 and CR1-1516 is the intermodular twist angle, and the SAXS data might not be of high enough resolution to distinguish between different values of this angle as the overall shape of the module pair would remain similar.

7.4 Intermodular re-orientation of C4BP12

The experimental structure together with the chemical shift perturbation data for M4 binding indicate that binding involves an intermodular reorientation of C4BP12. This is the first experimental evidence in support of intermodular flexibility being important for the ligand-binding properties of a mammalian RCA protein. The viral RCA, VCP, undergoes a conformational change upon heparin binding (Ganesh *et al.* 2004).

In the light of this information, it seems likely that such conformational flexibility could also be important for C4BP binding to the natural ligands, C4b and glycosaminoglycans. Indeed, residues implicated as critical for C4b and heparin binding by mutagenesis - Arg^{39}, Lys^{63},

Arg^{64} and His^{67} - are not clustered together in the structure of the free protein. These residues would be brought into juxtaposition by a twist between the modules, creating an interaction that is dominated by electrostatic forces. A similar hypothesis was suggested in the case of DAF23 as an explanation for the pattern of functionally critical mutants within the solution structure (Uhrinova *et al.* 2003). Moreover, the lack of consistency amongst intermodular angles within C3b/C4b-binding regions of the free RCAs (Section 4.6) could be explained if each RCA undergoes a conformational change upon binding, perhaps converging on a common C3b/C4b-bound orientation that is necessary for cofactor and/or decay-accelerating activity.

An interesting possibility is that a conformational change is induced by glycosaminoglycan binding - analogous to that observed for VCP (Ganesh *et al.* 2004) - which in turn might exert a positive effect on the regulatory activity of C4BP.

7.5 The C4BP12:M4-N complex

7.5.1 Spectra of the complex

The quality of the spectra of the C4BP12:M4-N complex recorded at low (20 μM) concentrations of C4BP12 was good. However, the complex did not behave well at higher concentrations. The ^{1}H-^{15}N HSQC spectra of the complex at 0.6 mM (Figure 7.2) were poor with most peaks broadened compared to the spectra recorded with a C4BP12 concentration of 20 μM.

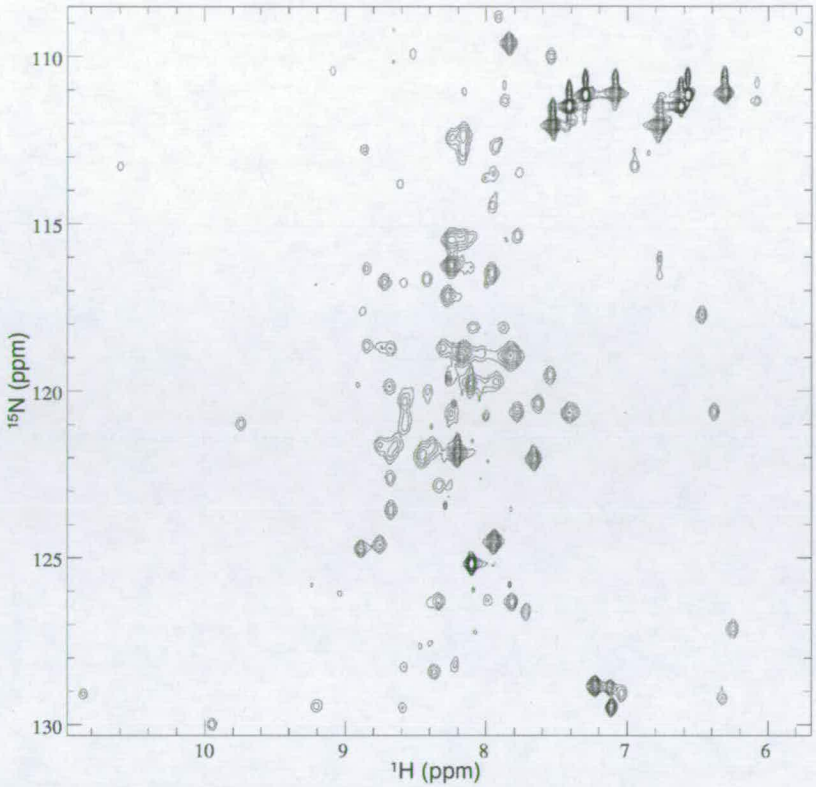

Figure 7.2 ^{1}H-^{15}N HSQC spectrum of 0.6 mM C4BP12:M4-N complex, 20 mM sodium phosphate buffer, pH 6.0, 37 °C.

An overlay of planes from the 3D HNCO spectrum of the C4BP12:M4-N complex is shown in Figure 7.3. The only cross-peaks that are present are from the N-terminus (Asn^{1}), the C-terminus (Glu^{126}, and the His_6 tag) and residues in the flexible (see section 4.6) hypervariable loop (Asp^{15}, Ile^{16}, Thr^{17}, Leu^{18}, Thr^{19} and Glu^{20}). The other cross-peaks in Figure 7.3 could not be assigned. The CBCA(CO)NH spectrum of the complex contained only cross-peaks from the C-terminal His_6-tag. The HNCO (the most sensitive of the triple-resonance experiments) spectra shown in Figure 7.3 only yielded cross-peaks from flexible residues that showed no or small shift changes upon M4-binding. This poor spectral quality was attributed to chemical exchange (the K_d for the interaction measured by ITC was 0.5 μM) and the poor solubility of the complex.

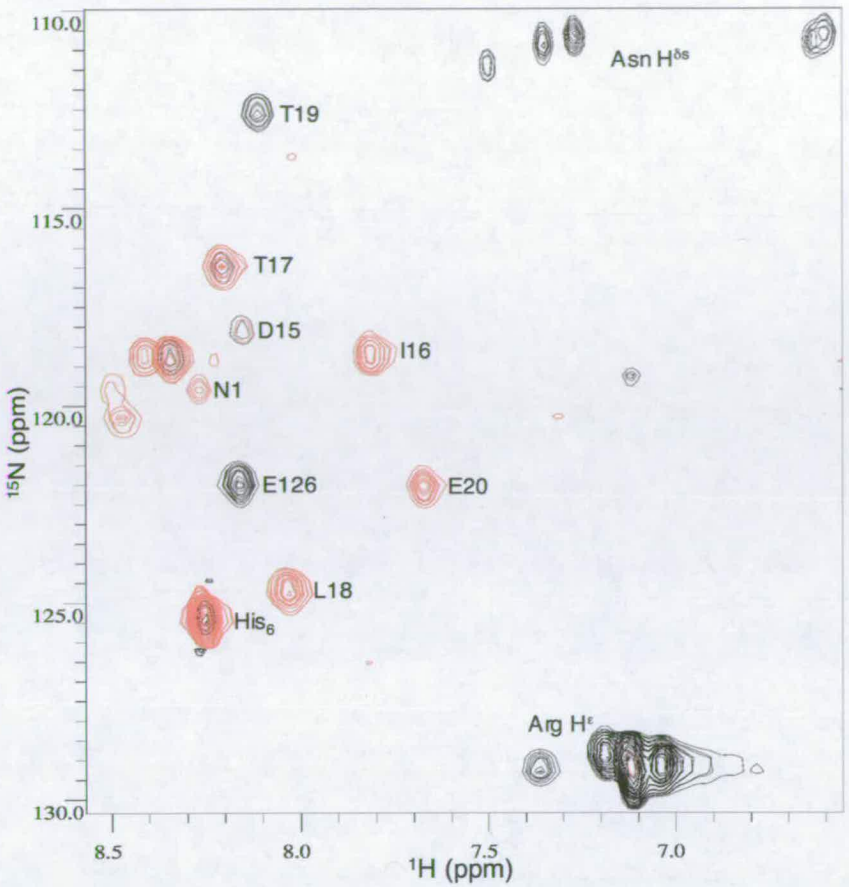

Figure 7.3 Overlay of ^{15}N-^{1}H planes of 3D HNCO spectrum recorded on 0.5 mM C4BP12:M4-N complex, 20 mM sodium phosphate/acetate mixture, pH 4.5, 37 °C.

Thus assignments of C4BP12 in the complex could not be made. Therefore, chemical shift changes for the cross-peaks in slow exchange could not be quantified. This limited the mapping studies described in Chapter 5. Perturbations of resonances in fast exchange could be tracked during the titration, while resonances in slow exchange were simply flagged as undergoing large chemical shift changes. Other methods that could be used to study the C4BP12:M4-N interaction are discussed in Sections 7.5.2 - 7.5.4.

7.5.2 Mapping with paramagnetism

Paramagnetic species have permanent magnetic moments (dipoles), even in the absence of an applied field. This generally occurs due to the presence of unpaired electrons. Attachment

of a paramagnetic tag to M4-N and the resulting paramagnetic relaxation enhancement (PRE) induced line broadening (Petros *et al.* 1990) on the ^{15}N-HSQC spectrum of C4BP12 in the C4BP12:M4-N complex would provide further information about the position of M4-N in the complex. In an alternative approach (Deschamps *et al.* 2005), nitroxide radicals could be added to solutions of both free C4BP12 and of the C4BP12:M4-N complex. The residues on the surface of the C4BP12 involved in M4-N would be protected in the C4BP12:M4-N complex from non-specific weak protein-radical interactions and thus changed PRE effects would be seen between free and complexed C4BP12. These residue specific differential PRE effects (ΔPRE) would provide complementary information to the chemical shift perturbation data. Both approaches could potentially help to discriminate between residues that are in direct contact with M4-N and residues that show NH chemical shift-perturbations resulting from intermodular re-orientation on M4-N binding.

7.5.3 Mapping with pseudocontact shifts

Attachment of a paramagnetic tag to M4-N could also be used to map the C4BP12:M4-N interface through pseudocontact shifts (Allegrozzi *et al.* 2000). The basis of these shifts is as follows: If the electronic magnetic dipole of the spin label does not stay aligned with the external magnetic field when the complex tumbles in solution then the electron-nuclear dipolar interaction will not average to zero. The residual electron-nuclear coupling that remains will result in shifts of the NMR resonances - known as pseudocontact shifts. These pseudocontact shift contain information on both distance (with a r^{-3} dependence) and orientation of the paramagnetic centre on the ligand with respect to the reporting protein. Attachment of EDTA-based lanthanide binding tags can also be used to generate weak alignment (Ikegami *et al.* 2004) enabling RDCs to be measured from the complex.

7.5.4 Mapping with cross saturation

Cross saturation (Takahashi *et al.* 2000) could also be used to map the contact site for M4-N onto C4BP12. This has the advantage over chemical shift peturbation based mapping that only protons close to the interface will be affected by saturation transfer from the donor to the acceptor molecule and thus decrease in intensity even if long range conformational changes occur. The use of this method would, however, require perdeuteration of the target protein (i.e C4BP12).

7.5.5 Structural studies of M4

The recent work on the structure of M4-N, both free and in complex with C4BP12 (André *et al.* 2006), suggested that residues Asn^{19}-Ser^{37} were in an α-helical conformation. This conclusion is based on secondary $^{13}C^{\alpha}$ and ^{13}CO chemical shifts and $\{^{1}H\}$-^{15}N NOE measurements. André *et al.* were unable to assign residues Asn^{13}-Asn^{19} due to missing cross-peaks. This region is nevertheless shown as structured in the model of M4-N (André *et al.* 2006) which was produced using SIR4 from *Saccharomyces Cerevisiae* (Murphy *et al.* 2003: PDB id = 1PL5) as a template. Residues Ala^{1}-Trp^{12} and Ala^{39}-Cys^{46} are not structured based on the random coil values of their $^{13}C^{\alpha}$ and ^{13}CO chemical shifts and the $\{^{1}H\}$-^{15}N NOE data. Ala^{1}-Trp^{12} are clearly structured in the model produced but this region of the model was not compared to the experimental data. Agreement between experimental and calculated values (from the model) for $^{1}D_{NH}$ dipolar couplings and R_2/R_1 ratios was suggested to confirm the coiled-coil model produced, although the validity of using a model that clearly differs from experimental data in certain parts is questionable. Also the very limited set of R_2/R_1 values and RDCs that could be measured means that any agreement between the experimental and calculated values is highly dependent on the errors in measuring the experimental values.

As cross-peaks for all 27 residues that were proposed to form the coiled-coil region in free M4-N disappeared on addition of C4BP12 (André *et al.* 2006) the C4BP-binding region of M4-N was proposed to be up to 35 Å long. The authors suggested that this would mean M4-N was capable of simultaneously contacting both modules which is consistent with the module deletion (Jenkins *et al.* 2006) and chemical shift-mapping studies (Chapter 4). Binding arrangements such as one with M4-N bound at the CCP1/2 interface almost perpendicular to the long axis of the module pair would also explain the chemical shift perturbations and also the similar line-widths of NH cross-peaks of residues in C4BP12 in the C4BP12:M4-N complex compared to free C4BP12. This arrangement would require a much shorter length for the C4BP12-binding region in M4-N. As deletion of seven residues in the HVR of the related M22 protein abolished C4BP binding (Berggård *et al.* 2001a), it is tempting to speculate that the C4BP-binding region in the HVR is much shorter than suggested by André *et al.*

7.6 C4BP and DNA-binding

C4BP is known to bind polyanions such as heparin, which is considered a model for glycosaminoglycans. DNA is, of course, a polyanion, and the question of whether DNA binding by C4BP is physiologically relevant must be addressed. DNA is released from apoptotic cells and in larger quantities from necrotic cells (Jahr *et al.* 2001). C4BP is known to localise to apoptotic (Kask *et al.* 2004) and necrotic (Trouw *et al.* 2005) cells.

First, the role of C4BP in apoptosis should be discussed. All cells have membrane-bound RCA proteins on their surface (i.e DAF, MCP and CR1) that act to down-regulate the classical pathway C3 convertase. This would seem to make the recruitment of C4BP to areas of apoptosis or necrosis superfluous. It is possible though that such recruitment is designed to achieve a level of redundancy sufficient to ensure that complement-mediated

inflammation and tissue destruction cannot occur even if the cell-surface RCAs are 'swamped' by a high level of complement activation (many intracellular components, including DNA, are potent activators of the complement system). It is also possible that levels of RCA expression on dying cells might be reduced compared to normal cells.

Next the potential role of DNA in the localisation of C4BP to areas of cell death can be considered. C4BP exploits its strongly-bound protein S component to bind to apoptotic cells via phosphatidylserine molecules (Webb *et al.* 2002). Cell-surface GAGs presumably provide further anchor sites for C4BP via the binding sites on the alpha-chains. Since C4BP is polymeric, occupation of several arms by GAGs would leave others free to scavenge any segments of DNA that had leaked from the apoptotic or necrotic cell. The perpendicular orientation of B-form DNA, relative to the long axis of C4BP12, implicated by the data-driven docking procedure is consistent with several alpha-chains binding simultaneously to one DNA molecule. This polyvalency would assure a slow off-rate and effective trapping. In this way the DNA would be prevented from diffusing away and activating complement in the fluid phase. So long as one alpha-chain of C4BP still remained unoccupied, this would leave a regulatory site free to prevent complement activation on the entrapped DNA molecule.

The DNA-binding site in C4BP coincides with the putative (on the basis of mutagenesis) binding site for heparin, which has been shown to displace DNA bound to C4BP (Trouw *et al.* 2005). Therefore, it would be useful to measure the relative affinities of C4BP12 for DNA and heparin. It has also been observed that DNA binds to the C-terminal modules of factor H (fH1920) (Figure 7.4) - and the binding appears to be much tighter than to C4BP12. A comparison of the affinities of C4BP12 and fH1920 for both DNA and heparin

would shed light on the interplay between these fluid-phase regulators of the classical and alternative pathways, and their ligands.

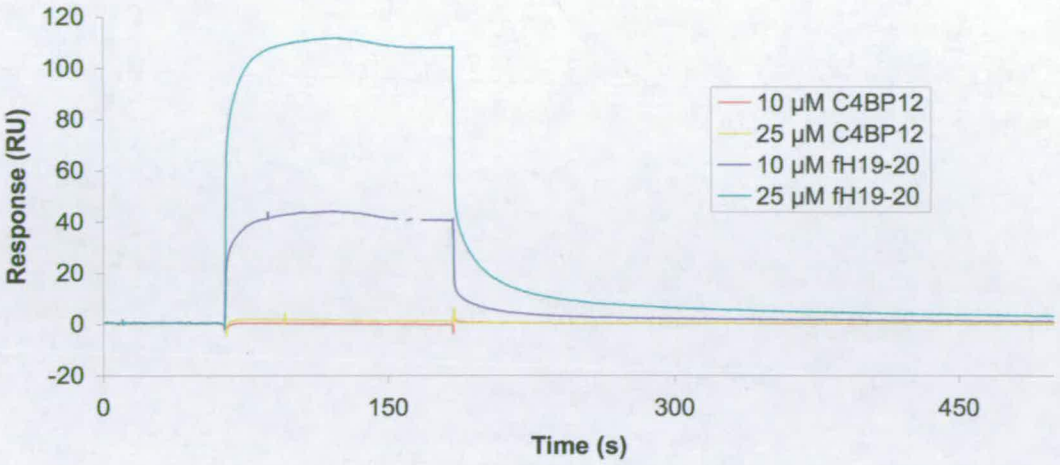

Figure 7.4 Preliminary surface plasmon resonance data for C4BP12 and fH19-20 binding to 20-bp dsDNA (sequence given in Chapter 6) coupled to a streptavidin chip via a biotin tag.

7.7 Site-directed mutagenesis based on the C4BP12 structure

One of the most obvious differences between the C4BP12 and the DAF23 crystal structures is the intermodular orientation (Figure 4.11). The orientation of CCP2 in C4BP12 exposes the Val^{108}-Val^{113} loop. This loop (see Figure 7.5), which is involved in the CCP2/3 interface in DAF (Lukacik *et al.* 2004), has a very different sequence in C4BP (QDRG) to that in DAF (SGSS). As this loop is critical to the C3b/C4b binding activities of CR1 sites 1 and 2 (Krych *et al.* 1994b; Krych *et al.* 1998b), it is likely to also be involved in C4b binding by C4BP. In particular Arg^{111} is exposed on the surface of all the structures in the low-energy ensemble. Since other surface-exposed, charged residues have been shown to be critical for C4b-binding, this residue seems likely to be the a key component of this loop. Prof Anna Blom has made the following mutants of these residues Q109A, D110N, R111Q and G112S and is currently testing their ability to bind C4b.

```
Name        SS β1       β2  HV-loop  β3   β4      β5       β6         β7    β8
C4BPA_02      RCRHPGELRNG-QVE---IKTDLSFGSQIEFSCSEGFFLIG--STTSRCEVQDRGVGWSHPLPQCEIV
DAF_03        SCPNPGEIRNG-QID---VPGGILFGATISFSCNTGYKLFG--STSSFCLISGSSVQWSDPLPECREI
CR1_02        SCRNPPDPVNG-MVH---VIKGIQFGSQIKYSCTKGYRLIG--SSSATCIISGDTVIWDNETPICDRI
CR1_09        SCKTPPDPVNG-MVH---VITDIQVGSRINYSCTTGHRLIG--HSSAECILSGNAAHWSTKPPICQRI
CR1_16        SCKTPPDPVNG-MVH---VITDIQVGSRINYSCTTGHRLIG--HSSAECILSGNTAHWSTKPPICQRI
```

Figure 7.5 Sequence alignment of CCP2 from C4BP a-chain, CCP3 from DAF and CCP2 (site 1), CCP9 (copy 1 site 2) and CCP16 (copy 2 of site 2) of CR1. The loop discussed in the text (comprising residues Val^{108}-Val^{113} in C4BP) is highlighted in bold text.

7.8 Structure of the C4b-binding site

The structure of C4BP12 represents two of the three modules necessary for complement regulation (Blom *et al.* 2001a). Therefore, the next target for structural studies on C4BP is to solve the structure of the CCP2-3 module pair (C4BP23) and to produce the structure of C4BP1-3 by overlaying the structures on CCP2, which is present in both constructs, in the same strategy that was used to produce the structure of the C3b-binding site of CR1 (Smith *et al.* 2002). This work is currently in progress. The ^{15}N-HSQC spectra of C4BP23 and C4BP12 are overlaid in Figure 7.6.

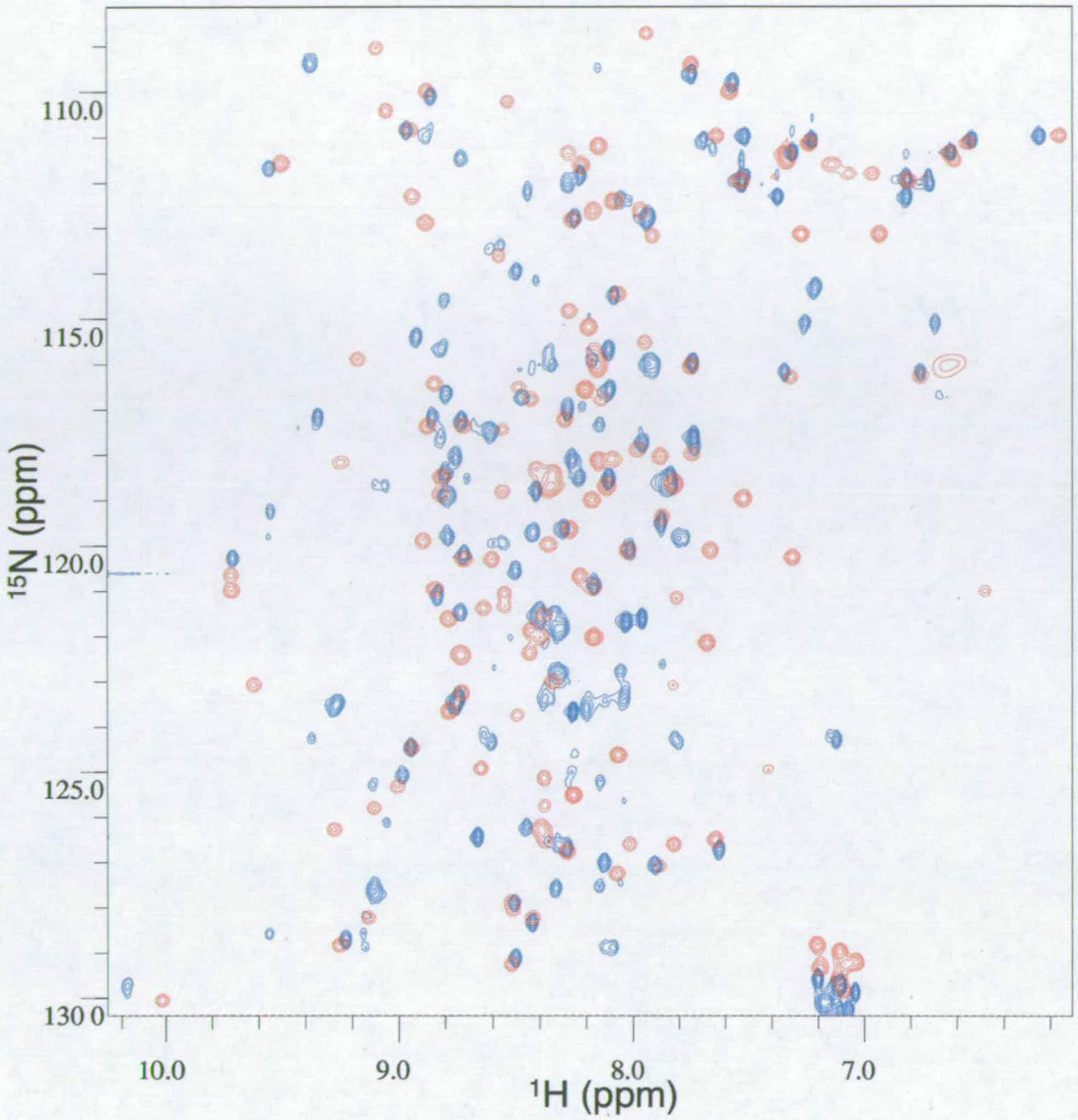

Figure 7.6 Overlay of ^{1}H-^{15}N HSQC spectra of C4BP12 (red) and C4BP23 (blue). Conditions for both samples were identical: 20 mM NaOAc buffer, pH 4.5, 37 °C

Comparison of the chemical shift changes between CCP2 in C4BP12 and C4BP23 show that the changes in shift are localised to the top and bottom of CCP2 (Figure 7.8) consistent with the removal of the CCP1/2 interface and addition of the C4BP23 interface. This suggests the structure of CCP2 is the same in both constructs.

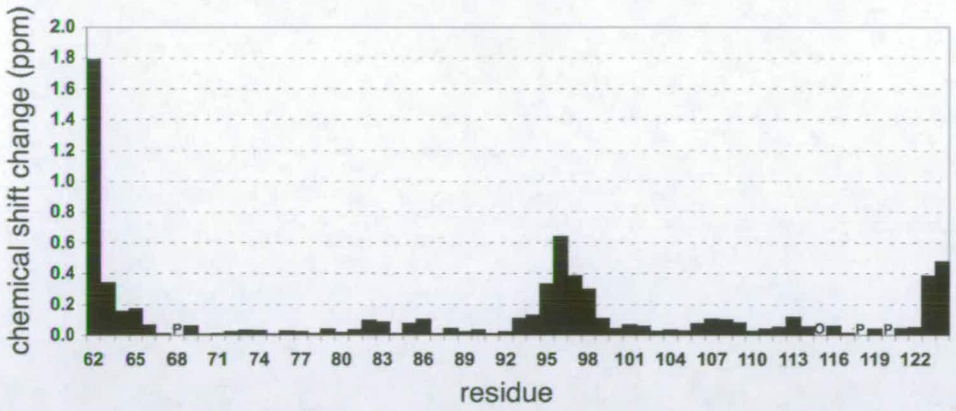

Figure 7.7 Combined chemical shift changes $[\sqrt{((\Delta\delta^{15}N/5)^2+(\Delta\delta^{1}H)^2)}]$ per residue between CCP2 in C4BP12 and C4BP23. P indicates proline residues and O indicates missing data.

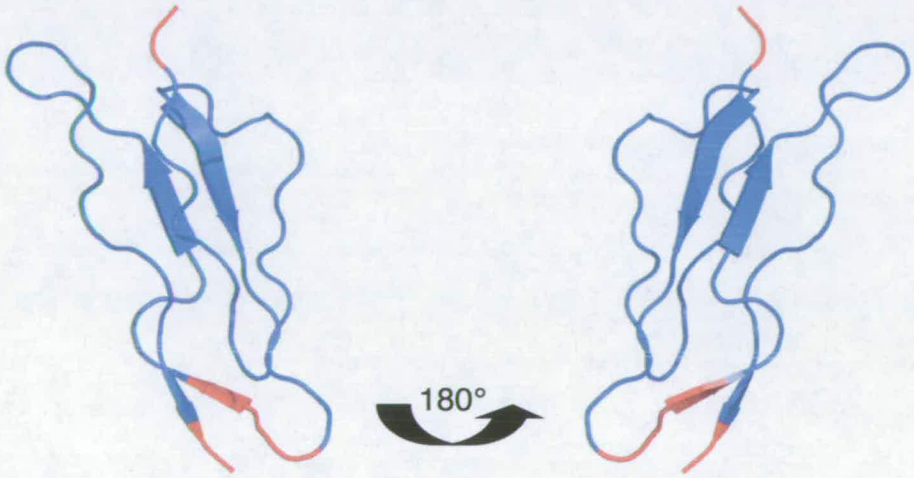

Figure 7.8 Chemical shift changes between CCP2 in C4BP12 and C4BP23 plotted onto the residues Tyr^{62}-Ile^{124} in the closest to mean structure of C4BP12. Residues with a combined chemical shift change > 0.2 ppm are coloured red.

Chapter 8

MATERIALS AND METHODS

8.1 Solution of the structure of C4BP12 using NMR spectroscopy

8.1.1 Sample preparation

Details of the preparation of u[^{15}N] and u[^{13}C;^{15}N]-labelled C4BP12 can be found in Chapter 3 and in (Jenkins *et al.* 2006).

8.1.2 Data collection

Details of the NMR spectra used to solve the structure of C4BP12 are given in Chapter 3.

8.1.3 Quadrature detection

During the acquisition period of an NMR experiment the oscillating magnetisation in the transverse plane generates an electric current in the receiver coil of the probe. The frequencies of the signals are measured as an offset from a reference frequency usually set to the centre of the spectrum. This measurement is either positive or negative depending on whether the signal frequency is greater or less than the reference frequency. In the quadrature detection approach the signal is split into two equal parts using a splitter. Each part is fed into a phase-sensitive detector. The first detector is fed with the reference frequency and the second is fed with a frequency 90° out of phase with respect to the first (Figure 8.1).

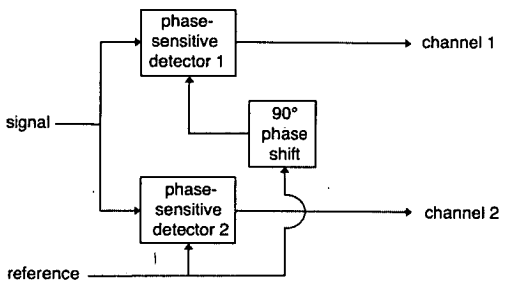

Figure 8.1 Block diagram of a typical quadrature detection system.

Quadrature detection works as follows: The signal $f(t)$ arising in the receiver coil from a vector precessing in the rotating frame with a frequency offset $+\Delta\omega$ from the reference frequency is:

$$f(t) = \cos\Delta\omega t \tag{8.1}$$

Using the identity $\cos(x) = \frac{1}{2}\left(e^{ix} + e^{-ix}\right)$ equation 8.1 can be written as:

$$f(t) = \tfrac{1}{2}\left(e^{i\Delta\omega t} + e^{-i\Delta\omega t}\right) \tag{8.2}$$

The Fourier transform (Section 8.1.6) of this is:

$$f(\omega) = \tfrac{1}{2}\left(\mathrm{Abs}(+\Delta\omega) + \mathrm{i\,Dis}(+\Delta\omega) + \mathrm{Abs}(-\Delta\omega) + \mathrm{i\,Dis}(-\Delta\omega)\right) \tag{8.3}$$

which gives resonance lines at $\pm\Delta\omega$ - i.e the sign of the precession frequency is not determined. However using quadrature detection the output from the unshifted channel is $\cos(\Delta\omega t)$ and that from the 90° phase shifted channel is (in complex notation) $i\sin(\Delta\omega t)$. The detected signal is therefore:

$$f(t) = \cos(\Delta\omega t) + i\sin(\Delta\omega t) \tag{8.4}$$

If the additional identity $\sin(x) = -\frac{1}{2}\left(e^{ix} - e^{-ix}\right)$ is used equation 8.4 becomes:

$$f(t) = \tfrac{1}{2}\left(e^{i\Delta\omega t} + e^{-i\Delta\omega t}\right) + \tfrac{1}{2}\left(e^{i\Delta\omega t} - e^{-i\Delta\omega t}\right) = e^{i\Delta\omega t} \tag{8.5}$$

The Fourier transform of equation 8.5 gives:

$$f(\omega) = \mathrm{Abs}(+\Delta\omega t) + \mathrm{i}\,\mathrm{Dis}(+\Delta\omega t) \tag{8.6}$$

In this way the sign of the precessing frequency has been determined and the absorption (Abs) and dispersion (Dis) components are the real and imaginary parts of the spectrum respectively (Figure 8.2).

8.1.4 Digital Sampling

The frequency of digital sampling of the time domain data is crucial as this determines the maximum spectral width in the experiment. If the continuous time domain signal is $x(t)$ and sampling is performed by measuring the value of the signal every Δt seconds the sampled signal is given by:

$$x[n] = x(n\Delta t) \qquad \text{with } n = 0,1,2,3,.... \tag{8.7}$$

The sampling frequency f_s is defined as the number of samples obtained in one second:

$$f_s = \frac{1}{\Delta t} \tag{8.8}$$

In order to reconstruct the original signal perfectly, the Nyquist-Shannon sampling theorem states that the sampling frequency must be more than twice the maximum frequency. Therefore, the frequency equal to one half of the sampling rate is the highest frequency that can be unambiguously represented by the sampled signal. This frequency f_N is known as the Nyquist frequency:

$$f_N = \frac{1}{2\Delta t} \tag{8.9}$$

Signals with frequencies greater than f_N can be observed in the digital signal but their frequency is ambiguous, i.e a component with frequency f cannot be distinguished from a component with frequency $2f_N+f$, $4f_N+f$, etc. This ambiguity is known as aliasing. Aliasing is often exploited in the indirect dimensions of multi-dimensional NMR experiments to 'fold' in outlying signals, thereby enabling the spectral width to be minimised and thus increasing the resolution without lengthening the experimental time.

8.1.5 Data processing

All NMR spectra were processed using AZARA (W. Boucher, Department of Biochemistry, University of Cambridge, UK). A 90° shifted sine-bell squared apodising window function (Section 8.1.8) and zero filling (Section 8.1.9) were applied to the FID before Fourier transformation (Section 8.1.6) . Maximum entropy reconstruction (Section 8.1.10) was used for both indirect dimensions of all 3D spectra recorded.

8.1.6 Fourier transform

Modern spectrometers collect data in the time domain. To convert this information into the frequency domain a Fourier transform is performed. The Fourier transform relates time-domain data $f(t)$ with frequency-domain data $f(\omega)$:

$$f(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt \qquad (8.10)$$

The Fourier transform of a decaying exponential function:

$$f(t) = e^{i\omega t} e^{-t/T_2} \qquad (8.11)$$

if it is assumed that for $t < 0, f(t) = 0$, is:

$$\mathrm{FT}\left\{e^{i\omega t}e^{-t/T_2}\right\} = \left[\mathrm{Abs}(\omega^+) + i\,\mathrm{Dis}(\omega^+)\right] \tag{8.12}$$

The Fourier transform of the related function:

$$\mathrm{FT}\left\{e^{-i\omega t}e^{-t/T_2}\right\} = \left[\mathrm{Abs}(\omega^-) + i\,\mathrm{Dis}(\omega^-)\right] \tag{8.13}$$

gives similar lineshapes at negative frequencies. Again Abs(ω) is the real part or absorption lineshape and i Dis(ω) is the imaginary or dispersion lineshape. As shown in Section 8.1.3, the function $e^{i\omega t}$ is the free-induction decay (FID) of the function cos(ωt) whose quadrature function is -sin(ωt). Therefore, equations 8.12 and 8.13 represent the Fourier transforms of signals detected using quadrature detection. These lineshapes are illustrated in Figure 8.2.

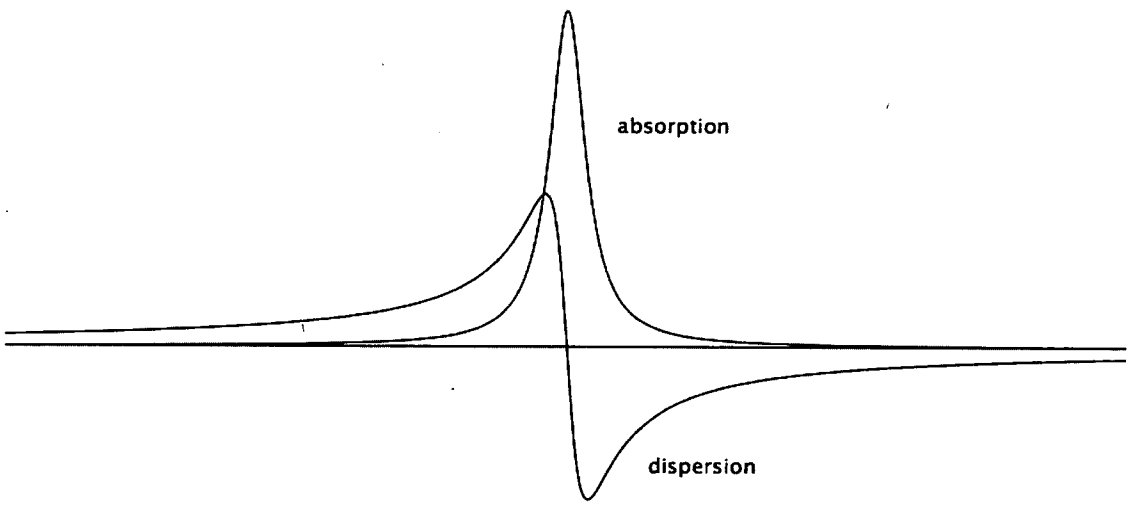

Figure 8.2 The absorption-mode (real) and dispersion-mode (imaginary) lineshapes.

8.1.7 Convolution with a box function

In order to remove the residual water signal from the spectrum the time domain data is convoluted with a box function at zero-frequency and the result subtracted from the data before Fourier transformation to give the spectrum. A box function has a value of 1 within the half-width and 0 outside. The typical half-widths used are 8-32 points. The effect of convolution with a box function of half-width 8 is shown in Figure 8.3.

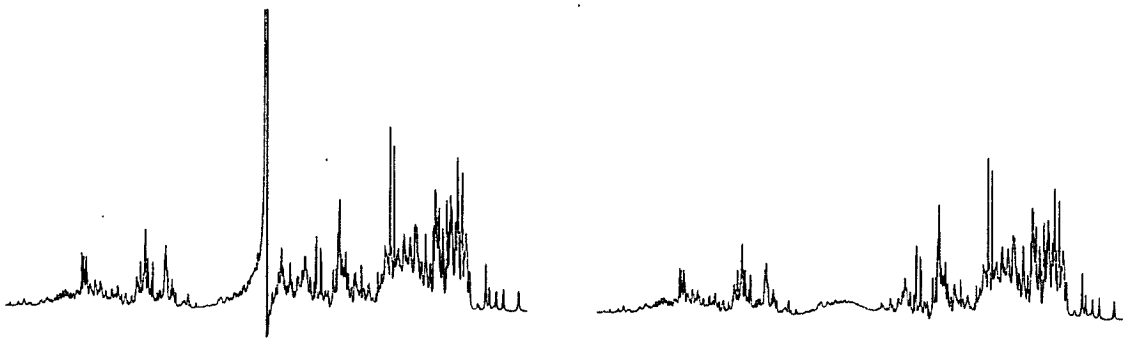

Figure 8.3 The effect of convoluting the data with a box function of half-width 8 is shown in the right-hand spectrum. The left-hand spectrum did not have this convolution applied.

8.1.8 Window functions

If the signal in a FID has not decayed to zero by the end of the acquisition period then lineshapes in the Fourier transformed spectrum will be convoluted with a sinc function ($\sin x/x$) as shown in Figure 8.4. These distortions are known as 'sinc wiggles' and their intensity increases with the degree of truncation. As the acquisition times in the indirect dimensions of 3D experiments are often short, truncation is common. The effects of truncation can be removed by applying a window function that apodises the data - i.e a function that decays smoothly to zero at the end of the FID. A window function is simply the convolution of the time-domain data with a simple (usually trigonometric) function prior to Fourier transformation. The shifted sine-bell squared function was used for all spectra recorded on C4BP12. This consists of a $\sin^2(x)$ function shifted by $\pi/2$ so that it has its

maximum value at the start of the FID. The result of applying this window function to the FID and the resulting lineshape after Fourier transform is shown in Figure 8.4.

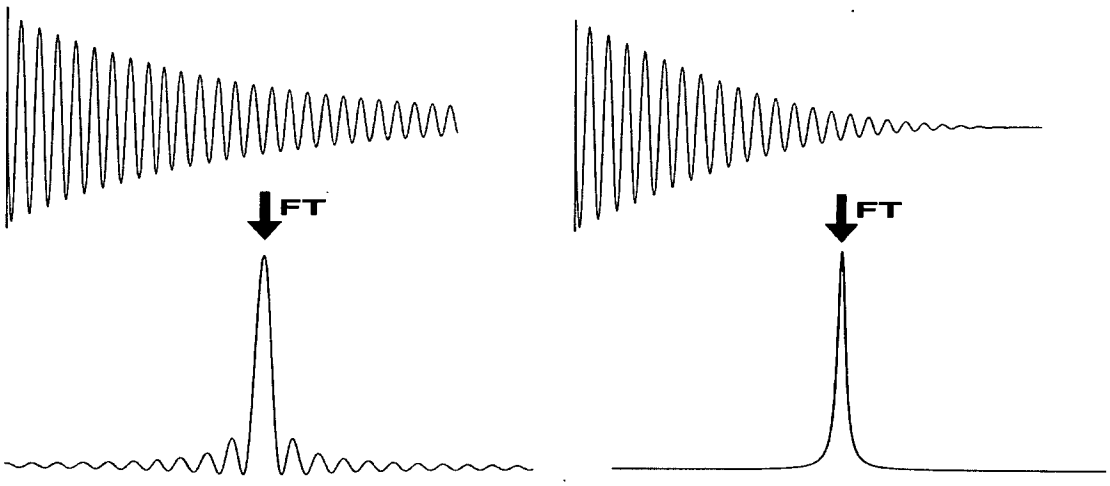

Figure 8.4 Effect of apodisation: left no window function; right $\sin^2(\omega+\pi/2)$ applied

8.1.9 Zero filling

The fast Fourier transform algorithms used to process the data require that the number of points in the input data is 2^n. By extending the FID by adding zeros - a process known as zero filling, the *digital* resolution of the data is increased. As this does not increase the acquisition time for the data, zero filling each dimension to the next power of 2 (thus doubling the size of that dimension) is routinely used in data processing.

8.1.10 Maximum entropy reconstruction

Maximum entropy (MaxEnt) reconstruction (Sibisi *et al.* 1984) can be used to increase the signal/noise ratio and resolution in the (usually truncated) indirectly detected dimensions of 3D spectra. Rather than computing the spectra directly from the time-domain data, MaxEnt reconstruction minimises the information content of a trial spectrum whilst ensuring its inverse discrete Fourier transform (DFT) agrees with the observed data to within the level of

error of the data. The agreement of the reconstruction with the data is measured using an unweighted χ^2 statistic:

$$C(\mathbf{f}) = \sum_{k=0}^{N-1} |m_k - d_k|^2 \tag{8.14}$$

where m_k is the mock data from the inverse DFT of the trial spectrum and d_k is the experimental data. The information content is measured using an entropy function resembling:

$$S(f) = -\sum_{i=0}^{M-1} \frac{|f_i|}{def} \log\left(\frac{|f_i|}{def}\right) \tag{8.15}$$

where M is the number of points in the spectrum and *def* is a scale factor. MaxEnt reconstruction maximises $S(f)$ subject to the constraint that $C(\mathbf{f}) \approx C_0$ where C_0 is an estimate of the noise level of the data thus determining the spectrum containing the least information (hence the fewest artefacts) that is consistent with the experimental data.

8.1.11 Example AZARA processing script

An example script used for processing a 3D spectrum in AZARA is shown below with an explanation of the commands used:

```
input ser.par                  input data parameter file
output mx.spc                  output data file

script_com 1                   commands for acquisition dimension
  complex                      data is complex
  conv_box 8                   convolution with box function to remove water
  avance 12 16                 required to correct for Bruker digital filter
  sinebell2 90                 90° shifted sinebell-squared window function
  zerofill 1                   zero fill so data is twice original size
  fft                          fast Fourier transform
  avance_phase                 phase  correction for Bruker digital filter
  phase 0 0                    phase correction (0 order) (1st order)
  reduce                       discard imaginary part of data
! upper 512                    keep only 1st 512  points (! = not used)
end_script

script_com 2 !This is 1H       1st indirectly detected dimension
  mask_ppmm                    ++-- filter as data collected using States-TPPI
  complex                      data is complex
  scale2 1 1 0.5 0.0           scale 1st data point by ½ to remove t1 noise
end_script

script_com 3 ! this is 13C     2nd indirectly detected dimension
  mask_ppmm                    ++-- filter as data collected using States-TPPI
  complex                      data is complex
  scale2 1 1 0.5 0.0           scale 1st data point by ½ to remove t1 noise
  conjugate                    calculate complex conjugate of data
end_script

maxent2_com 2 3                2D  MaxEnt  process dimensions 2 and 3
iter 50                        Number of iterations for Maxent algorithm
noise 1200                     Estimate of noise level in data
log m.log2                     log file for MaxEnt
rate 0.2                       step size for MaxEnt algorithm
dim 1                          1st Dimension to MaxEnt process
npts 256  !extra zerofill      zero fill
complex                        data is complex
phase 0 0                      phase correction in dimension 2

dim 2                          2nd Dimension to MaxEnt process
npts 128   !extra zerofill     zero fill
complex                        data is complex
phase 90 -180                  phase correction in dimension 3
end_maxent

script_com 1
 base_trig 8 0                 Baseline correction for acquisition dimension
end_script
```

8.1.12 Data analysis, assignment and structure calculation

Spectra were visualised and assigned using ANSIG (Kraulis 1989) as described in Chapter 2. Structures were calculated, as described in Chapter 3, using CNS (Brünger *et al.* 1998) with a protocol incorporating molecular dynamics based simulated annealing (MDSA) *(Nilges et al.* 1988a; Nilges *et al.* 1988b). Ambiguous distance restraints were incorporated into the structure calculation using the protocols derived by Nilges (Nilges 1995) and these were 'filtered' using ARIA (Nilges *et al.* 1997).

8.2 Isothermal titration calorimetry

Experiments were performed at 25 °C on a MicroCal VP-ITC microcalorimeter. The cell contained 0.01 mM C4BP12 (1.4 ml) and the syringe contained 0.15 mM M4-N. Both solutions were in 10 mM sodium phosphate buffer (pH 6.0). For titrations, one preliminary injection of 1 µl was made, followed by 19 injections of 10 µl with an injection speed of 0.5 µl/s. The stirring speed was 310 rpm and the delay time between injections was 3 minutes. A blank titration, i.e. injecting M4-N into buffer, was used to correct the M4-N:C4BP12 titration for heat of dilution effects. The titration was performed with the assistance of Margaret Nutley (University of Glasgow). Analysis of the data was performed using MicroCal Origin software.

8.3 Preparation of DNA oligonucleotides

8.3.1 Annealing

DNA oligonucleotides purchased from Sigma-Genosys were resuspended in annealing buffer (10 mM Tris, pH 7.8, 50 mM NaCl, 1 mM EDTA) and the concentration calculated from the A_{260} measurement using the extinction coefficient supplied by Sigma-Genosys.

Equimolar concentrations of the two oligos were mixed and 100 µl aliquots were dispensed into PCR tubes (600 µl volume). The tubes were placed in a thermal cycler and the following programme set: (i) heat to 95 °C and remain at 95 °C for 2 minutes, (ii) ramp-cool to 25 °C over a period of 45 minutes, (iii) cool to a storage temperature of 4 °C. After a brief spin in a microfuge to draw all moisture from the lid, the samples were pooled and stored at 4 °C prior to ethanol-precipitation.

8.3.2 Ethanol precipitation to concentrate the DNA

The 400 µl of dsDNA produced in the previous step was ethanol-precipitated, dried and resuspended in NMR buffer (20 mM sodium phosphate buffer, pH 6.0) as follows. 40 µl of 3.0 M NaOAc, pH 5.2 was added. Subsequently $MgCl_2$ to a final concentration of 10 mM was added along with 880 µl of cold (4 °C) 100% ethanol. After mixing the sample was stored at -20 °C for > 1 hour. The sample was spun at 13,000g for 30 minutes in a microfuge at 4 °C. The supernatant was decanted and 1 ml of 70% ethanol was added. After mixing the sample was spun at 13,000g for 2 minutes. The supernatant was again decanted and another 1 ml of 70% ethanol was added before mixing and a further spin at 13,000g for 2 minutes. The supernatant was decanted and the sample dried at room temperature. 200 µl of NMR buffer was added and the concentration calculated from the A_{260} measurement using the general extinction coefficient for dsDNA of 50 $(\mu g/ml)^{-1}cm^{-1}$.

8.4 NMR titration with M4-N

The peptide dimer, M4-N was expressed in *E. coli*, as described in (André *et al.* 2006). Dimerisation was achieved by oxidation of bacterial lysates with 20 mM $CuCl_2$ in 0.4 M NaCl, pH 8.0 (Morfeldt *et al.* 2001). The dimerised peptide was purified by ion-exchange chromatography and gel filtration, and was finally dialysed against doubly-distilled H_2O.

Lyophilised peptide was dissolved in NMR buffer (deuterated NaOAc, pH 6.0) to give 100 μM (dimer concentration). A 20 μM solution of C4BP12 in the same buffer was added and a series of ^{15}N-^{1}H HSQC spectra were subsequently recorded at M4-N concentrations of 100, 50, 25, and 10 μM (dimer concentrations), and compared to a reference spectrum of 20 μM C4BP12 with no M4-N present.

8.5 NMR titration with 10-bp DNA

A ^{15}N-^{1}H HSQC spectrum of 50 μM u[^{15}N]-labelled C4BP12 in 20 mM sodium phosphate buffer, pH 6.0 was recorded. Small volumes of DNA (1.99 mM concentration), prepared as described in Section 8.3, were added and a series of ^{15}N-^{1}H HSQC spectra were subsequently recorded at DNA concentrations of 25, 50, 100 and 200 μM.

BIBLIOGRAPHY

BIBLIOGRAPHY

Accardo, P., Sánchez-Corral, P., Criado, O., Garcia, E. and Rodríguez de Córdoba, S. (1996). "Binding of human complement component C4b-binding protein (C4BP) to Streptococcus pyogenes involves the C4b-binding site." J. Immunol. **157**(11): 4935-9.

Agostoni, A. and Cicardi, M. (1992). "Hereditary and acquired C1-inhibitor deficiency: biological and clinical characteristics in 235 patients." Medicine (Baltimore) **71**(4): 206-15.

Allegrozzi, M., Bertini, I., Janik, M., Lee, Y., Liu, G. and C, L. (2000). "Lanthanide-Induced Pseudocontact Shifts for Solution Structure Refinements of Macromolecules in Shells up to 40 Å from the Metal Ion." J Am Chem Soc **122**(17): 4154-4161.

André, I., Persson, J., Blom, A. M., Nilsson, H., Drakenberg, T., Lindahl, G. and Linse, S. (2006). "Streptococcal M Protein: Structural Studies of the Hypervariable Region, Free and Bound to Human C4BP." Biochemistry **45**(14): 4559-4568.

Arlaud, G. J., Gaboriaud, C., Thielens, N. M. and Rossi, V. (2002). "Structural biology of C1." Biochem Soc Trans **30**(Pt 6): 1001-6.

Arlaud, G. J. and Thielens, N. M. (1993). "Human complement serine proteases C1r and C1s and their proenzymes." Methods Enzymol **223**: 61-82.

Arlaud, G. J., Volanakis, J. E., Thielens, N. M., Narayana, S. V., Rossi, V. and Xu, Y. (1998). "The atypical serine proteases of the complement system." Adv Immunol **69**: 249-307.

Baker, N. A., Sept, D., Joseph, S., Holst, M. J. and McCammon, J. A. (2001). "Electrostatics of nanosystems: application to microtubules and the ribosome." Proc. Natl. Acad. Sci. USA **98**(18): 10037-41.

Ball, G. (2005). Structure Refinement and Dynamics of Proteins Using Residual Dipolar Couplings and NMR Relaxation Data. School of Chemistry. Edinburgh, University of Edinburgh. **Doctor of Philosophy**.

Ball, G., Meenan, N., Bromek, K., Smith, B. O., Bella, J. and Uhrin, D. (2006). "Measurement of one-bond 13Calpha-1Halpha residual dipolar coupling constants in proteins by selective manipulation of CalphaHalpha spins." J Magn Reson **180**(1): 127-36.

Barlow, P. N., Baron, M., Norman, D. G., Day, A. J., Willis, A. C., Sim, R. B. and Campbell, I. D. (1991). "Secondary structure of a complement control protein module by two-dimensional 1H NMR." Biochemistry **30**(4): 997-1004.

Berggård, K., Johnsson, E., Morfeldt, E., Persson, J., Stalhammar-Carlemalm, M. and Lindahl, G. (2001a). "Binding of human C4BP to the hypervariable region of M protein: a molecular mechanism of phagocytosis resistance in Streptococcus pyogenes." Mol. Microbiol. **42**(2): 539-51.

Berggård, K., Lindahl, G., Dahlbäck, B. and Blom, A. M. (2001b). "Bordetella pertussis binds to human C4b-binding protein (C4BP) at a site similar to that used by the natural ligand C4b." Eur. J. Immunol. **31**(9): 2771-80.

Bloch, F. and Siegert, A. (1940). "Magnetic Resonance for Nonrotating Fields." Phys Rev **57**: 522.

Blom, A. M., Berggård, K., Webb, J. H., Lindahl, G., Villoutreix, B. O. and Dahlbäck, B. (2000). "Human C4b-binding protein has overlapping, but not identical, binding sites for C4b and streptococcal M proteins." J. Immunol. **164**(10): 5328-36.

Blom, A. M., Kask, L. and Dahlbäck, B. (2001a). "Structural requirements for the complement regulatory activities of C4BP." J. Biol. Chem. **276**(29): 27136-44.

Blom, A. M., Kask, L. and Dahlbäck, B. (2003). "CCP1-4 of the C4b-binding protein alpha-chain are required for factor I mediated cleavage of complement factor C3b." Mol. Immunol. **39**(10): 547-56.

Blom, A. M., Rytkonen, A., Vasquez, P., Lindahl, G., Dahlback, B. and Jonsson, A. B. (2001b). "A novel interaction between type IV pili of Neisseria gonorrhoeae and the human complement regulator C4B-binding protein." J. Immunol. **166**(11): 6764-70.

Blom, A. M., Webb, J., Villoutreix, B. O. and Dahlbäck, B. (1999). "A cluster of positively charged amino acids in the C4BP alpha-chain is crucial for C4b binding and factor I cofactor function." J. Biol. Chem. **274**(27): 19237-45.

Bodenhausen, G. and Ruben, D. J. (1980). "Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy." Chem. Phys. Lett. **69**(1): 185.

Bork, P., Downing, A. K., Kieffer, B. and Campbell, I. D. (1996). "Structure and distribution of modules in extracellular proteins." Q. Rev. Biophys. **29**(2): 119-67.

Botto, M., Dell'Agnola, C., Bygrave, A. E., Thompson, E. M., Cook, H. T., Petry, F., Loos, M., Pandolfi, P. P. and Walport, M. J. (1998). "Homozygous C1q deficiency causes glomerulonephritis associated with multiple apoptotic bodies." Nat Genet **19**(1): 56-9.

Braun, W. (1987). "Distance geometry and related methods for protein structure determination from NMR data." Q Rev Biophys **19**(3-4): 115-57.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. and Karplus, M. (1983). "CHARMM: A program for macromolecular energy, minimization, and dynamics calculations." Journal of Computational Chemistry **4**(2): 187-217.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. and Warren, G. L. (1998). "Crystallography & NMR system: A new software suite for macromolecular structure determination." Acta. Crystallogr. D Biol. Crystallogr. **54 (Pt 5)**: 905-21.

Brünger, A. T., Clore, G. M., Gronenborn, A. M. and Karplus, M. (1986). "Three-dimensional structure of proteins determined by molecular dynamics with

interproton distance restraints: application to crambin." Proc Natl Acad Sci U S A **83**(11): 3801-5.

Carlsson, F., Berggård, K., Stalhammar-Carlemalm, M. and Lindahl, G. (2003). "Evasion of phagocytosis through cooperation between two ligand-binding regions in Streptococcus pyogenes M protein." J. Exp. Med. **198**(7): 1057-68.

Casasnovas, J. M., Larvie, M. and Stehle, T. (1999). "Crystal structure of two CD46 domains reveals an extended measles virus-binding surface." EMBO J. **18**(11): 2911-22.

Cavanagh, J., Fairbrother, W. J., Palmer, A. G. r. and Skelton, N. J. (1996). Protein NMR Spectroscopy, Principles and Practice. San Diego, Academic Press.

Cedervall, T., Johansson, M. U. and Akerstrom, B. (1997). "Coiled-coil structure of group A streptococcal M proteins. Different temperature stability of class A and C proteins by hydrophobic-nonhydrophobic amino acid substitutions at heptad positions a and d." Biochemistry **36**(16): 4987-94.

Chalaoux, F. R., O'Donoghue, S. I. and Nilges, M. (1999). "Molecular dynamics and accuracy of NMR structures: Effects of error bounds and data removal." Proteins: Struct., Funct., Genet. **34**(4): 453-463.

Clubb, R. T., Thanabal, V. and Wagner, G. (1992). "A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue 1HN, 15N, and 13C' chemical shifts in 15N---13C-labelled proteins." J. Magn. Reson. **97**(1): 213.

Colten, H. R. and Rosen, F. S. (1992). "Complement deficiencies." Annu Rev Immunol **10**: 809-34.

Cuff, J. A. and Barton, G. J. (2000). "Application of multiple sequence alignment profiles to improve protein secondary structure prediction." Proteins **40**(3): 502-11.

Dahlbäck, B., Smith, C. A. and Müller-Eberhard, H. J. (1983). "Visualization of human C4b-binding protein and its complexes with vitamin K-dependent protein S and complement protein C4b." Proc. Natl. Acad. Sci. USA **80**(11): 3461-5.

Dahlbäck, B. and Stenflo, J. (1981). "High molecular weight complex in human plasma between vitamin K-dependent protein S and complement component C4b-binding protein." Proc. Natl. Acad. Sci. USA **78**(4): 2512-6.

Deschamps, M. L., Pilka, E. S., Potts, J. R., Campbell, I. D. and Boyd, J. (2005). "Probing protein-peptide binding surfaces using charged stable free radicals and transverse paramagnetic relaxation enhancement (PRE)." J Biomol NMR **31**(2): 155-60.

Dominguez, C., Boelens, R. and Bonvin, A. M. (2003). "HADDOCK: a protein-protein docking approach based on biochemical or biophysical information." J Am Chem Soc **125**(7): 1731-7.

Driscoll, P. C., Gronenborn, A. M. and Clore, G. M. (1989). "The influence of stereospecific assignments on the determination of three-dimensional structures of proteins by nuclear magnetic resonance spectroscopy. Application to the sea anemone protein BDS-I." FEBS Lett **243**(2): 223-33.

Fang, Y., Xu, C., Fu, Y. X., Holers, V. M. and Molina, H. (1998). "Expression of complement receptors 1 and 2 on follicular dendritic cells is necessary for the generation of a strong antigen-specific IgG response." J Immunol **160**(11): 5273-9.

Fearon, D. T. (1979). "Regulation of the amplification C3 convertase of human complement by an inhibitory protein isolated from human erythrocyte membrane." Proc Natl Acad Sci U S A **76**(11): 5867-71.

Figueroa, J., Andreoni, J. and Densen, P. (1993). "Complement deficiency states and meningococcal disease." Immunol Res **12**(3): 295-311.

Folmer, R. H., Hilbers, C. W., Konings, R. N. and Nilges, M. (1997). "Floating stereospecific assignment revisited: application to an 18 kDa protein and comparison with J-coupling data." J Biomol NMR **9**(3): 245-58.

Frishman, D. and Argos, P. (1995). "Knowledge-based protein secondary structure assignment." Proteins **23**(4): 566-79.

Fujita, T., Gigli, I. and Nussenzweig, V. (1978). "Human C4-binding protein. II. Role in proteolysis of C4b by C3b-inactivator." J. Exp. Med. **148**(4): 1044-51.

Fujita, T. and Nussenzweig, V. (1979). "The role of C4-binding protein and beta 1H in proteolysis of C4b and C3b." J. Exp. Med. **150**(2): 267-76.

Ganesh, V. K., Smith, S. A., Kotwal, G. J. and Murthy, K. H. (2004). "Structure of vaccinia complement protein in complex with heparin and potential implications for complement regulation." Proc. Natl. Acad. Sci. USA **101**(24): 8924-9.

Garcia de Frutos, P. and Dahlbäck, B. (1994). "Interaction between serum amyloid P component and C4b-binding protein associated with inhibition of factor I-mediated C4b degradation." J. Immunol. **152**(5): 2430-2437.

Gigli, I., Fujita, T. and Nussenzweig, V. (1979). "Modulation of the classical pathway C3 convertase by plasma proteins C4 binding protein and C3b inactivator." Proc. Natl. Acad. Sci. USA **76**(12): 6596-600.

Grishaev, A., Wu, J., Trewhella, J. and Bax, A. (2005). "Refinement of multidomain protein structures by combination of solution small-angle X-ray scattering and NMR data." J Am Chem Soc **127**(47): 16621-8.

Grzesiek, S. and Bax, A. (1992a). "Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR." J. Am. Chem. Soc. **114**: 6291.

Grzesiek, S. and Bax, A. (1992b). "An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins." J. Magn. Reson. **99**(1): 201.

Grzesiek, S. and Bax, A. (1992c). "Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein." J. Magn. Reson. **96**(2): 432.

Grzesiek, S. and Bax, A. (1993a). "Amino acid type determination in the sequential assignment procedure of uniformly $^{13}C/^{15}N$-enriched proteins." J. Biomol. NMR **3**(2): 185.

Grzesiek, S. and Bax, A. (1993b). "Amino-Acid Type Determination in the Sequential Assignment Procedure of Uniformly C-13/N-15-Enriched Proteins." J. Biomol. NMR **3**(2): 185-204.

Grzesiek, S. and Bax, A. (1993c). "The Origin and Removal of Artifacts in 3D HCACO Spectra of Proteins Uniformly Enriched with 13C." J. Magn. Reson. Ser. B **102**(1): 103.

Guntert, P., Mumenthaler, C. and Wuthrich, K. (1997). "Torsion angle dynamics for NMR structure calculation with the new program DYANA." J Mol Biol **273**(1): 283-98.

Habeck, M., Rieping, W. and Nilges, M. (2006). "Weighting of experimental evidence in macromolecular structure determination." Proc Natl Acad Sci U S A **103**(6): 1756-61.

Hardig, Y. and Dahlback, B. (1996). "The amino-terminal module of the C4b-binding protein beta-chain contains the protein S-binding site." J. Biol. Chem. **271**(34): 20861-7.

Heiring, C. and Muller, Y. A. (2001). "Folding screening assayed by proteolysis: application to various cystine deletion mutants of vascular endothelial growth factor." Protein Eng. **14**(3): 183-8.

Henderson, C. E., Bromek, K., Mullin, N. P., Smith, B. O., Uhrin, D. and Barlow, P. N. (2001). "Solution structure and dynamics of the central CCP module pair of a poxvirus complement control protein." J. Mol. Biol. **307**(1): 323-39.

Herbert, A. P., Uhrin, D., Lyon, M., Pangburn, M. K. and Barlow, P. N. (2006). "Disease-associated sequence variations congregate in a polyanion recognition patch on human factor H revealed in three-dimensional structure." J Biol Chem **281**(24): 16512-20.

Hessing, M., Vlooswijk, R. A., Hackeng, T. M., Kanters, D. and Bouma, B. N. (1990). "The localization of heparin-binding fragments on human C4b-binding protein." J. Immunol. **144**(1): 204-8.

Hillarp, A. and Dahlback, B. (1988). "Novel subunit in C4b-binding protein required for protein S binding." J. Biol. Chem. **263**(25): 12759-64.

Hillarp, A. and Dahlbäck, B. (1990). "Cloning of cDNA coding for the beta chain of human complement component C4b-binding protein: sequence homology with the alpha chain." Proc. Natl. Acad. Sci. USA **87**(3): 1183-7.

Hillarp, A., Hessing, M. and Dahlbäck, B. (1989). "Protein S binding in relation to the subunit composition of human C4b-binding protein." FEBS Lett. **259**(1): 53-6.

Hus, J. C., Marion, D. and Blackledge, M. (2001). "Determination of protein backbone structure using only residual dipolar couplings." J Am Chem Soc **123**(7): 1541-2.

Hwang, T. L. and Shaka, A. J. (1995). "Water Suppression That Works - Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients." J. Magn. Reson. Ser. A **112**(2): 275-279.

Härdig, Y., Hillarp, A. and Dahlbäck, B. (1997). "The amino-terminal module of the C4b-binding protein alpha-chain is crucial for C4b binding and factor I-cofactor function." Biochem. J. **323 (Pt 2)**: 469-75.

Ikegami, T., Verdier, L., Sakhaii, P., Grimme, S., Pescatore, B., Saxena, K., Fiebig, K. M. and Griesinger, C. (2004). "Novel techniques for weak alignment of proteins in solution using chemical tags coordinating lanthanide ions." J Biomol NMR **29**(3): 339-49.

Jahr, S., Hentze, H., Englisch, S., Hardt, D., Fackelmayer, F. O., Hesch, R. D. and Knippers, R. (2001). "DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells." Cancer Res **61**(4): 1659-65.

Jarva, H., Ram, S., Vogel, U., Blom, A. M. and Meri, S. (2005). "Binding of the complement inhibitor C4bp to serogroup B Neisseria meningitidis." J. Immunol. **174**(10): 6299-307.

Jeener, J., Meier, B. H., Bachmann, P. and Ernst, R. R. (1979). "Investigation of exchange processes by two-dimensional NMR spectroscopy." J. Chem. Phys. **71**: 4546.

Jenkins, H. T., Mark, L., Ball, G., Persson, J., Lindahl, G., Uhrin, D., Blom, A. M. and Barlow, P. N. (2006). "Human C4b-binding protein, structural basis for interaction with streptococcal M protein, a major bacterial virulence factor." J Biol Chem **281**(6): 3690-7.

Johnsson, E., Andersson, G., Lindahl, G. and Heden, L. O. (1994). "Identification of the IgA-binding region in streptococcal protein Arp." J. Immunol. **153**(8): 3557-64.

Johnsson, E., Areschoug, T., Mestecky, J. and Lindahl, G. (1999). "An IgA-binding peptide derived from a streptococcal surface protein." J. Biol. Chem. **274**(21): 14521-4.

Johnsson, E., Thern, A., Dahlbäck, B., Hedén, L. O., Wikström, M. and Lindahl, G. (1996). "A highly variable region in members of the streptococcal M protein family binds the human complement regulator C4BP." J. Immunol. **157**(7): 3021-9.

Jokiranta, T. S., Jaakola, V. P., Lehtinen, M. J., Parepalo, M., Meri, S. and Goldman, A. (2006). "Structure of complement factor H carboxyl-terminus reveals molecular basis of atypical haemolytic uremic syndrome." Embo J **25**(8): 1784-94.

Kask, L., Trouw, L. A., Dahlbäck, B. and Blom, A. M. (2004). "The C4b-binding protein-protein S complex inhibits the phagocytosis of apoptotic cells." J. Biol. Chem. **279**(23): 23869-73.

Kay, L. E., Xu, G. Y., Singer, A. U., Muhandiram, D. R. and Formankay, J. D. (1993). "A Gradient-Enhanced HCCH-TOCSY Experiment for Recording Side-Chain 1H and 13C Correlations in H2O Samples of Proteins." J. Magn. Reson. Ser. B **101**(3): 333.

Kelley, L. A., MacCallum, R. M. and Sternberg, M. J. (2000). "Enhanced genome annotation using structural profiles in the program 3D-PSSM." J. Mol. Biol. **299**(2): 499-520.

Kirkitadze, M. D. and Barlow, P. N. (2001). "Structure and flexibility of the multiple domain proteins that regulate complement activation." Immunol Rev **180**: 146-61.

Kirkpatrick, S., Gelatt, C. D. and Vecchi, M. P. (1983). "Optimization by simulated annealing." Science **220**(4598): 671-680.

Korb, L. C. and Ahearn, J. M. (1997). "C1q binds directly and specifically to surface blebs of apoptotic human keratinocytes: complement deficiency and systemic lupus erythematosus revisited." J Immunol **158**(10): 4525-8.

Kraulis, P. J. (1989). "ANSIG: A Program for the Assignment of Protein ^{1}H 2D NMR Spectra by Interactive Graphics." J. Magn. Reson. **24**: 627-33.

Krych, M., Clemenza, L., Howdeshell, D., Hauhart, R., Hourcade, D. and Atkinson, J. P. (1994a). "Analysis of the functional domains of complement receptor type 1 (C3b/C4b receptor; CD35) by substitution mutagenesis." J Biol Chem **269**(18): 13273-8.

Krych, M., Clemenza, L., Howdeshell, D., Hauhart, R., Hourcade, D. and Atkinson, J. P. (1994b). "Analysis of the functional domains of complement receptor type 1 (C3b/C4b receptor; CD35) by substitution mutagenesis." J. Biol. Chem. **269**(18): 13273-8.

Krych, M., Hauhart, R. and Atkinson, J. P. (1998a). "Structure-function analysis of the active sites of complement receptor type 1." J Biol Chem **273**(15): 8623-9.

Krych, M., Hauhart, R. and Atkinson, J. P. (1998b). "Structure-function analysis of the active sites of complement receptor type 1." J. Biol. Chem. **273**(15): 8623-9.

Lamagni, T., Efstratiou, A., Vuopio-Varkila, J., Jasir, A. and Schalen, C. (2005). "The epidemiology of severe Streptococcus pyogenes associated disease in Europe." Euro. Surveill. **10**(9).

Lancefield, R. (1928). "THE ANTIGENIC COMPLEX OF STREPTOCOCCUS HÆMOLYTICUS: I. DEMONSTRATION OF A TYPE-SPECIFIC SUBSTANCE IN EXTRACTS OF STREPTOCOCCUS HÆMOLYTICUS." J. Exp. Med. **47**: 91-103.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M. (1993). "PROCHECK: a program to check the stereochemical quality of protein structures." Journal of Applied Crystallography **26**: 283-291.

Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R. and Thornton, J. M. (1996). "AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR." J Biomol NMR **8**(4): 477-86.

Leung, E., Blom, A. M., Clemenza, L. and Isenman, D. E. (2006). "The Complement Regulator C4b-Binding Protein (C4BP) Interacts with both the C4c and C4dg Subfragments of the Parent C4b Ligand: Evidence for Synergy in C4BP Subsite Binding." Biochemistry **45**(27): 8378-92.

Lindahl, G., Sjobring, U. and Johnsson, E. (2000). "Human complement regulators: a major target for pathogenic microorganisms." Curr. Opin. Immunol. **12**(1): 44-51.

Linge, J. P., Habeck, M., Rieping, W. and Nilges, M. (2004). "Correction of spin diffusion during iterative automated NOE assignment." Journal of Magnetic Resonance **167**(2): 334.

Linge, J. P., Williams, M. A., Spronk, C. A., Bonvin, A. M. and Nilges, M. (2003). "Refinement of protein structures in explicit solvent." Proteins **50**(3): 496-506.

Lukacik, P., Roversi, P., White, J., Esser, D., Smith, G. P., Billington, J., Williams, P. A., Rudd, P. M., Wormald, M. R., Harvey, D. J., Crispin, M. D., Radcliffe, C. M., Dwek, R. A., Evans, D. J., Morgan, B. P., Smith, R. A. and Lea, S. M. (2004). "Complement regulation at the molecular level: the structure of decay-accelerating factor." Proc. Natl. Acad. Sci. USA **101**(5): 1279-84.

Macura, S. and Ernst, R. R. (1980). "Elucidation of cross relaxation in liquids by two-dimensional N.M.R. spectroscopy." Molecular Physics **41**(1): 95.

Matsushita, M., Endo, Y. and Fujita, T. (2000). "Cutting edge: complement-activating complex of ficolin and mannose-binding lectin-associated serine protease." J Immunol **164**(5): 2281-4.

Matsushita, M. and Fujita, T. (1995). "Cleavage of the third component of complement (C3) by mannose-binding protein-associated serine protease (MASP) with subsequent complement activation." Immunobiology **194**(4-5): 443-48.

McGuffin, L. J., Bryson, K. and Jones, D. T. (2000). "The PSIPRED protein structure prediction server." Bioinformatics **16**(4): 404-5.

Meri, T., Blom, A. M., Hartmann, A., Lenk, D., Meri, S. and Zipfel, P. F. (2004). "The hyphal and yeast forms of Candida albicans bind the complement regulator C4b-binding protein." Infect. Immun. **72**(11): 6633-41.

Meri, T., Cutler, S. J., Blom, A. M., Meri, S. and Jokiranta, T. S. (2006). "Relapsing Fever Spirochetes Borrelia recurrentis and B. duttonii Acquire Complement Regulators C4b-Binding Protein and Factor H." Infect Immun **74**(7): 4157-63.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). "Equation of State Calculations by Fast Computing Machines." The Journal of Chemical Physics **21**(6): 1087.

Misasi, R., Huemer, H. P., Schwaeble, W., Solder, E., Larcher, C. and Dierich, M. P. (1989). "Human complement factor H: an additional gene product of 43 kDa isolated from human plasma shows cofactor activity for the cleavage of the third component of complement." Eur J Immunol **19**(9): 1765-8.

Molina, H., Holers, V. M., Li, B., Fung, Y., Mariathasan, S., Goellner, J., Strauss-Schoenberger, J., Karr, R. W. and Chaplin, D. D. (1996). "Markedly impaired humoral immune response in mice deficient in complement receptors 1 and 2." Proc Natl Acad Sci U S A **93**(8): 3357-61.

Montelione, G. T., Lyons, B. A., Emerson, S. D. and Tashiro, M. (1992). "An efficient triple resonance experiment using carbon-13 isotropic mixing for determining sequence-specific resonance assignments of isotopically-enriched proteins ." J. Am. Chem. Soc. **114**(27): 10974.

Morfeldt, E., Berggård, K., Persson, J., Drakenberg, T., Johnsson, E., Lindahl, E., Linse, S. and Lindahl, G. (2001). "Isolated hypervariable regions derived from streptococcal M proteins specifically bind human C4b-binding protein: implications for antigenic variation." J. Immunol. **167**(7): 3870-7.

Morris, G. A. and Freeman, R. (1979). "Enhancement of Nuclear Magnetic Resonance Signals by Polarization Transfer." J Am Chem Soc **101**: 760.

Murphy, G. A., Spedale, E. J., Powell, S. T., Pillus, L., Schultz, S. C. and Chen, L. (2003). "The Sir4 C-terminal coiled coil is required for telomeric and mating type silencing in Saccharomyces cerevisiae." J Mol Biol **334**(4): 769-80.

Murthy, K. H., Smith, S. A., Ganesh, V. K., Judge, K. W., Mullin, N., Barlow, P. N., Ogata, C. M. and Kotwal, G. J. (2001). "Crystal structure of a complement control protein that regulates both pathways of complement activation and binds heparan sulfate proteoglycans." Cell **104**(2): 301-11.

Nederveen, A. J., Doreleijers, J. F., Vranken, W., Miller, Z., Spronk, C. A., Nabuurs, S. B., Guntert, P., Livny, M., Markley, J. L., Nilges, M., Ulrich, E. L., Kaptein, R. and Bonvin, A. M. (2005). "RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank." Proteins **59**(4): 662-72.

Ngampasutadol, J., Ram, S., Blom, A. M., Jarva, H., Jerse, A. E., Lien, E., Goguen, J., Gulati, S. and Rice, P. A. (2005). "Human C4b-binding protein selectively interacts with Neisseria gonorrhoeae and results in species-specific infection." Proc. Natl. Acad. Sci. USA **102**(47): 17142-7.

Nicholson-Weller, A., Burge, J., Fearon, D. T., Weller, P. F. and Austen, K. F. (1982). "Isolation of a human erythrocyte membrane glycoprotein with decay-accelerating activity for C3 convertases of the complement system." J Immunol **129**(1): 184-9.

Nilges, M. (1993). "A calculation strategy for the structure determination of symmetric dimers by 1H NMR." Proteins **17**(3): 297-309.

Nilges, M. (1995). "Calculation of Protein Structures with Ambiguous Distance Restraints - Automated Assignment of Ambiguous Noe Crosspeaks and Disulfide Connectivities." J. Mol. Biol. **245**(5): 645-660.

Nilges, M., Clore, G. M. and Gronenborn, A. M. (1988a). "Determination of 3-Dimensional Structures of Proteins from Interproton Distance Data by Dynamical Simulated Annealing from a Random Array of Atoms - Circumventing Problems Associated with Folding." FEBS Lett. **239**(1): 129-136.

Nilges, M., Clore, G. M. and Gronenborn, A. M. (1988b). "Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations." FEBS Lett **229**(2): 317-24.

Nilges, M., Habeck, M., O'Donoghue S, I. and Rieping, W. (2006). "Error distribution derived NOE distance restraints." Proteins.

Nilges, M., Macias, M. J., Odonoghue, S. I. and Oschkinat, H. (1997). "Automated NOESY interpretation with ambiguous distance restraints: The refined NMR solution structure of the pleckstrin homology domain from beta-spectrin." J. Mol. Biol. **269**(3): 408-422.

Nilson, B. H., Frick, I. M., Akesson, P., Forsen, S., Bjorck, L., Akerstrom, B. and Wikstrom, M. (1995). "Structure and stability of protein H and the M1 protein from Streptococcus pyogenes. Implications for other surface proteins of gram-positive bacteria." Biochemistry **34**(41): 13688-98.

Nordström, T., Blom, A. M., Forsgren, A. and Riesbeck, K. (2004). "The emerging pathogen Moraxella catarrhalis interacts with complement inhibitor C4b binding protein through ubiquitous surface proteins A1 and A2." J. Immunol. **173**(7): 4598-606.

Norman, D. G., Barlow, P. N., Baron, M., Day, A. J., Sim, R. B. and Campbell, I. D. (1991). "Three-dimensional structure of a complement control protein module in solution." J. Mol. Biol. **219**(4): 717-25.

Ogata, R. T., Mathias, P., Bradt, B. M. and Cooper, N. R. (1993). "Murine C4b-binding protein. Mapping of the ligand binding site and the N-terminus of the pre-protein." J. Immunol. **150**(6): 2273-80.

Ottiger, M., Delaglio, F. and Bax, A. (1998). "Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra." J. Magn. Reson. **131**(2): 373-8.

Pascal, S. M., Muhandiram, D. R., Yamazaki, T., Formankay, J. D. and Kay, L. E. (1994). "Simultaneous Acquisition of 15N- and 13C-Edited NOE Spectra of Proteins Dissolved in H2O." J. Magn. Reson. Ser. B **103**(2): 197.

Perkins, S. J., Chung, L. P. and Reid, K. B. (1986). "Unusual ultrastructure of complement-component-C4b-binding protein of human complement by synchrotron X-ray scattering and hydrodynamic analysis." Biochem. J. **233**(3): 799-807.

Perkins, S. J., Haris, P. I., Sim, R. B. and Chapman, D. (1988). "A study of the structure of human complement component factor H by Fourier transform infrared spectroscopy and secondary structure averaging methods." Biochemistry **27**(11): 4004-12.

Permi, P., Heikkinen, S., Kilpelainen, I. and Annila, A. (1999). "Measurement of (1)J(NC') and (2)J(H(N))(C') couplings from spin-state-selective two-dimensional correlation spectrum." J. Magn. Reson. **140**(1): 32-40.

Petros, A. M., Mueller, L. and Kopple, K. D. (1990). "NMR identification of protein surfaces using paramagnetic probes." Biochemistry **29**(43): 10041-8.

Powell, M. J. D. (1977). "Restart procedures for the conjugate gradient method." Math. Program. **12**: 241-254.

Prasadarao, N. V., Blom, A. M., Villoutreix, B. O. and Linsangan, L. C. (2002). "A novel interaction of outer membrane protein A with C4b binding protein mediates serum resistance of Escherichia coli K1." J. Immunol. **169**(11): 6352-60.

Presanis, J. S., Kojima, M. and Sim, R. B. (2003). "Biochemistry and genetics of mannan-binding lectin (MBL)." Biochem Soc Trans **31**(Pt 4): 748-52.

Prompers, J. J., Groenewegen, A., Hilbers, C. W. and Pepermans, H. A. M. (1998). "Two-Dimensional NMR Experiments for the Assignment of Aromatic Side Chains in13C-labeled Proteins." J. Magn. Reson. **130**(1): 68.

Prota, A. E., Sage, D. R., Stehle, T. and Fingeroth, J. D. (2002). "The crystal structure of human CD21: Implications for Epstein-Barr virus and C3d binding." Proc. Natl. Acad. Sci. USA **99**(16): 10641-6.

Ram, S., Cullinane, M., Blom, A. M., Gulati, S., McQuillen, D. P., Monks, B. G., O'Connell, C., Boden, R., Elkins, C., Pangburn, M. K., Dahlbäck, B. and Rice, P. A. (2001). "Binding of C4b-binding protein to porin: a molecular mechanism of serum resistance of Neisseria gonorrhoeae." J. Exp. Med. **193**(3): 281-95.

Ramachandran, G. N., Ramakrishnan, C. and Sasisekharan, V. (1963). "Stereochemistry of polypeptide chain configurations." J Mol Biol **7**: 95-9.

Reid, K. B. and Day, A. J. (1989). "Structure-function relationships of the complement components." Immunol. Today **10**(6): 177-80.

Rieping, W., Habeck, M. and Nilges, M. (2005). "Inferential structure determination." Science **309**(5732): 303-6.

Robinson, J. H. and Kehoe, M. A. (1992). "Group A streptococcal M proteins: virulence factors and protective antigens." Immunol. Today **13**(9): 362-7.

Rost, B. (1996). "PHD: predicting one-dimensional protein structure by profile-based neural networks." Methods Enzymol. **266**: 525-39.

Sass, H. J., Musco, G., Stahl, S. J., Wingfield, P. T. and Grzesiek, S. (2001). "An easy way to include weak alignment constraints into NMR structure calculations." J Biomol NMR **21**(3): 275-80.

Scharfstein, J., Ferreira, A., Gigli, I. and Nussenzweig, V. (1978). "Human C4-binding protein. I. Isolation and characterization." J. Exp. Med. **148**(1): 207-22.

Schubert, M., Labudde, D., Oschkinat, H. and Schmieder, P. (2002). "A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on 13C chemical shift statistics." J. Biomol. NMR **24**(2): 149-54.

Schwalbe, R. A., Dahlback, B. and Nelsestuen, G. L. (1990). "Independent association of serum amyloid P component, protein S, and complement C4b with complement C4b-binding protein and subsequent association of the complex with membranes." J. Biol. Chem. **265**(35): 21749-57.

Seya, T., Turner, J. R. and Atkinson, J. P. (1986). "Purification and characterization of a membrane protein (gp45-70) that is a cofactor for cleavage of C3b and C4b." J Exp Med **163**(4): 837-55.

Shindyalov, I. N. and Bourne, P. E. (1998). "Protein structure alignment by incremental combinatorial extension (CE) of the optimal path." Protein Eng. **11**(9): 739-47.

Sibisi, S., Skilling, J., Brereton, R. G., Laue, E. D. and Staunton, J. (1984). "Maximum entropy signal processing in practical NMR spectroscopy." Nature **311**(5985): 446.

Sim, R. B. and DiScipio, R. G. (1982). "Purification and structural studies on the complement-system control protein beta 1H (Factor H)." Biochem. J. **205**(2): 285-93.

Sjoberg, A. P., Trouw, L. A., McGrath, F. D., Hack, C. E. and Blom, A. M. (2006). "Regulation of complement activation by C-reactive protein: targeting of the inhibitory activity of C4b-binding protein." J Immunol **176**(12): 7612-20.

Sklenar, V., Piotto, M., Leppik, R. and Saudek, V. (1993). "Gradient-Tailored Water Suppression for 1H-15N HSQC Experiments Optimized to Retain Full Sensitivity." J. Magn. Reson. Ser. A **102**(2): 241.

Smith, B. O., Mallin, R. L., Krych-Goldberg, M., Wang, X., Hauhart, R. E., Bromek, K., Uhrin, D., Atkinson, J. P. and Barlow, P. N. (2002). "Structure of the C3b binding site of CR1 (CD35), the immune adherence receptor." Cell **108**(6): 769-80.

Soares, D. C., Gerloff, D. L., Syme, N. R., Coulson, A. F., Parkinson, J. and Barlow, P. N. (2005). "Large-scale modelling as a route to multiple surface comparisons of the CCP module family." Protein Eng Des Sel.

Solomon, I. (1955). "Relaxation Processes in a System of Two Spins." Physical Review **99**(2): 559.

Suankratay, C., Mold, C., Zhang, Y., Lint, T. F. and Gewurz, H. (1999). "Mechanism of complement-dependent haemolysis via the lectin pathway: role of the complement regulatory proteins." Clin. Exp. Immunol. **117**(3): 442-8.

Summerfield, J. A., Sumiya, M., Levin, M. and Turner, M. W. (1997). "Association of mutations in mannose binding protein gene with childhood infection in consecutive hospital series." Bmj **314**(7089): 1229-32.

Szakonyi, G., Guthridge, J. M., Li, D., Young, K., Holers, V. M. and Chen, X. S. (2001). "Structure of complement receptor 2 in complex with its C3d ligand." Science **292**(5522): 1725-8.

Sánchez-Corral, P., Criado Garcia, O. and Rodríguez de Córdoba, S. (1995). "Isoforms of human C4b-binding protein. I. Molecular basis for the C4BP isoform pattern and its variations in human plasma." J. Immunol. **155**(8): 4030-6.

Takahashi, H., Nakanishi, T., Kami, K., Arata, Y. and Shimada, I. (2000). "A novel NMR method for determining the interfaces of large protein-protein complexes." Nat Struct Biol **7**(3): 220-3.

Taylor, P. R., Carugati, A., Fadok, V. A., Cook, H. T., Andrews, M., Carroll, M. C., Savill, J. S., Henson, P. M., Botto, M. and Walport, M. J. (2000). "A hierarchical role for classical pathway complement proteins in the clearance of apoptotic cells in vivo." J Exp Med **192**(3): 359-66.

Thern, A., Stenberg, L., Dahlbäck, B. and Lindahl, G. (1995). "Ig-binding surface proteins of Streptococcus pyogenes also bind human C4b-binding protein (C4BP), a regulatory component of the complement system." J. Immunol. **154**(1): 375-86.

Trouw, L., Nilsson, S., Goncalvez, I., Landberg, G. and Blom, A. M. (2005). "C4b-binding protein binds to necrotic cells and DNA, which limits DNA release and inhibits complement activation." J. Exp. Med. **201:** 1937-48.

Uhrinova, S., Lin, F., Ball, G., Bromek, K., Uhrin, D., Medof, M. E. and Barlow, P. N. (2003). "Solution structure of a functionally active fragment of decay-accelerating factor." Proc. Natl. Acad. Sci. USA **100**(8): 4718-23.

Villoutreix, B. O., Blom, A. M., Webb, J. and Dahlback, B. (1999). "The complement regulator C4b-binding protein analyzed by molecular modeling, bioinformatics and computer-aided experimental design." Immunopharmacology **42**(1-3): 121-34.

Vuister, G. W. and Bax, A. (1992). "Resolution enhancement and spectral editing of uniformly 13C-enriched proteins by homonuclear broadband 13C decoupling." J. Magn. Reson. **98**(2): 428.

Wang, A. C., Lodi, P. J., Qin, J., Vuister, G. W., Gronenborn, A. M. and Clore, G. M. (1994). "An Efficient Triple-Resonance Experiment for Proton-Directed Sequential Backbone Assignment of Medium-Sized Proteins." J. Magn. Reson. Ser. B **105**(2): 196.

Webb, J. H., Blom, A. M. and Dahlback, B. (2002). "Vitamin K-dependent protein S localizing complement regulator C4b-binding protein to the surface of apoptotic cells." J Immunol **169**(5): 2580-6.

Weis, J. J., Toothaker, L. E., Smith, J. A., Weis, J. H. and Fearon, D. T. (1988). "Structure of the human B lymphocyte receptor for C3d and the Epstein-Barr virus and relatedness to other members of the family of C3/C4 binding proteins." J Exp Med **167**(3): 1047-66.

Wiles, A. P., Shaw, G., Bright, J., Perczel, A., Campbell, I. D. and Barlow, P. N. (1997). "NMR studies of a viral protein that mimics the regulators of complement activation." J. Mol. Biol. **272**(2): 253-65.

Williams, P., Chaudhry, Y., Goodfellow, I. G., Billington, J., Powell, R., Spiller, O. B., Evans, D. J. and Lea, S. (2003). "Mapping CD55 function. The structure of two pathogen-binding domains at 1.7 A." J. Biol. Chem. **278**(12): 10691-6.

Wishart, D. S. and Sykes, B. D. (1994). "The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data." J. Biomol. NMR **4**(2): 171-80.

Wishart, D. S., Sykes, B. D. and Richards, F. M. (1992). "The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy." Biochemistry **31**(6): 1647-51.

Wooster, D. G., Maruvada, R., Blom, A. M. and Prasadarao, N. V. (2006). "Logarithmic phase Escherichia coli K1 efficiently avoids serum killing by promoting C4bp-mediated C3b and C4b degradation." Immunology **117**(4): 482-493.

Yamazaki, T., Forman-Kay, J. D. and Kay, L. E. (1993). "Two-dimensional NMR experiments for correlating carbon-13.beta. and proton.delta./.epsilon. chemical shifts of aromatic residues in 13C-labeled proteins via scalar couplings." J. Am. Chem. Soc. **115**(23): 11054 - 11055.

Zipfel, P. F., Jokiranta, T. S., Hellwage, J., Koistinen, V. and Meri, S. (1999). "The factor H protein family." Immunopharmacology **42**(1-3): 53-60.

Zweckstetter, M. and Bax, A. (2000). "Prediction of Sterically Induced Alignment in a Dilute Liquid Crystalline Phase: Aid to Protein Structure Determination by NMR." J Am Chem Soc **122**: 3791-2.

Zweckstetter, M., Hummer, G. and Bax, A. (2004). "Prediction of charge-induced molecular alignment of biomolecules dissolved in dilute liquid-crystalline phases." Biophys J **86**(6): 3444-60.

Appendix A

ASSIGNMENT OF C4BP12 CHEMICAL SHIFTS

Residue	Atom	Shift (ppm)	Residue	Atom	Shift (ppm)
Met^{0}	HA	4.07	Pro^{4}	CD	49.31
Met^{0}	HB2	1.96	Pro^{5}	HA	3.48
Met^{0}	HB3	1.96	Pro^{5}	HB2	0.24
Met^{0}	HG2	2.54	Pro^{5}	HB3	1.14
Met^{0}	HG3	2.54	Pro^{5}	HG2	0.92
Met^{0}	C	171.22	Pro^{5}	HG3	1.38
Met^{0}	CA	54.85	Pro^{5}	HD2	3.01
Met^{0}	CB	32.70	Pro^{5}	HD3	3.47
Met^{0}	CG	30.90	Pro^{5}	CA	60.70
Asn^{1}	H	8.27	Pro^{5}	CB	30.20
Asn^{1}	HA	4.91	Pro^{5}	CG	26.58
Asn^{1}	HB2	2.62	Pro^{5}	CD	50.07
Asn^{1}	HB3	2.76	Pro^{6}	HA	4.32
Asn^{1}	HD21	7.26	Pro^{6}	HB2	1.64
Asn^{1}	HD22	6.93	Pro^{6}	HB3	2.16
Asn^{1}	C	174.82	Pro^{6}	HG2	1.48
Asn^{1}	CA	53.04	Pro^{6}	HG3	1.99
Asn^{1}	CB	40.37	Pro^{6}	HD2	2.74
Asn^{1}	N	119.57	Pro^{6}	HD3	3.39
Asn^{1}	ND2	113.12	Pro^{6}	C	176.41
Cys^{2}	H	8.85	Pro^{6}	CA	61.65
Cys^{2}	HA	4.56	Pro^{6}	CB	31.78
Cys^{2}	HB2	2.24	Pro^{6}	CG	26.45
Cys^{2}	HB3	3.04	Pro^{6}	CD	50.21
Cys^{2}	C	175.94	Thr^{7}	H	8.29
Cys^{2}	CA	58.49	Thr^{7}	HA	3.82
Cys^{2}	CB	45.86	Thr^{7}	HB	3.86
Cys^{2}	N	116.36	Thr^{7}	HG2	1.05
Gly^{3}	H	8.53	Thr^{7}	C	173.75
Gly^{3}	HA2	3.96	Thr^{7}	CA	63.48
Gly^{3}	HA3	4.86	Thr^{7}	CB	68.89
Gly^{3}	C	170.70	Thr^{7}	CG2	21.73
Gly^{3}	CA	44.12	Thr^{7}	N	117.15
Gly^{3}	N	110.16	Leu^{8}	H	8.43
Pro^{4}	HA	4.65	Leu^{8}	HA	4.36
Pro^{4}	HB2	1.78	Leu^{8}	HB2	-0.09
Pro^{4}	HB3	2.44	Leu^{8}	HB3	1.24
Pro^{4}	HG2	2.06	Leu^{8}	HG	1.03
Pro^{4}	HG3	2.19	Leu^{8}	HD1	-0.29
Pro^{4}	HD2	3.64	Leu^{8}	HD2	0.07
Pro^{4}	HD3	3.92	Leu^{8}	C	176.84
Pro^{4}	CA	61.91	Leu^{8}	CA	52.77
Pro^{4}	CB	30.85	Leu^{8}	CB	42.88
Pro^{4}	CG	28.21	Leu^{8}	CG	26.19

Residue	Atom	Shift (ppm)	Residue	Atom	Shift (ppm)
Leu[8]	CD1	25.25	Pro[13]	CA	62.24
Leu[8]	CD2	22.66	Pro[13]	CB	31.49
Leu[8]	N	128.18	Pro[13]	CG	26.57
Ser[9]	H	8.57	Pro[13]	CD	49.93
Ser[9]	HA	4.03	Met[14]	H	8.23
Ser[9]	HB2	3.81	Met[14]	HA	4.09
Ser[9]	HB3	3.93	Met[14]	HB2	1.57
Ser[9]	C	173.86	Met[14]	HB3	1.85
Ser[9]	CA	59.41	Met[14]	HG2	2.24
Ser[9]	CB	63.44	Met[14]	HG3	2.24
Ser[9]	N	113.52	Met[14]	C	176.19
Phe[10]	H	5.81	Met[14]	CA	56.86
Phe[10]	HA	3.53	Met[14]	CB	33.79
Phe[10]	HB2	2.15	Met[14]	CG	32.27
Phe[10]	HB3	2.51	Met[14]	N	120.64
Phe[10]	HD1	6.66	Asp[15]	H	8.14
Phe[10]	HD2	6.66	Asp[15]	HA	4.56
Phe[10]	HE1	6.29	Asp[15]	HB2	2.48
Phe[10]	HE2	6.29	Asp[15]	HB3	2.70
Phe[10]	HZ	7.00	Asp[15]	C	173.72
Phe[10]	C	173.68	Asp[15]	CA	54.07
Phe[10]	CA	53.87	Asp[15]	CB	40.88
Phe[10]	CB	41.27	Asp[15]	N	118.07
Phe[10]	CD1	132.55	Ile[16]	H	7.81
Phe[10]	CD2	132.55	Ile[16]	HA	4.12
Phe[10]	CE1	131.39	Ile[16]	HB	1.90
Phe[10]	CE2	131.39	Ile[16]	HG12	1.09
Phe[10]	CZ	131.16	Ile[16]	HG13	1.35
Phe[10]	N	109.36	Ile[16]	HG2	0.83
Ala[11]	H	6.46	Ile[16]	HD1	0.78
Ala[11]	HA	4.91	Ile[16]	C	176.07
Ala[11]	HB	1.23	Ile[16]	CA	61.89
Ala[11]	C	173.30	Ile[16]	CB	38.86
Ala[11]	CA	51.71	Ile[16]	CG1	27.25
Ala[11]	CB	25.88	Ile[16]	CG2	17.58
Ala[11]	N	120.89	Ile[16]	CD1	13.49
Ala[12]	H	8.63	Ile[16]	N	118.45
Ala[12]	HA	4.61	Thr[17]	H	8.21
Ala[12]	HB	1.17	Thr[17]	HA	4.34
Ala[12]	CA	49.34	Thr[17]	HB	4.17
Ala[12]	CB	20.81	Thr[17]	HG2	1.15
Ala[12]	N	121.27	Thr[17]	C	174.54
Pro[13]	HA	3.12	Thr[17]	CA	61.45
Pro[13]	HB2	1.54	Thr[17]	CB	69.66
Pro[13]	HB3	1.63	Thr[17]	CG2	21.90
Pro[13]	HG2	1.80	Thr[17]	N	116.49
Pro[13]	HG3	1.80	Leu[18]	H	8.07
Pro[13]	HD2	3.26	Leu[18]	HA	4.42
Pro[13]	HD3	3.42	Leu[18]	HB2	1.62
Pro[13]	C	176.79	Leu[18]	HB3	1.69

Residue	Atom	Shift (ppm)	Residue	Atom	Shift (ppm)
Leu^{18}	HG	-	Arg^{22}	CA	54.49
Leu^{18}	HD1	0.89	Arg^{22}	CB	32.66
Leu^{18}	HD2	0.86	Arg^{22}	CG	26.68
Leu^{18}	C	177.45	Arg^{22}	CD	43.27
Leu^{18}	CA	55.48	Arg^{22}	N	120.26
Leu^{18}	CB	42.10	Arg^{22}	NE	84.34
Leu^{18}	CG	27.23	Phe^{23}	H	8.84
Leu^{18}	CD1	25.14	Phe^{23}	HA	4.67
Leu^{18}	CD2	23.78	Phe^{23}	HB2	2.72
Leu^{18}	N	124.58	Phe^{23}	HB3	3.03
Thr^{19}	H	8.09	Phe^{23}	HD1	7.31
Thr^{19}	HA	4.31	Phe^{23}	HD2	7.31
Thr^{19}	HB	4.29	Phe^{23}	HE1	7.21
Thr^{19}	HG2	1.12	Phe^{23}	HE2	7.21
Thr^{19}	C	174.78	Phe^{23}	HZ	-
Thr^{19}	CA	61.99	Phe^{23}	C	174.58
Thr^{19}	CB	69.33	Phe^{23}	CA	56.62
Thr^{19}	CG2	21.66	Phe^{23}	CB	42.22
Thr^{19}	N	112.44	Phe^{23}	CD1	132.65
Glu^{20}	H	7.68	Phe^{23}	CD2	132.65
Glu^{20}	HA	4.26	Phe^{23}	CE1	131.10
Glu^{20}	HB2	1.73	Phe^{23}	CE2	131.10
Glu^{20}	HB3	1.73	Phe^{23}	CZ	-
Glu^{20}	HG2	2.03	Phe^{23}	N	120.85
Glu^{20}	HG3	2.03	Lys^{24}	H	8.55
Glu^{20}	C	175.37	Lys^{24}	HA	4.39
Glu^{20}	CA	56.24	Lys^{24}	HB2	1.74
Glu^{20}	CB	30.11	Lys^{24}	HB3	1.76
Glu^{20}	CG	35.46	Lys^{24}	HG2	1.43
Glu^{20}	N	122.10	Lys^{24}	HG3	1.50
Thr^{21}	H	8.15	Lys^{24}	HD2	1.65
Thr^{21}	HA	4.31	Lys^{24}	HD3	1.65
Thr^{21}	HB	4.27	Lys^{24}	HE2	2.98
Thr^{21}	HG1	4.64	Lys^{24}	HE3	2.98
Thr^{21}	HG2	1.06	Lys^{24}	C	178.64
Thr^{21}	C	173.41	Lys^{24}	CA	56.13
Thr^{21}	CA	61.58	Lys^{24}	CB	33.54
Thr^{21}	CB	69.95	Lys^{24}	CG	24.64
Thr^{21}	CG2	21.68	Lys^{24}	CD	29.36
Thr^{21}	N	111.06	Lys^{24}	CE	41.93
Arg^{22}	H	7.32	Lys^{24}	N	120.99
Arg^{22}	HA	4.94	Thr^{25}	H	8.60
Arg^{22}	HB2	1.53	Thr^{25}	HA	3.52
Arg^{22}	HB3	1.53	Thr^{25}	HB	3.79
Arg^{22}	HG2	1.48	Thr^{25}	HG2	1.04
Arg^{22}	HG3	1.50	Thr^{25}	C	174.53
Arg^{22}	HD2	3.05	Thr^{25}	CA	65.66
Arg^{22}	HD3	3.11	Thr^{25}	CB	69.22
Arg^{22}	HE	7.19	Thr^{25}	CG2	21.84
Arg^{22}	C	174.90	Thr^{25}	N	120.23

Residue	Atom	Shift (ppm)	Residue	Atom	Shift (ppm)
Gly^{26}	H	9.18	Lys^{30}	CB	33.43
Gly^{26}	HA2	3.48	Lys^{30}	CG	-
Gly^{26}	HA3	4.37	Lys^{30}	CD	-
Gly^{26}	C	174.68	Lys^{30}	CE	41.97
Gly^{26}	CA	44.53	Lys^{30}	N	122.34
Gly^{26}	N	115.81	Tyr^{31}	H	8.40
Thr^{27}	H	7.87	Tyr^{31}	HA	4.81
Thr^{27}	HA	4.06	Tyr^{31}	HB2	2.51
Thr^{27}	HB	4.01	Tyr^{31}	HB3	2.65
Thr^{27}	HG2	1.51	Tyr^{31}	HD1	6.52
Thr^{27}	C	173.30	Tyr^{31}	HD2	6.52
Thr^{27}	CA	65.81	Tyr^{31}	HE1	6.22
Thr^{27}	CB	69.79	Tyr^{31}	HE2	6.22
Thr^{27}	CG2	21.89	Tyr^{31}	C	174.28
Thr^{27}	N	117.95	Tyr^{31}	CA	56.99
Thr^{28}	H	8.73	Tyr^{31}	CB	44.27
Thr^{28}	HA	5.43	Tyr^{31}	CD1	131.54
Thr^{28}	HB	3.95	Tyr^{31}	CD2	131.54
Thr^{28}	HG2	1.14	Tyr^{31}	CE1	117.77
Thr^{28}	C	173.79	Tyr^{31}	CE2	117.77
Thr^{28}	CA	60.87	Tyr^{31}	N	126.21
Thr^{28}	CB	71.14	Thr^{32}	H	8.94
Thr^{28}	CG2	21.38	Thr^{32}	HA	4.71
Thr^{28}	N	122.34	Thr^{32}	HB	3.96
Leu^{29}	H	9.27	Thr^{32}	HG2	1.10
Leu^{29}	HA	4.77	Thr^{32}	C	173.63
Leu^{29}	HB2	1.30	Thr^{32}	CA	59.47
Leu^{29}	HB3	1.62	Thr^{32}	CB	71.15
Leu^{29}	HG	1.47	Thr^{32}	CG2	20.31
Leu^{29}	HD1	0.08	Thr^{32}	N	112.20
Leu^{29}	HD2	0.76	Cys^{33}	H	8.43
Leu^{29}	C	176.36	Cys^{33}	HA	5.09
Leu^{29}	CA	53.12	Cys^{33}	HB2	2.79
Leu^{29}	CB	45.69	Cys^{33}	HB3	3.23
Leu^{29}	CG	26.66	Cys^{33}	C	175.80
Leu^{29}	CD1	26.18	Cys^{33}	CA	53.95
Leu^{29}	CD2	22.59	Cys^{33}	CB	38.50
Leu^{29}	N	126.21	Cys^{33}	N	122.21
Lys^{30}	H	8.75	Leu^{34}	H	7.39
Lys^{30}	HA	4.50	Leu^{34}	HA	4.40
Lys^{30}	HB2	1.67	Leu^{34}	HB2	1.18
Lys^{30}	HB3	1.85	Leu^{34}	HB3	1.60
Lys^{30}	HG2	-	Leu^{34}	HG	1.01
Lys^{30}	HG3	-	Leu^{34}	HD1	0.56
Lys^{30}	HD2	-	Leu^{34}	HD2	0.81
Lys^{30}	HD3	-	Leu^{34}	C	174.09
Lys^{30}	HE2	2.93	Leu^{34}	CA	54.24
Lys^{30}	HE3	2.93	Leu^{34}	CB	40.63
Lys^{30}	C	175.89	Leu^{34}	CG	27.65
Lys^{30}	CA	56.83	Leu^{34}	CD1	25.35

Residue	Atom	Shift (ppm)	Residue	Atom	Shift (ppm)
Leu^{34}	CD2	22.45	Arg^{39}	HB3	1.92
Leu^{34}	N	124.87	Arg^{39}	HG2	1.72
Pro^{35}	HA	4.34	Arg^{39}	HG3	1.80
Pro^{35}	HB2	1.79	Arg^{39}	HD2	3.18
Pro^{35}	HB3	2.28	Arg^{39}	HD3	3.18
Pro^{35}	HG2	1.95	Arg^{39}	HE	7.06
Pro^{35}	HG3	2.11	Arg^{39}	C	176.56
Pro^{35}	HD2	3.57	Arg^{39}	CA	57.55
Pro^{35}	HD3	3.77	Arg^{39}	CB	30.91
Pro^{35}	C	177.02	Arg^{39}	CG	26.98
Pro^{35}	CA	64.66	Arg^{39}	CD	43.99
Pro^{35}	CB	31.75	Arg^{39}	N	125.12
Pro^{35}	CG	28.15	Arg^{39}	NE	84.77
Pro^{35}	CD	50.61	Ser^{40}	H	8.44
Gly^{36}	H	8.28	Ser^{40}	HA	4.52
Gly^{36}	HA2	3.25	Ser^{40}	HB2	3.46
Gly^{36}	HA3	3.70	Ser^{40}	HB3	3.91
Gly^{36}	C	171.71	Ser^{40}	C	174.23
Gly^{36}	CA	44.92	Ser^{40}	CA	56.99
Gly^{36}	N	111.29	Ser^{40}	CB	64.46
Tyr^{37}	H	8.14	Ser^{40}	N	118.32
Tyr^{37}	HA	5.01	His^{41}	H	8.06
Tyr^{37}	HB2	2.63	His^{41}	HA	4.72
Tyr^{37}	HB3	3.41	His^{41}	HB2	3.12
Tyr^{37}	HD1	6.69	His^{41}	HB3	3.36
Tyr^{37}	HD2	6.69	His^{41}	HD2	-
Tyr^{37}	HE1	6.72	His^{41}	HE2	-
Tyr^{37}	HE2	6.72	His^{41}	C	174.63
Tyr^{37}	C	175.22	His^{41}	CA	55.76
Tyr^{37}	CA	56.64	His^{41}	CB	30.47
Tyr^{37}	CB	40.75	His^{41}	CD2	-
Tyr^{37}	CD1	133.15	His^{41}	CE2	-
Tyr^{37}	CD2	133.15	His^{41}	N	118.04
Tyr^{37}	CE1	118.18	Ser^{42}	H	8.16
Tyr^{37}	CE2	118.18	Ser^{42}	HA	4.37
Tyr^{37}	N	116.68	Ser^{42}	HB2	3.87
Val^{38}	H	9.25	Ser^{42}	HB3	3.94
Val^{38}	HA	4.35	Ser^{42}	C	174.27
Val^{38}	HB	1.72	Ser^{42}	CA	59.09
Val^{38}	HG1	0.53	Ser^{42}	CB	64.12
Val^{38}	HG2	0.56	Ser^{42}	N	115.73
Val^{38}	C	175.11	Thr^{43}	H	8.26
Val^{38}	CA	59.35	Thr^{43}	HA	4.21
Val^{38}	CB	34.84	Thr^{43}	HB	4.10
Val^{38}	CG1	20.55	Thr^{43}	HG2	1.17
Val^{38}	CG2	20.43	Thr^{43}	C	174.38
Val^{38}	N	118.08	Thr^{43}	CA	61.75
Arg^{39}	H	8.38	Thr^{43}	CB	69.18
Arg^{39}	HA	4.66	Thr^{43}	CG2	21.75
Arg^{39}	HB2	1.92	Thr^{43}	N	115.36

Residue	Atom	Shift (ppm)	Residue	Atom	Shift (ppm)
Gln^{44}	H	8.38	Cys^{48}	CA	55.08
Gln^{44}	HA	3.01	Cys^{48}	CB	39.27
Gln^{44}	HB2	-	Cys^{48}	N	127.78
Gln^{44}	HB3	-	Asn^{49}	H	9.06
Gln^{44}	HG2	-	Asn^{49}	HA	4.81
Gln^{44}	HG3	-	Asn^{49}	HB2	2.87
Gln^{44}	HE21	-	Asn^{49}	HB3	3.38
Gln^{44}	HE22	-	Asn^{49}	HD21	7.15
Gln^{44}	CA	53.41	Asn^{49}	HD22	7.15
Gln^{44}	CB	28.28	Asn^{49}	C	176.60
Gln^{44}	CG	-	Asn^{49}	CA	51.72
Gln^{44}	N	126.35	Asn^{49}	CB	39.28
Gln^{44}	NE2	-	Asn^{49}	N	110.36
Thr^{45}	H	6.53	Asn^{49}	ND2	111.60
Thr^{45}	HA	5.20	Ser^{50}	H	8.16
Thr^{45}	HB	3.87	Ser^{50}	HA	4.13
Thr^{45}	HG2	0.89	Ser^{50}	HB2	3.80
Thr^{45}	C	172.88	Ser^{50}	HB3	3.92
Thr^{45}	CA	58.52	Ser^{50}	C	174.48
Thr^{45}	CB	72.67	Ser^{50}	CA	60.94
Thr^{45}	CG2	21.90	Ser^{50}	CB	62.93
Thr^{45}	N	107.44	Ser^{50}	N	112.53
Leu^{46}	H	8.50	Asp^{51}	H	7.65
Leu^{46}	HA	4.88	Asp^{51}	HA	4.74
Leu^{46}	HB2	1.03	Asp^{51}	HB2	2.55
Leu^{46}	HB3	1.85	Asp^{51}	HB3	2.80
Leu^{46}	HG	1.41	Asp^{51}	C	175.85
Leu^{46}	HD1	0.00	Asp^{51}	CA	53.72
Leu^{46}	HD2	-0.15	Asp^{51}	CB	40.64
Leu^{46}	C	176.72	Asp^{51}	N	120.07
Leu^{46}	CA	55.51	Gly^{52}	H	7.93
Leu^{46}	CB	45.18	Gly^{52}	HA2	3.42
Leu^{46}	CG	31.81	Gly^{52}	HA3	3.69
Leu^{46}	CD1	24.64	Gly^{52}	C	173.75
Leu^{46}	CD2	21.40	Gly^{52}	CA	47.24
Leu^{46}	N	123.64	Gly^{52}	N	108.65
Thr^{47}	H	8.49	Glu^{53}	H	7.52
Thr^{47}	HA	5.58	Glu^{53}	HA	4.49
Thr^{47}	HB	3.79	Glu^{53}	HB2	1.75
Thr^{47}	HG2	0.98	Glu^{53}	HB3	1.90
Thr^{47}	C	173.15	Glu^{53}	HG2	2.06
Thr^{47}	CA	61.38	Glu^{53}	HG3	2.16
Thr^{47}	CB	71.77	Glu^{53}	C	174.93
Thr^{47}	CG2	20.85	Glu^{53}	CA	53.22
Thr^{47}	N	116.42	Glu^{53}	CB	31.18
Cys^{48}	H	8.52	Glu^{53}	CG	34.26
Cys^{48}	HA	3.98	Glu^{53}	N	118.95
Cys^{48}	HB2	1.33	Trp^{54}	H	8.40
Cys^{48}	HB3	2.38	Trp^{54}	HA	4.98
Cys^{48}	C	174.34	Trp^{54}	HB2	2.90

Residue	Atom	Shift (ppm)	Residue	Atom	Shift (ppm)
Trp^{54}	HB3	3.07	Asn^{57}	C	174.33
Trp^{54}	HD1	7.28	Asn^{57}	CA	51.20
Trp^{54}	HE1	10.00	Asn^{57}	CB	40.52
Trp^{54}	HE3	7.06	Asn^{57}	N	118.70
Trp^{54}	HZ2	6.94	Asn^{57}	ND2	111.44
Trp^{54}	HZ3	6.71	Thr^{58}	H	8.27
Trp^{54}	HH2	6.24	Thr^{58}	HA	3.58
Trp^{54}	C	176.83	Thr^{58}	HB	3.86
Trp^{54}	CA	56.41	Thr^{58}	HG2	0.53
Trp^{54}	CB	30.12	Thr^{58}	CA	62.90
Trp^{54}	CD1	127.09	Thr^{58}	CB	67.15
Trp^{54}	CE3	120.27	Thr^{58}	CG2	21.61
Trp^{54}	CZ2	114.22	Thr^{58}	N	114.77
Trp^{54}	CZ3	121.56	Phe^{59}	H	-
Trp^{54}	CH2	122.93	Phe^{59}	HA	4.65
Trp^{54}	N	121.95	Phe^{59}	HB2	2.93
Trp^{54}	NE1	129.93	Phe^{59}	HB3	3.57
Val^{55}	H	9.62	Phe^{59}	HD1	7.74
Val^{55}	HA	4.44	Phe^{59}	HD2	7.74
Val^{55}	HB	2.05	Phe^{59}	HE1	6.90
Val^{55}	HG1	0.88	Phe^{59}	HE2	6.90
Val^{55}	HG2	0.96	Phe^{59}	HZ	6.33
Val^{55}	C	176.05	Phe^{59}	CA	56.93
Val^{55}	CA	61.66	Phe^{59}	CB	38.39
Val^{55}	CB	33.87	Phe^{59}	CD1	134.37
Val^{55}	CG1	21.12	Phe^{59}	CD2	134.37
Val^{55}	CG2	20.98	Phe^{59}	CE1	131.03
Val^{55}	N	123.00	Phe^{59}	CE2	131.03
Tyr^{56}	H	7.81	Phe^{59}	CZ	129.54
Tyr^{56}	HA	4.99	Phe^{59}	N	-
Tyr^{56}	HB2	3.21	Cys^{60}	H	6.54
Tyr^{56}	HB3	3.65	Cys^{60}	HA	5.55
Tyr^{56}	HD1	6.82	Cys^{60}	HB2	2.99
Tyr^{56}	HD2	6.82	Cys^{60}	HB3	2.99
Tyr^{56}	HE1	6.65	Cys^{60}	CA	55.38
Tyr^{56}	HE2	6.65	Cys^{60}	CB	48.24
Tyr^{56}	C	173.40	Cys^{60}	N	118.53
Tyr^{56}	CA	56.52	Ile^{61}	H	9.22
Tyr^{56}	CB	39.24	Ile^{61}	HA	4.48
Tyr^{56}	CD1	133.56	Ile^{61}	HB	1.72
Tyr^{56}	CD2	133.56	Ile^{61}	HG12	0.90
Tyr^{56}	CE1	117.71	Ile^{61}	HG13	1.23
Tyr^{56}	CE2	117.71	Ile^{61}	HG2	0.72
Tyr^{56}	N	121.10	Ile^{61}	HD1	0.69
Asn^{57}	H	8.55	Ile^{61}	CA	59.52
Asn^{57}	HA	4.88	Ile^{61}	CB	41.90
Asn^{57}	HB2	2.67	Ile^{61}	CG1	26.16
Asn^{57}	HB3	2.67	Ile^{61}	CG2	17.43
Asn^{57}	HD21	7.30	Ile^{61}	CD1	13.22
Asn^{57}	HD22	6.58	Ile^{61}	N	118.07

Residue	Atom	Shift (ppm)	Residue	Atom	Shift (ppm)
Tyr^{62}	H	8.41	Cys^{65}	HA	4.11
Tyr^{62}	HA	3.92	Cys^{65}	HB2	2.31
Tyr^{62}	HB2	2.54	Cys^{65}	HB3	2.38
Tyr^{62}	HB3	2.77	Cys^{65}	C	173.55
Tyr^{62}	HD1	6.55	Cys^{65}	CA	60.59
Tyr^{62}	HD2	6.55	Cys^{65}	CB	41.72
Tyr^{62}	HE1	6.69	Cys^{65}	N	121.32
Tyr^{62}	HE2	6.69	Cys^{65}	H	8.55
Tyr^{62}	HH	8.38	Arg^{66}	H	8.36
Tyr^{62}	C	176.74	Arg^{66}	HA	4.26
Tyr^{62}	CA	58.83	Arg^{66}	HB2	1.77
Tyr^{62}	CB	38.47	Arg^{66}	HB3	1.89
Tyr^{62}	CD1	133.02	Arg^{66}	HG2	-
Tyr^{62}	CD2	133.02	Arg^{66}	HG3	-
Tyr^{62}	CE1	118.18	Arg^{66}	HD2	3.24
Tyr^{62}	CE2	118.18	Arg^{66}	HD3	3.24
Tyr^{62}	N	125.92	Arg^{66}	HE	7.17
Lys^{63}	H	7.81	Arg^{66}	C	176.26
Lys^{63}	HA	4.16	Arg^{66}	CA	55.38
Lys^{63}	HB2	1.25	Arg^{66}	CB	31.21
Lys^{63}	HB3	1.43	Arg^{66}	CG	-
Lys^{63}	HG2	1.01	Arg^{66}	CD	43.58
Lys^{63}	HG3	1.15	Arg^{66}	N	118.73
Lys^{63}	HD2	1.29	Arg^{66}	NE	84.87
Lys^{63}	HD3	1.46	His^{67}	H	8.94
Lys^{63}	HE2	2.67	His^{67}	HA	4.27
Lys^{63}	HE3	2.72	His^{67}	HB2	2.98
Lys^{63}	C	174.12	His^{67}	HB3	3.11
Lys^{63}	CA	55.88	His^{67}	HD2	7.17
Lys^{63}	CB	35.89	His^{67}	HE2	-
Lys^{63}	CG	26.05	His^{67}	C	175.80
Lys^{63}	CD	29.53	His^{67}	CA	56.74
Lys^{63}	CE	42.06	His^{67}	CB	30.32
Lys^{63}	N	123.01	His^{67}	CD2	118.42
Arg^{64}	H	7.72	His^{67}	CE2	-
Arg^{64}	HA	4.52	His^{67}	N	124.60
Arg^{64}	HB2	1.42	Pro^{68}	HA	4.31
Arg^{64}	HB3	1.62	Pro^{68}	HB2	1.32
Arg^{64}	HG2	1.34	Pro^{68}	HB3	1.66
Arg^{64}	HG3	1.34	Pro^{68}	HG2	1.02
Arg^{64}	HD2	3.10	Pro^{68}	HG3	1.50
Arg^{64}	HD3	3.10	Pro^{68}	HD2	2.46
Arg^{64}	HE	7.07	Pro^{68}	HD3	3.21
Arg^{64}	C	175.72	Pro^{68}	C	176.01
Arg^{64}	CA	54.88	Pro^{68}	CA	63.80
Arg^{64}	CB	33.05	Pro^{68}	CB	31.78
Arg^{64}	CG	26.42	Pro^{68}	CG	26.48
Arg^{64}	CD	43.76	Pro^{68}	CD	49.72
Arg^{64}	N	117.89	Gly^{69}	H	9.62
Arg^{64}	NE	85.38	Gly^{69}	HA2	3.74

Residue	Atom	Shift (ppm)	Residue	Atom	Shift (ppm)
Gly^{69}	HA3	4.23	Asn^{73}	CA	55.20
Gly^{69}	C	171.66	Asn^{73}	CB	36.37
Gly^{69}	CA	43.55	Asn^{73}	N	117.29
Gly^{69}	N	111.84	Asn^{73}	ND2	116.26
Glu^{70}	H	7.73	Gly^{74}	H	7.64
Glu^{70}	HA	4.13	Gly^{74}	HA2	4.11
Glu^{70}	HB2	1.77	Gly^{74}	HA3	4.33
Glu^{70}	HB3	1.80	Gly^{74}	C	172.02
Glu^{70}	HG2	2.04	Gly^{74}	CA	46.70
Glu^{70}	HG3	2.16	Gly^{74}	N	104.23
Glu^{70}	C	175.28	Gln^{75}	H	8.81
Glu^{70}	CA	55.72	Gln^{75}	HA	4.73
Glu^{70}	CB	30.70	Gln^{75}	HB2	1.78
Glu^{70}	CG	35.61	Gln^{75}	HB3	1.89
Glu^{70}	N	115.96	Gln^{75}	HG2	2.13
Leu^{71}	H	8.29	Gln^{75}	HG3	2.13
Leu^{71}	HA	4.31	Gln^{75}	HE21	-
Leu^{71}	HB2	0.75	Gln^{75}	HE22	-
Leu^{71}	HB3	1.54	Gln^{75}	C	173.56
Leu^{71}	HG	1.01	Gln^{75}	CA	54.93
Leu^{71}	HD1	0.59	Gln^{75}	CB	32.43
Leu^{71}	HD2	0.48	Gln^{75}	CG	33.69
Leu^{71}	C	174.89	Gln^{75}	NE2	-
Leu^{71}	CA	52.46	Gln^{75}	N	118.42
Leu^{71}	CB	43.42	Val^{76}	H	8.35
Leu^{71}	CG	26.79	Val^{76}	HA	4.31
Leu^{71}	CD1	26.55	Val^{76}	HB	2.00
Leu^{71}	CD2	24.04	Val^{76}	HG1	0.98
Leu^{71}	N	126.65	Val^{76}	HG2	0.49
Arg^{72}	H	8.52	Val^{76}	C	175.18
Arg^{72}	HA	3.82	Val^{76}	CA	61.46
Arg^{72}	HB2	1.56	Val^{76}	CB	32.69
Arg^{72}	HB3	1.61	Val^{76}	CG1	21.20
Arg^{72}	HG2	1.48	Val^{76}	CG2	21.20
Arg^{72}	HG3	1.53	Val^{76}	N	122.94
Arg^{72}	HD2	3.11	Glu^{77}	H	9.25
Arg^{72}	HD3	3.11	Glu^{77}	HA	4.42
Arg^{72}	C	177.35	Glu^{77}	HB2	1.80
Arg^{72}	CA	57.81	Glu^{77}	HB3	1.82
Arg^{72}	CB	29.24	Glu^{77}	HG2	1.97
Arg^{72}	CG	26.94	Glu^{77}	HG3	2.06
Arg^{72}	CD	43.21	Glu^{77}	C	174.67
Arg^{72}	N	129.18	Glu^{77}	CA	55.07
Asn^{73}	H	8.88	Glu^{77}	CB	30.02
Asn^{73}	HA	3.79	Glu^{77}	CG	34.17
Asn^{73}	HB2	0.97	Glu^{77}	N	128.83
Asn^{73}	HB3	2.13	Ile^{78}	H	8.51
Asn^{73}	HD21	7.31	Ile^{78}	HA	4.05
Asn^{73}	HD22	6.75	Ile^{78}	HB	1.79
Asn^{73}	C	172.73	Ile^{78}	HG12	1.38

Residue	Atom	Shift (ppm)	Residue	Atom	Shift (ppm)
Ile^{78}	HG13	1.38	Leu^{82}	HD2	0.88
Ile^{78}	HG2	0.68	Leu^{82}	C	175.81
Ile^{78}	HD1	0.66	Leu^{82}	CA	53.88
Ile^{78}	C	175.73	Leu^{82}	CB	40.72
Ile^{78}	CA	60.49	Leu^{82}	CG	26.62
Ile^{78}	CB	37.78	Leu^{82}	CD1	26.13
Ile^{78}	CG1	27.14	Leu^{82}	CD2	23.35
Ile^{78}	CG2	17.14	Leu^{82}	N	115.39
Ile^{78}	CD1	13.13	Ser^{83}	H	7.92
Ile^{78}	N	127.92	Ser^{83}	HA	4.25
Lys^{79}	H	7.89	Ser^{83}	HB2	3.74
Lys^{79}	HA	4.22	Ser^{83}	HB3	3.74
Lys^{79}	HB2	1.90	Ser^{83}	HG	4.63
Lys^{79}	HB3	1.90	Ser^{83}	C	173.85
Lys^{79}	HG2	1.32	Ser^{83}	CA	59.08
Lys^{79}	HG3	1.40	Ser^{83}	CB	64.51
Lys^{79}	HD2	1.55	Ser^{83}	N	113.13
Lys^{79}	HD3	1.55	Phe^{84}	H	8.17
Lys^{79}	HE2	2.79	Phe^{84}	HA	3.61
Lys^{79}	HE3	2.86	Phe^{84}	HB2	2.50
Lys^{79}	C	177.68	Phe^{84}	HB3	2.75
Lys^{79}	CA	57.99	Phe^{84}	HD1	6.71
Lys^{79}	CB	32.81	Phe^{84}	HD2	6.71
Lys^{79}	CG	24.91	Phe^{84}	HE1	7.35
Lys^{79}	CD	28.29	Phe^{84}	HE2	7.35
Lys^{79}	CE	41.98	Phe^{84}	HZ	-
Lys^{79}	N	127.02	Phe^{84}	C	174.78
Thr^{80}	H	8.95	Phe^{84}	CA	59.73
Thr^{80}	HA	4.44	Phe^{84}	CB	39.44
Thr^{80}	HB	4.27	Phe^{84}	CD1	131.48
Thr^{80}	HG2	1.12	Phe^{84}	CD2	131.48
Thr^{80}	C	173.26	Phe^{84}	CE1	131.46
Thr^{80}	CA	60.25	Phe^{84}	CE2	131.46
Thr^{80}	CB	69.84	Phe^{84}	CZ	-
Thr^{80}	CG2	22.30	Phe^{84}	N	120.81
Thr^{80}	N	110.77	Gly^{85}	H	8.55
Asp^{81}	H	8.74	Gly^{85}	HA2	3.02
Asp^{81}	HA	4.46	Gly^{85}	HA3	3.88
Asp^{81}	HB2	2.79	Gly^{85}	C	174.72
Asp^{81}	HB3	3.01	Gly^{85}	CA	44.29
Asp^{81}	C	175.13	Gly^{85}	N	117.36
Asp^{81}	CA	53.24	Ser^{86}	H	8.34
Asp^{81}	CB	41.28	Ser^{86}	HA	4.37
Asp^{81}	N	123.21	Ser^{86}	HB2	3.99
Leu^{82}	H	7.94	Ser^{86}	HB3	3.99
Leu^{82}	HA	4.66	Ser^{86}	C	172.42
Leu^{82}	HB2	1.65	Ser^{86}	CA	59.33
Leu^{82}	HB3	1.96	Ser^{86}	CB	64.48
Leu^{82}	HG	1.66	Ser^{86}	N	118.47
Leu^{82}	HD1	0.97	Gln^{87}	H	8.72

Residue	Atom	Shift (ppm)	Residue	Atom	Shift (ppm)
Gln^{87}	HA	5.56	Phe^{90}	CD1	132.17
Gln^{87}	HB2	1.88	Phe^{90}	CD2	132.17
Gln^{87}	HB3	1.99	Phe^{90}	CE1	130.65
Gln^{87}	HG2	2.16	Phe^{90}	CE2	130.65
Gln^{87}	HG3	2.21	Phe^{90}	CZ	128.95
Gln^{87}	HE21	7.54	Phe^{90}	N	121.52
Gln^{87}	HE22	6.80	Ser^{91}	H	8.26
Gln^{87}	C	175.55	Ser^{91}	HA	4.51
Gln^{87}	CA	54.43	Ser^{91}	HB2	3.78
Gln^{87}	CB	33.07	Ser^{91}	HB3	3.78
Gln^{87}	CG	35.06	Ser^{91}	HG	4.65
Gln^{87}	N	120.24	Ser^{91}	C	172.68
Gln^{87}	NE2	111.99	Ser^{91}	CA	57.23
Ile^{88}	H	8.82	Ser^{91}	CB	65.62
Ile^{88}	HA	4.71	Ser^{91}	N	112.73
Ile^{88}	HB	1.16	Cys^{92}	H	8.72
Ile^{88}	HG12	0.24	Cys^{92}	HA	5.26
Ile^{88}	HG13	0.42	Cys^{92}	HB2	2.43
Ile^{88}	HG2	0.48	Cys^{92}	HB3	2.95
Ile^{88}	HD1	-0.62	Cys^{92}	C	175.00
Ile^{88}	C	173.61	Cys^{92}	CA	53.56
Ile^{88}	CA	59.45	Cys^{92}	CB	41.14
Ile^{88}	CB	40.36	Cys^{92}	N	117.21
Ile^{88}	CG1	24.11	Ser^{93}	H	8.89
Ile^{88}	CG2	18.76	Ser^{93}	HA	4.29
Ile^{88}	CD1	12.80	Ser^{93}	HB2	3.53
Ile^{88}	N	118.74	Ser^{93}	HB3	3.96
Glu^{89}	H	7.82	Ser^{93}	C	172.52
Glu^{89}	HA	4.80	Ser^{93}	CA	59.08
Glu^{89}	HB2	1.69	Ser^{93}	CB	64.05
Glu^{89}	HB3	1.92	Ser^{93}	N	119.84
Glu^{89}	HG2	2.12	Glu^{94}	H	8.43
Glu^{89}	HG3	2.13	Glu^{94}	HA	4.14
Glu^{89}	C	173.84	Glu^{94}	HB2	1.94
Glu^{89}	CA	53.56	Glu^{94}	HB3	2.03
Glu^{89}	CB	32.98	Glu^{94}	HG2	2.32
Glu^{89}	CG	35.51	Glu^{94}	HG3	2.32
Glu^{89}	N	118.68	Glu^{94}	C	177.35
Phe^{90}	H	8.39	Glu^{94}	CA	58.51
Phe^{90}	HA	5.41	Glu^{94}	CB	29.15
Phe^{90}	HB2	2.54	Glu^{94}	CG	35.34
Phe^{90}	HB3	2.77	Glu^{94}	N	121.79
Phe^{90}	HD1	7.02	Gly^{95}	H	8.88
Phe^{90}	HD2	7.02	Gly^{95}	HA2	3.42
Phe^{90}	HE1	6.68	Gly^{95}	HA3	4.16
Phe^{90}	HE2	6.68	Gly^{95}	C	173.21
Phe^{90}	HZ	6.72	Gly^{95}	CA	44.76
Phe^{90}	C	174.50	Gly^{95}	N	112.82
Phe^{90}	CA	54.93	Phe^{96}	H	8.17
Phe^{90}	CB	43.21	Phe^{96}	HA	5.08

Residue	Atom	Shift (ppm)	Residue	Atom	Shift (ppm)
Phe^{96}	HB2	2.50	Ile^{99}	HB	1.72
Phe^{96}	HB3	3.22	Ile^{99}	HG12	1.05
Phe^{96}	HD1	6.71	Ile^{99}	HG13	1.32
Phe^{96}	HD2	6.71	Ile^{99}	HG2	0.81
Phe^{96}	HE1	7.13	Ile^{99}	HD1	0.70
Phe^{96}	HE2	7.13	Ile^{99}	C	175.16
Phe^{96}	HZ	-	Ile^{99}	CA	60.92
Phe^{96}	C	174.32	Ile^{99}	CB	38.90
Phe^{96}	CA	56.64	Ile^{99}	CG1	26.81
Phe^{96}	CB	42.18	Ile^{99}	CG2	17.17
Phe^{96}	CD1	131.65	Ile^{99}	CD1	13.10
Phe^{96}	CD2	131.65	Ile^{99}	N	128.13
Phe^{96}	CE1	131.36	Gly^{100}	H	8.44
Phe^{96}	CE2	131.36	Gly^{100}	HA2	3.60
Phe^{96}	CZ	-	Gly^{100}	HA3	4.47
Phe^{96}	N	118.94	Gly^{100}	C	172.02
Phe^{97}	H	9.72	Gly^{100}	CA	43.16
Phe^{97}	HA	4.85	Gly^{100}	N	116.74
Phe^{97}	HB2	2.87	Ser^{101}	H	8.15
Phe^{97}	HB3	2.90	Ser^{101}	HA	4.45
Phe^{97}	HD1	7.01	Ser^{101}	HB2	3.41
Phe^{97}	HD2	7.01	Ser^{101}	HB3	3.79
Phe^{97}	HE1	7.18	Ser^{101}	C	174.30
Phe^{97}	HE2	7.18	Ser^{101}	CA	58.09
Phe^{97}	HZ	-	Ser^{101}	CB	64.07
Phe^{97}	C	172.59	Ser^{101}	N	115.90
Phe^{97}	CA	56.12	Thr^{102}	H	8.21
Phe^{97}	CB	42.44	Thr^{102}	HA	4.28
Phe^{97}	CD1	132.19	Thr^{102}	HB	4.54
Phe^{97}	CD2	132.19	Thr^{102}	HG2	1.33
Phe^{97}	CE1	129.58	Thr^{102}	C	174.70
Phe^{97}	CE2	129.58	Thr^{102}	CA	63.47
Phe^{97}	CZ	-	Thr^{102}	CB	69.57
Phe^{97}	N	120.96	Thr^{102}	CG2	23.13
Leu^{98}	H	8.01	Thr^{102}	N	111.56
Leu^{98}	HA	4.56	Thr^{103}	H	7.57
Leu^{98}	HB2	1.39	Thr^{103}	HA	5.06
Leu^{98}	HB3	1.51	Thr^{103}	HB	3.99
Leu^{98}	HG	1.09	Thr^{103}	HG1	4.81
Leu^{98}	HD1	0.64	Thr^{103}	HG2	1.08
Leu^{98}	HD2	0.40	Thr^{103}	C	173.79
Leu^{98}	C	176.38	Thr^{103}	CA	59.49
Leu^{98}	CA	56.69	Thr^{103}	CB	73.29
Leu^{98}	CB	43.88	Thr^{103}	CG2	21.35
Leu^{98}	CG	28.41	Thr^{103}	N	109.88
Leu^{98}	CD1	26.67	Ser^{104}	H	8.05
Leu^{98}	CD2	25.59	Ser^{104}	HA	4.56
Leu^{98}	N	126.50	Ser^{104}	HB2	3.46
Ile^{99}	H	9.13	Ser^{104}	HB3	3.88
Ile^{99}	HA	4.22	Ser^{104}	C	171.99

Residue	Atom	Shift (ppm)	Residue	Atom	Shift (ppm)
Ser^{104}	CA	58.18	Gln^{109}	HA	4.42
Ser^{104}	CB	64.71	Gln^{109}	HB2	1.56
Ser^{104}	N	114.35	Gln^{109}	HB3	1.84
Arg^{105}	H	9.01	Gln^{109}	HG2	2.14
Arg^{105}	HA	5.61	Gln^{109}	HG3	2.14
Arg^{105}	HB2	1.65	Gln^{109}	HE21	7.69
Arg^{105}	HB3	1.70	Gln^{109}	HE22	6.13
Arg^{105}	HG2	1.33	Gln^{109}	C	175.08
Arg^{105}	HG3	1.40	Gln^{109}	CA	54.57
Arg^{105}	HD2	3.09	Gln^{109}	CB	30.50
Arg^{105}	HD3	3.09	Gln^{109}	CG	33.31
Arg^{105}	HE	7.02	Gln^{109}	N	127.19
Arg^{105}	C	175.03	Gln^{109}	NE2	111.00
Arg^{105}	CA	54.34	Asp^{110}	H	8.78
Arg^{105}	CB	34.52	Asp^{110}	HA	4.11
Arg^{105}	CG	27.65	Asp^{110}	HB2	2.61
Arg^{105}	CD	43.43	Asp^{110}	HB3	2.78
Arg^{105}	N	125.26	Asp^{110}	C	176.18
Arg^{105}	NE	84.71	Asp^{110}	CA	56.96
Cys^{106}	H	8.39	Asp^{110}	CB	39.85
Cys^{106}	HA	3.90	Asp^{110}	N	123.65
Cys^{106}	HB2	1.62	Arg^{111}	H	8.11
Cys^{106}	HB3	2.62	Arg^{111}	HA	4.31
Cys^{106}	C	173.25	Arg^{111}	HB2	1.68
Cys^{106}	CA	55.24	Arg^{111}	HB3	1.97
Cys^{106}	CB	38.81	Arg^{111}	HG2	1.54
Cys^{106}	N	126.17	Arg^{111}	HG3	1.54
Glu^{107}	H	9.10	Arg^{111}	HD2	3.12
Glu^{107}	HA	4.94	Arg^{111}	HD3	3.12
Glu^{107}	HB2	1.95	Arg^{111}	C	175.67
Glu^{107}	HB3	2.04	Arg^{111}	CA	56.09
Glu^{107}	HG2	2.27	Arg^{111}	CB	29.86
Glu^{107}	HG3	2.27	Arg^{111}	CG	27.10
Glu^{107}	C	174.76	Arg^{111}	CD	43.40
Glu^{107}	CA	54.15	Arg^{111}	N	118.67
Glu^{107}	CB	33.70	Gly^{112}	H	7.74
Glu^{107}	CG	34.45	Gly^{112}	HA2	3.96
Glu^{107}	N	131.22	Gly^{112}	HA3	4.01
Val^{108}	H	8.36	Gly^{112}	C	173.12
Val^{108}	HA	3.88	Gly^{112}	CA	44.96
Val^{108}	HB	1.86	Gly^{112}	N	109.33
Val^{108}	HG1	0.70	Val^{113}	H	8.15
Val^{108}	HG2	0.79	Val^{113}	HA	5.16
Val^{108}	C	176.02	Val^{113}	HB	1.89
Val^{108}	CA	64.63	Val^{113}	HG1	0.50
Val^{108}	CB	31.81	Val^{113}	HG2	0.42
Val^{108}	CG1	22.60	Val^{113}	C	175.52
Val^{108}	CG2	21.41	Val^{113}	CA	58.97
Val^{108}	N	119.89	Val^{113}	CB	34.58
Gln^{109}	H	8.06	Val^{113}	CG1	22.61

Residue	Atom	Shift (ppm)	Residue	Atom	Shift (ppm)
Val[113]	CG2	19.22	Pro[118]	HB2	2.01
Val[113]	N	115.95	Pro[118]	HB3	2.25
Gly[114]	H	8.88	Pro[118]	HG2	1.82
Gly[114]	HA2	3.71	Pro[118]	HG3	2.03
Gly[114]	HA3	3.86	Pro[118]	HD2	3.78
Gly[114]	C	171.28	Pro[118]	HD3	3.87
Gly[114]	CA	45.02	Pro[118]	CA	61.71
Gly[114]	N	109.92	Pro[118]	CB	32.67
Trp[115]	H	8.02	Pro[118]	CG	26.52
Trp[115]	HA	4.85	Pro[118]	CD	50.91
Trp[115]	HB2	2.79	Leu[119]	H	7.87
Trp[115]	HB3	3.51	Leu[119]	HA	3.36
Trp[115]	HD1	7.27	Leu[119]	HB2	1.21
Trp[115]	HE1	10.50	Leu[119]	HB3	1.33
Trp[115]	HE3	7.32	Leu[119]	HG	1.40
Trp[115]	HZ2	7.16	Leu[119]	HD1	0.56
Trp[115]	HZ3	6.85	Leu[119]	HD2	0.46
Trp[115]	HH2	6.77	Leu[119]	CA	53.56
Trp[115]	C	168.17	Leu[119]	CB	41.91
Trp[115]	CA	56.46	Leu[119]	CG	26.90
Trp[115]	CB	30.03	Leu[119]	CD1	25.55
Trp[115]	CD1	126.79	Leu[119]	CD2	24.84
Trp[115]	CE3	121.01	Leu[119]	N	119.27
Trp[115]	CZ2	113.98	Pro[120]	HA	4.60
Trp[115]	CZ3	122.93	Pro[120]	HB2	1.66
Trp[115]	CH2	123.87	Pro[120]	HB3	2.00
Trp[115]	N	119.98	Pro[120]	HG2	-
Trp[115]	NE1	127.79	Pro[120]	HG3	-
Ser[116]	H	9.73	Pro[120]	HD2	2.04
Ser[116]	HA	3.97	Pro[120]	HD3	2.19
Ser[116]	HB2	3.69	Pro[120]	C	172.64
Ser[116]	HB3	3.86	Pro[120]	CA	61.96
Ser[116]	HG	4.63	Pro[120]	CB	32.72
Ser[116]	C	174.78	Pro[120]	CG	-
Ser[116]	CA	60.43	Pro[120]	CD	48.28
Ser[116]	CB	62.77	Gln[121]	H	7.97
Ser[116]	N	120.63	Gln[121]	HA	4.38
His[117]	H	7.98	Gln[121]	HB2	1.78
His[117]	HA	5.34	Gln[121]	HB3	1.92
His[117]	HB2	3.13	Gln[121]	HG2	2.17
His[117]	HB3	3.36	Gln[121]	HG3	2.17
His[117]	HD2	-	Gln[121]	HE21	7.34
His[117]	HE2	-	Gln[121]	HE22	6.62
His[117]	C	172.87	Gln[121]	C	174.61
His[117]	CA	52.68	Gln[121]	CA	54.47
His[117]	CB	31.03	Gln[121]	CB	32.65
His[117]	CD2	-	Gln[121]	CG	34.05
His[117]	HE2	-	Gln[121]	N	112.71
His[117]	N	117.85	Gln[121]	NE2	111.38
Pro[118]	HA	4.58	Cys[122]	H	8.78

Residue	Atom	Shift (ppm)
Cys^{122}	HA	5.27
Cys^{122}	HB2	2.42
Cys^{122}	HB3	2.76
Cys^{122}	C	174.17
Cys^{122}	CA	53.75
Cys^{122}	CB	40.05
Cys^{122}	N	121.54
Glu^{123}	H	9.10
Glu^{123}	HA	4.80
Glu^{123}	HB2	1.47
Glu^{123}	HB3	1.66
Glu^{123}	HG2	2.20
Glu^{123}	HG3	2.28
Glu^{123}	C	175.35
Glu^{123}	CA	54.02
Glu^{123}	CB	32.23
Glu^{123}	CG	34.82
Glu^{123}	N	125.85
Ile^{124}	H	8.66
Ile^{124}	HA	3.60
Ile^{124}	HB	1.53
Ile^{124}	HG12	0.83
Ile^{124}	HG13	1.10
Ile^{124}	HG2	0.66
Ile^{124}	HD1	0.58
Ile^{124}	C	175.62
Ile^{124}	CA	62.28
Ile^{124}	CB	38.87
Ile^{124}	CG1	28.52
Ile^{124}	CG2	16.96
Ile^{124}	CD1	13.41
Ile^{124}	N	124.90
Leu^{125}	H	7.83
Leu^{125}	HA	4.19
Leu^{125}	HB2	1.28
Leu^{125}	HB3	1.28
Leu^{125}	HG	1.26
Leu^{125}	HD1	0.63
Leu^{125}	HD2	0.66
Leu^{125}	C	176.93
Leu^{125}	CA	55.52
Leu^{125}	CB	42.69
Leu^{125}	CG	27.05
Leu^{125}	CD1	25.11
Leu^{125}	CD2	24.16
Leu^{125}	N	126.54
Glu^{126}	H	8.18
Glu^{126}	HA	4.09
Glu^{126}	HB2	1.78
Glu^{126}	HB3	1.78
Glu^{126}	HG2	2.09
Glu^{126}	HG3	2.09
Glu^{126}	C	176.21
Glu^{126}	CA	56.65
Glu^{126}	CB	29.97
Glu^{126}	CG	35.69
Glu^{126}	N	121.97
His^{127}	H	8.33
His^{127}	HA	4.50
His^{127}	HB2	3.04
His^{127}	HB3	3.04
His^{127}	CA	55.47
His^{127}	CB	28.98
His^{127}	N	118.54

Appendix B

Jenkins, H. T., Mark, L., Ball, G., Persson, J., Lindahl, G., Uhrin, D., Blom, A. M, and Barlow, P. N. (2006) *J. Biol. Chem.* **281**(6):3690-7.