Studies of the structure and dynamics of the functional sites within complement receptor type 1

Gordon M. Black

A thesis submitted for the degree of Doctor of Philosophy

The University of Edinburgh
April 2004

Abstract

The complement system is part of our innate immune response and is tightly regulated to prevent damage to host cells. Complement receptor type 1 (CR1) is one of the main regulators of complement activation and is also the immune adherence receptor on erythrocytes, important for clearance of immune complexes from the bloodstream. CR1's functions arise from its ability to bind complement proteins C3b and C4b. CR1 is a multimodular glycoprotein (220 kDa) that is too large to study in its intact form by NMR. Recently, the solution structure of functional site 2, which consists of three contiguous complement control protein (CCP) modules, has been solved. However, the structural information alone does not complete the story as dynamic motions within CR1 are likely to have implications for its functions.

This thesis describes the assignment of ^{15}N and ^{1}H NMR data for the central CCP module (module 16) of functional site 2. The resonance assignments subsequently allowed the 3D solution structure to be determined and the Modelfree analysis of the module's isotropic dynamics. CR1~16 appears to show greater fast timescale flexibility on the alleged "binding face" of the module which was implicated by mutagenesis. The majority of the slow motions within the module also occur on this face. The structure and dynamics of the lone module, when compared with previous work on larger fragments of functional site 2, allowed assessment of the importance, for their structure and flexibility, of the context of CCP modules.

The thesis also describes NMR studies of a ^{13}C, ^{15}N and ^{1}H labelled sample of CR1 modules 2-3, which correspond to the C-terminal two thirds of functional site 1. Following data acquisition, the residue assignment of this double module (132 amino acid residues) was performed to near completion. In parallel, a homology-based model of the structure of modules 2-3 was built using the structure of site 2 as a starting point. The isotropic dynamics were analysed using Modelfree formalism. Module 2, like module 16, contains slow motion and a greater degree of fast timescale motion on its hypothesised binding face. The 2-3 junction shows no evidence of flexibility on the fast timescale. While loops on module 3 at the junction may be undergoing slow timescale motion, the rest of the junction, and the linker, show no evidence of slow timescale motion. Analysis of the raw relaxation data for CR1 module pair 16-17, which is known to possess some junction flexibility, provides a contrast. While the 16-17 pair show little flexibility on the fast timescale in the junction, there is evidence to suggest that slow motion may be occurring. This difference between the sites could be contributing to the functional differences of CR1 binding sites 1 and 2.

Abbreviations and symbols

C1 – C9	members of complement cascade numbered in order of their activation
C3a, C4a	anaphylatoxin fragments of C3 and C4
C3b, C4b	proteolysis activated fragments of C3 and C4
C4BP	C4b binding protein
CA	cofactor activity
cbEGF	calcium binding epidermal growth factor
CCP	complement control protein
CR1	complement receptor type 1
CR2	complement receptor type 2
CSA	chemical shift anisotropy
DAA	decay accelerating activity
DAF	decay accelerating factor
DD	dipole-dipole
fH	factor H
FID	free induction decay
γ_X	gyromagnetic ratio of spin X
hetNOE	steady state heteronuclear NOE
HSQC	heteronuclear single quantum coherence
I	nuclear spin quantum number
INEPT	insensitive nucleus enhancement by polarisation transfer
LHR	long homologous repeat
m	magnetic quantum number
M	bulk magnetisation
MAC	membrane attack complex
MASP	mannan-binding lectin associated serine protease
MBP	mannan-binding protein
MCP	membrane cofactor protein
MEM	maximum entropy method
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser enhancement
NOESY	nuclear Overhauser effect spectroscopy
ppm	parts per milion
PDB	protein databank
PDF	probability density function
R_1	longitudinal relaxation rate
R_2	transverse relaxation rate
R_{ex}	chemical exchange contribution to R_1
RCA	regulator of complement activation
RMD	restrained molecular dynamics
RMSD	root mean square deviation
S^2	generalised order parameter
S_f^2	order parameter for fast ps-ns timescale motions
S_s^2	order parameter for slow ps-ns timescale motions
SA	simulated annealing
SSE	sum squared error

T_1	longitudinal relaxation time
T_2	transverse relaxation time
τ_c	correlation time
τ_e	fast timescale correlation time, associated with S^2
τ_f	fast timescale correlation time, associated with S_s^2
τ_m	macromolecular correlation time
SCR	short consensus repeat
TOCSY	total correlation spectroscopy
TROSY	transverse relaxation optimised spectroscopy
VCP	*Vaccinia* virus complement protein
ω	Larmor frequency

Names and abbreviations for amino acids

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

- Mutations from amino acid X to Y at the position n are indicated by XnY.
- Constructs of protein fragments consisting of modules 1 and 2 are indicated by protein-name˜1-2.
- Single module 1, when within an 1-2 construct, is indicated by 1^2.

For my parents.

Thank you for all that you have done for me.

Acknowledgements

Thanks to:

Dr. Brian Smith, for teaching me processing, assignment and structure calculation and answering my frequent questions.

Dr. Krystyna Bromek-Burnside, for training me in Modelfree use and various help with relaxation data analysis.

Dr. Dušan Uhrín, for help with NMR theory, the spectrometers and for his neverending patience.

Juraj Bella, for help using the spectrometers.

Dr. Malgorzata Krych-Goldberg, Dr. Xuefeng Wang and Prof. John Atkinson, for expressing and purifying the CR1 samples.

Prof. Paul Barlow, my supervisor.

BBSRC.

Dr. Rosie Mallin, Graeme Ball and Eve Brook, for making room 208 good fun to work in. Everyone in room 44, for making the latter years fun.

My family, for support throughout.

Princess Anne, Vancouver, green beers and the bushbaby.

Katrina Tait, Lorna McLachlan, Alice Williams and Dave Sanders, for cards, laughs and making lunchtime the highlight of the day.

Special thanks to Robin Andrews, who managed to keep me sane.

Contents

List of Figures

Chapter 1

Complement and Complement receptors

1.1 The human immune system

Millions of years of evolution have provided human beings with a highly specific immune system capable of protecting us by destroying invading pathogens. The adaptive immune system works primarily via a triggered and targeted response e.g. the production of antibodies that recognise an antigen. However, a full and rapid antibody-mediated response requires prior knowledge of the pathogenic epitopes and is of limited use as a first line of defence against a novel threat. The complement system is part of the innate immune response and is more indiscriminate, able to rapidly attack and destroy any material it comes into contact with that is not suitably protected. The complement system also augments the adaptive immune system [80].

1.2 The complement system

1.2.1 Complement system pathways

The complement system involves over thirty different glycoproteins on cell surfaces and in the blood serum, and is initiated by one of three pathways: the classical, lectin and alternative pathways. A schematic diagram showing an overview of the complement cascade is shown in Figure 1.1. The main components of the cascade are the complement proteins and are named C1-C9.

The **Classical Pathway** is activated by the presence of antibody-antigen complexes that are produced when specific sites on multivalent antibodies allow them to bind to the surfaces of invading pathogens and cross-link them. A complement molecule known as C1q is then able to bind to the F_c portions (i.e. the domain not containing the antigen binding site) of the aggregated antibodies and instigate the complement cascade. Molecules C1r and C1s bind to C1q and this facilitates the cleavage of complement protein C4 into two fragments, C4a (about 9 kDa) and C4b (about 190 kDa) [92].

The **Lectin Pathway** is stimulated by specific carbohydrates on pathogenic surfaces and so avoids the requirement of antibodies. Instead, the lectin pathway requires the mannan-binding protein (MBP) which has a structure similar to that of C1q. In this case the MBP, and proteins analogous to C1r and C1s - called the MBP-associated serine proteases (MASP) - bind to lectins on the pathogens which then leads to the cleavage of C4 [48].

The **Alternative Pathway** is also antibody-independent. It is non-specific and is initiated by any surface not bearing the appropriate complement regulatory proteins. Since the activated components of the alternative pathway are constantly present at low levels in the serum, this pathway is said to be triggered, and it provides a rapid response mechanism to foreign matter. While the classical and lectin pathways produce the cleavage of C4, the alternative pathway relies on the cleavage of complement protein C3 to its two fragments - C3a (about 9 kDa) and C3b (about 176 kDa). In the blood serum C3 is constantly undergoing hydrolysis to form $C3(H_20)$, which has similar properties to fragment C3b. This can then bind with factor B, which allows the cleavage of factor B by factor D. The resulting complex, $C3(H_20)Bb$, is then able to begin cleavage of C3 to C3a and C3b. In this way, there is constantly a small level of C3b produced. Host cells have complement regulatory proteins on, or close to, their surfaces which bind key components of the cascade and prevent its amplification. Foreign cells do not, and in this way are vulnerable to the alternative pathway [92].

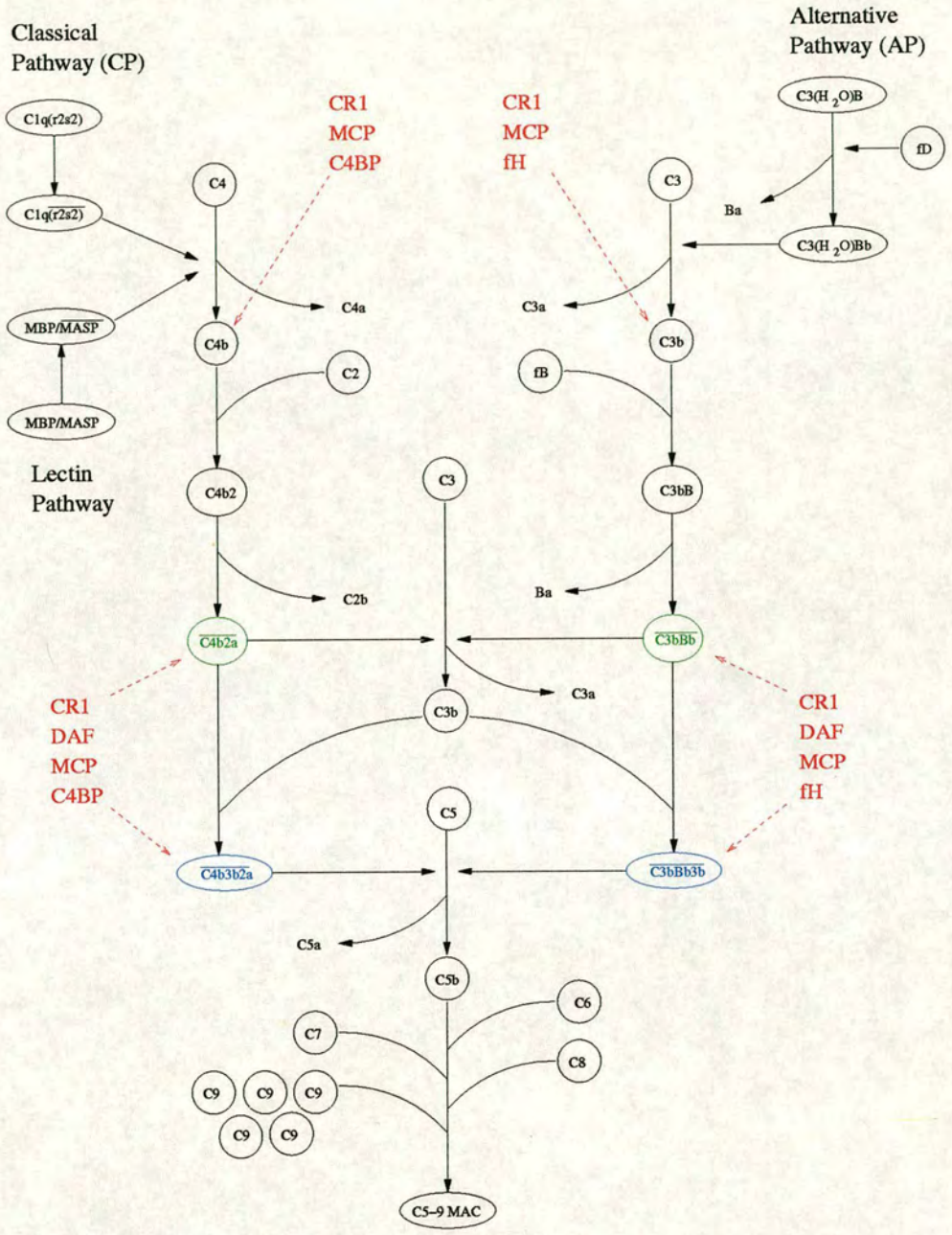


Figure 1.1: The complement system showing the three inter-linked pathways. The active enzymes are shown with a bar above their names. Complements proteins C1-C9 shown. fB, fD represents factor B and factor D. Bb represents a cleavage product of factor B. C3 convertases shown in green. C5 convertases shown in blue. Complement receptor proteins shown in red, along with arrows depicting where they interact with the cascade. Derived from [80, 105].

1.2.2 Functions of the complement system

The complement system has a variety of functions in protecting the host which can be broadly ascribed to three categories.

Opsonisation

Activation of all three pathways results in the production of C3b [76]. C3b contains a highly reactive thiol-ester group that allows it to covalently bind to other molecules (e.g. on the surface of a pathogen) that contain a hydroxyl or amino group [79]. Opsonisation is the use of C3b as a tag on pathogenic surfaces to induce their degradation via phagocytosis. When C3b binds to antibody-antigen complexes (again via the thiol-ester mechanism), these complexes can subsequently be attached to erythrocytes via the immune adherence receptor, complement receptor type 1. It is in this way that erythrocytes use opsonisation to remove these immune complexes to the liver and spleen [122]. Phagocytes therein have a complement receptor on their surfaces that can bind to the C3b molecule.

Inflammation

Inflammation is promoted by the complement proteins C3a, C4a and C5a, all of which are formed during the proteolytic cascade that occurs downstream from complement activation. These are the anaphylotoxins and are able to recruit and activate phagocytes to assist in pathogen destruction [47].

Membrane attack complex (MAC)

The continued production of C3b allows the formation of the C5 convertase that catalyses cleavage of C5 and therefore the production of C5b. The assemblage of C5b, C6, C7, C8 and multiple copies of C9, is a structure that extends a hydrophobic domain to allow it to enter the lipid bilayer of a pathogen's cell membrane. Once it has penetrated the bilayer, this structure, known as the membrane attack complex (MAC) forms a transmembrane pore that causes lysis of the invading cell [80].

1.3 Complement receptor proteins

1.3.1 Regulators of complement activation (RCA)

As the complement system is so aggressive in its defence of the body, extremely tight regulation is required to prevent harm occurring to host cells. Complement regulators are proteins capable of binding to specific components of the complement cascade and inhibiting the activation of complement.

Some of these proteins are members of a family known as the regulators of complement activation (RCA) which have related sequence, structure and function [44, 59]. The family includes complement receptors type 1 and 2 (CR1, 2), membrane cofactor protein (MCP), decay accelerating factor (DAF), factor H (fH), C4b-binding protein (C4BP) and others. The *Vaccinia* virus complement control protein (VCP) is homologous to RCA proteins. The RCA proteins themselves are targets for pathogenic organisms [82]. A schematic of the structures of the RCA family members is shown in Figure 1.2.

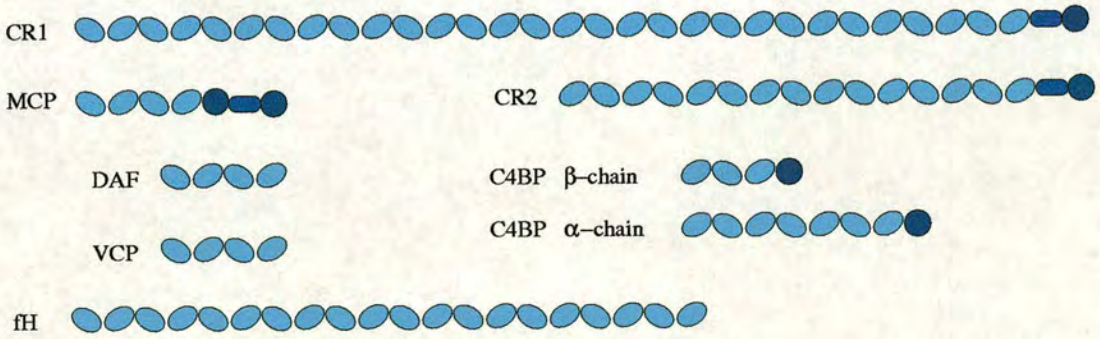


Figure 1.2: RCA proteins. Several members of the RCA family are shown schematically. *Vaccinia* virus complement control protein (VCP) is also shown. The pale blue ovals represent the CCP modules. Dark blue modules refer to non-CCP protein domains, frequently transmembrane or intracellular domains.

The RCA proteins are coded for by a gene cluster 1q32 on human chromosome 1 [140]. They can be present either in soluble form in the serum (e.g. DAF, factor H) or attached to the membrane of host cells (eg. CR1, MCP). All members of this family are predominantly composed of a single module-type, known as the complement control

protein (CCP) module. The RCA proteins are formed from varying numbers of CCP modules joined head-to-tail in contiguous fashion. In the case of membrane bound RCA proteins, transmembrane and cytoplasmic domains are located at the C-terminus.

1.3.2 Complement control protein (CCP) modules

The CCP module, also known as a short consensus repeat (SCR) or Sushi domain, is responsible for both the structure and function of the RCA proteins [59]. Each CCP module consists of approximately 60-70 amino acid residues that encompass a consensus sequence. There are two disulphide bonds absolutely conserved between pairs of cysteine residues in a 1-3, 2-4 formation. The sequence also includes a virtually invariant tryptophan residue located between the third and fourth cysteines. There are also several glycine and proline residues within the CCP sequence that are well conserved. In addition, some (generally) hydrophobic residues are frequently conserved or conservatively replaced within the sequence. Between neighbouring CCP modules is a short linker. For convenience, the linker is considered as lying between the fourth cysteine of one module and the first cysteine of the next. It is commonly of four amino acid residues in length, though it varies from three to eight residues [59, 42, 130].

It is also possible to make generalisations about the secondary and tertiary structure of CCP modules. CCP modules are slightly elongated, the dimensions being approximately 40 Å × 15 Å × 15 Å [59, 128]. The secondary structure of CCP modules is rich in anti-parallel β-strands aligned for the most part with the long axis of the module, forming a β-barrel-like structure. There are a maximum of eight short strands present, though in most cases one or more strands (most frequently strands A and C) are absent. Strands A and C form one small β-sheet close to the N-terminus, while strands E and H form a small C-terminal sheet. The remaining four strands form a larger sheet that covers one side of the module. The β-barrel contains a hydrophobic core made up of alkyl or aromatic amino acid side chains. The hydrophobic part of the invariant tryptophan indole ring is also buried in this hydrophobic core, as are the two disulphide bonds. Figure 1.3 shows an example of a CCP module tertiary structure.

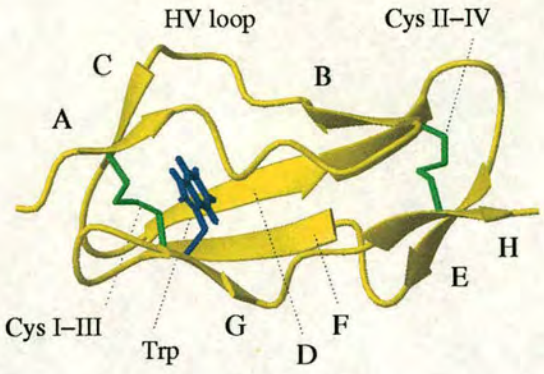


Figure 1.3: CCP module (VCP module 3) showing the conserved features. The eight β-strands are shown as arrows and labelled A-H using the CCP module convention. (Strand D is split into two portions to show which residues align with strands B and F.) The hypervariable loop is labelled HV loop. The two disulphide bonds are shown in green. The buried invariant tryptophan sidechain is shown in blue.

The stretch between β-strands B and C forms a hypervariable loop. This region has low sequence similarity between RCA proteins and is often of variable length. Figure 1.4 shows a sequence alignment for several members of the RCA family.

While there is high similarity in the primary, secondary and tertiary structures of CCP proteins, each RCA protein is functionally distinct. Differences between amino acid residues at key positions and differences in the loops (particularly the hypervariable loop) and turns presumably provides much of the variability within CCP modules that facilitates these distinct functions.

The RCA proteins form an integral part of our own immune systems. Therefore knowledge of the mechanisms whereby they perform their biological tasks could form the basis for design of treatments, for example, inflammation. Over twenty CCP module structures are now available (see section 1.3.3 below) and work is beginning on finding a way to reliably model the remaining unknown structures on the basis of sequence data [130]. Examples of RCA proteins which have had their structures modelled include MCP [86], C4BP [137, 136], DAF [74, 75] and factor H [114]. However, as important differences can be caused by small numbers of amino acid substitutions and the number of CCP structures currently available is still limited, the modelling procedure

```
          A        B    HV-loop   C     D        E          F        G        H
          *                                     *          *        +        *
CR1_1   GQCN APEWLPFARPT NLTDEF  EFPIGTYLNYECRPGYSG   RPFSIICLKNSVWTG  AKDRCRR
CR1_8   GHCQ APDHFLFAKLK TQTNAS  DFPIGTSLKYECRPEYYG   RPFSITCLDNLVWS  SPKDVCKR
fH_1    EDCNELPPR RNTEI  LTGSWSDQTYPEGTQAIYKCRPGYRSL  GNVIMVCRKG EWVALNPLRKCQK
DAF_1   GDCGLPPDVPNAQPA  LEGRT   SFPEDTVITYKCEESFVKIPGEKDSVTCLKGSQWSD  IEEFCNR
MCP_1   DACEEPP TFEAMEL  IGKPKP  YYEIGERVDYKCKKGYFYIPPLATHTICDRNHTWLP VSDDACYR
VCP_1   SCCT IPSRPINMKFKNSVETDANANYNIGDTIEYLCLPGYRKQKMGPIYAKCTG TGWT   LFNQCIK
C4BP_1  GNCG PPPTLSFAAPM DITLTE TRFKTGTTLKYTCLPGYVRS HSTQTLTCNSDGEWV   YNTFCIY
          C   P                    a  G    Y C   Y          C     W        C

          A        B   HV-loop   C     D        E          F        G        H
          *                                     *          *        +        *
CR1_2   KSCRNPPDPVNGMVHV   IK  GIQFGSQIK  YSCTKGYRL  IGSSSATCIISGD TVIWDNETPICDR
CR1_9   KSCKTPPDPVNGMVHV   IT  DIQVGSRIN  YSCTTGHRL  IGHSSAECILSGN AAHWSTKPPICQR
fH_2    RPCGHPGDTPFGTFTL   TG  GNVFEYGVKAVYTCNEGYQLLGEINYR EC    D TDGWTNDIPICEV
DAF_2    SCEVPTR LNSASLKQPYITQNYFPVGTVVE  YECRPGYRREPSLSPKTTCLQN    LKWSTAVEFCKK
MCP_2   ETCPYIRDPLNGQAVP  ANG  TYEFGYQMH  FICNEGYYL  IGEEILYCELKGS VAIWSGKPPICEK
VCP_2   RRCPSPRDIDNGQLDI    G  GVDFGSSIT  YSCNSGYHL  IGESKSYCELGSTGSMVWNPEAPICES
C4BP_2  KRCRHPGELRNGQVEI   KT  DLSFGSQIE  FSCSEGFFL  IGSTTSRCEVQDR GVGWSHPLPQCEI
          C  P    NG              G      a C  Ga L         C         W    P C

          A         B  HV-loop   C    D          E      F          G       H
          *                                     *          *        +        *
CR1_3   IPCGLPPTITNGDFIS TNRE  NFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSNDDQVGIWSGPAPQCIIP
CR1_10  IPCGLPPTIANGDFIS TNRE  NFHYGSVVTYRCNLGSRGRKVFELVGEPSIYCTSNDDQVGIWSGPAPQCIIP
fH_3    VKCLPVTAPENGKIVSSAMEPDREYHFGQAVRFVCNSG       YKIEGDEEMHCS  DD  GFWSKEKPKCVE
DAF_3   KSCPNPGEIRNGQIDV    PG  GILFGATISFSCNTG       YKLFGSTSSFCLISGS SVQWSDPLPECRE
MCP_3   VLCTPPPKIKNGKHTF SEVE  VFEYLDAVTYSCDPAPGPDP FSLIGESTIYCGDNS    VWSRAAPECKV
VCP_3   VKCQSPPSISNGRHNG  YED  FYTDGSVVTYSCNSG       YSLIGNSGVLCSGGE     WSD PPTCQI
C4BP_3  VKCKPPPDIRNGRHSG  EEN  FYAYGFSVTYSCDPR       FSLLGHASISCTVENETIGVWRPSPPTCEK
          C  P  I NG               G  V a C         a L G     C         WS   P C

          A         B HV-loop  C     D          E      F          G       H
          *                                     *          *        +        *
CR1_4   NKCTPP NVENGILVS DNRS LFSLNEVVEFRCQPGFVMKGPRRVKCQALN   KWEPELPSCSR
CR1_11  NKCTPP NVENGILVS DNRS LFSLNEVVEFRCQPGFVMKGPRRVKCQALN   KWEPELPSCSR
fH_4    ISCKSPD VINGSPIS QKII  YKENERFQYKCNMGYEYSERGDAVC TES   GWRP LPSCEE
DAF_4   IYCPAPPQIDNGIIQG ERDH  YGYRQSVTYACNKGFTMIGEHSIYCTVNNDEGEWSGPPPECRG
MCP_4   VKCRFP VVENGKQIS GFGK KFYYKATVMFECDKGFYLDGSDTIVCDSNS   TWDPPVPKCLKK
VCP_4   VKCPHP TISNGYLSS GFKR SYSYNDNVDFKCKYGYKLSGSSSSTCSPGN   TWKPELPKCVR
C4BP_4  ITCRKP DVSHGEMVS GFGP IYNYKDTIVFKCQKGFVLRGSSVIHCDADS   KWNPSPPACEP
          C  P    NG   S       a          C  Ga   G     C        W P  P C
```

Figure 1.4: Sequence alignments for the first four CCP modules of selected RCA proteins. Each protein contains a functional site within the first four modules. In the case of CR1, modules 8-11 also contain a functional site and these modules are also shown. Above each alignment, the β-strands *A-H* are labelled, as is the hypervariable loop ($HV-loop$). Disulphide linked cysteines and invariant tryptophans are marked with an * and + respectively. Below the alignments, positions which are conserved (or conserved in all cases shown except one) are indicated. Conservative replacements of aromatic residues are shown by an *a*. Derived from [59].

is not always accurate. Until such time as that approach is refined, structures must be determined experimentally, using either X-ray crystallography or nuclear magnetic resonance (NMR).

1.3.3 RCA structures to date

The RCA proteins for which structures have been determined so far are detailed below. RCA protein structures have been determined using X-ray crystallography and NMR spectroscopy via what is known as the "modular approach", i.e. studying one-, two- or three-module fragments because the intact protein is too flexible or large to be studied directly [5, 22].

The first CCP module structure to be solved was module 16 of factor H using homonuclear NMR spectroscopy. This was followed by determination of the structure of module 5 of the same protein, also using proton NMR [104, 6]. The first double CCP module structure to be solved was that of factor H, modules 15-16, solved in 1993 [7]. At present, the remainder of the factor H structure is unsolved. The structures of the Factor H 15-16 module pair is shown in Figure 1.5, along with double CCP module structures from CR1, DAF and MCP, solved subsequently.

The structure of VCP has been solved in its entirity (i.e. modules 1-4) by X-ray crystallography [97] and overlapping double module pairs 2-3 and 3-4 have also been solved using NMR spectroscopy [41, 143]. While the structures of the individual modules appeared similar using both methods, the nature of the junctions between the modules was found to differ depending on the method used. In the case of the X-ray structures, the module 2-3 junction appeared fixed in one position, whereas in the NMR-derived ensemble there was a variety of conformations and little contact between the main bodies of the two modules. In the NMR-derived structure of modules 3-4, the junction is better defined suggesting a lack of flexibility [142]. An analysis of the relaxation data for modules 3-4 confirms a lack of flexibility, on both fast (ps-ns) and slow (μs-ms) timescales, between modules 3 and 4 [15]. Relaxation data and protein backbone dynamics are discussed below in section 1.3.4

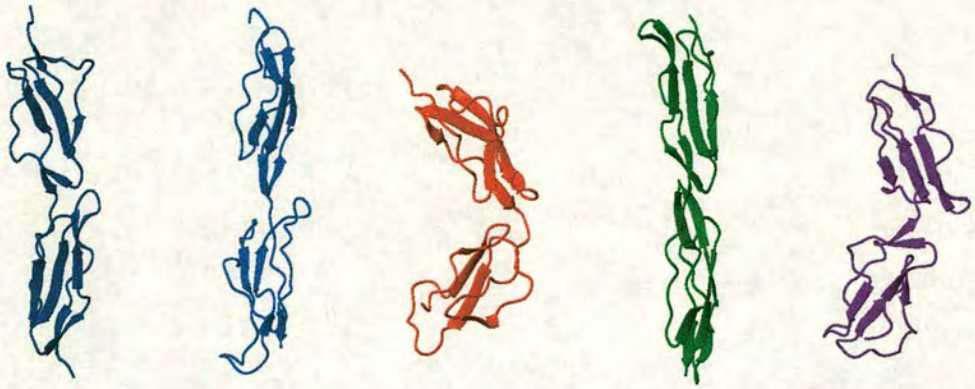


Figure 1.5: CCP module structures of selected RCA protein fragments. The CR1 structures were solved using ^{13}C,^{15}N NMR experiments. The factor H module pair was solved using homonuclear ^{1}H NMR. The MCP and DAF structures were solved using X-ray crystallography. β-strands are shown on all modules. The more N-terminal module is drawn at the top in each case.

The structure of modules 3-4 in DAF was solved using X-ray crystallography [144] while the structure of modules 2-3 was solved using NMR spectroscopy [135]. More recently, the structure of the whole of DAF (modules 1-4) has been solved using X-ray crystallography [87]. As in the case of VCP, the X-ray structure of the entire molecule implies a rigid, rod-like structure whereas the NMR structure (of modules 2-3) suggests a flexible junction.

The two N-terminal CCP modules of MCP have also been solved using X-ray crystallography [23]. The structure suggests rigidity between the two modules, although the lack of intermodular contact seems to contradict this [42].

The two N-terminal CCP modules of CR2 have been solved as a complex with their ligand C3d, again using X-ray crystallography [133]. These CR2 modules are unusual in that a hydrophobic interaction (presumably involving the non-conserved Trp residue W112) seems to allow the modules to fold up against one another into a "v-shape", with only module 1 in contact with the ligand. The remainder of the CR2 CCP mod-

ules are unsolved.

The structures of several CCP modules present in non-RCA proteins have also been solved. These are modules 2-3 of complement protein C1r [20, 19], module 2 of complement protein C1s [36] and modules 1-4 of apolipoprotein H (b2-GPI) [12]. All of these structures have been solved using X-ray crystallography.

CR1 forms the focus for this project. The three NMR-solved modules of CR1, modules 15-17 [128], are described in section 1.4.3 below. This has brought the number of solved CCP module structures to over twenty in recent years.

1.3.4 RCA dynamics

As well as determining the three-dimensional structures of CCP modules, insight into protein flexibility can be obtained through analysis of NMR relaxation data. Timescales of internal motions in proteins can range from subnanosecond (e.g. rotation of methyl groups) to the second (e.g. flipping of tyrosine rings) [33]. Various examples are known where protein mobility contributes to function [51, 54, 139]. In the bacterial response regulator protein SpoOF, slow (μs-ms timescale) motions have been detected that correspond with the residues previously known to be important functionally [33]. Similarly, in the bacterial nitrogen regulatory protein C (NtrC), only the region in which phosphorylation occurs contains the presence of slow (μs-ms timescale) motion [138]. This motion stops following the reaction, demonstrating that the phosphorylation (and therefore activation) of this domain involves internal protein dynamics. In the case of the bacterial chemotaxis pathway protein, CheW, protein-protein contacts involve motions on slow and also fast (ps-ns) timescales [38]. Motions on a variety of timescales are also seen in the acidic binding loop of the nematode anticoagulant protein (NAP) c2 [32].

Protein ^{15}N backbone dynamics are characterised by analysing ^{15}N relaxation times T_1 and T_2 (described in more detail in section 2.2.1 and section 2.2.2) and the heteronuclear NOE (described in more detail in section 2.3 and section 2.4.4) for the N−H

bond of each residue.

Low T_2 times with a non-lowered T_1 can indicate that slow, μs-ms timescale, motion (chemical exchange) is occurring around or within the backbone of that residue. If diffusional anisotropy is contributing to the relaxation of a residue, T_1 and T_2 can be pushed in opposite directions. The heteronuclear NOE is most sensitive to fast ps-ns timescale motion. It is, in general, a measure of flexibility, with low heteronuclear NOE values indicating high flexibility. Relaxation can be further interpreted as a more direct measure of the rigidity and the timescale of local motions within proteins [84, 85] and the method behind this is described in section 3.5.6.

In terms of studying ^{15}N dynamics from NMR relaxation data in this way, only five CCP modules have been successfully studied to date. These are the VCP overlapping module pairs 2-3 and 3-4 [15], MCP module 1 [106] and $GABA_B$ receptor CCP module 2 (unpublished work within this group, Dr. Stan Blein). With so few examples available, it is difficult to identify any dynamical elements which are considered characteristic of CCP modules. To obtain a clear picture of the sequence-structure-function relationships of these proteins, a combination of structural and dynamics data is required.

1.4 Complement receptor type 1 (CR1)

1.4.1 CR1 distribution

The RCA protein called complement receptor type 1 (CR1, or CD35) has been the focus of much study over the past decade. It plays a crucial role within the immune system of humans through its ability to bind complement proteins. Its ligands are both C3b and C4b, two of the most important proteins within the complement cascade [70, 68]. CR1 is expressed on the surface of a wide variety of cells within the body, including erythrocytes, neutrophils, monocytes, eosinophils, B-lymphocytes, follicular dendritic cells, podocytes and astrocytes [71]. Because erythrocytes are so numerous within the body, the erythrocyte-bound CR1 makes up about 90% of the total CR1 within the body. Each erythrocyte usually has over 500 copies of CR1 bound to its surface,

although this varies depending on the polymorphism of CR1, described in section 1.4.4. There is also a low level (about 30 ng/ml) of soluble CR1 found in human blood [150].

1.4.2 CR1 functions

CR1 contains three functional sites, which are described fully in section 1.4.3. Functional site 1 is able to bind complement cascade protein C4b and to a much lesser extent, C3b. Functional site 2, of which there are two copies within CR1, is able to bind both C4b and C3b. Of the binding capabilities of both sites, the binding of C3b by site 2 is by far the most efficient ligand binding activity within CR1. These binding capabilities lead to CR1's most important functions, of which there are three - immune complex clearance, cofactor activity (CA) and decay accelerating activity (DAA).

Immune complex clearance (immune adherence)

The adaptive immune response produces antibodies which are able to bind to and cross-link foreign antigens, forming aggregates, thus reducing the ability of antigenic particles to damage host cells and tissues. The aggregates become coated or "opsonised" by C3b molecules. These in turn allow the CR1 molecules on erythrocytes to bind to the immune-complexes via functional site 2. The complexes are subsequently cleared to the liver and spleen by the erythrocytes where they are engulfed by monocytes. Due to this function, CR1 is know as the immune adherence receptor. Monocyte-bound CR1 is also able to bind to nearby opsonised pathogens in order to facilitate phagocytosis of the pathogens directly.

Decay accelerating activity (DAA)

The C3 and C5 convertases, shown as C4b2a, C4b3b2a, C3bBb and C3bBb3b in Figure 1.1, play a vital role in the complement system by facilitating the breakdown of C3 and C5 so as to propagate the cascade. CR1 plays a part in regulating the complement system by accelerating the decay of both the C3 and C5 convertases. The presence of CR1 on host cells therefore blocks the local formation of C3b and C5b and so reduces opsonisation and also reduces the production of membrane attack complexes, thereby

protecting the immediate vicinity [46].

Site 1 plays a larger part than site 2 in the DAA functions of CR1. For the decay of C3 convertases (C4b2a and C3bBb), site 1 provides the majority of the function in terms of both the classical and alternative pathways, with site 2 having only one fifth of this potency. The C3 convertase decay of CR1 is further increased when both functional sites are present. Site 1 is similar to RCA protein DAF in that the functional site has very limited C3b-binding activity but has high DAA with respect to the C3 convertases. It is possible that DAF and CR1 site 1 have a higher affinity for the convertases as opposed to the C3b/C4b ligands alone [72].

Regarding the C5 convertases (C4b3b2a and C3b3b3b), site 1 alone has only limited DAA and site 2 alone has none. Only in the case where both site 1 and site 2 are present in CR1 is this function fully displayed.

Cofactor activity (CA)

CR1 is able to act as a cofactor in the degradation of both C3b and C4b. This cleavage can occur when C3b and C4b are free in the serum or when they are part of the C3 or C5 convertases. C3b is cleaved to C3c and C3dg, and C4b is cleaved to C4c and C4d. These degradations are irreversible and result in inactivation of the C3b/C4b ligands and/or the convertases containing C3b or C4b. The cleavage is carried out by the proteolytic enzyme factor I. This CA inhibits progress of the cascade and protects the immediate vicinity from the aggressive complement system. The products of CR1 promoted cleavage of C3b and C4b are ligands for other complement receptors, for example C3dg which is the substrate of CR2.

The two copies of functional site 2 contribute more to CA than does site 1. Site 2 has highly potent CA for the breakdown of C3b and a more moderate ability to aid the breakdown of C4b. Such CA capability makes the copies of site 2 similar in function to the RCA protein MCP. Site 1, on the other other hand, has very limited C4b CA and has no CA in the case of C3b [68]. Table 1.1 below summarises the functions of the functional sites within CR1.

Function	Site 1	Site 2
C4b-binding	Intermediate	Intermediate
C3b-binding	Low	High
CA	Low	High
DAA	High	Low

Table 1.1: General distribution of functions between functional sites in CR1

1.4.3 CR1 structure

The most common phenotype of CR1 (previously known as type A or F, now known as CR1*1) consists of just under 2000 amino acid residues with a total mass of 220 kDa. The C-terminal section, which consists of approximately 43 residues, is found within the cytoplasm of the cell to which CR1 is attached. Preceding this is a transmembrane domain of about 25 residues. There is a much larger extracellular, N-terminal domain made up (in CR1*1) of thirty contiguous CCP modules linked together.

Within the extracellular domain (where module 1 is at the N-terminus and module 30 is adjacent to the cell membrane) there are four long homlogous repeats (LHRs), A-D, each consisting of seven CCP modules and showing 60-100% sequence identity [64]. The two C-terminal modules, 29 and 30, do not form part of an LHR. The sequence alignment of CR1 is shown in Figure 1.6, in which each module is compared to the other three modules in equivalent positions within respective LHRs.

CR1 is glycosylated and contains several N-glycosylation sites. Examples include residues N115, N537 and N987 in modules 2, 9 and 16 respectively (see Figure 1.6).

In phenotype CR1*1, the three functional sites are found within the N-terminal portions of the first three LHRs A, B and C. The majority of the functions are provided by modules 1-3, 8-10 and 15-17, but the presence of the next CCP module in each case (4, 9 and 18) does confer some increase in activity [63]. The three apparently non-functional modules in each LHR (5-7, 12-14 and 19-21) might act as spacers between

```
                *                                *               *     +     *
CR1_1     0 GQCNAPEWLPFARPTNLTDEFEFPIGTYLNYECRPGYSGRPFSIICLKNSVWTGAKDRCRR
CR1_8   450 GHCQAPDHFLFAKLKTQTNASDFPIGTSLKYECRPEYYGRPFSITCLDNLVWSSPKDVCKR
CR1_15  900 GHCQAPDHFLFAKLKTQTNASDFPIGTSLKYECRPEYYGRPFSITCLDNLVWSSPKDVCKR
CR1_22      GHCKTPEQFPFASPTIPINDFEFPVGTSLNYECRPGYFGKMFSISCLENLVWSSVEDNCRR

                *                          9    *               *       + 2     *
CR1_2    61 KSCRNPPDPVNGMVHVIKGIQFGSQIKYSCTKGYRLIGSSSATCIISGDTVIWDNETPICDR
CR1_9   521 KSCKTPPDPVNGMVHVITDIQVGSRINYSCTTGHRLIGHSSAECILSGNAAHWSTKPPICQR
CR1_16  961 KSCKTPPDPVNGMVHVITDIQVGSRINYSCTTGHRLIGHSSAECILSGNTAHWSTKPPICQR
CR1_23      KSCGPPPEPFNGMVHINTDTQFGSTVNYSCNEGFRLIGSPSTTCLVSGNNVTWDKKAPICEI

                *                                *                   *           +       *
CR1_3   123 IPCGLPPTITNGDFISTNRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSNDDQVGIWSGPAPQCIIP
CR1_10  573 IPCGLPPTIANGDFISTNRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSNDDQVGIWSGPAPQCIIP
CR1_17 1023 IPCGLPPTIANGDFISTNRENFHYGSVVTYRCNLGSRGRKVFELVGEPSIYCTSNDDQVGIWSGPAPQCIIP
CR1_24      ISCEPPPTISNGDFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQVGVWSSPPPRCIST

                *                          *               *     +     *
CR1_4       NKCTPPNVENGILVSDNRSLFSLNEVVEFRCQPGFVMKGPRRVKCQALNKWEPELPSCSR
CR1_11      NKCTPPNVENGILVSDNRSLFSLNEVVEFRCQPGFVMKGPRRVKCQALNKWEPELPSCSR
CR1_18      NKCTPPNVENGILVSDNRSLFSLNEVVEFRCQPGFVMKGPRRVKCQALNKWEPELPSCSR
CR1_25      NKCTAPEVENAIRVPGNRSFFSLTEIIRFRCQPGFVMVGSHTVQCQTNGRWGPKLPHCSR

                *                          *               *     +     *
CR1_5       VCQPPPDVLHAERTQRDKDNFSPGQEVFYSCEPGYDLRGAASMRCTPQGDWSPAAPTCEV
CR1_12      VCQPPPDVLHAERTQRDKDNFSPGQEVFYSCEPGYDLRGAASMRCTPQGDWSPAAPTCEV
CR1_19      VCQPPPEILHGEHTPSHQDNFSPGQEVFYSCEPGYDLRGAASLHCTPQGDWSPEAPRCAV
CR1_26      VCQPPPEILHGEHTLSHQDNFSPGQEVFYSCEPSYDLRGAASLHCTPQGDWSPEAPRCTV

                *                          *               *           +       *
CR1_6       KSCDDFMGQLLNGRVLFPVNLQLGAKVDFVCDEGFQLKGSSASYCVLAGMESLWNSSVPVCEQ
CR1_13      KSCDDFMGQLLNGRVLFPVNLQLGAKVDFVCDEGFQLKGSSASYCVLAGMESLWNSSVPVCEQ
CR1_20      KSCDDFLGQLPHGRVLFPLNLQLGAKVSFVCDEGFRLKGSSVSHCVLVGMRSLWNNSVPVCEH
CR1_27      KSCDDFLGQLPHGRVLLPLNLQLGAKVSFVCDEGFRLKGRSASHCVLAGMKALWNSSVPVCEQ

                *                          *                   *           +       *
CR1_7       IFCPSPPVIPNGRHTGKPLEVFPFGKAVNYTCDPHPDRGTSFDLIGESTIRCTSDPQGNGVWSSPAPRCGIL
CR1_14      IFCPSPPVIPNGRHTGKPLEVFPFGKAVNYTCDPHPDRGTSFDLIGESTIRCTSDPQGNGVWSSPAPRCGIL
CR1_21      IFCPNPPAILNGRHTGTPSGDIPYGKEISYTCDPHPDRGMTFNLIGESTIRCTSDPHGNGVWSSPAPRCELSVRA
CR1_28      IFCPNPPAILNGRHTGTPFGDIPYGKEISYACDTHPDRGMTFNLIGESSIRCTSDPQGNGVWSSPAPRCELSVP

                *                          *               *     +     *
CR1_29      AACPHPPKIQNGHYIGGHVSLYLPGMTISYTCDPGYLLVGKGFIFCTDQGIWSQLDHYCKE
CR1 30      VNCSFPLFMNGISKELEMKKVYHYGDYVTLKCEDGYTLEGSPWSOCOADDRWDPPLAKCTS
```

Figure 1.6: Sequence alignments for the 30 CCP modules within CR1. Modules are arranged to compare the seven different positions within each LHR. Functional site N-terminal residues are numbered. Cysteines and invariant tryptophans are marked with * and + respectively. The *2* shows the glycosylation site within module 2 and the *9* shows the glycosylation site in modules 9 and 16. For a more detailed alignment of the functional sites of CR1, see previous Figure 1.4.

functional sites and facilitate cooperative binding. A schematic structure of CR1 is shown in Figure 1.7.

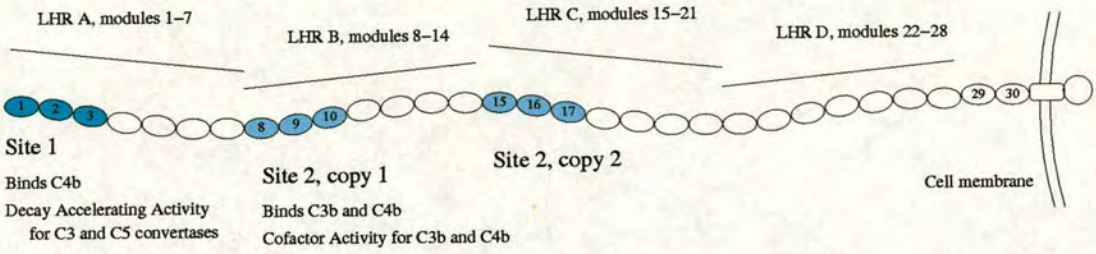


Figure 1.7: Schematic of the structure of CR1, CCP modules shown as beads. Functional site 1 coloured cyan, copies of functional site 2 coloured light blue. Long homologous repeats labelled A-D. Modules 29 and 30, which do not form part of an LHR, are numbered.

Modules 8-10 and 15-17 have an identical primary structure, with the exception of three amino acids residues: T1010, L1056 and R1059 become Ala, Pro and Gly respectively, in modules 8-10. Sequence identity between modules 1, 2 and 3 and 15, 16 and 17 is 56%, 68% and 96% respectively. The majority of the functional differences between sites 1 and 2 therefore comes from the two N-terminal modules in each site, although the third module is still required for function.

Recently, the structure of the second copy of functional site 2 of CR1, i.e. modules 15-17, has been solved using NMR spectroscopy. This was completed by studying two overlapping double module pairs: 15-16 and 16-17 [128]. Using these pairs as a basis, the structure of the whole functional site was determined by overlapping the 15-16 and 16-17 pairs on the shared module 16. This was completed using software Modeller [124, 123] which verified that the structure had realistic stereochemistry. Figure 1.8 shows the lowest energy 3D-structure of CR1, site 2.

The structure determination indicated that there is a more extensive contact between modules 15 and 16 than between 16 and 17. There were 94 unambiguous NOEs identified from modules 15 and 16 to the 15-16 linker, whereas there were only 68 NOEs from modules 16 and 17 to the 16-17 linker. There were also 13 intermodular NOEs identified in the 15-16 construct and only four intermodular NOEs in the 16-17 con-

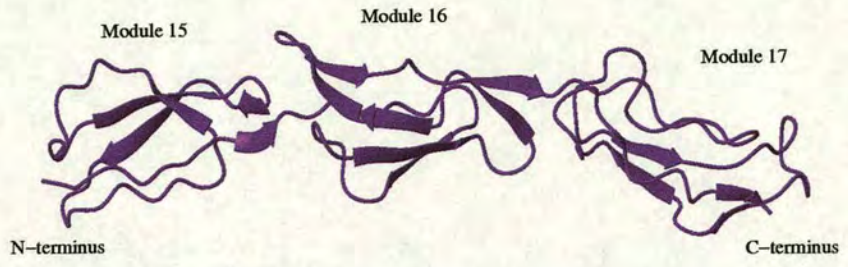


Figure 1.8: Structure of functional site 2 (i.e. CCP modules 15-17) of CR1, made using the overlapping double module fragments 15-16 and 16-17. β-strands are shown.

struct [89]. Module melting studies showed that the 16-17 junction melts first, followed by the 15-16 junction [62, 60]. The higher backbone atom RMSD (root mean square deviation) obtained within the ensemble of NMR-derived structures for module 17 was 1.10 Å. For modules 15 and 16 respectively, RMSD was 0.68 and 0.73 Å, suggesting that a higher degree of flexibility resides within module 17.

Analysis of the ^{15}N relaxation data for both the 15-16 and 16-17 pairs was attempted to provide a description of the dynamics within the fragments. However, the data could not be fitted to the software employed for this purpose. Although the fitting is still ongoing, there is currently no detailed analysis of the dynamics of any of the CCP modules in CR1. The raw relaxation data alone indicated that loop regions and termini were likely to be flexible and that some residues within the constructs were undergoing slow timescale motion. This data is discussed in section 6.2.

The elongated structure of site 2 is approximately 105 Å × 15 Å × 15 Å, giving an axial ratio of around 7:1. In reality this could be notably smaller as the molecule might tilt at the junctions. The influence of neighbouring modules is also difficult to assess. Hydrodynamic data for CCPs 15-17 had previously predicted an axial ratio of approximately 6:1. Sedimentation studies of site 1 showed modules 1-3 to have an axial ratio of around 5.2:1 [30]. Melting of modules 1-3 appears to take place as a single transition, unlike the case of modules 15-17 [24]. A possible explanation for this is that

the junctions in CR1~1-3 are structured while the modules themselves are somewhat flexible [62]. This would be a distinct difference between the two functional sites that could have ramifications for their differences in function.

1.4.4 CR1 polymorphism

Within humans, there are several genetic polymorphisms of CR1. Polymorphisms of relative molecular mass result in CR1 having a variety of sizes with different numbers of LHRs [145]. Table 1.2 summarises the different forms of CR1.

Name	Old name	Mass (Da)	No. of LHRs	Frequency
CR1*1	A or F	220,000	4	83%
CR1*2	B or S	250,000	5	15%
CR1*3	C or F′	190,000	3	1%
CR1*4	D	280,000	6	<1%

Table 1.2: Polymorphisms of mass in CR1.

The variation within CR1 structure and repetition of functional sites likely resulted from recombination of sections of DNA to produce multiple copies of the same polypeptide chain [45].

A second form of polymorphism involves the number of CR1 molecules expressed on erythrocytes. There are two alleles for this polymorphism, high (H) and low (L). Homozygotes for the L allele express fewer than 200 CR1 molecules per erythrocyte, whereas homozygotes for H express three of four times this many. Heterozygote HL individuals show an intermediate number of CR1 molecules.

One further polymorphism relates to CR1 being able to bind to a membrane protein expressed on erythrocytes infected by *Plasmodium falciparum*, the most lethal strain of the malaria parasite [1]. Following infection, CR1 causes healthy red blood cells to attach to the infected ones in what is known as "rosetting" [73]. The rosettes can adhere to endothelial cells which line the walls of capillaries as they connect to form venules, which can reduce blood flow [120]. CR1 carries two Knops blood group antigens on

LHR D with two possible phenotypes for each antigen. McC(a+) and McC(b+) form one pair of phenotypes while Sl(a+) and Sl(a−) form the other pair. McC(b+) and Sl(a−) have been shown to be in extremely low frequency (< 1%) within Caucasian populations but present in around 50% of African and African-American populations [95]. African and African-American individuals have long been known to be more resistant to malaria than Caucasians. It is believed that these two Knops blood group phenotypes may be involved in reducing binding of CR1 to infected erythrocytes and therefore reducing rosetting. Indeed, in some malaria endemic areas of Melanesia, populations show high numbers of healthy individuals who are deficient in erythrocyte CR1 [26]. Another study has shown that the CR1 binding site involved in rosetting is within the copies of functional site 2 on LHRs B and C and that rosetting can occur in the absence of C3b [121]. At present the exact rosetting mechanism is unknown.

1.4.5 CR1 mutagenesis data

As the fundamental structural framework of individual CCP modules is conserved, functional differences between active sites must arise from specific sequence differences at key points. Extensive mutagenesis to map the residues which are required for functionality in CR1 has been completed by collaborators Dr. Malgorzata Krych and Dr. John Atkinson (University of Washington Medical School, St. Louis).

Site 1/Site 2 replacements

There have been various studies involving the replacement of residues in one site with the equivalent residue from another site, to study the functional consequences that result from these changes [70, 68, 132]. As the third modules in functional site 1 and site 2 are almost identical in sequence, this mutagenesis approach has been confined to the first two modules in each functional site. Only copy 1 of site 2 (modules 8-10) has been used in this work, although as there are only three amino acid point changes between the two copies of site 2 and no functional differences, this data should be largely applicable to both sites. Table 1.3 summarises the mutagenesis data, which are then more fully explained. The locations of the residues on a schematic of the functional sites are shown in Figure 1.9.

Mutations which reduce site 1 C4b-binding and DAA	
CCP 1	G35E
CCP 2	R64K, N65T, Y94H
Mutations which confer C3b-binding to site 1	
CCP 1	T14K [R12-E21], N29K
CCP 2	D109N, E116K
Mutations which reduce site 2 C3b-binding and CA	
CCP 8 (15)	Y487S* (Y937)
CCP 9 (16)	N559D (N1009), K566E (K1016) [N559-Q571 (N1009-Q1021)]

Table 1.3: Residues found by site-directed mutagenesis to be implicated in providing the differences in function between the two functional sites in CR1. All mutations replace the native residue with the equivalent residue from the other functional site. Multiple residues replaced together are shown in square brackets. For site 2 residues, the mutagenesis was carried out on modules 8-9, though the equivalent residues in modules 15-16 are shown in brackets. *In the case of Y487S, this mutation was found to reduce only C3b-binding.

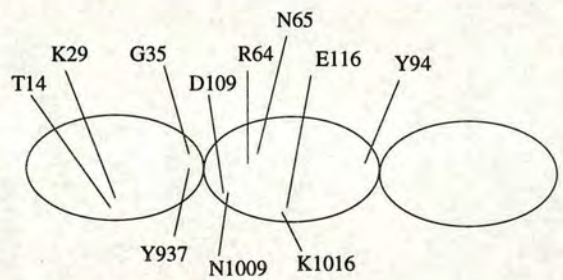


Figure 1.9: Residues which caused a change in function when mutated to the equivalent residue from functional site 1 to site 2 (and vice versa) in CR1. The structure of site 2 was used as a template for determining approximate residue location in both sites 1 and 2. Site 1 residues are labelled above the triple module schematic, site 2 residues below (using site 2 copy 2 numbering).

In the case of functional site 1, four residues were identified in modules 1-2 as being required for C4b-binding according to this strategy. The mutations G35E, R64K, N65T and Y94H (in each case replacing the native site 1 residue with the corresponding residue from site 2) were all found to markedly reduce C4b-binding, therefore highlighting these residues as potentially crucial to C4b-binding in site 1. The C4b-binding capability is similar between the two sites, although the binding mechanism within the different sites may be dependent on different characteristics of the CCP modules

involved.

In some cases, replacing site 1 residues with the equivalents from site 2 created gain-of-function within site 1. The mutations T14K, N29K, D109N and E116K each provided site 1 with C3b-binding, showing the equivalent residues as being potentially crucial to the binding function of site 2. Regarding mutation T14K, it was found that C3b-binding was further improved by replacing the entire R12-E21 peptide (RPTNL*T*DEFE) in CCP 1 by residues K462-D473 (KLKTQ*T*NASD) from module 8. (The only common residue between these stretches is T17/T467, shown in italics.) Therefore in this case residue K464 appears to be important in providing the binding function but that entire region, the hypervariable loop, plays a part.

As site 1 is responsible for the decay accelerating activity (DAA), mutagenesis of this site has also been employed to probe the residues responsible [72]. Replacement of the previously mentioned four residues - G35E, R64K, N65T and Y94H - with the analogous residues from site 2 reduced DAA, as well as reducing C4b-binding. DAA was increased by the mutations E6D, N29K, S37Y, D109N and both D109N and E116K simultaneously. The mutation of R12-E21 to K462-D473 (i.e. replacing the hypervariable loop, as detailed above) also increased DAA as well as C3b/C4b-binding.

Site 2 has also had been the subject of extensive mutagenesis involving the replacement of native residues with the equivalent amino acids from site 1. In site 2 copy 1 (modules 8-10), mutations Y487S, N559D and K566E greatly reduced C3b-binding in these constructs, highlighting their importance in the functionality of site 2. Mutating the whole stretch from N559-Q571 (i.e. from NAAHWSTKPPICQ to DTVIWDNETPICD) further reduced C3b-binding. This peptide stretch encompasses β-strand G and most of strand H, as well as part of the FG-loop.

As site 2 already possesses C4b-binding capability equal to that of site 1, it was not possible to produce mutagenesis experiments to measure any transfer of the binding ability from site 1 to site 2.

Site 2 is also crucial to the cofactor activity (CA) of CR1. The above mutations which reduced the C3b-binding also strongly reduced the CA. The N559-Q571 replacement removed CA altogether.

Following an observation that the function of CR1 site 2 was abolished if CCP 10 was replaced by CCP 3 of complement receptor type 2 (CR2), mutagenesis involving substituting residues from module 3 (CR2) into CCP 10 of CR1 was also undertaken [119, 69]. Three mutations were found to reduce function in CCP10: Y596P reduced C3b- and C4b-binding and both T589E and R591V reduced only C3b-binding. (These three residues are equivalent to Y146, T139 and R141 in CR1~17.)

Structure guided mutagenesis of site 2

While the mutagenesis strategy described above explores which individual residues or groups of residues cause the differences between the two functional sites, structure guided mutagenesis has also been undertaken following the structure determination of site 2, copy 2. Various residues in modules 15-17 which were deemed likely to be involved in function were mutated. This included mutations to CCP 17, as the third CCP module does play an important part in the function of CR1 active sites. Table 1.4 details the mutations to modules 15-17 that introduced a major loss of C3b and/or C4b-binding. Figure 1.10 shows the locations of these residues on the structure of site 2.

Module	Mutation resulting in major loss of binding
CCP 15	K912E, K914E, R933E
CCP 16	K964E
CCP 17	T1039E, R1041V

Table 1.4: Point mutations which cause a loss of function in CR1 site 2 copy 2. Previously mentioned residues not included.

In module 15, mutation S927Y caused some loss of binding. In module 17, mutations Y1046P caused some loss of C4b-binding and R1053S caused some loss of C3b-binding.

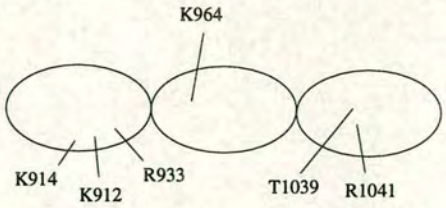


Figure 1.10: Structure guided mutagenesis of CR1 site 2. A loss in site 2 binding function resulted when the residues shown were mutated to the equivalent residue from site 1. The residues are labelled on a schematic of the structure of site 2.

Summary of residues important in CR1 function

Table 1.5 summarises the above mutagenesis data to highlight the residues implicated in providing the activity of functional sites 1 and 2 in CR1. Figure 6.15 shows the residues important to function in site 2 mapped onto the structure of modules 15-17.

Module	Residue	Function
CCP 1	G35	C4b, DAA
CCP 2	R64	C4b, DAA
	N65	C4b, DAA
	Y94	C4b, DAA
CCP 8/15	K464/K914	C3b
	S477/S927	C3b
	K479/K929	C3b
	Y487/Y937	C3b/C4b, CA
CCP 9/16	K514/K964	C3b
	N559/N1009	C3b/C4b, CA
	K566/K1016	C3b/C4b, CA
CCP 10/17	T589/T1039	C3b
	R591/R1041	C3b
	Y596/Y1046	C4b
	R603/R1053	C3b

Table 1.5: Residues required for functional activity in CR1 binding sites.

Analysis of this mutagenesis data in light of the known 3D-structure of modules 15-17 has provided insight into the binding mechanism of functional site 2. The residues thought to be important for binding in modules 15 and 16 are largely located on one face of the molecule and mainly consist of positively charged residues. In the case of another RCA protein C4BP, positive residues are also implicated in binding [10].

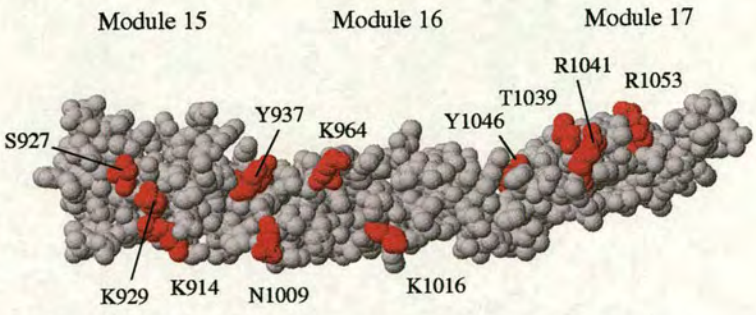


Figure 1.11: The structure of site 2 with residues involved in the function of site 2 shown in red.

The residues required in module 17 are again mainly located on one face, but this face is twisted around by approximately 60° with respect to the putative binding face of modules 15-16 (in the lowest energy structure from the NMR-derived ensemble of structures). Therefore it is suggested that when binding C3b or C4b, a conformational change takes place that involves twisting of the 16-17 linker to allow alignment of binding residues [128]. The 16-17 junction appears to have the requisite flexibility to accomplish this.

While many of the key residues are positively charged, tyrosine residues Y937 and Y1046 are also implicated in function. It is possible that while an electrostatic attraction is involved in aligning the ligand with the correct face of the binding site, a hydrophobic interaction between aromatic residues on CR1 and similarly hydrophobic residues on C3b/C4b enthalpically favour the binding. In this way, requirement of the binding mechanism of CR1 site 2 for hydrophobic interactions may be similar to that proposed for DAF [87]. These Tyr residues are located at intermodular junctions, however, therefore they may also be playing a structural role. The binding mechanism of CR1 site 1 is considered further (section 6.3.2) in the light of the structural and dynamics data obtained in the current study.

1.4.6 Therapies using CR1

As CR1 regulates the complement cascade in a variety of ways, it is of interest from a therapeutic perspective. Reducing the activity of complement in autoimmune dis-

orders is one such potential therapy. The modifications of binding function that can be attained through mutagenesis also show the potential for carefully engineering the activity of CR1-based therapies.

A soluble form of CR1, dubbed sCR1, has been developed with therapeutic applications in mind. It is identical to native CR1, but missing the transmembrane portion and the cytoplasmic tail. Clinical trials using sCR1 have already begun on treating acute lung injury and respiratory distress syndrome patients, and it is apparently safe [152, 81]. It has been shown to reduce immune-mediated RBC lysis and could be used in blood transfusions [149].

1.5 Project aims

Although sCR1, a soluble construct of CR1, has recently been crystallised [50], there has so far been no progress made on structure determination using X-ray diffraction. Nuclear magnetic resonance (NMR) provides an alternative method of structure determination that requires a soluble construct and can be performed when the protein is in its physiological environment. However, the requirement to assign NMR spectra prior to NOE-based structure calculation puts restrictions on the size of proteins that can be studied using NMR: current methods (e.g. using TROSY [111, 112]) involving labelling of the sample with NMR active isotopes extends this limit to about 40 kDa. In the case of larger proteins which contain many identical subunits, this limit is further raised. [37].

The study of backbone dynamics is an aspect of NMR that is becoming increasingly prevalent in work undertaken on proteins [9, 2, 151, 52]. While a protein's structure provides appropriate binding surfaces, the frequency and amplitude of motions within the structure can also play an important part. NMR relaxation data provides information about both the type and rate of motions occurring in a protein, from the macromolecular level down to that of single bonds.

The structure of CR1 modules 15-17 was determined by studying two overlapping

fragments (one of modules 15-16 and the other of modules 16-17) using NMR. A construct of the entire functional site was produced, but the NMR spectra obtained for this were not of high quality. Studying the site as two overlapping fragments followed by computational reconstruction was chosen as the most likely way of determining the structure of the whole site.

Analysis of the relaxation data for these pairs was attempted using the Modelfree approach [90, 109]. Though successful in the case of CCP modules VCP 2-3 and 3-4, MCP 1 and $GABA_B$ receptor CCP 2, the spatial and timescale parameters used by Modelfree could not be confidently fitted to the CR1 data. Prior to the current study there had been no in depth analysis of the relaxation data of any CCP module within CR1.

The starting point of the current work was a focus on CCP module 16 of CR1, the central module of the second copy of functional site 2. It was intended to use NMR data to determine the 3D-structure of the lone module for comparison with the structure in the context of a larger fragment and to study the dynamics present within the molecule. As no dynamics data was available for CR1, the results generated could be useful in interpreting the available mutagenesis data. Comparison of the single module with module 16 when part of a module pair would be useful in examining the effects on motion of the absence or presence of neighbouring modules. Assessment of any changes in structure and flexibility brought about by differing contexts is crucial given that proteins composed of multiple CCP modules are, in general, being studied by looking at single, double and triple module fragments.

Following this study of isolated CR1 module 16, it was intended to initiate determination of the structure and flexibility of CR1, functional site 1. Site 1, which binds C4b, has remained less well characterised and its structure determination is the next logical step in fully understanding the sequence-structure-function relationships within this important human immune system protein. Samples of the site 1 triple module (modules 1-3) and also module pairs 1-2 and 2-3 were available from our collaborators in St. Louis. The most promising of these, and most likely to lead to a completed

structure and interpretable dynamics, emerged as the 2-3 module pair. Acquisition and assignment of spectra, followed by analysis of relaxation data, were accomplished for this sample.

Chapter 2

Introduction to NMR theory and experiments

2.1 Introduction to NMR theory

2.1.1 Nuclear spins and vector diagrams

Nuclei with non-zero spin posses both angular momentum and a magnetic moment. When placed in a static magnetic field, the nuclei immediately start precessional movement around the external magnetic field. A dynamic equilibrium is created with the lower energy states being more occupied than the higher energy states. For a nucleus of spin I, there are $m = 2I + 1$ possible orientations that the nucleus may occupy, where m is the magnetic quantum number. m can take the value of I, $I - 1$, ..., $-I$. Protons, for example, have spin $I = \frac{1}{2}$ and m can have one of two values, either $+\frac{1}{2}$ or $-\frac{1}{2}$. In a physical sense, these can be thought of as describing alignment either parallel (α state, corresponding to $m = +\frac{1}{2}$) or anti-parallel (β state, corresponding to $m = -\frac{1}{2}$) to the field. The energies of these two quantised alignments are different, the separation between the two energy levels (ΔE) being:

$$\Delta E = \frac{h}{2\pi}\gamma_H B_0 \tag{2.1}$$

where h is Planck's constant, γ_H is the gyromagnetic ratio of protons and B_0 is the magnitude of the static field [43].

ΔE is small, so the energy (and therefore frequency) of electromagnetic radiation

required to cause a transition between the two states is relatively low. Typically, radiowaves in the MHz frequency range are used, the frequency given by:

$$\nu = \frac{B_0}{2\pi}\gamma \tag{2.2}$$

ν is known as the Larmor frequency and is dependent on both the nuclei in question and the strength of the external applied field. As the spinning nuclei have a magnetic moment, they can be thought of as precessing (with Larmor frequency) around the direction of the applied static field in the classical way that an object with angular momentum precesses around a gravitational field. Figures 2.1 and 2.2 show the precession of the α and β state nuclei.

Figure 2.1: Nucleus in magnetic field in α (parallel) state

Figure 2.2: Nucleus in magnetic field in β (anti-parallel) state

At thermal equilibrium, the populations of nuclei in the two states are governed by a Boltzmann distribution:

$$\frac{N_\beta}{N_\alpha} = e^{(-\frac{\Delta E}{kT})} \tag{2.3}$$

where N represents the numbers of spins in each state, k is the Boltzmann constant and T is the temperature in Kelvin.

Since the energy gap between the states is very small, the populations are almost equal, but the slight excess in state α provides a net magnetisation parallel to the external field. (At 13.8 T and Larmor frequency of 600 MHz, out of a population of one million spins there are only approximately ninety more in the α state than in the β state.) The sum of the precessing magnetic moments can be thought of as providing a bulk magnetisation (M_0) along the z axis, and as the parallel component (produced by the α states) is larger than the anti-parallel, this produces net magnetisation in

the same direction as the applied field. This is depicted in the vector diagram model shown in Figure 2.3.

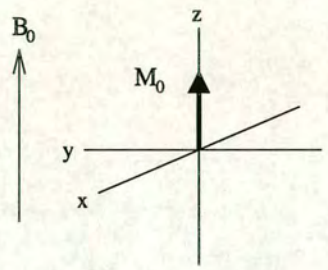


Figure 2.3: Vector diagram of net magnetisation on z axis

In the past, submitting a sample to a range of frequencies could be achieved by irradiation with a continuous wave of radiofrequency, the frequency changing over time as it was applied. However, a time advantage (known as the Felgett advantage) can be obtained by irradiating samples with a short radiofrequency pulse that can excite nuclei within a broad range of Larmor frequencies [29].

2.1.2 The effect of a radiofrequency pulse

Radiofrequency coils of NMR probes are designed to produce a linearly oscillating magnetic field along an arbitrary axis of the xy plane, which is perpendicular to the external magnetic field. This field can be visualised as two independent vectors rotating in opposite directions in the xy plane, as shown in Figure 2.4.

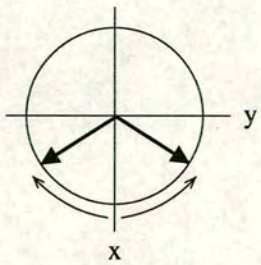


Figure 2.4: Vector diagram of pulse on x axis oscillating in xy plane

The resultant of these two components can be thought of as fluctuating between being on the x axis then on the $-x$ axis, with periods of 0 intensity intervening. Things can now be simplified if we imagine our viewpoint moving at Larmor frequency in time with the precession of the spins. This viewpoint, known as the rotating frame, now effectively no longer contains the static field B_0 as its effects (producing precession of spins around the z axis) are no longer noticeable. The net magnetisation on the z axis remains, as its position is no different when viewed from the rotating frame. However, for spins not precessing at exactly the same frequency as our viewpoint, there is a residual field experienced which produces some precessional motion even in the rotating frame. The two components of the applied pulse are now very different: one appears static as it is precessing with the same speed and direction as the viewer (B_1), whereas the other is now seen to be precessing at twice the Larmor frequency, as it was travelling in the opposite direction.

For the duration of the pulse B_1 acts like a torque upon magnetisation M_0, rotating it from the z axis towards the $-y$ axis. M_0 rotates through the angle θ defined by:

$$\theta = \gamma B_1 t \tag{2.4}$$

where t is the length of time the pulse is applied for.

If the pulse is switched on for the appropriate length of time, M_0 can be made to rotate through any angle. For example, following what is known as a $\frac{\pi}{2}$ pulse, the bulk magnetisation can be made to lie along the $-y$ axis, while B_1 has disappeared due to the pulse being switched off, as shown in Figure 2.5.

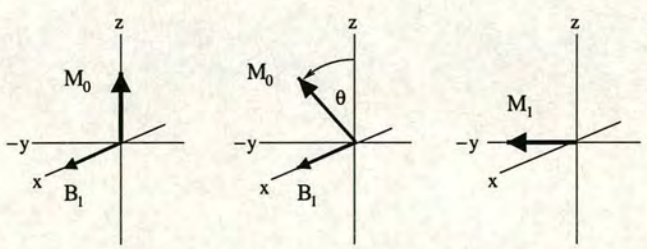


Figure 2.5: Vector diagrams showing effect of a $\frac{\pi}{2}_x$ pulse to the net magnetisation

Once in the xy plane, the magnetisation (now labelled M_1) reverts back to precessing around the external magnetic field at Larmor frequency. This induces oscillating electric current within the wire of the same coils used to produce the pulse, and thus is detectable. To visualise this, we must switch back from the rotating frame to the laboratory frame. In the laboratory frame, the magnetisation M_1 is precessing around the external magnetic field B_0 and can be described as two perpendicular oscillating components that are $\frac{\pi}{2}$ out of phase with each other, as shown in Figure 2.6. One is defined as a cosine, the other as a sine wave. Both decay over time due to relaxation as the magnetisation returns to its equilibrium position along the z axis.

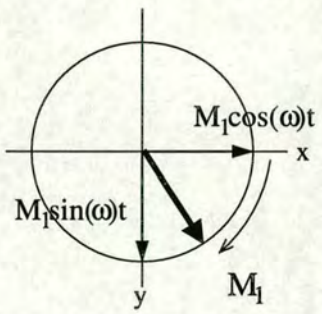


Figure 2.6: Vector diagram showing M_1 as two observable perpendicular oscillating components

The signal is detected over a period of time and recorded as a free induction decay, or FID, a schematic example of which is shown in Figure 2.7. As interest lies in the frequency of the individual spins, a Fourier transform is performed on the FID to change the time domain data to frequency domain data. The method of Fourier transform is shown in the equation below:

$$f(\omega) = \int_0^\infty f(t) e^{i\omega t} \, \mathrm{d}t$$

where $f(\omega)$ is the function of frequency and $f(t)$ the function of time. In practical terms the intergration is performed over a finite range. The Fourier transform produces an NMR spectrum which can then be used to interpret the system under study. A schematic of the transform is shown in Figure 2.7.

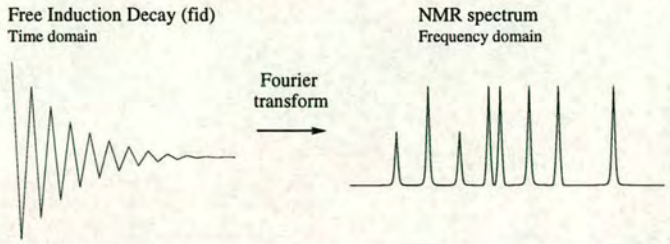


Figure 2.7: Fourier transform. The FID is Fourier transformed to the frequency domain NMR spectrum which can then be interpreted to identify individual spins.

As the pulse and the receiver have different phases, the detected signal is a mixture of the cosine and sine components which in their pure forms, following Fourier transformation, produce absorption and dispersion lines, respectively. These line forms are shown in Figure 2.8. The spectrum can be manipulated to vary the proportion of cosine and sine components during a process called phasing. In this way the signals can be converted to a pure absorption mode with narrow spectral linewidths.

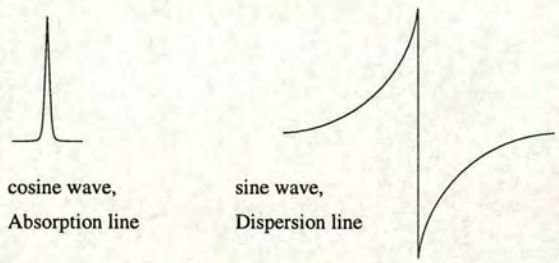


Figure 2.8: Absorption and dispersion lines

2.1.3 Product operator formalism and phase coherence transfer

Vector diagrams are an extremely useful way of describing what happens to the magnetisation during NMR experiments, but as they work from a classical basis there are areas in which they are deficient. Product operator (PO) formalism is a convenient alternative as the effect on the system of events (i.e. radiofrequency pulses or evolution due to either chemical shift or coupling constants) can be represented by rotations of operators, or products of operators, within "operator space". A full description of PO

formalism is beyond the scope of this thesis [34, 43, 35].

PO formalism is, however, useful when describing a crucial aspect of multi-dimensional NMR - the transfer of magnetisation (and therefore also the information within) between coupled spins. In this way, when the operator state of a population of spins has been affected by one of the above-mentioned events, the magnetisation can then be transferred to another population of spins, by use of the appropriate radiofrequency pulses. In this way, for example, it is possible to acquire magnetisation that correlates the chemical shift of each nitrogen atom to the chemical shift of its attached proton. This capability has allowed the creation of complex multi-dimensional NMR experiments which are able to acquire highly specific pieces of information concerning molecules. The basic NMR experiments are detailed in section 2.4 and the application of the most useful multi-dimensional experiments involved in macromolecular NMR are described in methods section 3.3.

2.2 Relaxation

The induced coherence of spins produced during NMR experiments is not a permanent one. The NMR signal gradually decreases over time due to the system returning to its thermal equilibrium state. There are two main ways by which this relaxation of signal occurs.

2.2.1 Longitudinal relaxation time, T_1

This relaxation time describes the speed at which the equilibrium bulk magnetisation, M_0, is repopulated on the z axis following the radiofrequency pulse moving it to a non-equilibrium state (e.g. the xy plane or the $-z$ axis). Equation 2.5 links the relaxation time T_1 to the equilibrium magnetisation M_0 and the current z axis magnetisation, M_z. Figure 2.9 depicts this return to the equilibrium state of the z axis in terms of vector diagrams.

$$\frac{dM_z}{dt} = -\frac{(M_z - M_0)}{T_1} \tag{2.5}$$

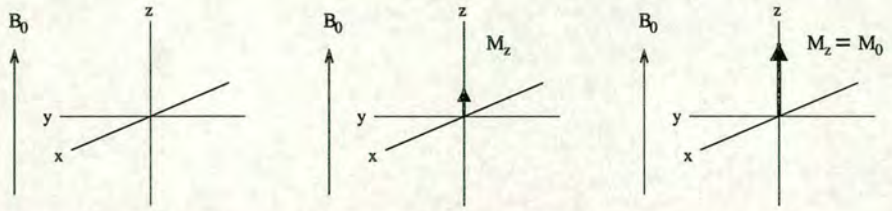


Figure 2.9: T_1 relaxation mechanism. The left diagram shows the state of the bulk magnetisation directly following a $\frac{\pi}{2}$ pulse (M_z=0). After a certain time period the repopulation of the z axis has begun. The final diagram depicts the system back at thermal equilibrium.

This relaxation is caused by interaction with the surrounding environment (lattice), hence one of the names for the constant that quantifies this relaxation is known as the spin-lattice relaxation time. Is it mostly governed by fluctuating dipole-dipole interactions produced by the rotation of molecules in solution. As molecules tumble, their individual spins experience fluctuating magnetic fields produced by nearby spins. This is because of the spatial anisotropy of magnetic dipoles and the fact that each dipole (at least in the short term) does not change its orientation with respect to the external magnetic field. This produces a constantly changing magnetic field around each spin. When the frequency at which these local magnetic fields fluctuate is comparable to that of the spin transitions which govern NMR, these fields allow the dispersion of energy i.e. relaxation. This effect can be compared to the action of the radiofrequency pulses applied in the xy plane, only this time they are local and act only on one spin at a time. Relaxation is therefore closely related to the speed at which the molecules are tumbling in solution i.e. the correlation time, τ_c.

2.2.2 Transverse relaxation time, T_2

Transverse relaxation time can be thought of as the reduction of the measurable coherent signal in the xy plane. As the fluctuating magnetic fields of nearby spins change temporarily, the static magnetic field B_0 experienced by individual spins is slightly altered for a short period of time. This process can also be compared to the application of a short radiofrequency pulse, this time acting along the z axis. Equation 2.6 below shows the relationship between the xy plane magnetisation and transverse relaxation

time, T_2.

$$\frac{dM_{xy}}{dt} = -\frac{M_{xy}}{T_2} \tag{2.6}$$

This relaxation mechanism leads to a loss of phase coherence within the net magnetisation or a "spreading out" of the precessing spins, as shown in Figure 2.10.

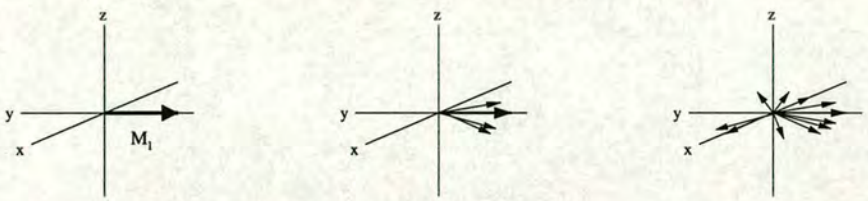


Figure 2.10: T_2 relaxation mechanism. Viewed in the rotating frame, the first diagram shows the net magnetisation, M_1, on the y axis. Following some time, the spins have begun to fan out, and several are no longer in phase with the main pack. The net magnetisation has dropped in intensity, as shown by the thinner arrow depicting it. In the final diagram, many spins have become dephased from the pack and the intensity of the xy magnetisation is dropping to zero.

Transverse relaxation is also known as spin-spin relaxation as the fluctuating magnetic fields most likely originate in nuclear spins, although other sources such as the unpaired electrons of paramagnetic species provide an even more powerful source of relaxation. As the spins do not change their position with respect to the z axis during the course of transverse relaxation, no energy is exchanged between them and the lattice. This is in contrast with the T_1 relaxation. Since both relaxation mechanisms rely so heavily on molecular motion, relaxation analysis can be used as a probe of dynamics.

2.2.3 Relaxation times and correlation times

The way in which relaxation times T_1 and T_2 varying with respect to molecular correlation times is shown in Figure 2.11. For T_1, there is a parabolic relationship between relaxation and correlation times, giving a minimum T_1 when correlation times are between 10^{-10} and 10^{-9} seconds (at 600 MHz). For T_2, the relaxation time is inversely proportional to the correlation time.

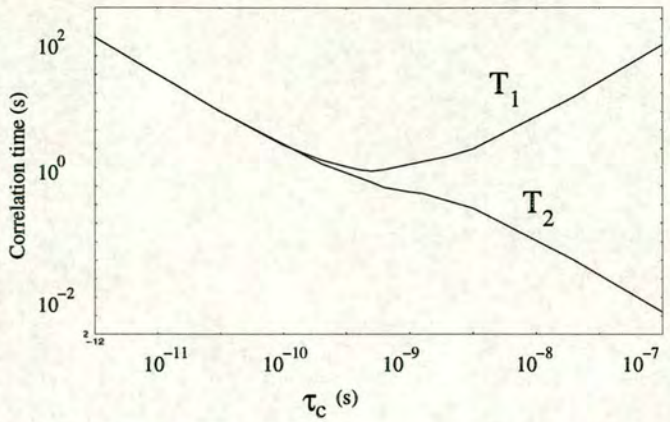


Figure 2.11: How relaxation times vary with correlation time, Larmor frequencies around 600 MHz.

2.3 Nuclear Overhauser enhancement

An NMR-derived structure determination of a protein (or protein domain) is reliant on distance-related information obtained from NMR spectra. The most commonly used method to obtain distance information relates to the phenomenon known as the nuclear Overhauser effect (NOE). This phenomenon provides clues as to the mutual proximity of pairs of atoms. Figure 2.12 depicts the populations of two homonuclear spins at different chemical shifts (and not coupled through covalent bonds).

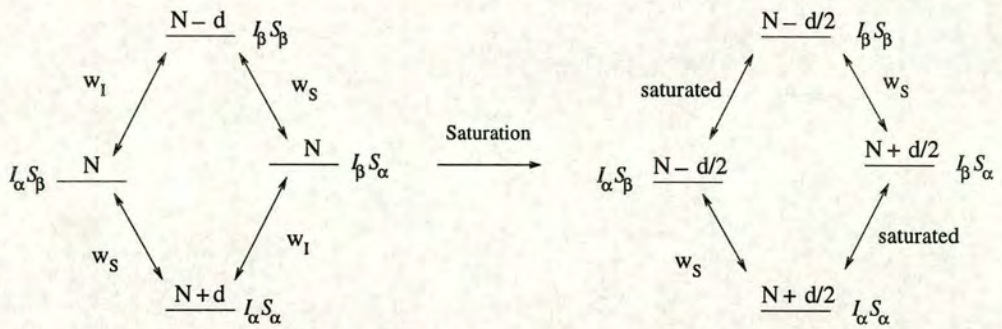


Figure 2.12: Possible single quantum transitions between two spins, I and S, coupled by dipolar coupling. α and β refer to the two spin states of the nuclei. N is the population of spins, altered by amount d in two of the cases. $w_{I,S}$ are the single quantum transitional probabilities between I and S spin states. Follow saturation of spin I, the populations change as shown. Derived from [115].

The four arrows on each half of Figure 2.12 depict the possible routes in which transi-

tions that alter the populations can occur, each being a change in the alignment of the spin of either nucleus I or S. Each of these involves a single quantum change (either α to β or β to α) and can be measured by NMR spectroscopy. Following a saturation of the spin I transition (which is achieved by applying a weak radiofrequency pulse at the Larmor frequency of spin I), the population difference on either side of the spin S transitions is $(N + \frac{d}{2}) - (N - \frac{d}{2}) = d$.

After a system is perturbed (for example by a pulse of radiofrequency) it will try to compensate by restoring itself to its natural thermal equilibrium state. Following the saturation of the transitions of spin I there is a change in the intensity of the spin S peak in the spectrum. It is this change in intensity following the perturbation of the spin system that is known as the nuclear Overhauser effect (NOE) and it occurs due to cross relaxation. Figure 2.13 shows the two possible cross relaxation mechanisms which can occur and cause this intensity change.

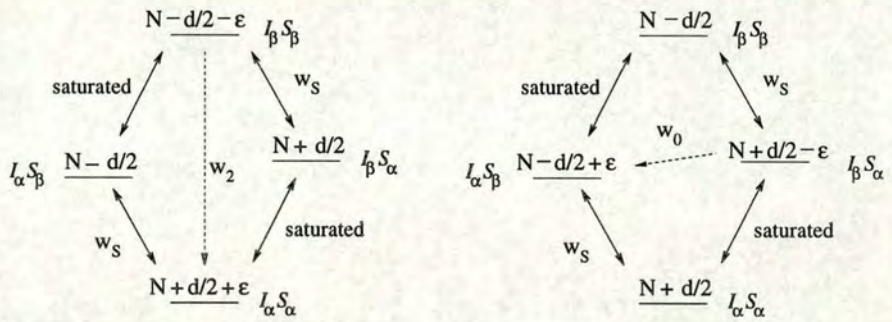


Figure 2.13: Cross relaxation following perturbation to a spin system coupled by dipolar coupling. α and β refer to the two spin states of the nuclei. N is the population of spins, altered by amount d and/or ε. w_0 and w_2 are the cross relaxation probabilities. Derived from [115].

Relaxation along the w_0 and w_2 pathways is referred to as cross relaxation as it involves simultaneous changes of spins states of both spins. Although they are "forbidden" transitions in the conventional sense that they cannot be directly excited by a radiofrequency pulse or cause a directly detectable NMR signal, they can still occur in the context of relaxation. The NOE arises as w_0 and w_2 allow changes to the populations on either side of the S transitions. These changes alter the rate of transition and

therefore the signal intensity, while maintaining the saturated state of the I transitions.

For example, relaxation via w_2 increases the $I_\alpha S_\alpha$ population at the expense of the $I_\beta S_\beta$ population. This changes the population difference along w_S from d (in the simple saturated case shown in Figure 2.12) to $(N+\frac{d}{2}+\varepsilon)-(N-\frac{d}{2})=d+\varepsilon$. The increase in signal intensity brought about by this larger population difference is the NOE. The w_2 mechanism, which leads to a positive NOE, is predominant in small molecules. For large molecules, w_0 is dominant, leading to a drop in population difference from d to $(N+\frac{d}{2})-(N-\frac{d}{2}+\varepsilon)=d-\varepsilon$. The drop in population difference causes a drop in NMR signal intensity and is therefore a negative NOE.

The mechanism for producing NOEs is the dipolar coupling between the two nuclei in question. As the molecule rotates, the two spins maintain their orientation with respect to the external magnetic field, and therefore the local magnetic field experienced by each of these two spins change. The fluctuating magnetic fields between the two nuclei allow the transition from one state to another i.e. the exchange of magnetisation between the spins. The rate of cross relaxation, σ_{IS}, is governed by both the w_2 and w_0 transitions, as shown below:

$$\sigma_{IS} = w_{2IS} - w_{0IS} \tag{2.7}$$

where w_{2IS} and w_{0IS} are the double and zero quantum transition probabilities. Obviously the extent to which the spins can interact via their fluctuating magnetic fields is related to the internuclear distance. The rate of cross relaxation, and therefore the intensity of the NOE, can be more formally stated, as shown:

$$\sigma_{IS} = \frac{h^2\mu_0^2\gamma^4\tau_c}{40\pi^2 r_{IS}^6}\left(\frac{6}{1+4\omega_0^2\tau_c^2}-1\right) \tag{2.8}$$

where σ_{IS} is the rate of cross relaxation, μ_0 is the permittivity of a vacuum, γ is the gyromagnetic ratio for that nucleus, τ_c is the correlation time of the molecule, r_{IS} is the internuclear distance and ω_0 is the Larmor frequency [98].

The rate at which cross relaxation can occur is inversely proportional to the inter-

nuclear distance to the power six, and so the detection of an NOE is evidence that a pair of nuclei are close enough in space to interact. The stronger the NOE, the closer together the interacting nuclei. The information obtained from NOEs thus provides the cornerstone of structure determination by NMR spectroscopy. A two-dimensional NOESY (nuclear Overhauser effect spectroscopy) spectrum contains crosspeaks indicating the cross relaxation occurring between two nuclei. The existence of a crosspeak shows that the two protons are close in space, the usual maximum distance for an NOE to arise being approximately 6 Å. However, the complete assignment (and therefore appreciation of which atoms are close in space) of NOESY spectra can only be completed after the resonances themselves have been assigned.

2.4 Experiments

2.4.1 Basic multi-dimensional NMR experiments

The general scheme of a two-dimensional NMR experiment is shown in Figure 2.14 and consists of five parts. In the relaxation section, the system is at (or returning to) its equilibrium state. The preparation period is where the spin system is prepared for the evolution period. This could be as simple as a single $\frac{\pi}{2}$ pulse (in the case of many homonuclear experiments) or it could be more complicated, for example the application of an INEPT pulse sequence [94] (in heteronuclear experiments).[1] During the evolution period τ_1, either chemical shifts (in correlated experiments) or coupling constants (in J-modulated experiments) are allowed to evolve. Following this, the mixing period is when phase coherences are transferred between spins. In a homonuclear experiment, for example, the magnetisation is moved to a new set of protons. Depending on the nature of the experiment, either scalar interactions (COSY or TOCSY, see sections 2.4.2 and 2.4.3 below) or NOEs (NOESY experiments) are utilised to achieve this transfer. In a heteronuclear experiment, the magnetisation is moved from carbon or nitrogen to proton in a reverse INEPT step. The final stage is the acquisition (τ_2), where the signal is detected by the coils within the NMR probe. Fourier transformation of a series of FIDs in two perpendicular directions produces a two-dimensional spectrum.

---

[1] The INEPT sequence, or insensitive nuclei enhancement by polarisation transfer allows the transfer of magnetisation containing chemical shift information from protons to heteronuclei.

This concept is easily extended to 3D- and 4D-experiments by including additional evolution and mixing periods into pulse sequences.

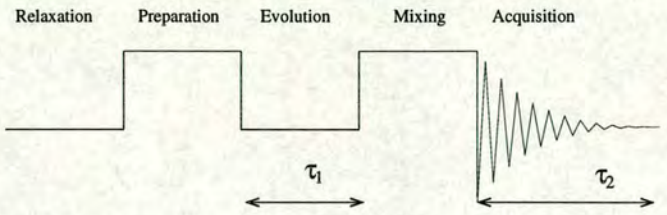


Figure 2.14: Schematic of general pathway of multi-dimensional NMR experiments. The periods τ_1 and τ_2 are when chemical shift labelling occurs.

A variety of NMR experiments have been developed over the years that provide specific pieces of information. The basic experiments used in NMR are described next.

2.4.2 Correlated spectroscopy (COSY)

Correlated spectroscopy involves the transfer of anti-phase[2] magnetisation between (typically homonuclear) spins that are connected by covalent bonding of a geminal or vicinal nature. Coupling constants between protons in proteins separated by four or more bonds are small and produce no crosspeaks in COSY spectra. In this way it can be used to determine the connectivity of atoms within molecules [3, 8].

Due to the limited amount of information available in a COSY spectrum, it is not as often used as the other experiment types. However, used in conjunction with a TOCSY (see section 2.4.3 below), it is extremely useful for clarifying the connectivity of sidechain protons (for example, distinguishing between a residue's H_β and H_γ protons when they have similar chemical shifts).

A consequence of this experiment having anti-phase character in both dimensions is that the COSY crosspeaks contain both positive and negative lines. This can cause problems in overlapped regions of COSY spectra due to partial cancellation of spectral

[2] Anti-phase magnetisation can only be fully described by product operator formalism and involves the use of imaginary numbers. In terms of vector models, it can be imagined as two vectors on the same axis but π out of phase with each other.

lines. This becomes increasingly critical for larger proteins with large linewidths. A section from a COSY spectrum containing several crosspeaks is shown in 2.15.

Figure 2.15: Section of COSY spectrum showing several peaks. Positive contours shown in black, negative contours shown in blue-grey.

2.4.3 Total correlation spectroscopy (TOCSY)

Total correlation spectroscopy (equivalent to cross-polarisation in solid state NMR) allows the passage of in-phase magnetisation mediated by scaler interactions up to a distance of five or six bonds. The mixing period of this experiment consists of a spin-lock period.

During this spin-lock period, the energies of the possible transitions between spins I and S are equalised, allowing the transfer of magnetisation from one spin to the other. (It can also be thought of as the Larmor frequencies of the precessing spins having been equalised as the net magnetisation now precesses around the axis to which the lock was applied.) This will provide a spectrum in which the chemical shifts of protons within a spin system are correlated. The information such a spectrum contains is a useful tool in assigning the backbone and sidechains of small proteins [14].

Using the COSY and TOCSY spectra in conjunction provides a useful way to further determine connectivity. For example, a TOCSY spectrum will link an H_α proton in a long sidechain to the H_β, H_γ, H_δ and possibly H_ε protons. Because the chemical shifts of H_β and H_γ protons are often similar, the COSY spectrum will allow them to be differentiated, as an interaction from the H_α will only go as far as the H_β protons.

As the size of the proteins increases, the efficiency of the TOCSY transfer decreases because it is mediated by relatively small $^{1}H-^{1}H$ coupling constants. In such cases the use of larger $^{13}C-^{13}C$ coupling constants becomes necessary in order to achieve the assignment of protein resonances (see section 2.4.5 below).

2.4.4 Nuclear Overhauser spectroscopy (NOESY)

NOESY sequences require a perturbation to the system to promote cross relaxation between pairs of nuclei within about 6 Å of one another due to the previously mentioned nuclear Overhauser effect. Usually this consists of the application of two $\frac{\pi}{2}$ pulses to the nuclei of concern which move the magnetisation from the z axis to the $-z$ axis. A mixing period is then allowed in which polarisation is transferred between nuclei via cross relaxation. This type of spectrum can therefore be used to correlate any two nuclei that are close in space [88].

NOESY spectra form the cornerstone of NMR-derived structures and provide the vast majority of the experimental restraints used in the structure calculations.

2.4.5 Isotopic labelling

The first NMR-derived protein structures were determined using only homonuclear proton spectra i.e. acquiring information from only the protons [13]. However, a far greater amount of information can be gathered by also using other nuclei, nitrogen and carbon in particular. The NMR active isotopes ^{15}N and ^{13}C are naturally present at only 0.37 and 1.1% respectively.[3] By expressing proteins from hosts which have been grown in media containing only ^{15}N and ^{13}C as sources of nitrogen and carbon it is possible to produce samples with up to 99% isotopic labelling.

This provides two main benefits for an NMR-derived structure determination. Firstly, it allows dispersion of the resonances: no longer is all the chemical shift information of interest found within a single frequency region (that of proton). By including ^{15}N

[3] ^{14}N (99.6%) has spin 1 and due to fast quadrupolar relaxation is not suitable for protein NMR.

and ^{13}C it is possible to create experiments which differentiate between protons not only on the basis of their own chemical shift, but also the chemical shift of a nearby through-bond heteronuclear spin.

Secondly, bonds involving ^{15}N and ^{13}C have much larger coupling constants than those between protons. Figure 2.16 shows the average coupling constants in a protein. The transfer of magnetisation of coupled spins is proportional to the inverse of J, the coupling constant. So in the case of ^{15}N and ^{13}C, the large coupling constants provide efficient transfer of magnetisation with less time available for relaxation to occur.

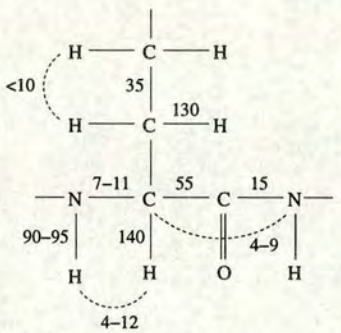


Figure 2.16: Coupling constants within a protein

Isotopic labelling has lead to a large number of multi-dimensional NMR experiments with a variety of purposes. The experiments used in this work are detailed in methods section 3.3.

2.4.6 Heteronuclear single quantum coherence (HSQC)

This experiment involves the transfer of magnetisation between covalently bonded heteronuclei, one of which is a proton. A proton has a relatively high gyromagnetic ratio compared to the nucleus of the heavy atom ($\gamma_{1H}/\gamma_{13C} = 4$, $\gamma_{1H}/\gamma_{15N} = 10$). This means that a larger proportion of the protons contribute to the net magnetisation, compared with the ^{15}N and ^{13}C. In order to take advantage of this, the ^{15}N and ^{13}C magnetisation that undergoes chemical shift labelling during τ_1 originates on protons. An INEPT step at the start of the experiment transfers magnetisation from protons to the other nucleus where it is chemical shift labelled. A reverse INEPT step then

brings the magnetisation back to protons for chemical shift labelling and detection [93].

When working with isotopically labelled proteins, the ^{15}N,^{1}H-HSQC is always the starting point. This experiment can be run on relatively dilute samples and provides an excellent way of determining the quality of the protein sample. As it correlates the chemical shifts of any (non-tertiary) amine nitrogen to its proton, it provides a single peak per residue as each has one amide nitrogen present in the polypeptide backbone. Proline residues are absent from the ^{15}N,^{1}H-HSQC due to their lack of amide proton. Amide groups in residue sidechains (Asn, Gln) will provide extra peaks for those residue types, as will the guanidino and amine groups of Arg and Lys and the indole ring N−H of Trp.

Chapter 3

Materials and methods

3.1 Complement receptor type 1 constructs

Collaborators Dr. Malgorzata Krych-Goldberg, Dr. Xuefeng Wang and Prof. John Atkinson (University of Washington Medical School, St. Louis) completed the expression and purification of the protein constructs used in this project.

3.1.1 CR1~16 construct

An ^{15}N-labelled construct of CR1~16 (i.e. module 16 of CR1 in isolation) was expressed in *Pichia pastoris* using the plasmid pSG5. It consisted of residues K961-P1024 of the native sequence along with a four residue EAEA N-terminal section derived from the *P. pastoris* signal peptide, giving a total construct size of 68 residues. The mutation N987T was carried out to remove the possibility of N-glycosylation. The sample of CR1~16 was made up as a 1 mM solution in 25 mM sodium phosphate buffer, pH 6.0, for acquisition of spectra.

Initially, to verify that the sample was folded and a viable candidate for structure elucidation using NMR spectroscopy, a one-dimensional ^{1}H spectrum was acquired (not shown). On the basis of chemical shift dispersion and linewidth it was determined that the sample was suitable. All spectra required for both chemical shift assignment and relaxation data analysis were acquired by Dr. Dušan Uhrín and Dr. Brian Smith (University of Edinburgh), using a 600 MHz Varian INOVA NMR spectrometer, prior

to the onset of this project.

3.1.2 CR1 site 1 constructs

A variety of double- and triple-module constructs were expressed representing CR1 functional site 1 (modules 1-3). All constructs were expressed in *Pichia pastoris* using the plasmid pSG5. The intermodular linker between each module in site 1 contains 4 residues. The CR1~1-2 constructs contained the entire native sequence from Q1-R122, as well as four non-native N-terminal residues, EAEA, from the cloning procedure. This construct included two of the linker residues following the fourth cysteine of module 2. The mutations N15T and N115T were carried out to remove potential N-glycosylation sites. The CR1~2-3 constructs contained the native sequence from K61-I192, which included two residues of the 1-2 linker residues and two of the 3-4 linker residues. Again, the N-terminus was preceded by EAEA, and the mutation N115T was present. Complete site 1 (modules 1-3) constructs contained Q1-I192, again with EAEA at the N-terminus and the above mutations. Table 3.1 details all the constructs made.

Construct	Unlabelled?	^{15}N labelled?	^{13}C^{15}N labelled?
CR1~1-2	Yes	Yes	No
CR1~2-3	Yes	Yes	Yes
CR1~1-3	Yes	Yes	No

Table 3.1: Site 1 constructs provided.

Preliminary inspections of the unlabelled samples showed that CR1~2-3 had much better chemical shift dispersion in comparison to CR1~1-2, particularly in the amide proton region, hinting that there may be a problem with the CR1~1-2 sample. This was confirmed using the ^{15}N labelled samples: the ^{15}N^{1}H-HSQC of CR1~1-2 (see Figure 3.1) showed a plateau of signal at the random coil shift from amide protons indicating that either one of the modules was unfolded or a portion of the protein molecules were unfolded.

The CR1~2-3 ^{15}N,^{1}H-HSQC showed good peak dispersion and only a small amount of blurring around the random coil shift and is shown in section 5.1.2. The triple

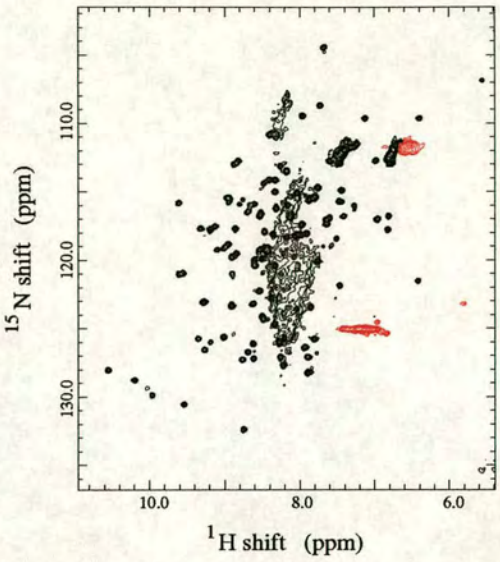


Figure 3.1: ^{15}N,^{1}H-HSQC of CR1~1-2. Acquired on a 600 MHz Bruker AVANCE NMR spectrometer. Dimension 1 - ^{1}H. 512 complex points, sweep width 4000 Hz. Dimension 2 - ^{15}N. 182 complex points, sweep width 2431.5 Hz. Temperature 37°C, pH 6.0. Black peaks are positive, red peaks are negative due to aliasing (which is described in section 4.1).

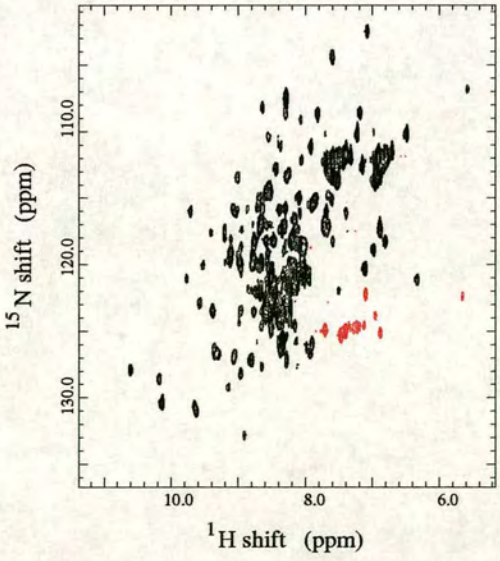


Figure 3.2: ^{15}N,^{1}H-HSQC of CR1~1-3. Acquired on a 800 MHz Bruker AVANCE NMR Spectrometer. Dimension 1 - ^{1}H. 1024 complex points, sweep width 4882.8 Hz. Dimension 2 - ^{15}N. 64 complex points, sweep width 2838.1 Hz. Temperature 37°C, pH 6.0. Black peaks are positive, red peaks are negative due to aliasing (which is described in section 4.1).

module sample ^{15}N,^{1}H-HSQC (see Figure 3.2) indicated a similar, but lesser, folding problem and a large amount of peak overlap that would hamper assignment. It was subsequently decided that solving the structure of modules 2-3 would be the first task in working towards the complete site.

The ^{15}N labelled sample of CR1~2-3 was diluted to 0.7 mM protein in 20 mM sodium phosphate buffer, pH 6.0, in a normal NMR tube. The $^{13}C^{15}N$ sample was diluted to 0.8 mM in 20 mM potassium phosphate buffer, pH 6.0, this time in a 330 μl Shigemi NMR tube due to limited availability.

3.2 Processing

The scripts used to process NMR data acquired using both Varian and Bruker spectrometers are detailed below.

3.2.1 Scripts used for processing Varian data

The FIDs collected on the spectrometer require Fourier transformation, as well as other modifications to provide spectra with highest possible resolution. AZARA software [11] was used for processing of the NMR data. Table 3.2 shows an example of the processing scripts used, in this case to obtain the ^{15}N,^{1}H-HSQC spectrum.

Processing of each dimension acquired in the spectrum requires a section of script detailing the commands to be applied. The commands used are described below. The input file contains the parameters used in acquisition as well as a link to the FID itself. The output file is the Fourier transformed spectrum.

Interlace 2 was used when the experiment involved sensitivity enhancement. Summations of the various signals obtained from phase cycling were required to produce comprehensible acquisition data. The "2" refers to which dimension of the acquisition requires the interlacing - in this case it was the second dimension, nitrogen.

```
input                fid.par
output               spc

interlace 2

script_com 1
    complex
    conjugate
    conv_box 16
    sinebell2 90
    zerofill 1
    fft
    phase 92 58
    upper 1024
    reduce
end_script

script_com 2
    complex
    conjugate
    sinebell2 90
    zerofill 1
    fft
    phase 0 -180
    reduce
end_script
```

Table 3.2: AZARA processing script for Varian data. Script for processing a ^{15}N,^{1}H-HSQC.

Complex initiated the processing of each dimension and indicated to the AZARA software that the data involved was complex.

Conjugate informed AZARA that the complex conjugate of the data was to be used.

Conv_box was used in the processing of the first dimension only and created a convolution (that is, a multiplication) of the FID with a box shaped function, the box being overlaid on the frequency of water. The frequency range within the box was multiplied by 0, and the remainder of the spectrum was untouched. This therefore discarded the water signals and retained the rest of the spectrum. The "16" described the number of points in the half width of the function.

Sinebell2 assisted in removing "sinc wiggles". During zerofilling (see below) a harsh step is created in the FID where the end of the signal has been truncated. The in-

tegration of such a step would produce $(1/x)\sin x$ character giving rise to distortion at the base of each peak. In processing both dimensions, the FID is multiplied by a sinebell squared function that smoothed this step and therefore prevented this apodisation from occurring. The "90" refers to the phase of the sine wave (in degrees) applied.

Zerofilling involved truncating the FID after it had relaxed to the extent that it contained no data that can be usefully interpreted from amongst the noise. The removed part of the FID was replaced with a series of zero intensity data points. This produced an effectively higher sampling number and helped to increase the resolution of the spectrum produced. The "1" dictates that one doubling of the number of points is performed.

Fft fast Fourier transformed the data, changing the time domain of the FID into the frequency domain of the spectrum.

Phasing ensured that the data was entirely in absorption line (i.e. cosine) form to avoid overlap of peaks causing cancellation of signal, as would happen if the peaks appeared partially or completely in dispersion mode. The phasing was set manually using AZARA. The numbers refer to the constant and variable phase corrections to be made to each dimension.

Upper selected only those points acquired in the FID up to the specified number, the excess being discarded. This truncated the spectrum to remove any redundant parts, for example the portion of the ^{15}N,^{1}H-HSQC below 4 ppm were no peaks occur.

Reduce transformed all the complex data into real data for display purposes.

3.2.2 Scripts used for processing Bruker data

These scripts are very similar in form and function to those for Varian data, but require a few additional commands.

Avance 12 24 was added between any `conv_box` and `sinebell2` command in the first dimension. This command performed the first part of a transform required specifically by Bruker AVANCE data.[1]

Avance_phase appeared between the `fft` and `phase` commands and performed the second part of the transform required of Bruker AVANCE data.[2]

Mask_ppmm appeared before the `complex` command in those dimensions requiring it. It formed a $++--$ multiplication for phase cycling, that is it multiplied the third and fourth points of every set of four points by -1.

3.2.3 Maximum entropy method (MEM) processing

This method was designed to improve resolution in the indirectly detected dimensions of 3D experiments. MEM works by randomly producing a wide selection of spectra, reverse Fourier transforming them and then comparing the FIDs obtained to that of an actual data set. In most cases there is no match, but where there is significant correlation these spectra are considered. The spectrum that produced the closest matching FID fit is then used. In the case of more than one spectrum producing a well fitting FID, the spectrum containing the smallest amount of information and still producing the correct FID is used, as it is less likely to contain extraneous peaks. The minimum information spectrum by definition has the maximum entropy, hence the name [78].

To use this method, first a normal 3D script has all commands for the 2^{nd} and 3^{rd} dimensions commented out, with the exception of `complex` and `conjugate`. The FID is then part-processed using this script. Next, a plane cutting through the first dimension (that is, a "2 3 plane" whose axes are dimensions 2 and 3) which contains several peaks is selected and extracted from the part-processed spectrum. To perform MEM

---

[1] As part of the Bruker digital filtering, there is an apparent delay at the start of the acquisition. The data must therefore be shifted left by a certain number of points, and then phase corrected. This command informs the FID to remove the requisite number of points and also gives the magnitude of the phase correction to be performed. The two numbers alongside the command (called DECIM and DSPFVS) refer to the spectral width and the Bruker filtering version being used.

[2] It links the `avance` command to the `phasing` command to insure the phase correction for the spectrum is correct.

on this plane and to optimise the parameters a script called "maxent" is created, which usually takes a form similar to the example shown in Table 3.3.

```
input          m.plane360.par
output         mem360

maxent2_com 1 2
    iter 20.0
    noise 25.0
    log m.log
    rate 0.2

    dim 1
    npts 512
    complex
    phase 49 -180

    dim 2
    npts 128
    complex
    phase 0 -180
end_maxent
```

Table 3.3: AZARA maximum entropy method (MEM) processing script for Varian data. Script for processing a ^{15}N,^{1}H-HSQC.

Maxent2_com specified how many dimensions are to undergo MEM, while the numerals allocate labels to these dimensions.

Iter dictated the maximum number of iterative cycles allowed in the attempt to reach convergence.

Noise gave an estimate of the background noise of the spectrum. The standard deviation of the background noise (i.e. an area in the plane containing no peaks) was calculated using the AZARA software. The noise was then calculated from this and the number of complex points in each dimension.

$$\text{Noise} = \frac{\text{Standard deviation}}{\sqrt{\text{npts1}* \times \text{npts2}*}} \tag{3.1}$$

Log command wrote a log file detailing the progress of the fit to ensure the spectrum's processing was successful and complete.

Rate was a variable within the algorithm that could be used for optimisation if convergence was not achieved, and was set at 0.2 [11].

For each dimension which underwent the maximum entropy method, there were three further commands.
Npts specified the new number of points that the dimension would contain following MEM. This was set to double the original number of points in the dimension.

Complex and Phase performed the same functions as previously described, and the phasing values were retained from the normal processing.

This "maxent" script was then used to complete the processing of dimensions 2 and 3 of the selected plane. Following convergence, the peaks were significantly sharper and intense than when processed without MEM, although the noise level in the spectrum was also increased. When the transformed plane was of suitable quality, the original part-processing script was appended with the maxent script at the bottom. The only change needed was that the `maxent2_com` command should then be followed by "2 3" to ensure that both indirect dimensions underwent MEM.

3.2.4 Contouring

While AZARA software is applicable to sophisticated processing, its assignment capabilities are limited. A second software package, ANSIG [67], is designed for this specific purpose. It allows viewing of the various planes of 3D experiments and also allows confident assignments to be attached to crosspeaks. In order to load the spectrum into ANSIG, a contour file describing the peaks within the spectrum is initially required. For a 2D spectrum a single file was sufficient, but for 3D spectra two contour files were required.

In the case of 3D datasets, contours describing the spectra were created using AZARA. For example the contours for the "1 2" dimensions were created - i.e. the spectrum was

represented by a set of planes whose axes were dimensions 1 and 2, with each plane describing a different frequency in the 3^{rd} dimension. Contours were drawn on each plane representing the intensity of the NMR signal at the appropriate chemical shifts. For each 3D spectrum, two sets of contour planes (viewed from orthogonal angles) were required to provide a complete description of the data. The base contour level, as well as the multiplier between contour levels, required to give the "best" looking spectrum was noted for each view. A contour script was then written for each view, an example of which is given in Table 3.4.

```
input                    spc.par.ref
output                       12.cnt
ppm_range 1                5.0 11.3
dims 1 2

levels          6666          -6666
levels          9333          -9333
levels         18292         -18292
levels         25609         -25609
levels         35852         -35852
levels         50192         -50192
levels         70269         -70269
levels         98377         -98377
levels        137727        -137727
```

Table 3.4: AZARA contour file script. Script for producing contours for a ^{15}N,^{1}H-HSQC.

As this was a ^{15}N spectrum, the `ppm_range` command was included to produce contours only for the selected ppm range of 5.0−11.3 ppm. This prevented bulky, unnecessary contours being created for the water signal at around 4.6 ppm. This script produced up to ten contours for each peak in the spectrum setting them at the arbitrary intensity levels listed.

3.3 Assignment

The appearance of a crosspeak in a multidimensional NMR spectrum is evidence of an interaction between two nuclei within the system under observation. In order to extract any meaningful information from NMR data, each contributor to each interaction observed must be identified. To achieve this, each nucleus (protons being the most

critical) needs to have a chemical shift assigned to it. This is achieved in a step-wise process known as assignment. Only with a complete, or very near complete assignment can the results be interpreted with certainty.

3.3.1 ^{15}N experiments for assignment

For CR1~16, six spectra were collected for the purpose of structure determination. All spectra were acquired at room temperature and a pH of 6.0 by Dr. Dušan Uhrín and Dr. Brian Smith (University of Edinburgh), using a 600 MHz Varian INOVA NMR spectrometer prior to the onset of this project. These are listed in Table 3.5.

Spectrum	Dims	Complex points collected	Spectral widths (Hz)	Mixing time (ms)
^{15}N,^{1}H-HSQC	2	1024 x 24	8000 x 936	-
^{15}N HSQC-TOCSY	3	1024 x 120 x 24	8000 x 7199.4 x 936	69.5
^{15}N HSQC-NOESY	3	1024 x 120 x 24	8000 x 7199.4 x 936	150
TOCSY	2	1024 x 512	8000 x 8000	38
NOESY	2	1024 x 512	8000 x 8000	100
COSY	2	1376 x 512	8000 x 8000	-

Table 3.5: Spectra acquired on CR1~16. ^{15}N was dimension 2 in the HSQC and dimension 3 in both 3D spectra. Dims refers to the number of dimensions in each spectrum.

While the NOESY spectra provide the distance related information required for the structure calculation, when using ^{15}N data they are also of use for assignment purposes.

3.3.2 Backbone assignment in an ^{15}N-labelled sample

Where only an ^{15}N-labelled sample is available (as was the case with CR1~16 in the current study), the 3D experiments ^{15}N HSQC-TOCSY and ^{15}N HSQC-NOESY contain most of the information used to complete backbone assignments. As its name suggests, the ^{15}N HSQC-TOCSY consists of an ^{15}N,^{1}H-HSQC with a TOCSY period added into the pulse sequence. This allows the correlation of the chemical shifts of an amide nitrogen, amide proton and also those protons within five or six bonds of the amide proton and within the same spin-system. The ^{15}N HSQC-NOESY is similar

although instead of through-bond TOCSY information, it correlates the shifts of any protons close in space ($< \sim 5.5$ Å) to the amide proton. Both of these 3D spectra can be visualised as an ^{15}N,^{1}H-HSQC with the indirectly detected proton dimension (either TOCSY or NOESY) at right angles to the plane, making a cube (Figure 3.3).

The identity of the ^{15}N,^{1}H-HSQC crosspeaks can be determined primarily by studying the TOCSY dimension of the ^{15}N HSQC-TOCSY for each peak. An "^{15}N plane" through the 3D spectrum can be extracted showing amide proton shifts on one axis and all proton shifts on the other for a given ^{15}N chemical shift.

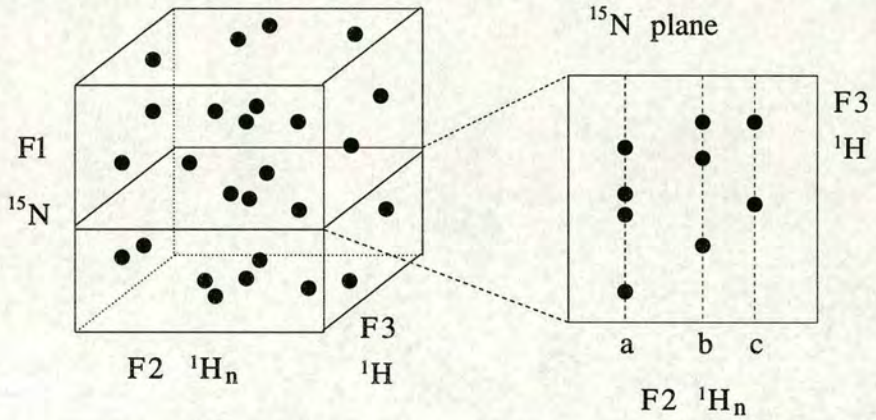


Figure 3.3: ^{15}N plane from a 3D ^{15}N HSQC-TOCSY. The "strip" (shown as a vertical dashed line) for each amide group, *a*, *b* and *c*, shows the chemical shifts of atoms within TOCSY range of that amide group.

Extending in the indirectly detected proton dimension (F3) from each amide proton crosspeak on the diagonal is a "strip" containing various crosspeaks, the number and chemical shift of which give indications as to which residue type that amide group belongs to. For example, glycine residues are distinctive because of the presence of two peaks in the H_α chemical shift region (see Figures 3.4 and 3.5).

However, many residues contain sidechains that frequently only display the H_α and H_β protons in the ^{15}N HSQC-TOCSY, whether because they are the only other protons present in the spin-system, e.g. Ser, Cys, Asn, Asp, or because a tertiary carbon interrupts the TOCSY magnetisation transfer e.g. Phe, Trp, Tyr and His. Some of these

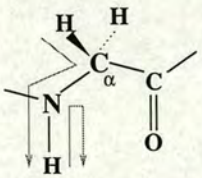


Figure 3.4: TOCSY magnetisation pathway in glycine residue. Glycine residue showing the path of magnetisation from side chain protons to amide proton, then labelling nitrogen before returning to amide proton

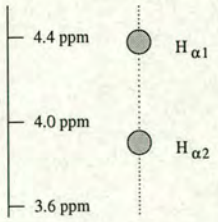


Figure 3.5: H_α proton peaks of a glycine residue

residue types can be putatively identified on the basis of the chemical shifts of their H_β protons.

Following some preliminary residue type determination, the the information in the indirectly detected ^1H dimension of the ^{15}N HSQC-NOESY may be used to help link together residues. Typically, in a β-strand, the proximity of an H_α of one residue to the backbone N−H of the next residue allows a strong NOESY interaction to occur, therefore giving rise to an intense crosspeak [117, 116].

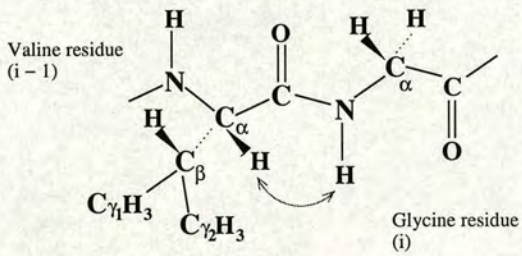


Figure 3.6: Sequential Val and Gly residues with NOE between residue (i) amide proton and residue ($i-1$) H_α

So in this example, where a Gly residue is preceded by a Val (see Figure 3.6), in the NOESY strip corresponding to the N−H of the Gly there appear not only crosspeaks at the chemical shift of the Gly, but also crosspeaks that match the H_α, H_β and H_γ protons of the preceding Val (see Figure 3.7). These strips can be compared to each other for possible matches using ANSIG which displays, in turn, every strip containing a crosspeak of similar chemical shift to the one being examined. In this way the backbone amide N−H groups of each residue can be assigned sequentially.

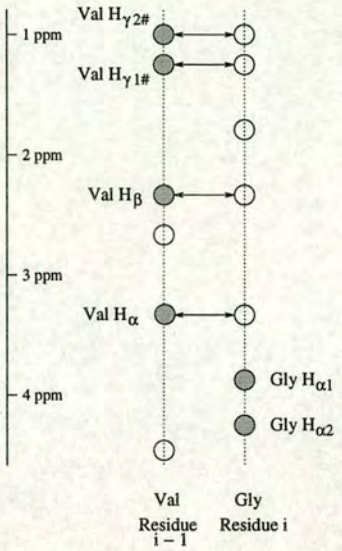


Figure 3.7: NOESY crosspeaks linking a glycine residue with the preceding valine residue. NOESY crosspeaks shown as white circles, TOCSY crosspeaks shown as grey circles. The four proton shifts within the Val residue produce NOESY crosspeaks in the strip for the succeding Gly residue.

3.3.3 Sidechain assignment in an ^{15}N-labelled sample

The aforementioned process allows the assignment of each of the ^{15}N,^{1}H-HSQC peaks to a particular residue in the sequence. The next step is to assign the sidechain atoms. When using ^{15}N-edited data, the ^{15}N HSQC-TOCSY contains chemical shift information on spin-systems that include amide groups. H_α protons can usually be assigned relatively easily for each residue, as in all cases bar glycine there is a single H_α and its shift is generally distinct from any other protons. (But in the case of Ser residues, the H_β shift can be at a higher ppm than the H_α shift.) For H_β and H_γ protons there is a precedent for the position of the shifts relative to one another, and so tentative assignment may be carried out. However, in several cases (e.g. Lys, Arg, Pro) the shifts of H_β and H_γ protons are very similar and cannot be distinguished from ^{15}N HSQC-TOCSY data alone. A second deficiency of the ^{15}N HSQC-TOCSY is that, as the experiment is 3D, the extended acquisition time may allow relaxation of some of the more remote side chains signals to zero, thus removing their chemical shift information.

Conclusive assignment of side chains can be achieved using 2D homonuclear TOCSY and COSY spectra. As the homonuclear TOCSY is not ^{15}N-edited, the shorter acquisition time required allows less opportunity for relaxation and so more intense signals relating to H_γ, H_δ and H_ε protons. However, as there is a large amount of information contained within a spectrum of only two dimensions, overlap does occur frequently. Figure 3.8 shows the 2D homonuclear TOCSY for CR1~16.

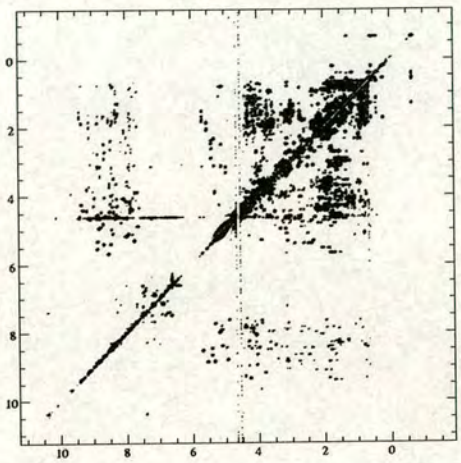


Figure 3.8: 2D Homonuclear TOCSY spectrum for CR1~16. Sweep widths 8000 × 8000 Hz. Complex points collected 1024 × 512.

Problems can occur if the shifts of COSY-linked protons are so similar that the cross-peaks get lost in the diagonal, or if parts of the homonuclear spectra are too over-crowded and overlapped due to the large number of proton shifts.

3.3.4 Proline assignment in an ^{15}N-labelled sample

As proline residues form imide bonds they do not produce peaks in a ^{15}N,^{1}H-HSQC spectrum. This prevents their assignment using the above described methods as the ^{15}N HSQC-TOCSY will not contain any information relating to them, and the ^{15}N HSQC-NOESY will only contain proline chemical shifts in the indirectly detected proton dimension of residues close in space to prolines. To assign the proline, a connection to a more easily identifiable residue must be found. This is commonly a NOESY interaction from the H_α or H_β protons of the proline to the N–H of the following residue. From here the homonuclear TOCSY will allow identification of the other

proline protons, and the COSY will help to differentiate between them.

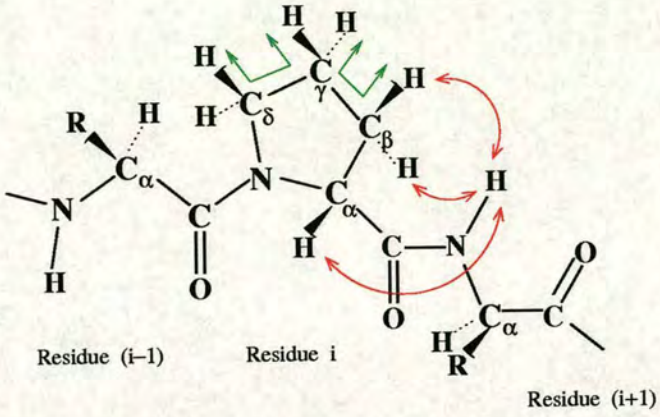


Figure 3.9: Assignment of prolines (residue *i*) initially through NOESY interaction (red) between proline H_α and following residue's (residue *i+1*) N–H group. COSY interaction (green) then allows assignment of protons around the proline ring.

3.3.5 Aromatic ring assignment in an ^{15}N-labelled sample

In the case of residues whose sidechain contains an aromatic ring, the assignment method is different. As a group, the ring protons can be easily identified because they appear in the chemical shift range of 6.0–8.0 ppm [91]. In homonuclear NOESY, TOCSY and COSY spectra, the protons in separate residues are distinguishable by the symmetry patterns. For instance, phenylalanine rings normally have three chemically different ring protons (see Figure 3.10); these three diagonal crosspeaks give rise to a symmetric pattern containing six crosspeaks. This is providing that the ring is able to rotate or "flip" freely as is normally observed in proteins. Otherwise the magnetic environment of each pair of protons would be different, leading to five separate chemical shifts.

Tyrosine's two pairs of chemically equivalent protons leads to a symmetrical pattern of two crosspeaks and the four distinct indole ring protons on tryptophan sidechains leads to a symmetrical pattern of twelve crosspeaks.

The assignment of tryptophan sidechains is made easier by the ability to see the $N_{\varepsilon 1}$

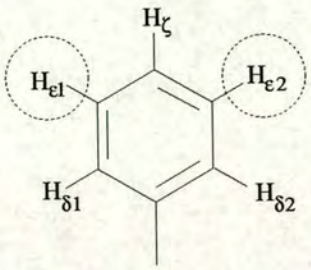


Figure 3.10: Phenylalanine sidechain ring. One pair of chemically equivalent protons are shown circled.

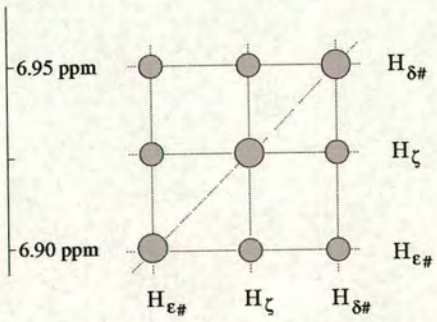


Figure 3.11: Schematic of crosspeaks caused by Phe aromatic ring in 2D homonuclear spectra.

nitrogen in a ^{15}N,^{1}H-HSQC spectrum. The attached proton is often found at an unusually high chemical shift which can aid its identification. From the $N_{\varepsilon 1}$ shifts, the ^{15}N HSQC-NOESY and -TOCSY often clearly show the $H_{\delta 1}$ proton. This provides a useful starting point in the assignment of Trp residues. From there the homonuclear COSY and NOESY experiments are used to assign the rest of ring. The $H_{\varepsilon 1}$ proton shows a strong NOESY interaction to the H_{ζ_2} proton. The assignment is completed by following the COSY/TOCSY interactions around the ring, from one neighbouring proton to another, as shown in Figure 3.12.

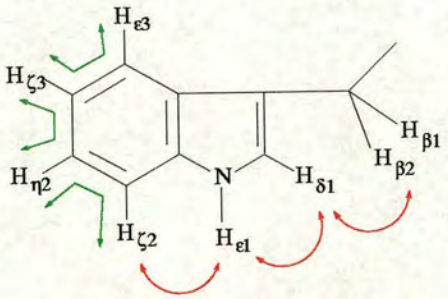


Figure 3.12: Tryptophan residue sidechain showing the protons. The red arrows show the strong NOESY interactions that will occur between the H_β, $H_{\delta 1}$, $H_{\varepsilon 3}$ and the $H_{\zeta 3}$ protons. The assignment can be continued around the indole ring using a COSY experiment shown in green.

In the cases of the other aromatic residues, observation of the prominent NOESY interaction between the H_β protons and the H_δ protons on the ring is often a good way to start assigning each ring.

3.3.6 ^{13}C,^{15}N experiments for assignment

Ideally a protein, or section of protein, being studied by NMR would have both ^{13}C and ^{15}N labelling, though the increased cost of making double-labelled samples means that this is not always so. ^{13}C,^{15}N labelling allows the structure of much larger proteins to be solved by NMR and makes backbone and sidechain assignment much easier as carbon atoms are more widespread within proteins than nitrogen. A double-labelled sample of CR1~2-3 was available for the current study.

As with ^{15}N-labelled constructs, acquisition of a ^{15}N,^{1}H-HSQC and ^{15}N HSQC-TOCSY and -NOESY are essential. Homonuclear spectra are not absolutely required for assignment as most of the information contained within them will be present in the ^{13}C data in a much more accessible form. However, the homonuclear NOESY can provide extra restraints and so was collected in this case. Table 3.6 lists the experiments that were collected on the double-labelled CR1~2-3 sample. The spectra below were each acquired using an 800 or one of two 600 MHz Bruker AVANCE NMR spectrometers (AVA800, AVA600 or BIO600) or a 600 MHz INOVA Varian spectrometer (VAR600).

3.3.7 Backbone assignment in a ^{13}C,^{15}N-labelled sample

CBCA(CO)NH, CBCANH spectra

The ^{13}C,^{15}N experiments were acquired in complementary pairs whose information content is mutually enhanced when the two spectra are compared. CBCANH and CBCA(CO)NH are one such complementary pair [96]. In CBCANH, the chemical shifts of each amide group are correlated with not only the C_β and C_α chemical shifts of their own sidechain, but also with the C_β and C_α of the preceding residue's sidechain. In CBCA(CO)NH the amide shifts are correlated with only the preceding residues C_β and C_α chemical shifts. Figure 3.13 shows the pathway of the magnetisation during these experiments. Comparison of these spectra allow the linkage of an N−H group of residue i to the C_α and C_β carbons of the $(i-1)$ residue with a good degree of confidence.

Spectrum	Spectro-meter	Dims	Complex points collected	Spectral widths (Hz)
^{15}N CR1~2-3 sample: ^{15}N and homonuclear spectra				
^{15}N,^{1}H-HSQC	AVA800	2	1024 x 64	8992.8 x 2184.2
^{15}N HSQC-TOCSY	VAR600	3	512 x 90 x 32	8000 x 1315.0 x 7200
^{15}N HSQC-NOESY	VAR600	3	512 x 90 x 32	8000 x 1315.0 x 7200
TOCSY	AVA600	2	1024 x 256	8992.8 x 7399.9
NOESY	AVA600	2	1024 x 256	8992.8 x 7399.9
CR1~1-2 ^{15}N,^{1}H-HSQC	AVA600	2	512 x 192	8012.8 x 2431.5
^{13}C^{15}N CR1~2-3 sample: ^{13}C^{15}N spectra				
CBCA(CO)NH	AVA800	3	1024 x 80 x 64	11160.7 x 2184.4 x 15105.7
CBCANH	AVA800	3	1024 x 80 x 64	9615.4 x 2184.4 x 14084.5
HBHA(CO)NH	AVA800	3	1024 x 48 x 112	11160.7 x 2184.4 x 5842.0
HBHANH	AVA800	3	1024 x 48 x 64	11160.7 x 2184.4 x 5842.0
(H)C(CO)NH-TOCSY	AVA600	3	512 x 32 x 64	8389.3 x 1637.1 x 11312.2
H(C)(CO)NH-TOCSY	AVA600	3	512 x 35 x 64	7788.2 x 1637.1 x 4378.8
HNCO	AVA800	3	512 x 48 x 80	11160.7 x 2184.4 x 3220.4
HN(CA)CO	AVA800	3	512 x 40 x 64	11160.7 x 2184.4 x 3220.4
^{13}C^{15}N CR1~2-3 sample: ^{13}C spectra				
^{13}C,^{1}H-HSQC	AVA600	2	2048 x 256	7183.9 x 6485.1
HCCH-TOCSY	BIO600	3	1024 x 64 x 128	8389.3 x 6487.2 x 4379.2
Aromatic HSQC	AVA800	2	2048 x 248	12019.2 x 8049.9
(HB)CB(CGCD)HD	AVA800	2	1024 x 96	9615.4 x 5634.6
(HB)CB(CGCDCE)HE	AVA800	2	1024 x 64	9615.4 x 4025.0

Table 3.6: Spectra acquired on the samples of CR1~2-3. Dims refers to the number of dimensions in each experiment.

HBHA(CO)NH, HBHANH spectra

The HBHANH and HBHA(CO)NH are analogous to the CBCANH and CBCA(CO)NH, but in their cases the magnetisation begins on (and labels) the H_β and H_α protons before migrating to the attached carbon atoms and then continuing in the same fashion as the CBCANH and CBCA(CO)NH acquisitions.

HNCO, HN(CA)CO spectra

Another two pair of complementary backbone spectra are the HNCO and the HN(CA)CO [55, 49]. HNCO correlates the amide group shifts with the carbonyl carbon shift of

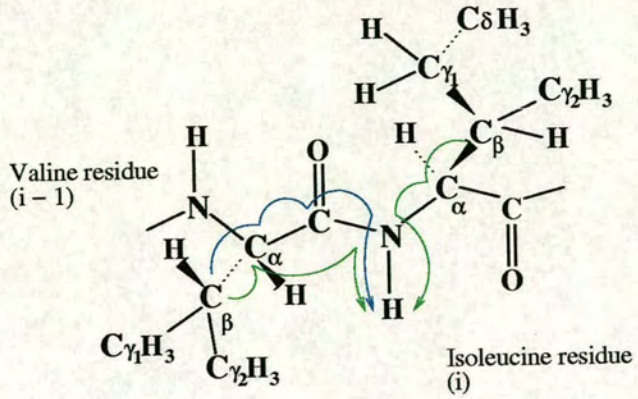


Figure 3.13: Pathways of CBCA(CO)NH (blue) from preceding valine residue ($i-1$) to isoleucine residue amide group, and also CBCANH (green) magnetisation from both valine and isoleucine carbon atoms to isoleucine amide group.

the preceding residue. HN(CA)CO correlates the amide group with both the carbonyl carbon of the current residue as well as the preceding residue. Figure 3.14 shows the pathway of the magnetisation during these experiments.

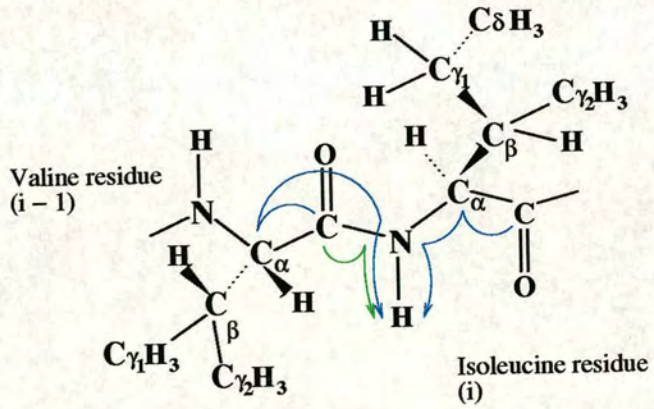


Figure 3.14: Pathways of HNCO (green) from preceding valine residue ($i-1$) to isoleucine residue amide group, and also HN(CA)CO (blue) magnetisation from both valine and isoleucine C_α and carbonyl carbons to the isoleucine amide group.

3.3.8 Sidechain assignment in a $^{13}C,^{15}N$-labelled sample

H(C)(CO)NH-TOCSY, (H)C(CO)NH-TOCSY spectra

A TOCSY mixing period is found in all sidechain $^{13}C,^{15}N$ experiments. In (H)C(CO)NH-TOCSY, the magnetisation moves down from the sidechain carbon atoms, through the carbonyl carbon and onto the amide group of the following residue. H(C)(CO)NH-TOCSY is very similar although the magnetisation begins on, and labels, the sidechain protons before moving to the carbon atoms to which they're bound, then continuing as in the (H)C(CO)NH-TOCSY. Figure 3.15 shows the pathway of the magnetisation during these experiments.

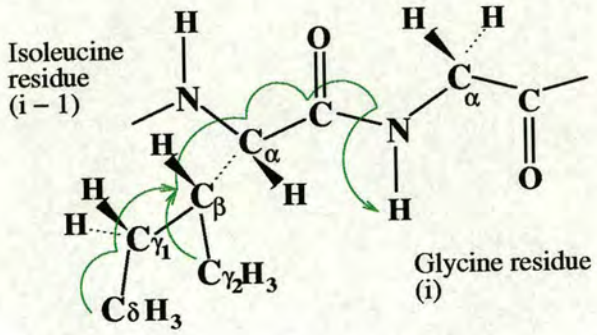


Figure 3.15: (H)C(CO)NH-TOCSY magnetisation pathway from carbon atoms of preceding isoleucine residue to glycine residue amide group

HCCH-TOCSY spectrum

In the other $^{13}C,^{15}N$ sidechain experiments the magnetisation remains on a single residue for the duration of the pulse sequence. One further essential sidechain experiment is the HCCH-TOCSY [58]. This experiment involves less selective acquisition of data as it simply transfers magnetisation from protons to their attached carbons, on to other carbon atoms in a TOCSY mixing period, and then back to proton. This gives information not obtained in the ^{15}N HSQC-TOCSY as here detecting protons is not dependent on the proximity of their carbon atoms to the residue's amide group. This gives a fuller picture of the sidechain protons, including those which can be hard to detect by other methods. The HCCH-TOCSY contains a large amount of information

and is useful for verifying long aliphatic side chains and proline protons.

3.3.9 Aromatic sidechain assignment in a $^{13}C,^{15}N$-labelled sample

There are also a set of 2D aromatic experiments that are very useful for assigning the protons and carbons in the rings of His, Phe, Trp and Tyr residues. A $^{13}C,^{1}H$-HSQC of the aromatic region chemical shifts can be used as a starting point by correlating the shifts of any aromatic region carbon with its attached proton. In conjunction with this, the (HB)CB(CGCD)HD and (HB)CB(CGCDCE)HE [148] experiments can then be used to "move" around the ring. The (HB)CB(CGCD)HD experiment transfers magnetisation from H_β protons to their carbons and then from there through neighbouring carbons and around the aromatic ring. The signal is then selectively transferred back to the H_δ ring protons. This correlates the C_β shift with the H_δ shifts. The (HB)CB(CGCDCE)HE works in a similar fashion but correlates the C_β shift with the H_ε shifts. Provided the C_β shifts of each aromatic residue is known, the ring protons can be assigned. Overlap can cause problems as a protein typically contains several aromatic residues and ^{1}H and ^{13}C shifts may be similar.

In the case of tryptophan sidechain indole protons, the assignment is made easier by the previously mentioned $N_{\varepsilon 1}$ nitrogen that appears in the $^{15}N,^{1}H$-HSQC which can be a useful starting point.

3.3.10 NOESY data in $^{13}C,^{15}N$ samples

In the case of ^{15}N data, the NOESY spectra - specifically the ^{15}N HSQC-NOESY - plays a crucial role in the assignment of the backbone, as well as supplying the distance restraint information that will make the structure calculation possible. However, because NOESY crosspeaks can be produced by a variety of different interacting nuclei, if possible it would be preferrable not to use them for assignment purposes at all. With $^{13}C,^{15}N$ labelled proteins, the assignment can be completed with great reliability, and therefore the NOESY information is not required for the assignment process. The ^{13}C HSQC-NOESY is extremely useful in conjunction with the ^{15}N HSQC-NOESY because including ^{13}C resonances provides an additional set of restraints that can be

used in the structure calculation. It works in the same was as a ^{15}N HSQC-NOESY by correlating the shifts of any carbon atom to the attached proton and then to any nearby proton [110]. It can be visualised as a ^{13}C,^{1}H-HSQC with a third dimension showing the shifts of any protons in the vicinity to the carbon-bound protons.

Following the assignment of the majority of the backbone and sidechain, assignment of the NOESY spectra can begin. The assignment of many of the crosspeaks will be trivial and for those which are not, the ANSIG software can highlight possibilities with the correct chemical shift. During the process of inspecting the NOESY spectra, a proportion of those nuclei that remained unassigned from through-bond experiments may become assigned.

The results of the assignment process for CR1~16 and CR1~2-3 are detailed in sections 4.1 and 5.1.

3.4 Structure calculation

Calculation of an accurate and precise 3D solution structure of a protein, or a protein domain, is one of the many applications of current NMR technology. Following full, or near-complete, assignment of the NOEs for a protein, they can be transformed to internuclear distance restraints which provide the basis for the determination of macromolecular structures.

3.4.1 Data handling

In the current study, each NOE, whether assigned to a pair of contributing atoms or not, was integrated by volume to give it an intensity. The linewidths defining the area surrounding the crosspeak to be integrated were constant: in this work 0.03 ppm for proton dimensions and 0.3 ppm for ^{13}C and ^{15}N dimensions [127]. Although the use this area did not result in integration of the entirity of the peak, it did give an accurate relative intensity of each crosspeak and it was chosen because it minimises errors due to peak overlap.

A complete table of chemical shifts for the protein was then compiled. All chemical shift assignments from the through-bond and NOESY experiments were collated. In the case of a nucleus being assigned in more than one spectra, an average shift was calculated. Any cases where one nucleus had been assigned with two relatively different chemical shifts were flagged to highlight possible assignment errors.

3.4.2 Distance restraint generation

A script from AZARA [11] was used for each NOESY spectrum to compare that spectrum's data with the complete chemical shift list. The NOEs fell into one of two categories: unambiguous and ambiguous. The unambiguous NOEs were those which were given assignments in each dimension manually, using ANSIG, and about which there was no uncertainty as to the nuclei producing the interaction. These unambiguous NOEs and their assignments were left unchanged by the script. The ambiguous NOEs were those which were not manually assigned in every dimension, or completely unassigned. The ambiguous NOEs underwent an automatic assignment procedure. By finding ambiguous NOESY crosspeaks which had chemical shifts values that matched a manually-made assignment, the script matched the unambiguous crosspeaks to possible assignments [99]. The net result was two output files per spectrum: one containing the unambiguously assigned crosspeaks and one containing the ambiguous crosspeaks along with all realistic assignments that could represent the nuclei responsible for each ambiguous NOE.

The strength of the NOE is inversely proportional to the distance separating the two interacting partners.

$$\text{NOE intensity} \quad \alpha \quad r^{-6} \tag{3.2}$$

where r is the distance between the nuclei in question.

Each NOE intensity was scaled to a certain distance range or bracket, and therefore interpreted to represent a specific maximum distance between the interacting nuclei. Table 3.7 shows the distance bound ascribed to each intensity. The first column of numbers shows the NOE intensity (in arbitrary units) required to fit into that bracket.

The second column of numbers shows the maximum distance threshold ascribed to that NOE intensity. For instance, if an NOE peak had an intensity greater than 0.1 (i.e. in the range of $0.1 < \text{NOE} \leq 1.3$) then it was presumed to be occurring over a distance of no greater than 5.0 Å.

Intensity	**Distance**
2.8>	2.7 Å
1.3-2.8	3.3 Å
0.1-1.3	5.0 Å
<0.1	6.0 Å

Table 3.7: Upper distance bounds obtained from relative NOE intensities (arbitrary units).

Only an upper bound was used to prevent errors. It could well be that a weak NOE actually represented a short distance, because the intensity had been reduced, for example by spin diffusion (dispersion of the net magnetisation). It was much less likely that a strong peak arose from a longer distance interaction.

This method therefore used a large number of low specificity restraints. Having each NOE relate to a distance range rather than a fixed distance allowed more flexibility when it came to producing structures that fitted the data. In this way a large number of local structural possibilities were allowed, but due to the large number of restraints used, convergence was achieved. Only a structure that consistently fitted all (or very nearly all) of the restraints was considered to be consistent with the experimental data.

3.4.3 Potential energy input files

An empirically-derived definition of how atoms behave makes possible the production of a realistic structure from the experimental data. This information, known as the force field, includes bond lengths and angles, stereochemistry, steric interactions and electrostatics. In the current work, the standard force field was used - the information in this force field having originated from high resolution X-ray crystallography [83].

The force field contained information describing the atoms present in each L-amino acid residue, their element type, atom "chemical type" (i.e. specified what other atoms were bound to an atom of that "type", for example in Glycine the C_α carbon has two protons attached), the charge on each atom type, which atoms were bonded and where dihedral[3] and improper angles[4] existed. It also detailed the changes to amino acids when they form disulphide and peptide bonds. The force field also described the typical bond lengths, angle values and improper and dihedral angle values for the bonds that can be found within a protein, as well as the strength of any non-bonded (van der Waals) interactions that occur between the various atom types within a protein.

A file containing the backbone sequence of the protein was used together with the force field to generate a molecular template file (MTF) for the protein. The disulphide bridges which were known to exist were defined in the structure here.

3.4.4 Calculation

The structure calculation used in this work employed restrained molecular dynamics (RMD) which solve Newton's equations of motion. Restraints were applied to a model of the structure and the potential energy function on each atom was calculated. Following this, the atoms were given masses and velocities, which were dependent on the temperature of the system, and all forces produced from the potential energy were calculated and applied to the atoms. At a specified time later, the position of each atom was noted and then this process was repeated iteratively. This method was able to circumvent possible problems associated with the structures becoming trapped in local energy minima, as opposed to the overall global minimum.

This worked in tandem with simulated annealing (SA), which involves theoretically heating the protein and then allowing it to slowly cool as one does when annealing

---

[3] Dihedral angles refer to any pair of bonds linked by one further bond, for example the amide N–H and the C_α–H_α bond.

[4] Improper angles are the angles between any two bonds, when they are not covalently linked. In this way they are similar to a dihedral angle except the 2nd and 3rd atoms of the four involved are not covalently bound. The planarity of aromatic rings, for example, is maintained by the presence of improper terms.

metals or glasses to harden them. The high temperature provided kinetic energy to the system (which determined the velocities of the atoms during the RMD) and allowed the exploration of conformational space. At various points during the SA protocol, periods of Powell energy minimisation were undertaken. This calculated the potential energy gradient on each atom then moved the atoms and noted any changes to the potential energy. Moves were accepted if the potential energy dropped. This process was continued iteratively until no further drops in energy occurred and therefore minimisation was deemed to have been achieved. The restrained molecular dynamics were continued for the length of the SA protocol.

This method of calculation is standard for NMR-derived protein structure determination and is described fully in the literature [101, 102, 17] and summarised in reviews [100, 103]. Figure 3.16 shows a basic overview of the process by which the structure calculation is performed.

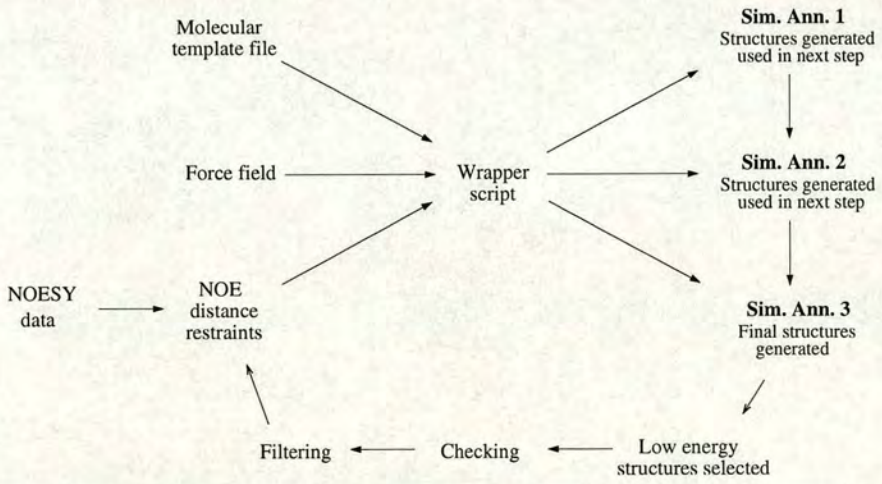


Figure 3.16: Flow diagram showing the key points of the structure calculation method used for CR1~16.

Three inputs were required: the MTF file, the force field and the NOE distance restraints. The main body of the calculation was done by CNS (Crystallography and NMR system) protocols [18, 16]. Three periods of simulated annealing were performed, each involving the structure being heated to 2000 K and slowly cooled to 100 K. A script known as the wrapper script controlled the course of the calculation by feeding

the input files into the correct CNS protocols. Following the generation of structures, they were analysed to select the lowest energy (i.e. most accurate) which were then compared to the NOE restraints. Processes known as checking and filtering (described fully in sections 3.4.6 and 3.4.7 below) allowed the alteration of the NOE restraints files. This whole process was termed "round 1". The calculation was then repeated in its entirity, using the new restraints, in "round 2". To obtain a final set of low energy, relatively precise structures, 6 rounds of calculations were required.

One difference between the protocols used in this work and those given in the literature was the use of prochiral swapping during the periods of simulated annealing. For cases where pairs of methylene protons on the same carbon atom had different chemical shifts, the previously random assignment was moved from one proton to the other and the energy of each possibility was noted. The lower energy version of each prochiral centre was then selected to achieve the more appropriate chemical shift assignments for each of these protons.

3.4.5 Analysis

Following round 1, the structures were analysed to determine which were accurate enough to be included in influencing the next round of calculation. The first step involved determining the potential energy of each generated structure.

The energy of the structures can be thought of as having two parts: empirical and experimental. Firstly, there were the empirical terms which described the energy of different contributions to the molecule as a function of its coordinates in space. This included the energy contributions of each bond and angle, and also van der Waals interactions.

There was also an effective (or experimental) term obtained from the closeness of fit of the restraints. In this work the term was provided entirely by the fit of the NOE restraints. The total energy was given as:

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

where *bond* refers to the energy of the covalent bonds; *angle* refers to bond angle; *dihe* refers to dihedral angles; *impr* refers to improper angle; *vdw* refers to van der Waals interactions and *noe* refers to the contribution from the NOE restraints.

A simple harmonic (i.e. quadratic) function described deviations from the norm of bond lengths and angles. Dihedral and improper angles were described by a similar function. The van der Waals term contained only a repulsive force which was proportional to r^{-12} (where r was the distance between the two atoms).

The energy of NOE restraints was calculated from:

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

where L and U denote the lower and upper bounds (i.e. the distance bracket) assigned to a certain NOE distance and D was the actual distance between the nuclei in that structure. In the case where an NOE has been assigned to more than one set of contributors (during the automatic assignment section), an effective distance (R_{eff}) was calculated using:

$$R_{eff} = (\sum_{ij} R_{ij}^{-6})^{-1/6} \tag{3.5}$$

where R_{ij} is the distance (in Å) between the atoms in each contributing pair. This distance was then used to determine the energy for that restraint. This is known as sum averaging.

In general, the greater the NOE energy term, the greater the number of restraints that have been violated in the structure. A file containing those restraints not satisfied within each structure was generated with each structure file. The violation threshold was set to 0.3 Å, i.e. any restraint for which the measured distance in that structure was more than 0.3 Å larger than the upper distance bound was written to the violations file.

A low overall energy indicates realistic covalent chemistry and good agreement with the experimental data. Such a structure is likely to be relatively accurate. The structures were ranked in order of NOE energy and studied for a significant jump in either NOE

or total energy which separated the structures into different groups. Figure 3.17 shows the NOE and total energy of the CR1˜16 structures during round 1 of the calculation. Table 3.8 lists the NOE and total energy of structures 17-21 (ranked by NOE energy).

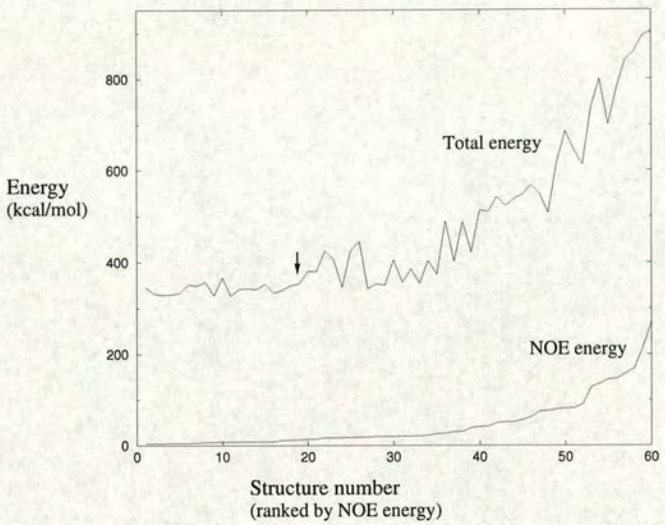


Figure 3.17: NOE and total energy values of CR˜16 structures following a single round of calculation. The cut-off in total energy is shown by an arrow. There was no change in NOE energy at this cut-off.

Structure number	**Total energy kcal mol^{-1}**	**NOE energy kcal mol^{-1}**
17	339.6	11.4
18	349.6	11.9
19	353.4	13.2
20	381.1	14.2
21	380.7	14.5

Table 3.8: Energy cut-off selected between the 19^{th} and 20^{th} CR1˜16 structures after one round of calculation. The cut-off is shown as a horizontal line.

A large jump in total energy (from 353.4 to 381.1 kcal mol^{-1}) occurred between the 19^{th} and 20^{th} structures. Therefore the first 19 structures were selected and were used in refining the restraints before round 2 of calculation, as described in sections 3.4.6 and 3.4.7.

3.4.6 Filtering

A process called filtering was undertaken in an attempt to achieve the most accurate set of assignments. An ambiguous assignment can lead to the intensity of a single NOE being divided between several possible contributors. When preparing the restraints for the subsequent round of calculation, filtering is the term used to describe the process that compares the restraints to the structures selected from the analysis of the previous round. In this way it was possible to determine how realistic was the proportion of the NOE intensity attributed to each candidate pair of nuclei. Two nuclei far apart in space are unlikely to be contributing much, if anything, to an NOE restraint. Filtering is the removal of very minor contributors to NOE crosspeaks which were unlikely to be actually contributing anything. A filtering level was set to 0.99 in the first round and when this proportion of each NOE intensity has been accounted for, CNS added no further possible assignments to that restraint. For example, if the below restraint was filtered at 0.99, the final contributors, which make up only 0.6% and 0.3% of the total respectively, would be removed and this crosspeak could only represent interaction between two pairs - D968:P967 and D968:V974.

Restraint	**Assignment 1**	**Assignment 2**	**Distance**	**Proportion of intensity**
207	968 ASP HN	967 PRO HB2	3.33	0.959
207	968 ASP HN	974 VAL HB	5.97	0.0291
207	968 ASP HN	966 PRO HG1	7.64	0.00659
207	968 ASP HN	973 MET HB1	8.69	0.00305

Table 3.9: Illustration of data used for filtering of restraints.

As the distances set for these two excluded interactions were both over 7.5 Å, the removal of these weak, unlikely contributions would provide more accurate structures in the next round as the structures will no longer try to accommodate them and were no longer penalised for failing to do so.

As the structures became more accurate over subsequent rounds, the filtering level was slowly lowered to remove a higher proportion of unlikely contributors. For the

first two rounds, filtering was set at 0.99. For rounds three and four it was reduced to 0.98 and finally to 0.95 in rounds five and six.

3.4.7 Checking

Following the filtering step, the restraints were then checked. This involved comparing the restraints from each spectrum and looking for duplication. If a certain through-space interaction was represented by a peak in more than one NOESY spectrum, the intensity of each was compared and the restraint with the higher intensity was kept while the redundant lower intensity restraint was removed. This prevented overemphasis of some restraints due to multiplicity.

3.4.8 Subsequent rounds

After both the filtering and checking steps were completed, the set of restraints was employed in round 2 of the structure calculation that followed the same route as the first. Following the second round, the structures were analysed to find the most realistic set, against which the original restraint files were filtered and checked. A third round of calculations were then run, and so on. Decent convergence of the structures had occurred following the 6^{th} round. The structures produced in the 7^{th} round were so similar to those of the 6^{th} round that the restraints file used to produce each set had not been altered significantly during checking and filtering.

3.5 Dynamics and Modelfree v4.1

3.5.1 Relaxation and protein dynamics

While structural information concerning proteins is relatively straightforward to obtain from NMR or X-ray diffraction methods, this does not provide the complete picture. As previously mentioned (section 1.3.4), characterisation of the mobility of proteins is important to understanding of function. The motion of molecules (section 2.2) is intrinsically linked to relaxation mechanisms, and therefore the study of relaxation gives insight into dynamics. The relaxation of an NMR signal is governed by dipole-dipole (DD) interactions between nuclei, chemical shift anisotropy (CSA) and quadrupolar

interactions. Quadrupolar interactions only play a part when deuterium, or other nuclei with a quadrupolar nature, are present, therefore only DD interactions and CSA are of interest in ^{15}N- or ^{13}C,^{15}N-labelled proteins.

3.5.2 Relaxation data

Heteronuclear NMR relaxation data for each residue - that is relaxation times T_1, T_2 and the heteronuclear NOE (hetNOE) - provide information on the overall motion of the protein and internal mobility of residues. In the current work, ^{15}N relaxation was investigated for both CR1˜16 and CR1˜2-3. The "Modelfree" formalism [84, 85, 25] is a useful method of determining information describing both the amplitude and timescale of motions within a macromolecule, providing detail at the level of individual bonds. Only sections of the protein which have assigned N−H resonances can be usefully studied as it must obviously be known which set of crosspeaks belong to which residue in order to interpret the data.

3.5.3 Acquisition of data

T_1 and T_2 are determined from sets of spectra each based on the ^{15}N,^{1}H-HSQC [57, 56] Spectra were acquired with varying delays during which relaxation can take place. Analysis of the drop in intensity over time allowed the calculation of T_1 and T_2. The relaxation delays of the T_1 and T_2 spectra collected on both CR1˜16 and CR1˜2-3 are shown in Table 3.10. Duplicate spectra were acquired for the relaxation delays shown with asterisks.

Protein sample	**Experiment**	**Number collected**	**Relaxation delay (ms)**
CR1˜16	T_1	9	12.2*, 132.1*, 273.8, 600.8* and 1091.3*
CR1˜16	T_2	10	16*, 32, 64*, 128, 196* and 286*
CR1˜2-3	T_1	6	20, 100, 400, 600, 700 and 900
CR1˜2-3	T_2	6	15.9, 31.9, 63.7, 95.6, 111.6 and 127.5

Table 3.10: Relaxation delays for T_1 and T_2 relaxation data spectra acquired for CR1˜16 and CR1˜2-3. Relaxation delays with an asterisk had duplicate spectra acquired.

The CR1~16 experiments were acquired on a Varian 600 MHz INOVA NMR spectrometer by Dr. Dušan Uhrín and Dr. Brian Smith (University of Edinburgh). The CR1~2-3 experiments were acquired on a Bruker 600 MHz AVANCE NMR spectrometer. The spectra were processed using AZARA [11] as described in section 3.2.

3.5.4 T_1 and T_2 data

The method for determining ^{15}N T_1 values from the acquired spectra is described below. Extracting T_2 values from the relevant spectra is an identical procedure.

The intensities of each T_1 ^{15}N,^{1}H-HSQC crosspeak were measured using the integration method described for NOE crosspeaks in section 3.4.1, using the same linewidths. This gave a set of decaying intensities for each residue. In the case of overlapped ^{15}N,^{1}H-HSQC peaks, reliable intensities could not be determined and calculation of T_1 was not possible. Peak errors associated within each spectrum were determined using background noise. To do this, twenty "empty" areas of background containing no crosspeaks were selected and integrated using the standard linewidths. The magnitudes of these integrated values were then averaged to give an error associated with every peak in that spectrum. The intensity and associated error of the crosspeak for each residue, from each of the T_1 spectra, were then collated. This resulted in a file for each residue. A file listing the amino acid residue sequence was also required.

An in-house written program (by Dr. Krystyna Bromek-Burnside) called `fit_gauss` was then used to fit the set of crosspeaks for each residue to the below exponential decay.

$$\mathrm{I}(t) = \mathrm{I}(0) \times e^{-t/T_1} \tag{3.6}$$

where t is the relaxation delay, $\mathrm{I}(t)$ is the crosspeak intensity at time t and $\mathrm{I}(0)$ is the crosspeak intensity when $t = 0$. Figure 3.18 shows the intensity decay for residue H975 T_2.

This program produced an output file listing the T_1 value, and associated error, of each ^{15}N. These relaxation times are most commonly quoted as relaxation rates R_1

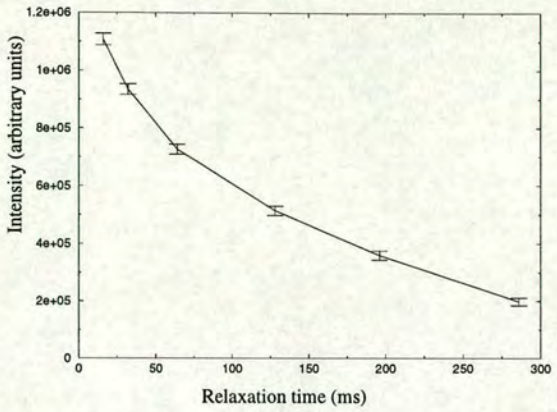


Figure 3.18: ^{15}N T_2 exponential decay for residue H975.

and R_2 i.e. the inverse of T_1 and T_2. Equations 3.7 and 3.8 show the relationship of the relaxation rates to the spectral density (J).

$$R_1 = \overbrace{\frac{d^2}{4}\Big[J(\omega_H - \omega_N) + 3J(\omega_N)\Big]}^{\text{DD contribution}} + \overbrace{c^2 J(\omega_N)}^{\text{CSA contribution}} \tag{3.7}$$

$$R_2 = \overbrace{\frac{d^2}{8}\Big[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H)\Big]}^{\text{DD contribution}} + \overbrace{\frac{c^2}{6}\Big[4J(0) + 3J(\omega_N)\Big]}^{\text{CSA contribution}} + R_{ex} \tag{3.8}$$

where $d = \mu_0 h \gamma_N \gamma_H \left(r_{NH}^{-3}\right) / \left(8\pi^2\right)$, $c = \omega_N \Delta\sigma / \sqrt{3}$ i.e. d relates to the DD contribution and c to the CSA contribution. μ_0 is the permeability of free space; h is Planck's constant; $\gamma_{N,H}$ are the gyromagnetic ratios of ^{15}N and ^{1}H; r_{NH} is the length of the N−H bond; $\omega_{N,H}$ are the Larmor frequencies of ^{15}N and ^{1}H and $\Delta\sigma$ is the chemical shift anisotropy (CSA) of ^{15}N (assuming an axially symmetric chemical shift tensor).

From equations 3.7 and 3.8 it can be seen that both DD interactions and CSA determine the relaxation rates. R_{ex} is a term showing the contribution to R_2 from μs-ms timescale motion, also known as chemical exchange. In the case of protein motions this refers to intramolecular conformational changes occurring within the protein, and in other molecules it could refer to the movement of the nuclei during a chemical reaction.

3.5.5 Heteronuclear NOE data

For the measurement of the hetNOE for each residue, two ^{15}N,^{1}H-HSQCs were recorded. The pulse sequence for one of these, the saturated spectrum, started with a 5 s period which contained 3 s of proton spin saturation. While the proton spin was saturated, an NOE developed between proton and ^{15}N, thus reducing the intensity of each residue's crosspeak. The pulse sequence to produce the second spectrum, the reference spectrum, contained a 5 s delay with no saturation. The peaks in both spectra were integrated and their intensities compared. This steady state NOE is given by:

$$\eta = \frac{[\mathrm{I} - \mathrm{I}(0)]}{\mathrm{I}(0)} \tag{3.9}$$

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

However, the convention during relaxation data analysis is to state the NOE as a simple ratio:

$$\mathrm{NOE} = 1 + \eta = \frac{\mathrm{I}}{\mathrm{I}(0)} \tag{3.10}$$

The ratio of saturated intensity over reference intensity is calculated for each residue that had an assigned, non-overlapped ^{15}N,^{1}H-HSQC peak. Errors were determined and quoted for each crosspeak using the background noise intensity of the spectra, as described above for relaxation times. Equation 3.11 shows how the hetNOE relates to the spectral density function.

$$NOE = 1 + \left(\frac{d^2}{4R_1}\right)\left(\frac{\gamma_H}{\gamma_N}\right)\Big[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)\Big] \tag{3.11}$$

Terms are the same as for equations 3.7 and 3.8. From equation 3.11 it can be seen that DD interactions, and not CSA, is largely involved in determining hetNOE values.

3.5.6 Modelfree theory

Inspection of relaxation data alone is limited in what it can tell us about the internal and global motions of a protein. Modelfree [90, 109] was devised as a way of interpreting relaxation data in a way that is applicable to the "real world", so as to provide insights into the amplitude and timescale of protein motions at the level of individual residues. Whether the motion of the macromolecule is isotropic or anisotropic plays a

part in determining the function and primarily the isotropic case will be detailed.

The correlation function $C(t)$ describes the motion of the molecule as a function of time. If the overall and internal motions are independent i.e. their timescales are suitably different, then the correlation function can be divided into two parts: one which describes the overall motion and another which describes internal motion within the macromolecule.

$$C(t) = C_o(t) \cdot C_I(t) \tag{3.12}$$

$C(t)$ denotes the total correlation function, $C_o(t)$ the correlation function of the overall macromolecular motions and $C_I(t)$ the correlation function of the internal motions. The spectral density function, $J(\omega)$, relates to this total correlation function through the Fourier transform 3.13.

$$J(\omega) = 2\int_0^\infty (cos\omega t)C(t)dt \tag{3.13}$$

where ω is the Larmor frequency for the nucleus in question and t is the time. The overall correlation function can be described as an exponential dependent on both time t and τ_m, the macromolecular correlation time.

$$C_o(t) = \frac{1}{5}e^{-t/\tau_m} \tag{3.14}$$

The internal correlation function can be described as a sum of exponentials involving τ_i, the internal correlation times.

$$C_I(t) = \sum_{i=0} a_i e^{-t/\tau_i} \tag{3.15}$$

At time ∞, all possible motions will have occurred and all possible spatial positions occupied, and so $C_I(\infty)$ can be equated with a generalised order parameter, S.

$$C_I(\infty) = S^2 \tag{3.16}$$

When an internal motion is totally isotropic (i.e. the vector between interacting nuclei can take on any orientation), $S = 0$. Totally restricted motions have $S = 1$. S^2 satisfies the inequality $0 \leq S^2 \leq 1$ and is used as a measure of the amplitude of motions.

Below is shown the simplest approximation for $C_I(t)$, where the superscript A denotes an approximation.

$$C_I^A(t) = S^2 + (1 - S^2)e^{-t/\tau_e} \tag{3.17}$$

τ_e, the effective correlation time, is determined by ensuring the integrals of $C_I(t)$ and $C_I^A(t)$ are equal. The equations displaying total correlation function and from that, spectral density function, are shown below.

$$C(t) = \frac{1}{5}S^2 e^{-t/\tau_m} + \frac{1}{5}(1 - S^2)e^{-t/\tau} \tag{3.18}$$

where

$$\tau^{-1} = \tau_m^{-1} + \tau_e^{-1} \tag{3.19}$$

$$J(\omega) = \frac{2}{5}\left[\frac{S^2\tau_m}{1 + (\omega\tau_m)^2} + \frac{(1 - S^2)\tau}{1 + (\omega\tau)^2}\right] \tag{3.20}$$

If the motion occurs in the extreme narrowing limit, $\tau = \tau_e$ and by comparison with the equations that show R_1, R_2 and hetNOE to be functions of the spectral density function, it is possible to solve the equation.

Therefore for this formalism to provide meaningful results, there are two conditions that must be met. (1) The overall motion must be isotropic. (2) The internal motions must be sufficiently faster than the overall motion ($\tau_e << \tau_m$) and lie in the extreme narrowing limit. If the internal motions are much slower than the overall motion (i.e. chemical exchange motion), Modelfree can again be used, although the generalised order parameter generated will relate only to the fast internal motions. In the case of internal and overall motions being comparable, Modelfree is unable to separate the two and fails to provide a valid description of these motions [84, 85].

Modelfree derives its name from the fact that when order parameters and correlation times are calculated from relaxation data, they do not specify motion following any model - the values are relevant to any model of motion to which they fit [84, 85]. Recently a fully automated version of Modelfree (FAST-Modelfree) has been described [21].

3.5.7 Global correlation time

To use Modelfree, an overall correlation time for the molecule must first be estimated. This can be estimated from R_2/R_1 ratios, provided that only fast motion, distinct from the overall macromolecular motion, is contributing to the ratios. Firstly, all residues with a hetNOE less than 0.6 were excluded as they were likely to be flexible. The mean R_2/R_1 of all remaining residues was calculated, and all residues with a R_2/R_1 not within one standard deviation of the mean were also excluded. A program called `r2r1_tm` (available from the Palmer group website [108]) was then used to estimate an overall correlation time, τ_m, for the protein. Table 3.11 shows the global correlation times calculated for both CR1 constructs.

Construct	τ_m **(ns)**
CR1~16	3.75
CR1~2-3	8.42

Table 3.11: Macromolecular correlation times for CR1 constructs.

3.5.8 Modelfree input files

In order to prepare the input files for the Modelfree software [90, 109], an in-house script was used (courtesy of Dr. Krystyna Bromek). The R_1, R_2 and hetNOE data provide the basis for the inputs required by Modelfree. The inputs include the spin relaxation data for each amide as well as the spectrometer field strength for protons in MHz. Upper and lower bounds for the fitting of each parameter are also required. The length of the amide N−H bond was taken for this work to be 1.02 Å and the chemical shift anisotropy was taken as −172 ppm [2]. The gyromagnetic ratio for ^{15}N was given as $-2.71 \times 10^7\ [Ts]^{-1}$. The global correlation time for the protein was also supplied to Modelfree.

3.5.9 Modelfree fitting

Interpretation of relaxation data using Modelfree was achieved by fitting the spectral density function parameters S^2 and τ_e to the relaxation data. There were four possible parameters that can be fitted to the data depending on which timescale was being studied - two order parameters and two correlation times.

- S_s^2, the order parameter. If a distinction between fast and slow motion was made, this order parameter describes slow motions.
- S_f^2, the order parameter for fast motions. If no distinction between different timescales of motion was made then this was set to 1.
- τ_e, the fast (ps-ns) timescale motions correlation time.
- R_{ex}, the contribution to R_2 from the slow (μs-ms) timescale motions which are also called chemical exchange.

The order parameter was set to $S^2 = S_s^2 \times S_f^2$. If there was no distinction between fast and slow motion then $S = S_s$. An algorithm fitted the internal motion parameters to the relaxation data by least-squares regression. The sum of the errors between the real relaxation data and the relaxation data determined by back-calculating from the fitted motion parameters, gave an estimate of the quality of the fit. The sum of squared errors (SSE) was described as below.

$$\chi^2 = \sum_{i=1}^{N} SSE(i) = \sum_{i=1}^{N} \sum_{j=1}^{M} \left[\frac{(R_{1ij} - \hat{R}_{1ij})^2}{\sigma_{R_1ij}^2} + \frac{(R_{2ij} - \hat{R}_{2ij})^2}{\sigma_{R_2ij}^2} + \frac{(NOE_{ij} - \hat{NOE}_{ij})^2}{\sigma_{NOEij}^2} \right] \tag{3.21}$$

where R_{1ij}, R_{2ij} and NOE_{ij} are the relaxation values for the i^{th} and j^{th} spins and $\hat{R}_{1ij}$, $\hat{R}_{2ij}$ and $\hat{NOE}_{ij}$ are the values back-calculated from the fitted internal motion parameters. The σ values are the errors associated with the relaxation data. N defines the number of spins in total and M the number of different magnetic field strengths used in acquiring the data [90, 109].

A series of Monte Carlo simulations of the data were generated by randomly adding noise terms to the relaxation data and redoing the calculations. The χ^2 value over

the set of simulations was used as an indication of the quality of the fit, and to discriminate between adequate and inadequate fitting to individual residues. Firstly, a confidence level was selected - 0.05 for both the CR1˜16 and CR1˜2-3 work. When the χ^2 value of a residue was lower than the value of the (1−[confidence level]) × 100% of the distribution of the simulations, then the fit was taken to be adequate. For the residues that failed this test, fitting of a more complex parameter set was required to define their internal motions. The Monte Carlo simulations were also used to estimate the error in the fitted parameter values.

There are five models combining the motional parameters which can be fitted to the data in turn.

- Model 1: S^2
- Model 2: S^2 and τ_e
- Model 3: S^2 and R_{ex}
- Model 4: S^2, τ_e and R_{ex}
- Model 5: S_f^2, S_s^2 and τ_e

Models 1 and 2 assume that motion on the slow timescale is negligible. Model 1 assumes that all of the internal motion is very fast, on a timescale of less than 20 ps. Model 5 assumes that motion described by S_f^2 is on this timescale, while S_s^2 describes a motion on the the hundreds of ps timescale. τ_e in model 5 therefore relates to S_s^2, and is sometimes called τ_f.

Figure 3.19 shows the process of assigning models to each residue. The residues not fitted by model 1 were fitted next by model 2, and again those residues with SSE values lower than the 0.95 cut-off were selected as having this model fit them. However, as a second fitting parameter had been added, the reduction in SSE may simply have been a response to the extra degree of freedom to which the parameters could fit. An F test was carried out between models 1 and 2 to determine if any improvement in the

fit was statistically significant, generating an F statistic for each spin:

$$F = \left[\frac{d_2}{(d_1 - d_2)}\right] \frac{SSE_1(i) - SSE_2(i)}{SSE_2(i)} \tag{3.22}$$

where d_1 and d_2 are the degrees of freedom in each model [90, 109].

A second confidence level was selected - in this work 0.2 - and if the F statistic was larger than the (1−[confidence level]) cut-off, fitting was taken to be a significant improvement and a more suitable model was deemed to have been found for that residue. Spins for which parameters were not fitted using either 1 or 2 were fitted with model 3 and tested in a similar way to that used for model 2. For those residues not fitted by any of the first three models, a re-check of model 1 was carried out before attempting to fit the parameters of models 4 and 5. This was done to avoid false fitting of the more complex cases when in fact the simpler one-parameter case would be more appropriate. The SSE of each spin not fitted by model 1 was studied, and those deemed to be reasonably close (in the current work, the χ^2 was to be within three times the distribution value at the 1−[confidence level]), were selected as being fitted by model 1.

Following this, the spins still not fitted by a model were fitted with model 4, and those with an SSE of zero were accepted. Finally, remaining spins were fitted using model 5 using the same selection criteria. Any spins remaining were deemed to be "not fittable" by any of models. One explanation of unfittable residues is that they are displaying internal motion that is within the intermediate timescale (ns-μs) which Modelfree cannot simulate.

3.5.10 Anisotropic fitting

The case described previously is appropriate to interpretation of relaxation data from constructs in which the global protein motion is isotropic i.e. the protein tumbles as if it were a sphere. The majority of proteins however have anisotropic structures: for example, CCP modules typically have dimensions of approximately 40 Å x 15 Å x 15 Å [59, 128]. An analysis that takes into account the three dimensional protein structure, and therefore also the contributions of diffusional anisotropy to the relaxation,

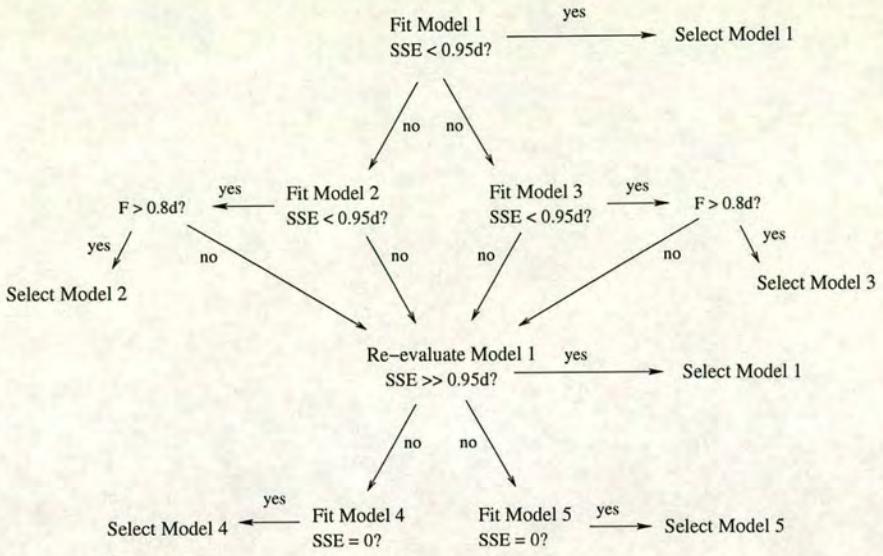


Figure 3.19: Flowchart showing an overview of the Modelfree parameter fitting. Each spin is tested against five models to see if any can fit the relaxation data and describe the internal motions of that residue. d refers to the distribution of the Monte Carlo simulations.

would be even more useful. For this, a reliable, NMR-derived or X-ray crystallographic structure of the protein is required.

Fully anisotropic diffusion is characterised by its diffusion tensor $\mathbf{D}$, with three distinct D_i values (D_{xx}, D_{yy} and D_{zz}) which can be translated to five correlation times. In the axially symmetric case, there are three distinct elements in $\mathbf{D}$, and three correlation times fully describe the diffusion. These are frequently reported as an overall correlation time and two $\mathbf{D}$ elements: $D_{\parallel}$ ($= D_{zz}$) and $D_{\perp}$ ($= D_{xx} = D_{yy}$). The internal correlation function involves the orientation angle θ of the N−H bond with respect to the main axis of the diffusion tensor $\mathbf{D}$. A second angle, ϕ, is required to fully describe the position of the N−H bond vector and the $\mathbf{D}$ elements. In the familiar isotropic case all of the $\mathbf{D}$ elements are identical, the dependency on the orientation of the main axis becomes irrelevant and only one correlation time is required to describe the diffusion.

To visualise the directions of the diffusion of the molecule it was useful to first use a program called `pdbinertia` (available from the Palmer group website [108]) to calculate the principal moments of inertia from a PDB-format (protein database) file

describing the atomic coordinates of the protein. It outputs a new PDB file with the centre of mass located at the origin and the moments of inertia rotated to align with the cartesian axes describing the coordinates.

A program called `r2r1_diffusion` (also available from the Palmer group website [108]) was then used to determine the diffusion tensor from the relaxation data and the respective orientations of the N−H bonds within the protein structure. Its inputs were the R_2/R_1 ratios for each residue (and associated errors), the output structure from `pdbinertia` and an initial estimate for the value of the isotropic diffusion tensor, $\mathbf{D}_{iso}$. The $\mathbf{D}_{iso}$ estimate was calculated by $1/(6\tau_c)$ using the previously determined global correlation time. The relaxation data was fitted by $\mathbf{D}_{iso}$, $D_{||}/D_{\perp}$, θ and ϕ by again selecting the lowest determined χ^2 values.

Fitting of the axially symmetric anisotropic diffusion tensor was attempted for both CR1˜16 and CR1˜2-3 (using a homology model of the module pair, detailed in section 5.3). For CR˜16, no improvement on the previous isotropic fit was detected. These results are discussed in section 4.6. In the case of the CR1˜2-3 models, an improvement was detected. However, as there was no reproducibility, and as the intermodular orientation within the 2-3 models cannot be presumed to be accurate, anisotropic fitting was not completed for CR1˜2-3. The isotropic results, and the drawnbacks due to anisotropy not being accounted for, are described in full and discussed in section 5.6.

3.6 Modeller v6.0

Modeller [123] is a program able to generate high quality three-dimensional models of structures by comparative modelling based on similar structures previously determined from NMR or X-ray data.

3.6.1 Comparative modelling

In cases where an X-ray crystallography or NMR-derived structure of a protein is not available, theoretical structure modelling may be helpful. Comparative modelling is one such theoretical method. It relies on the fact that proteins which have a high

sequence similarity will also have high structure similarity. A protein of unknown structure can therefore be modelled, in some cases, using the solved structures of related proteins which share high sequence similarity. A level of sequence similarity greater than 40% is usually sufficient for Modeller to generate a reasonably accurate model [131]. CCP modules have sequence similarity (or "are homologous") within their family and are therefore, in some cases, candidates for comparative modelling. While comparative modelling will not, at present, produce a result with accuracy as high as NMR or X-ray structures, the quality of the structure produced is increasing as more experimentally determined structures become available as templates.

3.6.2 Modeller 6.0 method

The standard Modeller method was used [124], the fundamentals of which are described below. Modeller has three stages leading to the prediction: sequence alignment, restraint determination and restraint application. In addition to the structures of related proteins, Modeller requires an accurate manual sequence alignment as a starting point. It is important that any amino acid residue insertions and deletions are taken into account so that the correct residues in each sequence are regarded as equivalent. Small variations in alignment can produce large variations in predicted structure. For the use of Modeller to model CCP modules as described in the current work - predicting modules 2-3 using modules 16-17 as a template - the number of template residues in each functional site is identical and there were no problems in obtaining an alignment.

Following the determination of a suitable alignment, spatial restraints to be imposed upon the modelled sequence are extracted from the template structures. The larger the number of structures used, the better the predicted model will be. The restraints to be applied to each modelled residue are determined as probability density functions (PDFs) that describe the likelihood of certain conformations occurring within that residue. For example, the sidechain dihedral angle χ_1[5] can be predicted from each residue position using (and is therefore dependent on) the backbone ϕ and ψ angles and the residue type. Each PDF, therefore, is a description of the frequency of occur-

[5] This is the dihedral angle between the backbone and the $C_\alpha - C_\beta$ bond in amino acids

rence of each value for that variable, dependent on other variables. Firstly, each bond or dihedral angle variable to be restrained is described as a PDF dependent on other variables within the structures. Combined local PDFs will produce a molecular PDF which describes the probability of finding a structure with each possible combination of conformations. Applying this molecular PDF to the sequence will therefore give the most probable structure.

An extended polypeptide is used as the starting point, with only the short range sections of the molecular PDF (ie. those which describe an intraresidue feature) being applied. The more long range sections are added incrementally until the whole molecular PDF has been applied. Further steps are then completed using the output from the previous step as a starting point. A set of structures are predicted by using a variety of starting conformations.

Chapter 4

CR1, module 16

4.1 Assignment

Backbone assignments in module 16 were completed primarily using the strong NOESY crosspeak from the H_α of the $(i-1)$ residue to the amide proton of residue (i). This method is detailed in section 3.3.2. Figure 4.1 shows an example of the sequential assignment. Sidechain assignments were completed using the method in section 3.3.3.

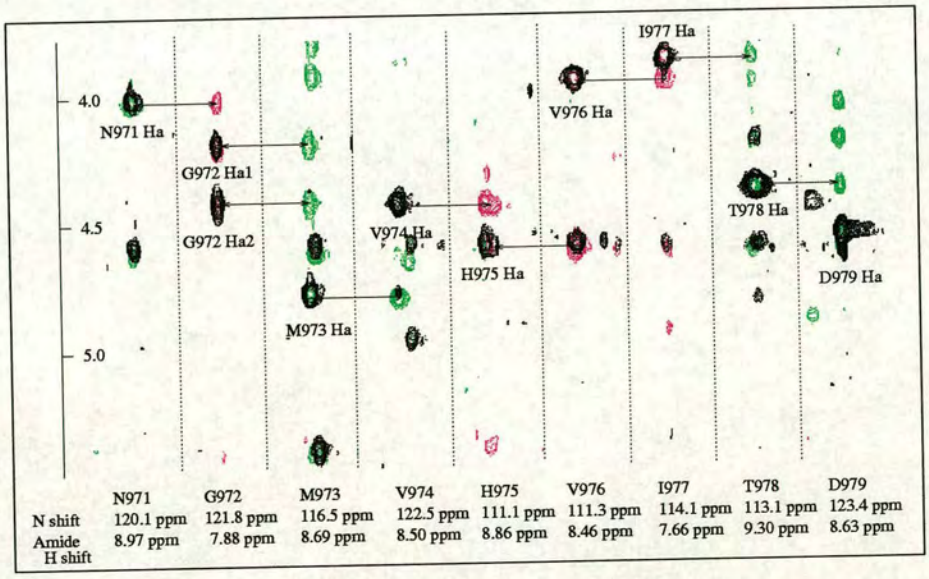


Figure 4.1: Strips from ^{15}N HSQC-TOCSY (black) and -NOESY (green/pink) spectra of sequential residues N971-D979. The presence of a strong NOESY crosspeak from the H_α of the $(i-1)$ residue to the amide proton of residue (i) suggests these are sequential residues.

The assignment of both backbone and sidechain atoms of module 16 was accomplished to near-completion. Figure 4.2 shows the assigned ^{15}N,^{1}H-HSQC. The overlapped, boxed section is magnified in Figure 4.3. The red peaks are negative due to aliasing. This means that their frequencies are outwith the range selected by the experiment and so they are detected at false shifts. G972 has an amide ^{15}N shift of 105.91 ppm. All other aliased peaks have genuine ^{15}N chemical shifts higher than 125 ppm. True chemical shifts for the aliased residues are contained within the assignment list in Appendix A.

Out of 64 native residues, 35 were fully assigned with respect to backbone and sidechain ^{15}N and ^{1}H atoms. They were: C963, P966-V974, V976-I977, D979-G983, I986, S989-T991, L996-G998, A1002-S1007, A1011, W1013, C1020-Q1021, I1023. (Non-native A960 was also fully assigned, while E57-E59 were not.)

Ten further serine and threonine residues were effectively fully assigned with the exception of their OH group proton (H_γ for serine and H_{γ_1} for threonine): S962, T965, T978, S984, T987, S1000-S1001, T1010, S1014-T1015. One further threonine, T992, had no H_α assignment as the chemical shift of this proton probably overlapped with the residual solvent water signal. An assignment for the OH group proton of the single tyrosine Y988 was also missing. If these serine, threonine and tyrosine residues are included, this gives a total of 47 residues out of 64 that were fully assigned.

Three out of the six proline residues remained partially unassigned: in P1017 the H_α and H_βs were not assigned while in P1018 the H_γs and H_δs remained unassigned. Figure 4.4 shows the positions within the sidechain of these unassigned protons. This set of seven protons could not be assigned by identifying crosspeaks from either neighbour within the chain as it appeared that within each of these prolines the chemical shifts of the H_βs and H_γs were very similar. This mean that any crosspeaks between the two in the COSY and TOCSY spectra were hidden in the diagonal. In P1024 the H_α and H_βs also remained unassigned. As it was the C-terminal residue, an assignment based on (i) to $(i-1)$ crosspeaks was not possible.

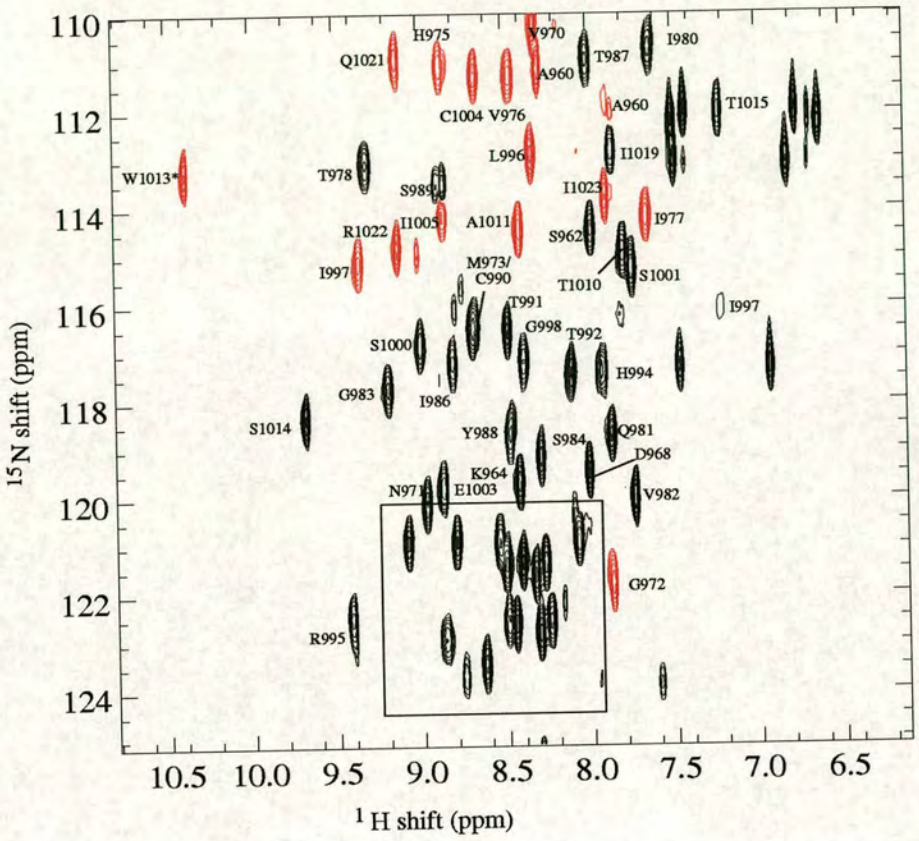


Figure 4.2: Assigned ^{15}N,^{1}H-HSQC of CR1~16. Acquired on a 600 MHz Varian INOVA NMR spectrometer. Dimension 1 (x axis) - ^{1}H. 1024 complex points, sweep width 8000 Hz. Dimension 2 (y axis) - ^{15}N. 24 complex points, sweep width 936 Hz. Black peaks are positive, red peaks are negative due to aliasing (as described in section 4.1). True ^{15}N chemical shifts for the aliased residues can be found in Appendix A.

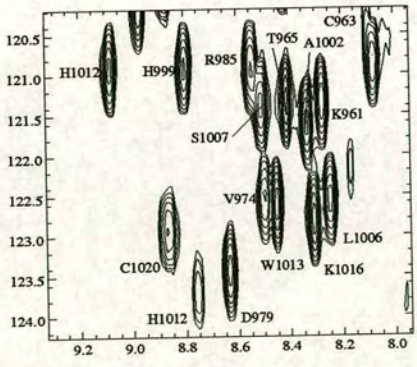


Figure 4.3: Enlarged region of assigned ^{15}N,^{1}H-HSQC of CR1~16

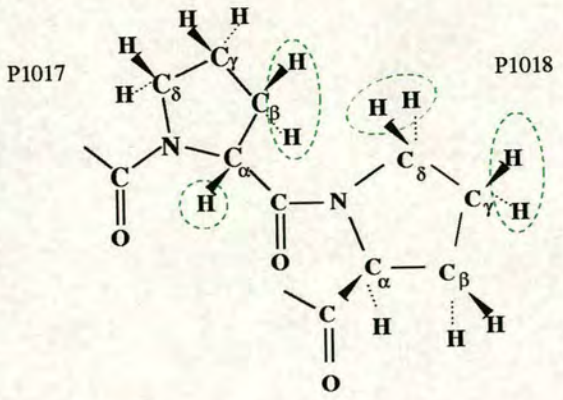


Figure 4.4: Unassigned protons in P1017-P1018. The H_α and H_βs of P1017 remained unassigned, as did the H_γs and H_δs of P1018. Unassigned protons are circled.

The two N-terminal residues of the signal secretion peptide (E957-A958) remained totally unassigned and of the next residue (E959) only the amide nitrogen and amide proton were assigned. None of these three residues gave rise to a visible ^{15}N,^{1}H-HSQC crosspeak and this was ascribed to the highly flexible nature of this section.[1]

Three further residues were missing an identifiable ^{15}N,^{1}H-HSQC peak: G993, G1008 and N1009. An assignment of each of these residues, with the exception of the amide nitrogen and amide proton of each, was however possible using the ^{15}N HSQC-NOESY. The lack of any signal arising from the amide groups of these residues implicates exchange of the amide protons with solvent water protons or exchange broadening.

All four histidines were partially unassigned, since they were missing assignments of N_{δ_1}, H_{δ_1} and N_{ε_2} in the imidazole rings. For H994 the H_{ε_1} protons were also unassigned. Figure 4.5 shows the positions of these unassigned protons in a histidine sidechain.

In all three lysines (K961, K964, K1016) the side chain NH_2 (N_ζ, H_ζ) group was not assigned. In residue K1016 the side chain H_ε protons also remained unassigned. All

---

[1] Mobile residues occupy different conformations within solution. In the case of motions faster than the NMR timescale, multiple conformations are averaged to a single crosspeak. If the motions are much slower than NMR, a separate crosspeak can be sampled for each distinct conformation. In the cases where the motion is intermediate, on a similar timescale to NMR, the averaging can be of so many points during the motion that the crosspeak is reduced or absent. This is known as exchange broadening.

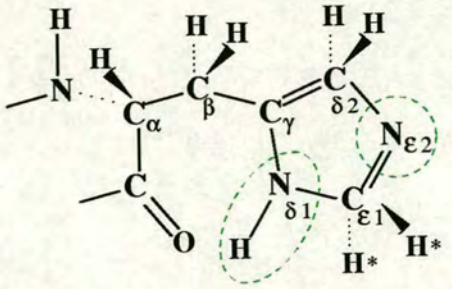


Figure 4.5: ^{15}N and ^{1}H atoms unassigned on His residues. The $H_{\varepsilon 1}$ protons are marked with an asterisk.

three arginines (R985, R995, R1022) were similarly missing guanidino-group assignments, specifically N_ε, H_ε and both branching NH_2 groups. The exception is R995 in which N_ε and H_ε were assigned. It is not unexpected that these protons could not be assigned as they are usually exposed to the solvent water and highly labile.

One of the seven isoleucines (I1019) was missing an assignment of the δ-CH_3 group protons. There were no visible peaks that could correspond to these protons, which suggests that they could be overlapped. The γ-CH_3 protons crosspeaks for this residue were intense, so it is possible that the H_δ protons are at the same chemical shift.

A complete assignment list for CR1~16 can be found in Appendix A.

4.2 Structure

The structure calculation method is described in detail in section 3.4. A total of 885 NOEs were provided to CNS [18, 16] for the structure calculation. Out of these, 757 were unambiguously assigned.

4.2.1 Energies

A final set of 60 structures were determined and Figure 4.6 shows the NOE restraint energy and total energy for each structure. The NOE energy alone is illustrated with a magnified scale in Figure 4.7. Considering the ensemble as a whole, several "cut-off" points may be discerned in Figures 4.6 and 4.7, each implying an end to a section of

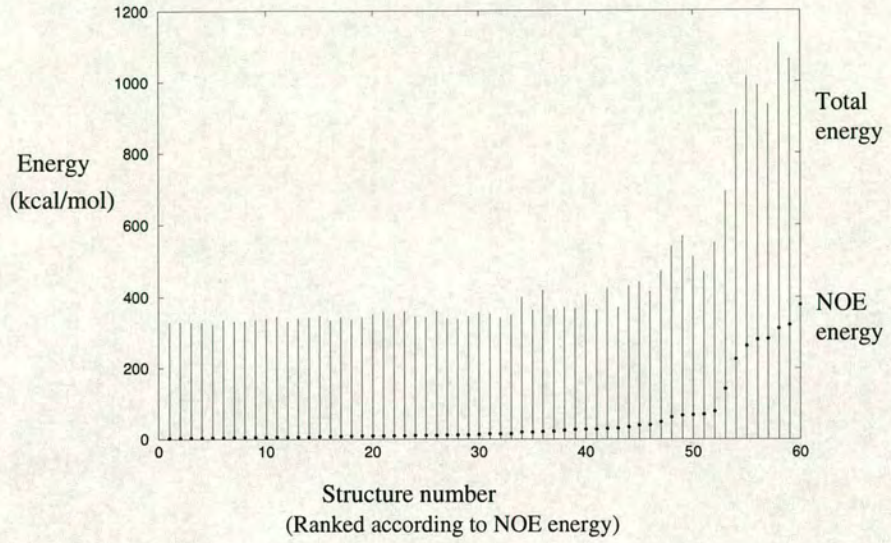


Figure 4.6: NOE and total energies of the ensemble of 60 CR1~16 structures. Total energy is shown by bars, NOE energy is shown by dots.

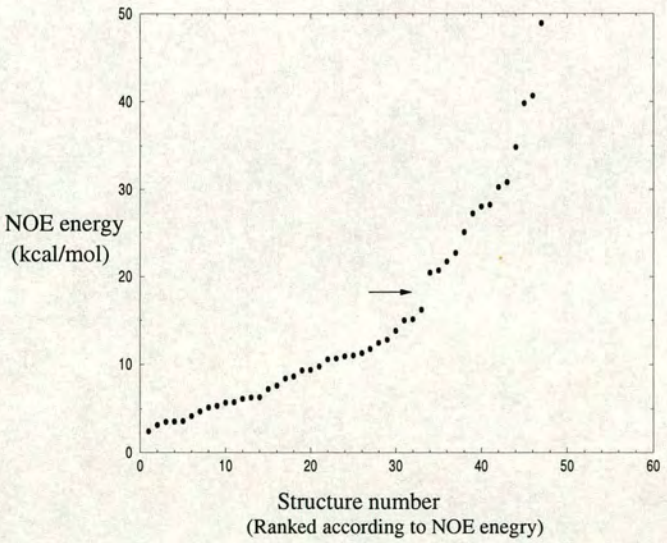


Figure 4.7: NOE energies of the CR1~16 structures. The scale is expanded to illustrate differences between the lower NOE energy structures. The NOE energies of structures 50-60 is shown above in Figure 4.6. The jump in NOE energy between the 33^{rd} and 34^{th} structures is shown by an arrow.

consistency between structures. The 34^{th} structure (in rank order according to NOE energy) shows an increase in NOE energy from the previous 33 structures. This is accounted for in the structure by a noticeable change between it and the previous structures: the section D968-V970 of the AB loop (see Figure 4.8 for loop nomenclature) takes on a distinctly different conformation in the 34^{th} structure. The 48^{th} (refine_14.pdb) and 53^{rd} (refine_37.pdb) structures show more noticeable increases in both NOE and total energy with respect to the structures preceding them in the series. In the case of refine_14.pdb the change is ascribed to residues T965-P967 taking on a different conformation compared to the previous structures - this region (often β-strand A in CCP modules) is pushed further out from the hydrophobic core of the module. In the case of refine_37.pdb, the C-terminus adopts a radically different conformation from S1014 onwards and is clearly distinct from preceding structures in the series. The structures of these higher energy cases were deemed a poorer representation of the biological reality.

While the first 33 structures of the series do show convergence (backbone atom RMSD of 1.05 Å for residues C963-C1020 inclusive), the NOE energy in this set increases by a factor of around seven from the lowest to highest. In an effort to select a smaller, more accurate group to represent the structure, the 20 structures with lowest NOE energy were selected arbitrarily, despite there being only a small jump in both NOE and total energy between the 20^{th} and 21^{st} structures. For these 20 structures, shown in Figure 4.8, the NOE energy ranges from 2.4 to 9.4 kcal/mol and the total energy ranges from 322.8 to 351.8 kcal/mol. This set has been deposited in the Protein Databank, ID number 1PPQ [27]. Ranking all calculated structures in terms of total energy rather than NOE energy produced few differences in the set of structures selected.

4.2.2 Secondary structure

Figure 4.9 depicts a cartoon, produced using MOLMOL [65, 147], of the lowest NOE energy structure out of the set of 20. The structure has six sections of β-strand, as determined by PROCHECK [77] and according to the Kabsch and Sander criteria [53]. In the figures strand F is shown as two adjoining strands to highlight the division be-

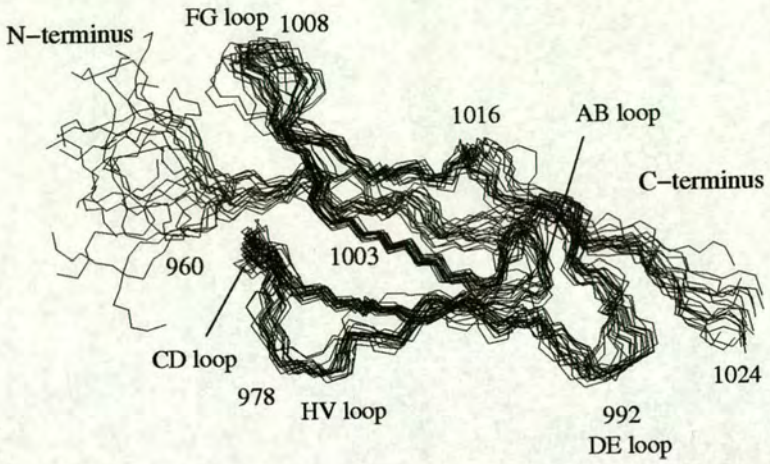


Figure 4.8: Overlay of 20 lowest NOE energy CR1~16 structures. Only backbone atom traces are shown. Backbone atoms of the structures are overlaid on those of the structure closest to the mean. Backbone RMSD for all residues C963-C1020 is 0.94 Å. Loops are labelled AB, CD, DE and FG, using the standard CCP module convention (section 1.3.2). Hypervariable loop is labelled HV. Selected residues numbered. These structures are available from the Protein Databank, ID number 1PPQ [27].

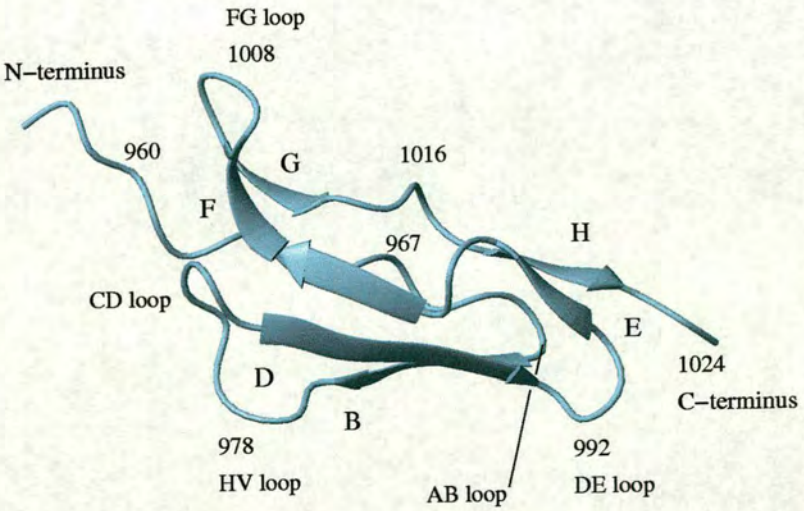


Figure 4.9: Lowest energy CR1~16 structure highlighting secondary structure elements as determined by PROCHECK [77]. β-strands and loops are labelled using the standard CCP module convention (section 1.3.2). Selected residues are numbered.

tween residues hydrogen-bonded to strand D on one hand, and strand G on the other. The residues comprising each strand are shown in Table 4.1.

Strand	Residues
B	M973-H975
D	S984-C990
E	R995-I997
F	S1000-I1005
G	H1012-S1014
H	I1019-Q1021

Table 4.1: Residues comprising the β-strands of module 16. Strands A and C, frequently found in other CCP modules, are not present in CR1~16.

This pattern of secondary structure is typical of CCP modules, and indeed these strands have been labelled according to the convention (see section 1.3.2) which recognises that there are a maximum of eight potential strands in all known CCP module structures, labelled A-H. In the case of module 16, strands A and C are not present, although they appear in regions equivalent to S962-K964 and Q981-I983 in some other CCP modules. The failure of residues equivalent to these to adopt a conformation that can be classified as β-strand is, however, quite commonplace amongst CCP modules [41, 128, 23].

The strands are mostly oriented with the long axis of the molecule and form a β-barrel type structure. Strands B, D, F and G form one anti-parallel sheet while strands E and H form a second, smaller one. The barrel structure encompasses the largely hydrophobic core and the six-membered ring of the partially buried Trp sidechain.

MOLMOL [65, 147] provided a Ramachandran analysis which is shown in Figure 4.10. The Ramachandran analysis (this time using PROCHECK [77]) is summarised in Table 4.2. With over 95% of residues in the most favoured or additionally allowed regions, this set of structures may be considered to have relatively good stereochemistry by NMR standards.

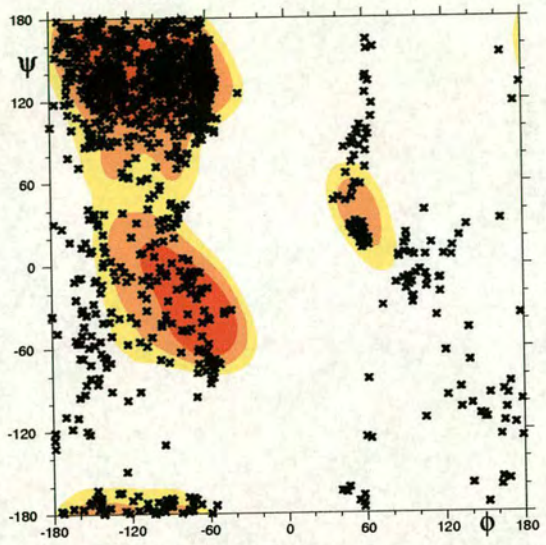


Figure 4.10: Ramachandran plot [113] of the top 20 CR1~16 structures. The colour change from red, orange, yellow to white shows decreasing favourability of regions. Each residue of each structure is shown by a cross. Made using MOLMOL [65, 147].

Residue type	**Number**	**Percentage**
Residues in most favoured regions	704	64.0%
Residues in additional allowed regions	344	31.3%
Residues in generously allowed regions	39	3.5%
Residues in disallowed regions	13	1.2%

Table 4.2: Ramachandran analysis of residues of the twenty lowest NOE energy structures of CR1~16.

4.2.3 NOEs

Table 4.3 shows the distribution of unambiguous inter-residue NOEs and Figure 4.11 shows NOEs per residue.

Restraint Type	**Number**
Total NOEs	885
Total Unambiguous NOEs	757
Total Ambiguous NOEs	128
For Unambiguous NOEs	
Intra-residue	348
Sequential	192
Short range, $2 \leq \|i\text{-}j\| \leq 4$	36
Long range, $\|i\text{-}j\| > 4$	181
Total NOE Violations (ensemble of 20) > 0.3 Å	1

Table 4.3: NOE statistics for CR1~16. i and j represent the residue number within the sequence. Short range NOEs occur between residues 2-4 positions away in the sequence. Long range NOEs occur between residues greater than four positions away in the sequence.

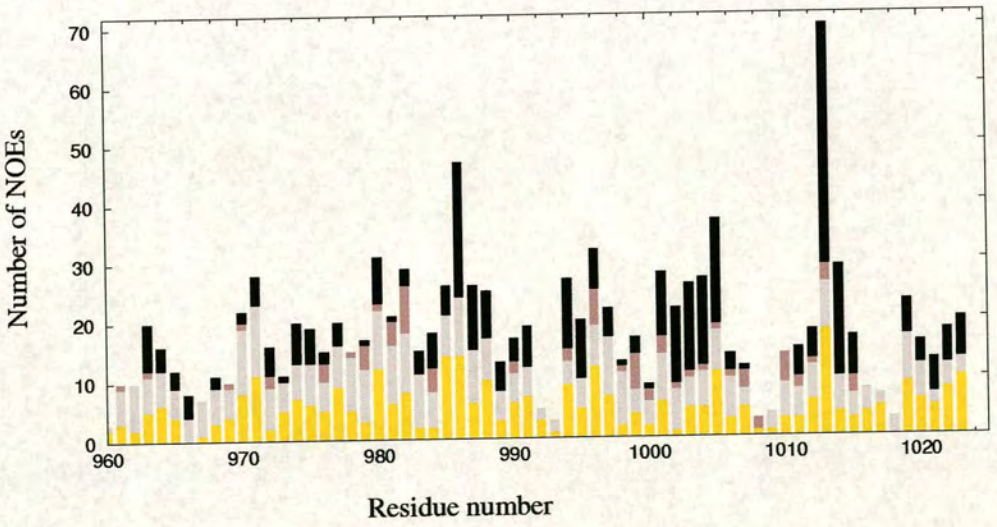


Figure 4.11: Breakdown of the number of NOEs per residue in CR~16. Intra-residue in yellow, sequential in grey, short range in brown and long range in black.

Residues I986 and W1013 had the greatest total number of NOEs presumably due

to their large size and because of the nature of their hydrophobic sidechains. Several additional core residues showed a high proportion of long range NOEs. Examples include C963 and C1004, since the disulphide bond formed between these two residues is buried deep within the N-terminal portion of the module. All residues in the section S1001-C1004 of strand F show a higher than average number of long range NOEs as this strand is sandwiched between strands D and G, as shown in Figure 4.12.

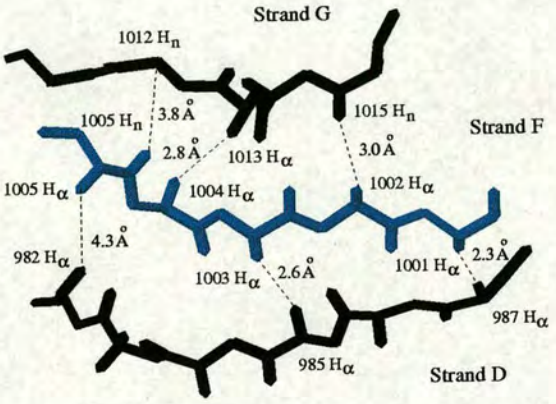


Figure 4.12: NOEs between strands D, F and G in CR1~16. Only backbone atoms, amide protons and H_αs are shown.

Residues L996 and H999 show a large number of short range (to residues 2-4 positions away) NOEs as their backbones form turns in the bulge directly following strand E. This brings them in close contact to residues following and preceding them in the sequence.

4.2.4 Accuracy and precision

In general the number of NOE violations displayed by the ensemble of calculated structures is low, suggesting that the structures mostly satisfy the data well. Indeed, the selected family of 20 structures exhibits only a single violation greater than 0.3 Å - the NOE from the C1020 amide proton to I1019 H_β - in just one structure. The intensity of the peak placed it in the less than 3.30 Å range but the distance calculated in that structure was 3.38 Å. However, the intensity was very close to the borderline

that would have placed the restraint in the <5.0 Å bracket. There is no question of this assignment being incorrect and the fact that only a single violation was generated shows that these structures are relatively accurate by NMR standards.

While the accuracy of the structure determination was relatively high, the structures produced were not of relatively high precision. Attempts were made to "tighten" the restraints by altering the intensity ranges that specify which distance category each restraint falls into. The intensity thresholds were lowered in the hope that the structures would be "pulled together" by the consequently shorter distance restraints. This tightening of restraints, however, did not result in a decrease of the RMSD value and simply increased the number of violations produced. (The intensity bounds scaled to each NOE are detailed in methods section 3.4.2.)

A second attempt was made at increasing precision by specifically calibrating the NOE intensity to a greater number of distance categories during each round of structure calculation. Unlike the usual method of having four classes of NOE strength, the calibration used the r^{-6} dependency of the NOE to scale each intensity to one of many maximum distances from 2.2, 2.3, 2.4 etc. up to 4.0 Å with one additional 5.0 Å category. As was observed when the restraints were tightened, this calibration method increased the number of violations produced per structure and did not improve the RMSD.

The backbone atom and C_α RMSD values for the 20 selected CR1~16 structures are shown in Table 4.4. Figure 4.13 shows the C_α RMSD per residue.

Residues	**Backbone RMSD**	**C_α RMSD**
K961-P1024	1.03 Å	1.09 Å
C963-C1020	0.94 Å	0.99 Å

Table 4.4: Backbone atom and C_α RMSD of CR1~16. Values are quoted for the entire native sequence (K961-P1024) and for the sequence from the first to fourth Cys residues, inclusive (C963-C1020).

In all four calculations of RMSD, refine_10.pdb, the 4^{th} lowest in NOE energy, was

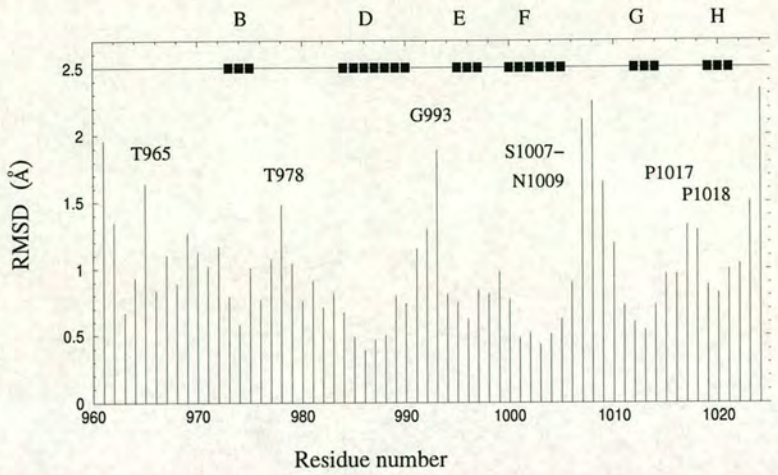


Figure 4.13: C_α RMSD per residue. The β-strands of the module 16 structure are shown as black rectangles and labelled using the CCP module convention. Several residues showed a relatively high C_α RMSD.

found to be the closest structure to the mean. Excluding the flexible loop consisting of residues S1007-T1010 reduced the backbone atom RMSD to 0.83 Å and the carbon atom RMSD to 0.88 Å. In both of these cases refine_43.pdb, the 17^{th} lowest in NOE energy, was the closest to the mean.

Residues that have a relatively high RMSD between structure include T965, T978, G993, S1007-N1009, P1017 and P1018. The dynamic motions that T965 and T978 undergo are described fully in section 4.5 and could be responsible for the low precision of these residues. G993 is most likely undergoing solvent exchange or intermediate timescale motion as it had no peak present in the ^{15}N,^{1}H-HSQC. Although the H_α peaks were assigned using NOEs to the following residue, the amide proton and nitrogen for G993 were never assigned which goes some way to explaining the lack of definition around this residue.

FG loop region S1007-N1009 was largely unassigned, even in the ^{15}N,^{1}H-HSQC, and so it is not unexpected that it has so high an RMSD. The dynamics of this section are detailed in section 4.5.

As described in section 4.1, both P1017 and P1018 contained protons that could not be

assigned (H_α and H_βs in P1017 and H_γs and H_δs in P1018) and this probably explains the lack of definition in the RMSD of these residues.

4.3 Structure comparison with double modules

The mean backbone atom RMSD for module 16 (residues C963-C1020) as part of the 15-16 (i.e $^{15}16$) and 16-17 (i.e. 16^{17}) ensembles (24 structures in each) was 0.73 Å and 0.76 Å respectively, compared with 0.94 Å for the ensemble of lone CR1~16 structures. The convergence of the calculated structures of module 16 was therefore significantly better in the context of the double module constructs. This was partially due to the inclusion of ^{13}C NOESY-derived distance restraints that were available for the ^{13}C, ^{15}N-labelled double modules. For the isolated module 16 structures, a total of 885 distinct NOEs were used. For modules 15-16, 2869 NOEs were used, while in 16-17, 2331 NOEs were used.

The structures of module 16 alone were very similar in structure to that of module 16 in the contexts of the double modules. Comparing lone module 16 to module $^{15}16$, the backbone atom RMSD of the two ensembles combined was 1.33 Å, showing the high similarity of the two versions of module 16. Comparing lone 16 and 16^{17}, the backbone atom RMSD for these two ensembles combined was again relatively low - 1.36 Å. An overlay of lone CR1~16 with $^{15}16$ and 16^{17} is shown in Figure 4.14.

The largest contribution to the poorer convergence of the ensemble of lone module 16 structures comes from loop regions that would be proximal to an adjoining module in the native protein. In the double modules, $^{15}16$ is more poorly converged at the C-terminal end where module 17 is "missing". Similarly, module 16^{17} is more poorly converged at the N-terminal end where module 15 is absent. In the case of module 16 alone, both "ends" of the molecule are missing neighbouring modules, presumably leading to a decrease in conformational stability of the loops near the junctions. In 16^{17}, the module 16 DE loop appears to be pulled in, relative to isolated module 16, towards the linker region and close to the FG loop of module 17. In a similar, but more pronounced way, the FG loop in module $^{15}16$ has moved, relative to lone module 16,

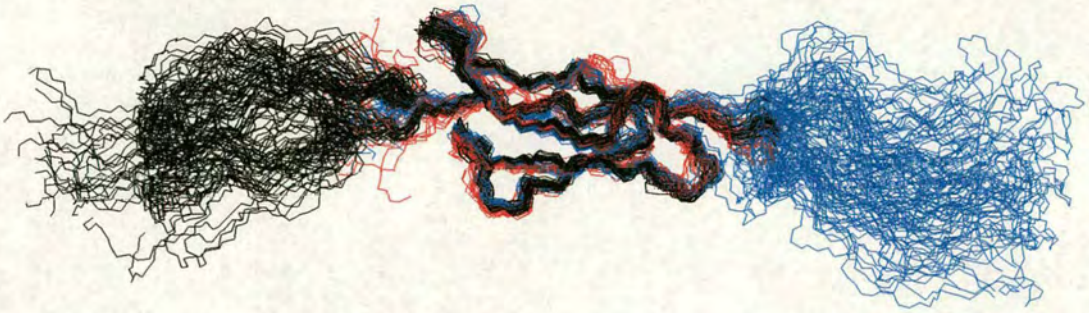


Figure 4.14: Comparison of the structures of module 16, modules 15-16 and modules 16-17. Backbone atom traces of the three ensembles are shown. CR1~15-16 (black), lone module 16 (red) and CR1~16-17 (blue) are overlaid on the lowest NOE energy lone module 16 structure.

so it occupies a position close to the DE loop of module 15, as shown in Figure 4.15.

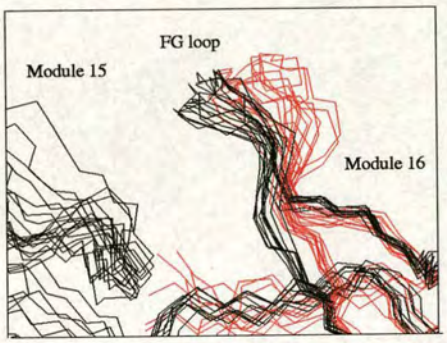


Figure 4.15: The effect of module 15 on the FG loop of module 16. When module 15 is present (ten structures of CR1~15-16, shown in black), the conformation of the FG loop of CR1~16 changes compared with the lone module (ten structures, shown in red).

Such an observation might have been anticipated given that the 15-16 junction is quite well structured. The smaller effect at the C-terminal end of module 16 is also consistent with the less well-structured junction proposed to exist between modules 16 and 17 [128, 62].

4.4 Relaxation data

4.4.1 Omissions

The dynamics of module 16 were investigated by analysing the ^{15}N T_1, T_2 and heteronuclear NOE relaxation data using the Modelfree software program [90, 109]. The methods used are described in detail in section 3.5.2. Several residues could not be analysed for the following reasons. Residues M973 and C990 were severely overlapped in the spectra and could not be integrated with any certainty. For several residues there was no assigned ^{15}N,^{1}H-HSQC peak, and these could not therefore be analysed. Specifically, these were G1008-N1009 in the flexible FG loop and G993 in the DE loop. In the non-native N-terminus, E957-E959 were not assigned in the ^{15}N,^{1}H-HSQC, while A960 did not have any visible peaks in the T_1 and T_2 spectra. Excluding the above residues and the six proline residues, complete ^{15}N relaxation data was available for 53 of the residues in CR1~16. Interpretation of raw relaxation data is described in section 1.3.4.

4.4.2 T_1 relaxation data

Figure 4.16 shows the T_1 relaxation data for module 16. The T_1 values were remarkably consistent. The mean was 455 ± 32 ms, and all but three residues fell within this range. S962 and D968 had T_1 times of 497 ± 7 ms and 501 ± 7 ms, respectively. K961, the most N-terminal native residue, had a T_1 time of 636 ± 12.

4.4.3 T_2 relaxation data

Figure 4.17 shows the T_2 relaxation data for module 16. The mean T_2 value (± standard deviation) was 196 ± 64 ms. Again, there was consistency among the majority of the residues and only five residues fell outwith the mean ± one standard deviation. Residues I977, A1011, T1015 and I1023 all had high T_2 values. I980 was the only residue to show a much lower than average T_2, which could indicate that chemical exchange timescale motion is occurring. None of these residues occur within β-strands.

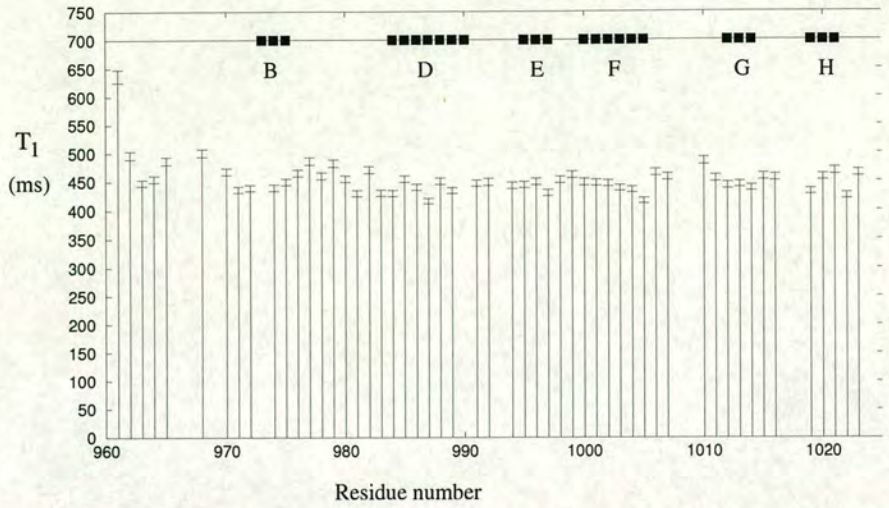


Figure 4.16: T_1 values for CR1~16. Error bars are shown for each relaxation time. The β-strands of the module 16 structure are shown as black rectangles and labelled using the CCP module convention.

4.4.4 Heteronuclear NOE data

The heteronuclear NOE data for CR1~16 can be compared to that for modules $^{15}16$ and 16^{17}. Figure 4.18 shows the heteronuclear NOE data for lone module 16, as well for modules $^{15}16$ and 16^{17}.

The mean ($\pm$ standard deviation) heteronuclear NOE for lone module 16 was 0.66 $\pm$ 0.10. By comparison, for $^{15}16$ it was 0.71 $\pm$ 0.05 and for 16^{17} it was 0.70 $\pm$ 0.11. The three residues in lone module 16 which had hetNOE values which were more than one standard deviation below the mean were all located in loops or near the termini (K961, S1007 and I1023). In general, modules 16, $^{15}16$ and 16^{17} follow a similar trend. Some differences, like in the cases of T992, H994 and T1010, could be accounted for by the absence of the neighbouring modules. For example, in T992 (which is found in the DE loop of module 16 and is part of the 16-17 junction) the hetNOE value is similar in modules 16 and $^{15}16$ as module 17 is absent in both cases. In the 16^{17} construct, the hetNOE value is higher, suggesting higher rigidity due to the stabilising presence of module 17.

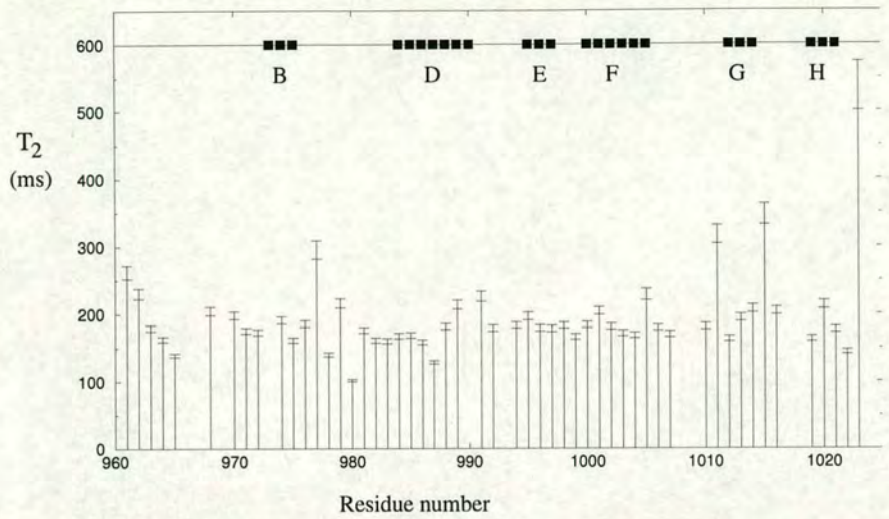


Figure 4.17: T_2 values for CR1˜16. Error bars are shown for each relaxation time. The β-strands of the module 16 structure are shown as black rectangles and labelled using the CCP module convention.

4.5 Isotropic dynamics

4.5.1 Isotropic Modelfree results

Of the residues eligible for Modelfree analysis, seven could not be isotropically fitted to one of the five models. These were I977, S989, T991, I1005, A1011, T1015 and I1023. An inability to fit Modelfree parameters to the relaxation data could indicate that intermediate timescale motion (which Modelfree is not designed to deal with) may be occurring. A second possibility, in the case of an isotropic fitting, is that diffusional anisotropy may be contributing to the relaxation data but not being taken into account by Modelfree. This left 46 residues fitted by the Modelfree analysis. Of these, 34 were fitted by the simplest model, S^2 alone. This includes the stretch of 11 residues from H994-C1004. Four residues required the fitting of S^2, τ_e and R_{ex}. Seven residues required the fitting of S^2 and R_{ex} A single residue, C963, was fitted by S^2 and τ_e.

4.5.2 S^2 values

The values of S^2 ranged from 0.47 (K961) – 0.87 (G983). However, the lower bound derived from an outlier in the distribution (residue K961, which is close to the N-terminus) and the second lowest S^2 was 0.72 (T965). The S^2 values are plotted in

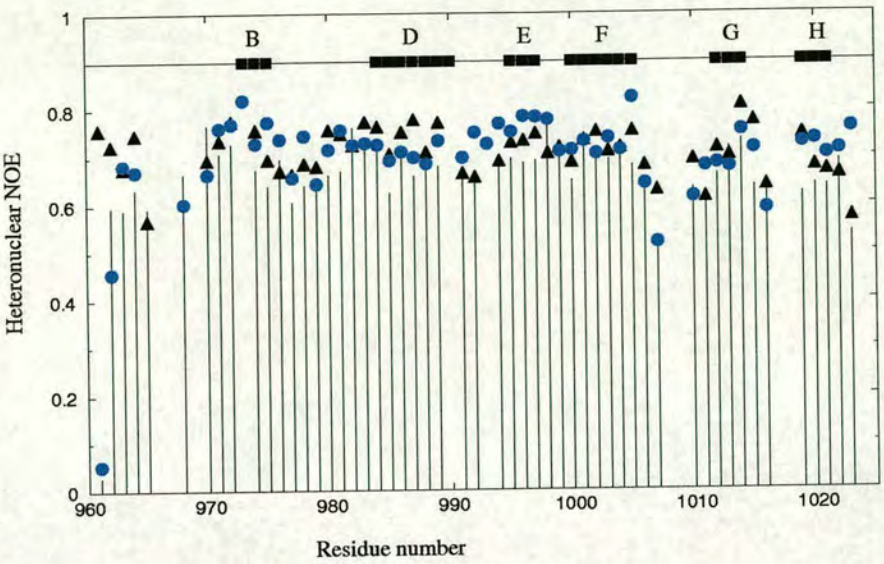


Figure 4.18: Comparison of heteronuclear NOEs in $^{15}16$ (black triangles), 16 alone (black bars) and 16^{17} (blue circles). The β-strands of the module 16 structure are shown as black rectangles and labelled using the CCP module convention.

Figure 4.19.

The average S^2 of all fitted residues was 0.81 $\pm$ 0.06 and was 0.82 $\pm$ 0.04 for residues from first to fourth Cys, inclusive. The average S^2 of residues found in β-strands was 0.83 $\pm$ 0.02 and the average S^2 of the non-β-strand residues was 0.78 $\pm$ 0.08. A statistical Student's t test to compare the β-strand and non-β-strand populations showed that the S^2 values are significantly different with over 95% confidence. The S^2 values as a whole reveal the module as having, in general, a high degree of fast timescale rigidity. S^2 values for the β-strands are given in Table 4.5.

Other examples of CCP modules show wider ranges of S^2 and often lower mean S^2. MCP˜1 had a mean S^2 of 0.79 $\pm$ 0.10 [106], $GABA_B$ CCP 2 had a mean S^2 of 0.85 $\pm$ 0.11 (work within this group, results unpublished). VCP˜2-3 had a mean S^2 of 0.69 $\pm$ 0.10 over both modules, while VCP˜3-4 and a mean S^2s of 0.81 $\pm$ 0.06 [15]. (This difference between these VCP pairs was possibly believed to be an effect of the greater flexibility within the 2-3 junction compared to the 3-4 junction.) In each of these cases a larger number of residues required the fitting of the two- or three-parameter models compared with CR1˜16.

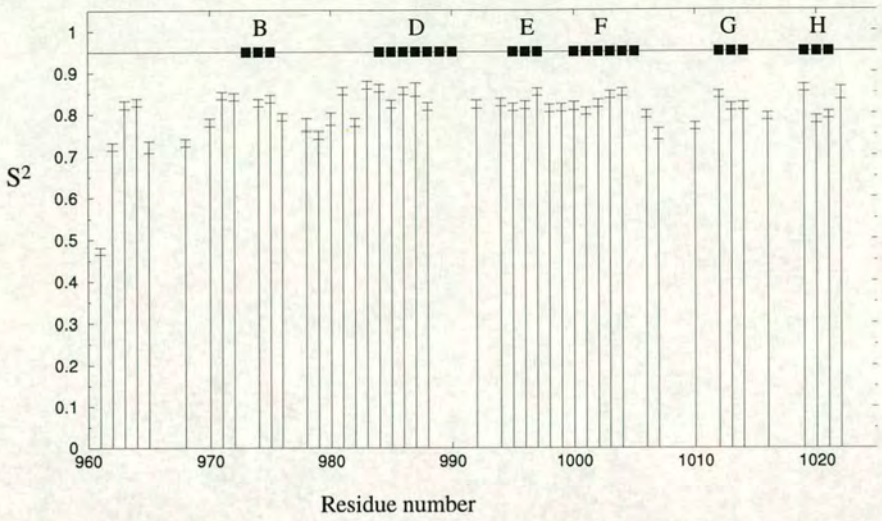


Figure 4.19: S^2 values of CR1~16. Error bars are shown for each S^2 value. The β-strands of the module 16 structure are shown as black rectangles and labelled using the CCP module convention.

Strand	**Residues**	**Residues fitted**	**Mean S^2 +/- sd**
B	M973-H975	2	0.83, 0.84
D	S984-C990	5	0.84 $\pm$ 0.02
E	R995-I997	3	0.83 $\pm$ 0.02
F	S1000-I1005	5	0.83 $\pm$ 0.02
G	H1012-S1014	3	0.82 $\pm$ 0.02
H	I1019-Q1021	3	0.81 $\pm$ 0.04

Table 4.5: S^2 value of β-strands

Consistently high S^2 values within protein domains implies rigidity only on the fast timescale. The transforming growth factor β (TGFβ) domain is an example in which high S^2 values are found throughout the domain, but functionality seems linked to slow timescale chemical exchange motions on the binding face [28].

4.5.3 Binding face of CR1~16

From extensive mutagenesis studies it has been shown that residues required for binding function are found on one face of each module in site 2. For modules 15 and 16, the binding faces are contiguous and for reference are termed the "front" of the site.

In module 17 the binding face is twisted relative to the other two modules. Positively charged amino acids appear to have the greatest importance in retaining binding function, implying a major electrostatic contribution to the interaction. The structure of module 16 is shown in Figure 4.20 with its amino acid residues sidechains coloured according to charge.

The distribution within the sequence of the more flexible residues shows that all residues with an S^2 value less than 0.8 are to be found within the N-terminal (K961-V982) or C-terminal (L1006-Q1021) parts. This means that one side of the module, the section containing strands D-F, is the more rigid of the two, whereas the other face, containing strands B, G and H, is the more flexible. The S^2 values are mapped on the structure shown in Figure 4.21.

The mutagenesis data show that there are three residues present in module 16 which are important for C3b- or C4b-binding functions: K964, N1009 and K1016, as summarised in section 1.4.5. (N1009 and K1016 are also required for cofactor activity in site 2.) All three of these occur on the more flexible face of the module. In Figure 4.20 it can be seen that residues K964 and K1016 are highlighted as two of the positive residues and both clearly appear on the face shown in the top figure. Residues N1009 is found in the FG loop (also labelled in Figure 4.20) and is also present in the positive, flexible face. Summarising the S^2 data, module 16 can be divided into two broad sections: a rigid, structural section of little direct activity, and a C3b- and C4b-binding face which is more flexible on the ps-ns timescale and contains several positively charged residues.

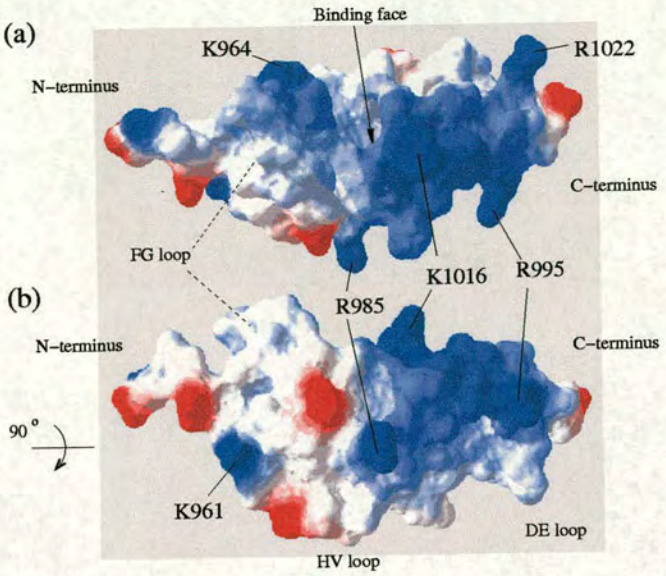


Figure 4.20: Module 16 showing the electrostatic surface. Positive residues (Lys, Arg) are shown in blue, negative residues (Glu, Asp) are shown in red. Residues implicated in the binding function (K964 and K1016) form a positively charged binding area. Positive residues R985 and R995 are not known to contribute directly to function and are not located on the binding face. Drawn using Swiss-PDB-Viewer [39].

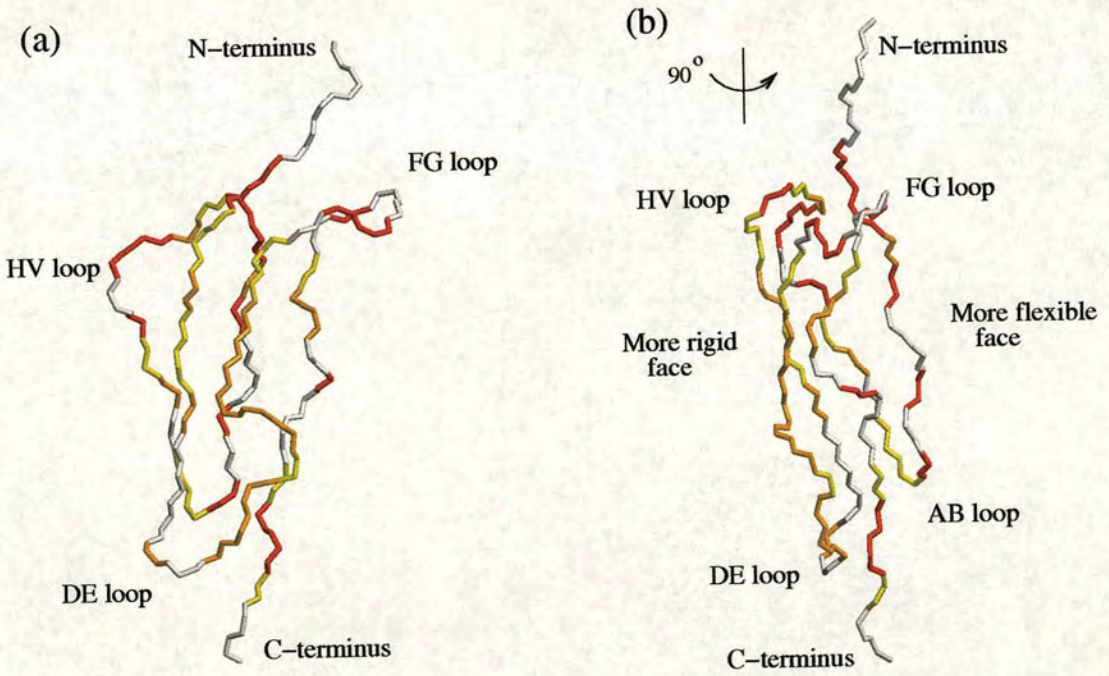


Figure 4.21: CR1~16 showing variation in S^2 values. Views rotated by 90°. showing the variation in S^2 values. Red depicts areas of high flexibility ($S^2 < 0.8$); orange corresponds to areas of moderate flexibility ($0.8 \leq S^2 < 0.83$); yellow shows rigid areas ($S^2 \geq 0.83$); white depicts residues for which no S^2 value could be determined.

4.5.4 Chemical exchange

Eleven residues required the fitting of a chemical exchange term: K961, T965, T978, I980, V982-G983, I986-T987, H999, S1007 and R1022. This data is plotted in Figure 4.22. The values fitted range from $0.41-4.32\ s^{-1}$.

In the native protein, residue 987 is an asparagine, but in the construct studied it had been mutated to a threonine to prevent glycosylation. Discounting this residue, which may well undergo different motion in the native protein, the other residues with the exception of I986 and H999 once again appear on the "front", binding face. The binding face therefore, as well as being more flexible on a fast timescale, also undergoes a wider range of motions as it includes this slower timescale (μs-ms) chemical exchange.

Module 16 chemical exchange had previously been qualitatively studied using the *Barbato et al.* method [4] for the module when it was part of the CR1~15-16 and 16-17 double module pairs. The results using the *Barbato et al.* method for double module constructs, and the results of the Modelfree analysis of lone module 16, are compared in Table 4.6.

Figure 4.23 highlights the residues on module 16 fitted with a chemical exchange term by Modelfree. Figure 4.24 highlights the residues implied by the *Barbato et al.* method to be undergoing chemical exchange within module 16 when part of module pairs 15-16 or 16-17.

The comparison shows some agreement between the different methods and residues I980, G983 and T987 are selected by all three methods. The region T991-R995 gives R_{ex} values only in 16^{17}. This is explained as the DE loop in which it occurs is at the junction of modules 16 and 17 - the DE loop residue sidechains come within 4 Å of both the linker and the FG loop sidechains - see Figures 4.25 and 4.26. Therefore in this region, the 16^{17} chemical exchange data is more likely to be representative of the motion occurring in native protein than that for $^{15}16$ or 16 alone.

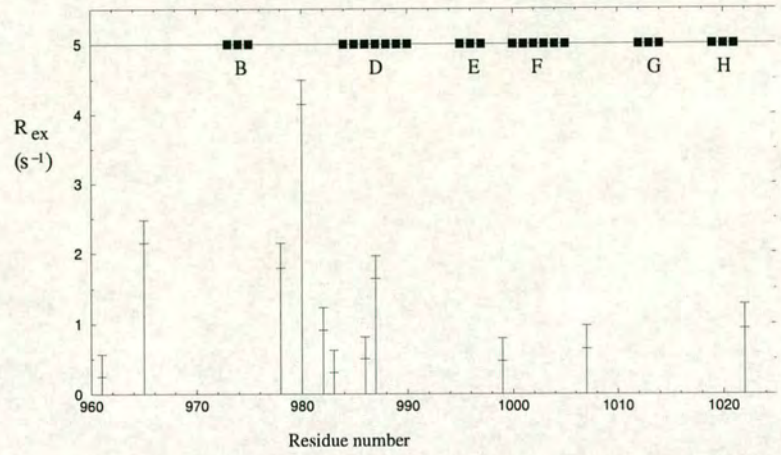


Figure 4.22: R_{ex} values of CR1~16. Error bars are shown for each exchange term. The β-strands of the module 16 structure are shown as black rectangles and labelled using the CCP module convention.

^{15}**16** (Barb.)	**16** (MF.)	**16**^{17} (Barb.)	^{15}**16** (Barb.)	**16** (MF.)	**16**^{17} (Barb.)
K961	K961 (0.41)			No fit	T991
		S962			T992
	T965 (2.31)				G993
G972					H994
		M973			R995
	T978 (1.97)	T978	G998		
I980*	I980 (4.32)	I980		H999(0.63)	
Q981			S1001		S1001
V982	V982 (1.07)			S1007 (0.81)	
G983*	G983 (0.47)	G983	H1012		
S984		S984	S1014		
I986	I986 (0.66)				I1019
T987*	T987 (1.80)	T987		R1022 (1.10)	
Y988				No fit	I1023

Table 4.6: Modelfree R_{ex} values of module 16 residues (shown as MF.) compared to residues selected by the qualitative *Barbato et al.* method (shown as Barb.) of R_{ex} measurement on module pairs 15-16 and 16-17. Residues I980, G983 and T987 are labelled as they are the only residues identified by all three methods. Residues which could not be fitted by Modelfree are labelled "No fit".

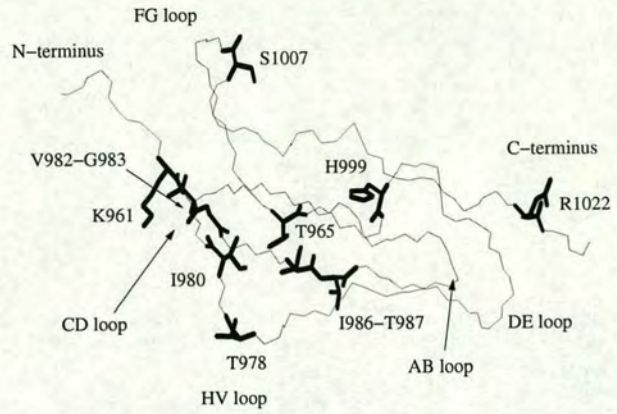


Figure 4.23: R_{ex} residues in module 16, as determined by Modelfree.

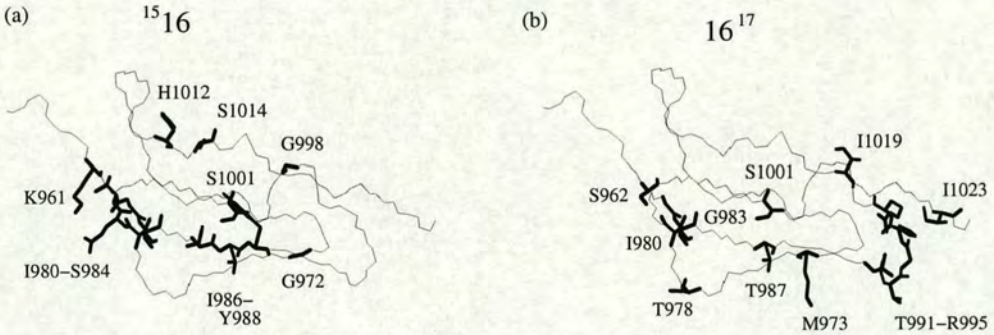


Figure 4.24: R_{ex} residues in module 16 when part of a 15-16 or 16-17 construct. The results determined by the qualitative *Barbato et al.* method.

In the CD loop region of module 16, two extra residues (Q981 and S984) only appear to undergo chemical exchange in the presence of module 15.

Chemical exchange data is therefore sensitive to the steric environment. If the amount of free space available is changed it alters the timescale of motion undergone. In the case of the DE loop of module 16 (and to a much lesser extent in the CD loop of module 16 as well) described above, the chemical exchange only takes place in the more crowded version of module 16 with a neighbouring module present. When the nearby residues in modules 15 and 17 are absent, the number of residues undergoing chemical exchange is reduced. It could be that this motion occurs over the whole junction in the case of modules 16-17 and this possibility is discussed further in section 6.2. Only I980, G983 and T987, the latter being a mutant in all three constructs, consistently

displays the chemical exchange motion.

Strand C is often located in the region of CCP modules equivalent to residues Q981-G983 in module 16. The slow timescale motion occurring in residues I980, G983 and possibly others in that region of module 16 may be responsible for preventing the formation of the AC β-sheet.

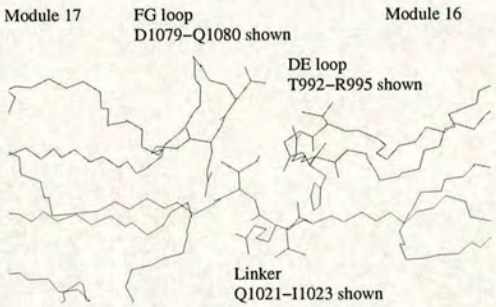


Figure 4.25: The DE loop of module 16 at the module 16-17 interface. The DE loop of module 16 is in close contact with the linker region and the FG loop of module 17. The heavy atoms of T992-R995, Q1021-I1023 and D1079-Q1080 are shown.

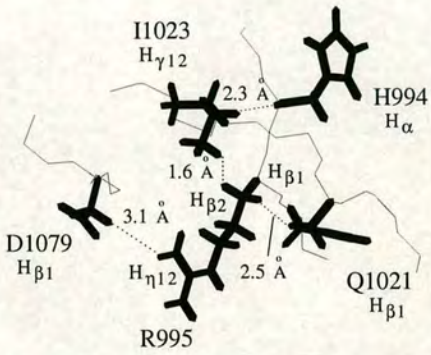


Figure 4.26: Contacts between the module 16 DE loop and the module 17 FG loop. Short distance contacts exist between residue sidechains in the DE loop and residues in module 17. The absence of module 17 in the lone module 16 constructs alters the dynamic motion of these residues.

4.5.5 Correlation times

Five of the residues required an internal correlation time to be fitted. The values ranged from 50 $\pm$ 22 to 105 $\pm$ 20 ps. Four of these required a chemical exchange term (K961, N965, K978 and S1007) although C963 did not. Again, their distribution was on the "flexible" face of the module.

4.6 Anisotropic dynamics

Use of Modelfree with reference to a 3D-structure allows the anisotropic motions to be studied. A Modelfree fitting assuming axially symmetric diffusion was carried out

incorporating 3D-structures, and using an F test to determine if any improvements in fit were purely a consequence of the inclusion of the structure. The structures used included those with the five lowest NOE energies, and those with the five lowest total energies. This totalled seven separate structures and included the closest-to-the-mean structure.

Two structures passed the F test but produced radically different $D_{||}/D_{\perp}$ values: 1.47 and 0.79. These results could not be reconciled because one suggests a ratio of length-to-width of 3:2, and the other, 4:5.

In the absence of consistent results, a further attempt at fitting the data to structures was made. The poor precision of the NOE-derived structures used was considered a possible explanation for the lack of fitting, since anisotropic behaviour has been more often successfully characterised in X-ray derived CCP module structures [106]. X-ray structures generally have higher precision than NMR structures. There is no X-ray structure for module 16 but the structure of module 16 within the double modules (15-16 and 16-17) had a higher precision. The fitting was therefore repeated using the three lowest NOE energy and three lowest total energy structures of both $^{15}16$ and 16^{17} (again, these included the closest-to-mean structures).

Three of these structures produced fits: one from the $^{15}16$ ensemble gave a $D_{||}/D_{\perp}$ of 5.26, which was obviously not consistent with CCP module structures. Two from the 16^{17} ensemble gave $D_{||}/D_{\perp}$ ratios of 1.29 and 1.43, respectively. These were more likely to be realistic, but cannot be accepted in the absence of reproducible results.

Generally, the implication from these results is that module 16 tumbles isotropically in solution. This does not seem, however, to be a realistic proposition due to the obviously non-spherical nature of the module's structure. The explanation of the inability to fit anisotropic character to this data probably lies in the lack of precision in the structures, leading to inaccurate orientation of the amide bond vectors. This has been the case previously with another CCP module - MCP˜1. The NMR-derived structure of this module could not be fitted anisotropically, though when using the X-ray structure is

was successful [106]. However, comparison of the isotropic and anisotropic Modelfree fittings of MCP~1 showed good agreement.

A plot of T_1/T_2 against θ (the angle between the N−H bond and the diffusion tensor of the molecule) can be an indicator of anisotropy. Figure 4.27 shows this for the anisotropic fitting of the lowest energy CR1~16 structure.

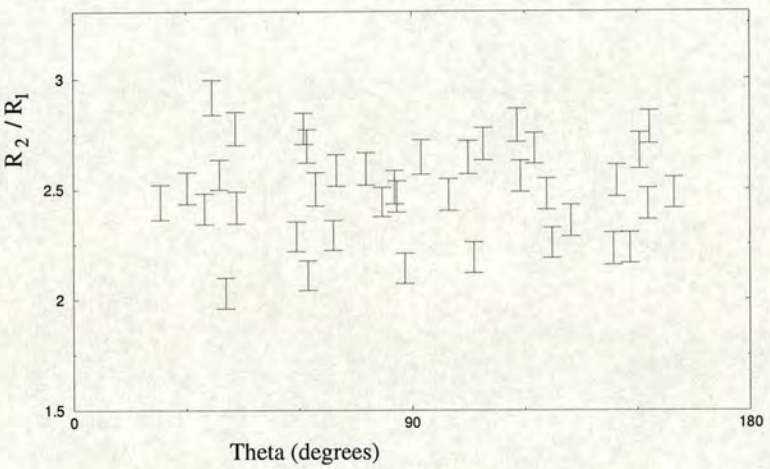


Figure 4.27: Variation of θ for CR1~16 residues dependent on T_1/T_2 ratio. A prolate symmetric top molecule would display a broad "u-shape" whereas an oblate molecule would display an upsidedown "u-shape". CR1~16 shows little anisotropy by this method.

In the case of CR1~16, this indicates little anisotropy. However, residues undergoing exchange motion had not been eliminated from the anisotropic analysis and may be skewing the fitting.

4.7 Conclusion

The assignment of the ^{15}N and ^{1}H resonances of CR1~16 has allowed the determination of its solution structure based on NOE-derived distance restraints. The structures generated were of relatively high accuracy but comparatively poor precision. The 3D-structure was shown to be largely independent of the context of the module and agreed well with the structure of module 16 when it is part of the 15-16 or 16-17 double modules. Conformations of loops and turns within CR1~16 and lying close to its

intermodular junctions were shown, however, to be dependent on context.

The assignment allowed the analysis of the isotropic dynamics of module 16 using Modelfree. Module 16 shows relatively reduced variation in flexibility over the sequence when compared to other CCP modules, which tend to display a greater range of S^2 values and require a greater range of models during fitting. While CR1~16 can be thought of as having a generally rigid scaffold structure, the presence of a variety of motions could assist in explaining its binding functions. In summary, the module can be thought of as having two faces with differing dynamic behaviour. One is relatively more rigid without fast (ps-ns) or slow (μs-ms) motions. The other has a wider range of dynamic motions as well as being more flexible. All but two of the ten native residues which shows slow timescale chemical exchange motion are located on this face. From mutagenesis data, it is believed that it is this dynamically varied and mobile face which is largely involved in the binding of ligands C3b and C4b. The slow motions could be involved in a conformational change that enables hydrophobic sidechains of CR1~16 to make specific contacts with similar residues on the ligands.

Chapter 5

CR1, modules 2-3

Following the completion of the structure determination of CR1 functional site 2 (copy 2), attention turned to the unique functional site 1. Sequence identity between modules 1, 2 and 3, and modules 15, 16 and 17, is 56%, 68% and 96% respectively. While this high degree of sequence similarity is likely to indicate a high degree of similarity between the structures of the two sites, there are several reasons why site 1 remains an important candidate for structure determination.

The structural basis for the differences in function between the two sites, while presumably moderated by key residues at the respective binding faces, still remains a mystery. While the overall fold of CCP modules is conserved, the structures of loop regions are hard to predict reliably and there is, as yet, no robust means of modelling these potentially crucial aspects of structure[1]. Furthermore, although a homology-based model may shed light on structure at the module level, another very important structural aspect would not be represented accurately: intermodular orientation. The skew, twist and tilt angles of modules with respect to each other are impossible to model reliably but they have major ramifications for function. For example, as seen in site 2, based on a combination of mutagenesis and structural evidence, it was hypothesised that the binding faces of the three modules in this site might be made into a contiguous surface by a rotation at the 16-17 junction [128]. The only way in which to obtain reliable information, in solution, on intermodular orientation and flexibility is through

[1] Module 3 is an exception since it is 96% identical to module 17 and homology modelling of module 3 would be accurate.

detailed NMR studies where intermodular NOEs can be assigned and used in structure determination. For these reasons, experimental work on the solution structure of CR1 functional site 1 employing NMR was undertaken.

5.1 Backbone assignment

5.1.1 $^{13}C^{15}N$ spectra

The majority of the backbone assignment was completed using two pairs of complementary spectra: primarily the CBCA(CO)NH, CBCANH pair, but also the HBHA(CO)NH, HBHANH pair. The HNCO and HN(CA)CO pair were then used to confirm the assignments. These techniques are described in more detail in section 3.3.7. Figures 5.1, (a)-(n), show the CBCA(CO)NH, CBCANH assignments trails for the CR1˜2-3 construct. In the strips from the CBCA(CO)NH spectrum, crosspeaks between the amide of residue i and the C_α and C_β of the preceding (i.e. residue $i-1$) residue are visible. In the CBCANH spectra strips, crosspeaks between the C_α and C_β and the N−H of both the preceding and the current residue (i.e. residues $i-1$ and i) are visible.

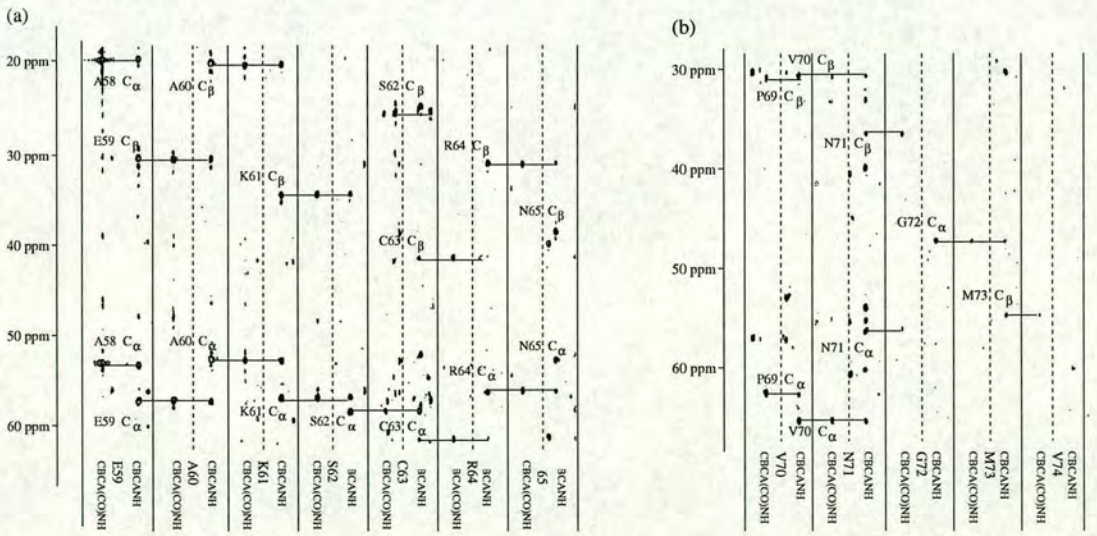
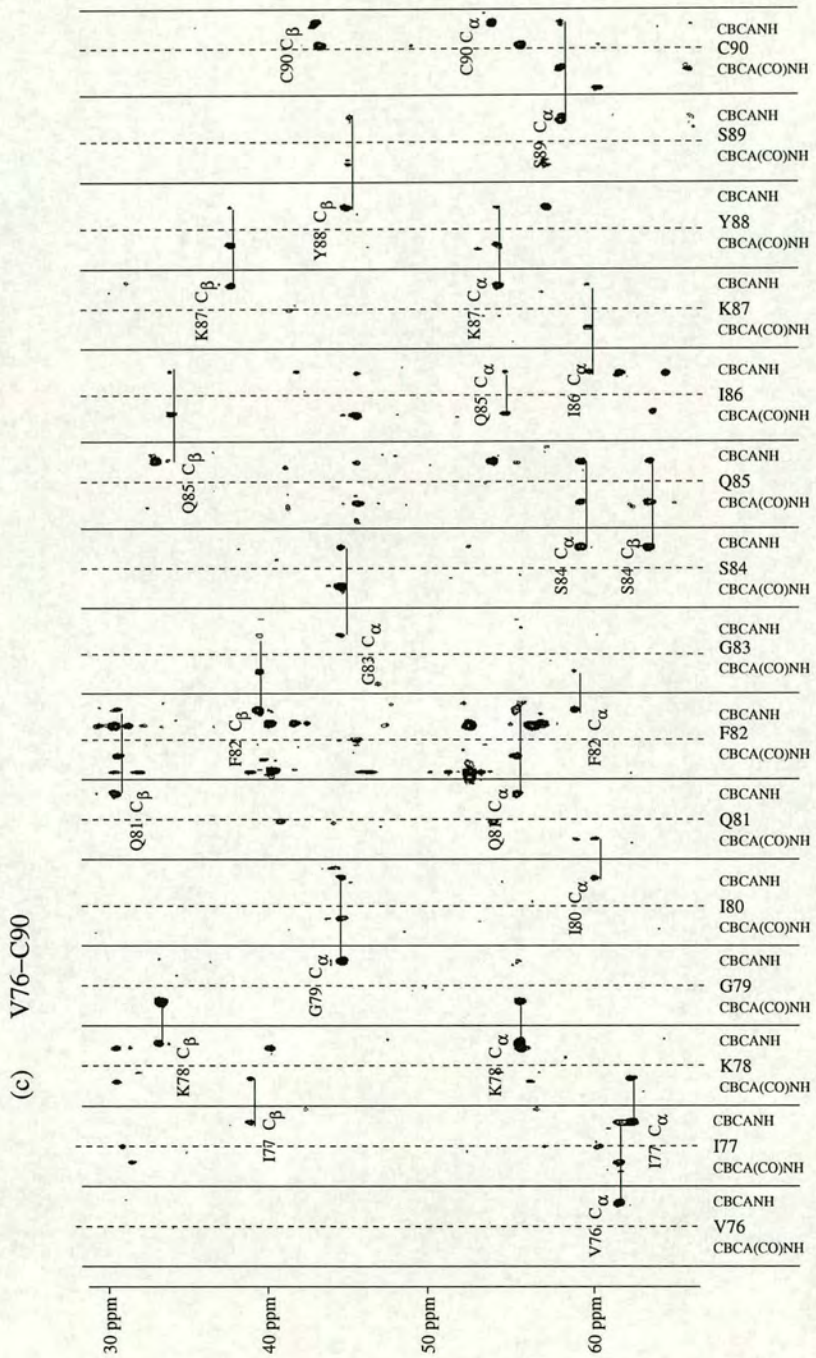
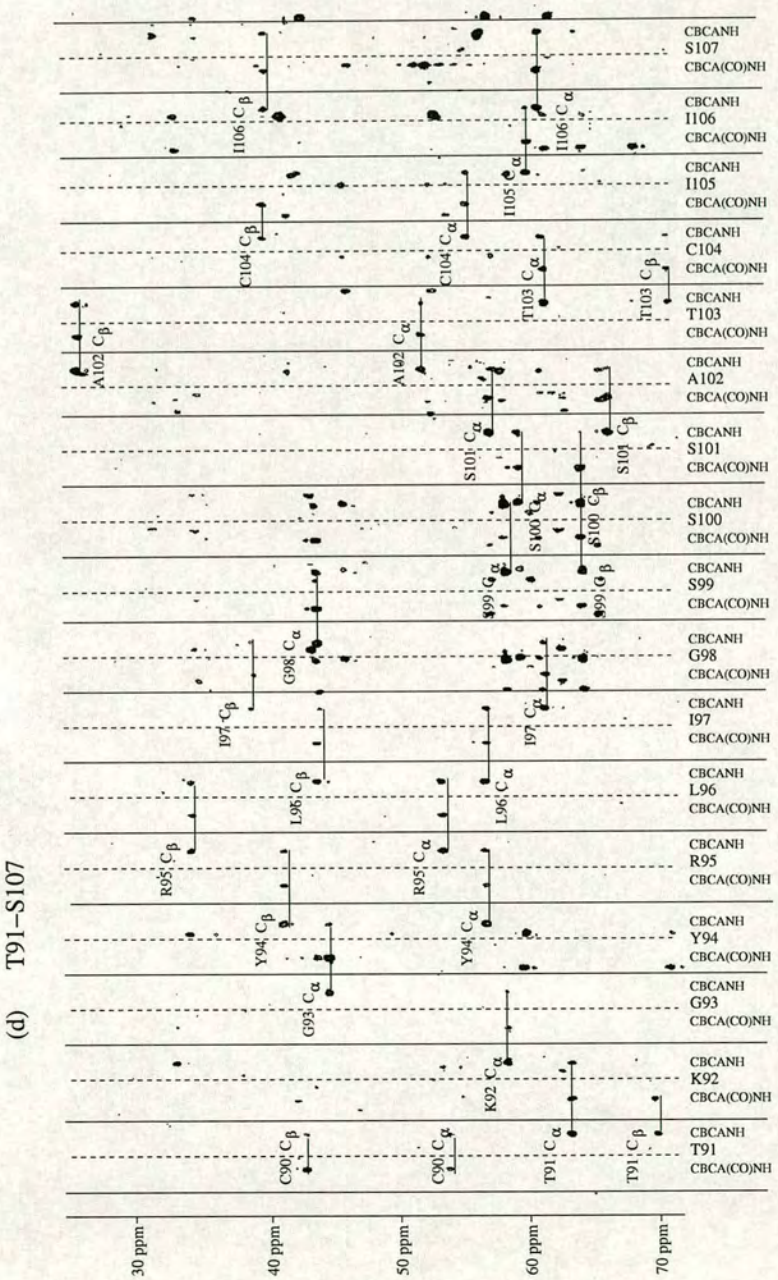
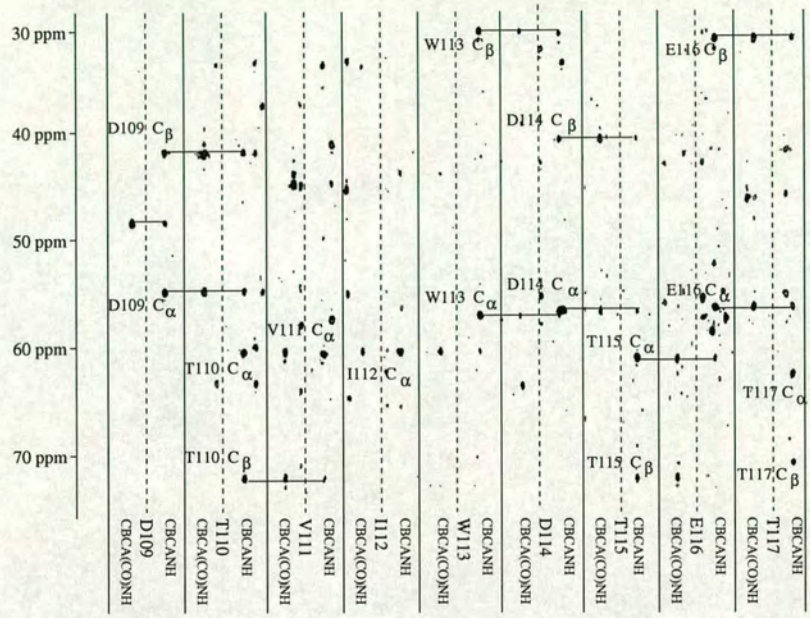


Figure 5.1: CBCA(CO)NH and CBCANH strips for CR1˜2-3. The CBCA(CO)NH strip shows the C_α and C_β of the preceding residues, while the CBCANH strip also shows this and the C_α and C_β of the current residue. Figures (a) and (b) show residues E59-N65 and V70-V74. Figures (c)-(n) show assignments of further residues as indicated.

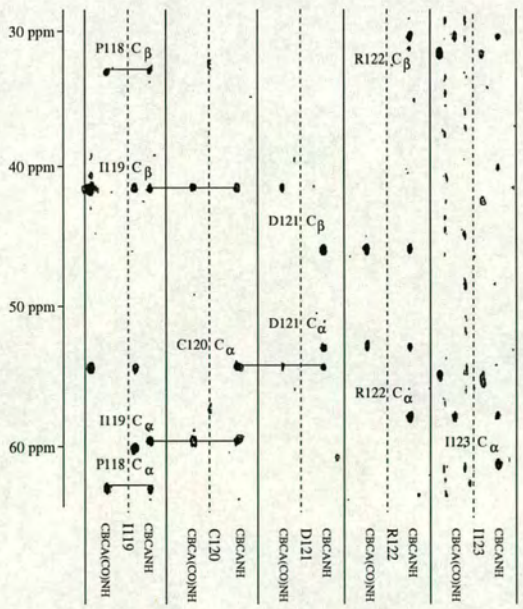

(c) V76–C90
CBCANH
C90
CBCA(CO)NH
S89
Y88
K87
I86
Q85
S84
G83
F82
Q81
I80
G79
K78
I77
V76
30 ppm
40 ppm
50 ppm
60 ppm


(d) T91–S107
CBCANH
S107
CBCA(CO)NH
I106
I105
C104
T103
A102
S101
S100
S99
G98
I97
L96
R95
Y94
G93
K92
T91
30 ppm
40 ppm
50 ppm
60 ppm
70 ppm

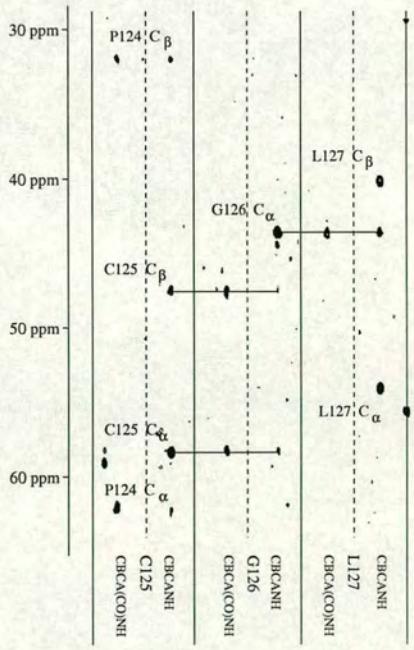
(e) D109–T117



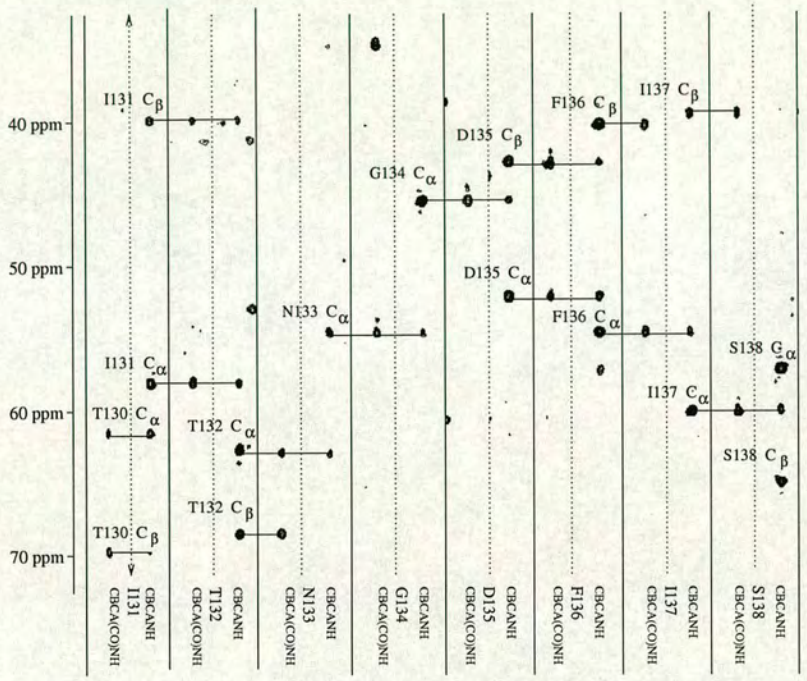
(f) I119–I123



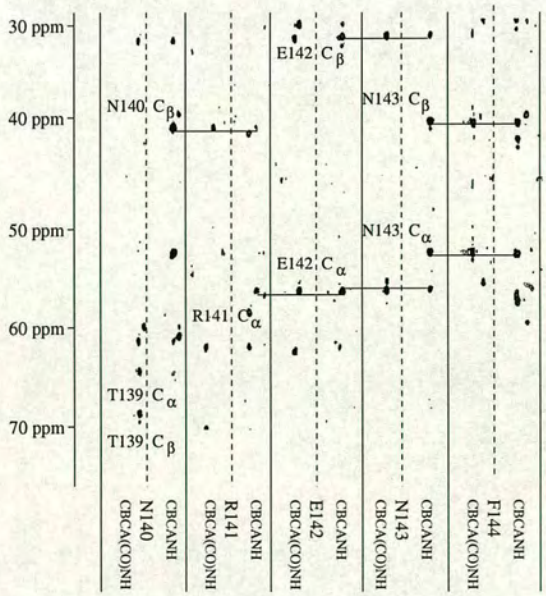
(g) C125–L127



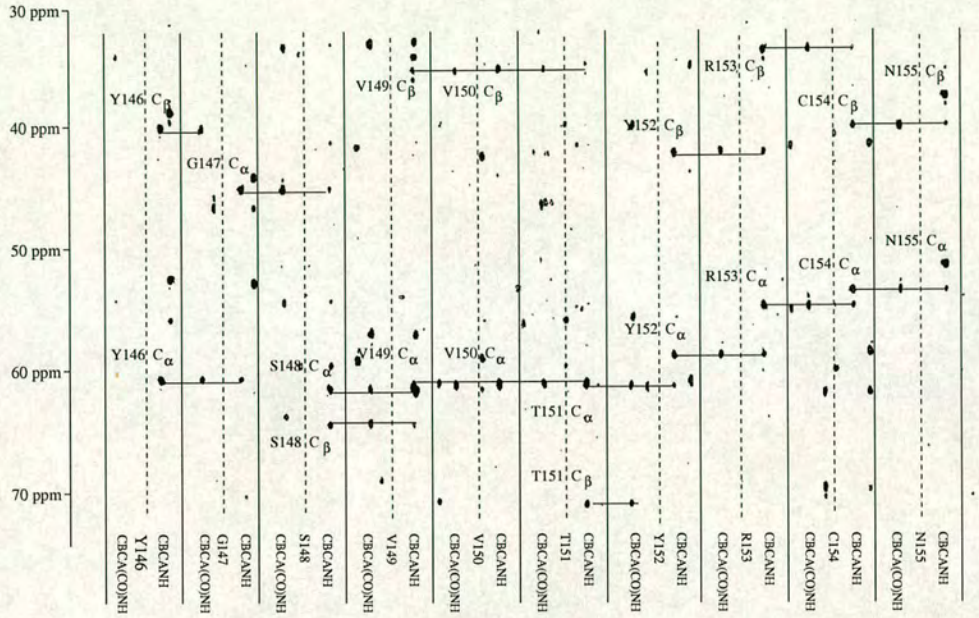
(h) I131–S138



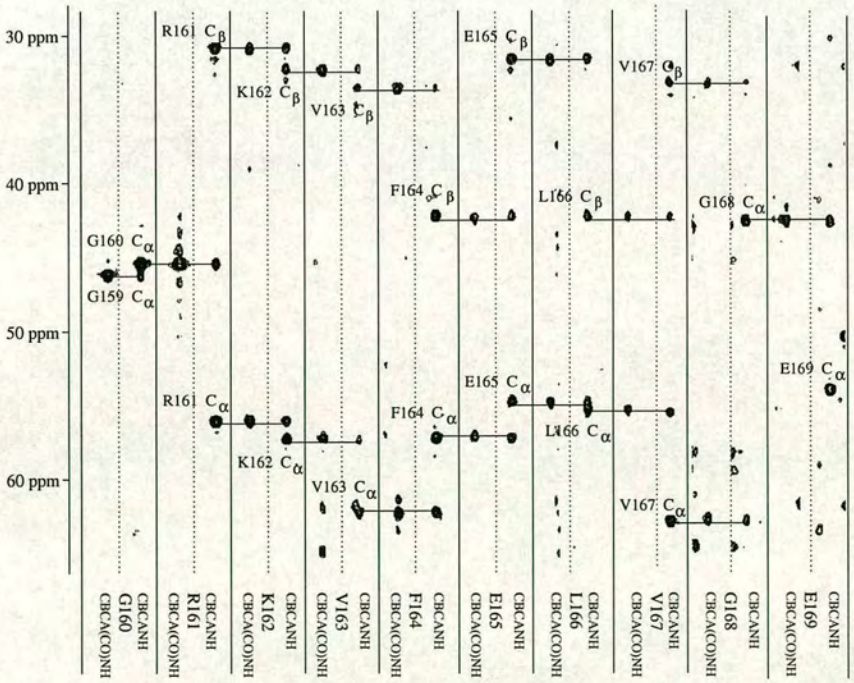
(i) N140–F144



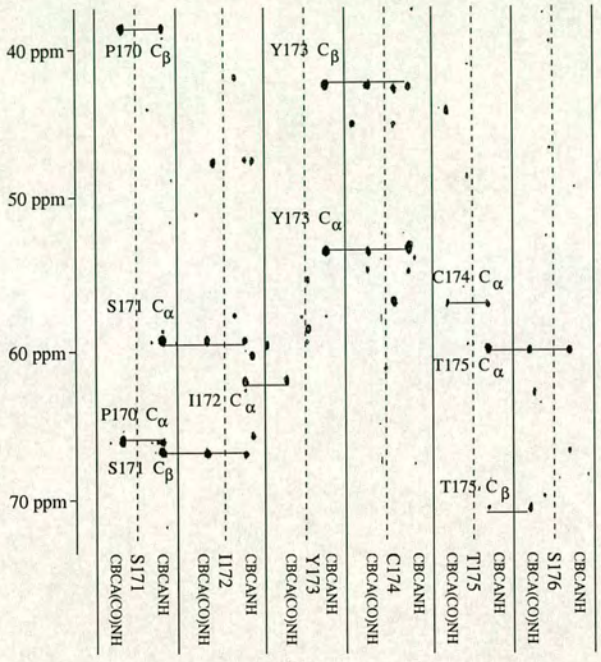
(j) Y146–N155



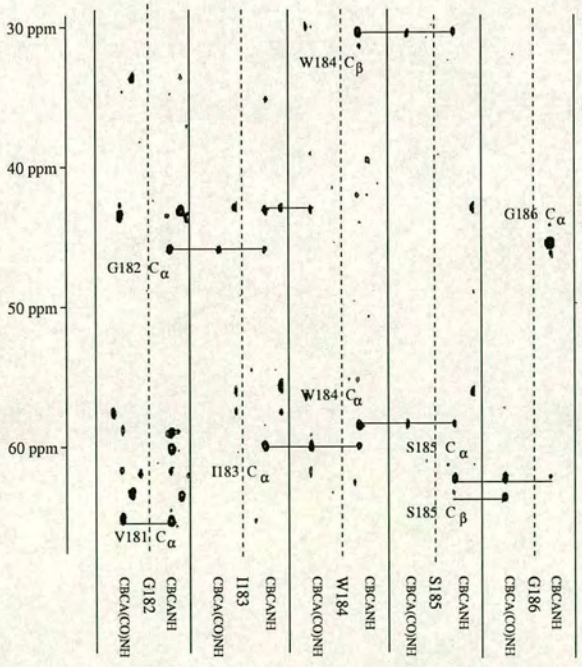
(k) G160–E169



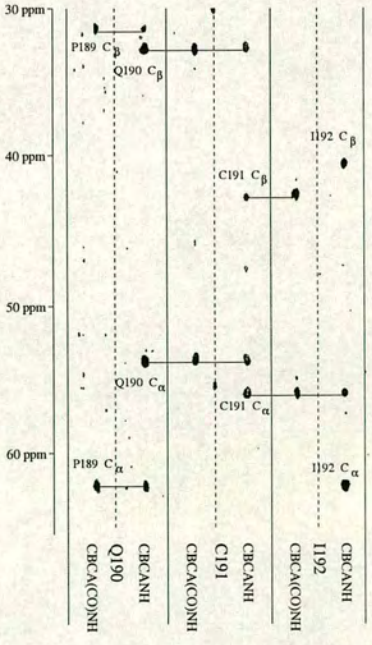
(l) S171–S176



(m) G182–G186



(n) Q190–I192



5.1.2 Extent of backbone assignment

Of the 132 native residues in the CR1˜2-3 constructs, 11 are prolines and do not have backbone amides which could be assigned. This leaves 121 residues, 112 of which were assigned. The assigned ^{15}N,^{1}H-HSQC, which illustrates the amide nitrogen and proton shifts for CR1˜2-3, is shown in Figure 5.2. The centre section contains a large amount of overlap and has been expanded and is shown in 5.3.

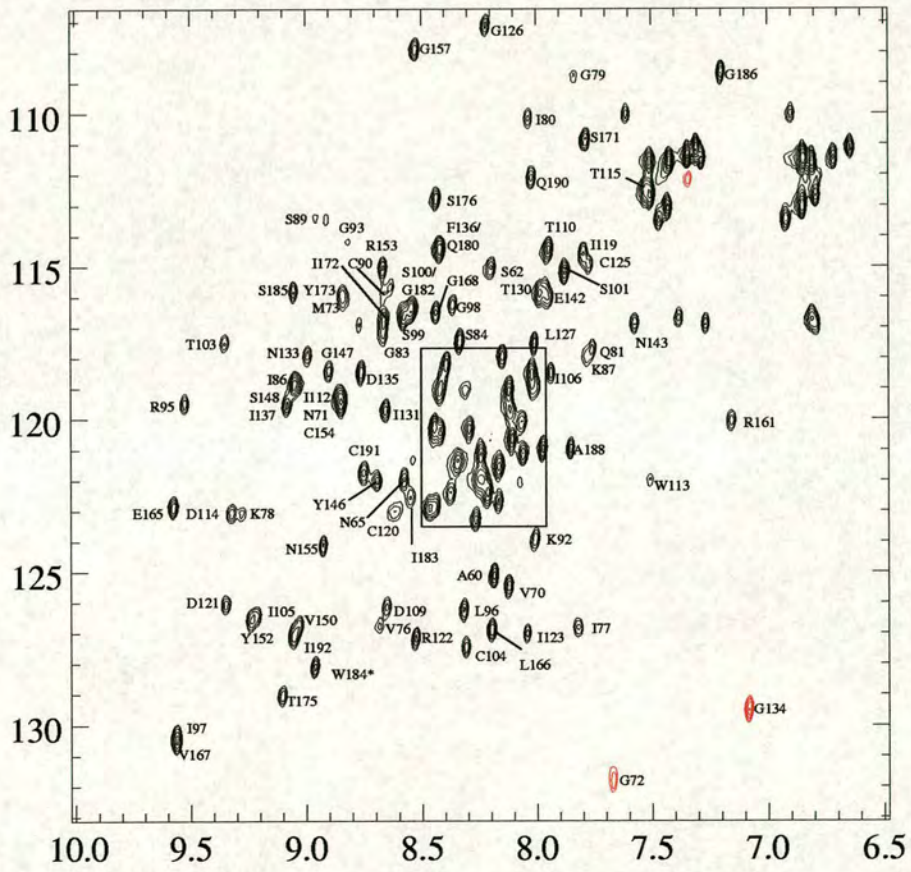


Figure 5.2: Assigned ^{15}N,^{1}H-HSQC of CR1˜16. Acquired on a 800 MHz Bruker AVA NMR spectrometer. Dimension 1 (x axis) - ^{1}H. 1024 complex points, sweep width 8992.8 Hz. Dimension 2 (y axis) - ^{15}N. 64 complex points, sweep width 2184.2. Black peaks are positive, red peaks are negative due to aliasing (as described in section 4.1). True ^{15}N chemical shifts for the aliased residues can be found in Appendix B.

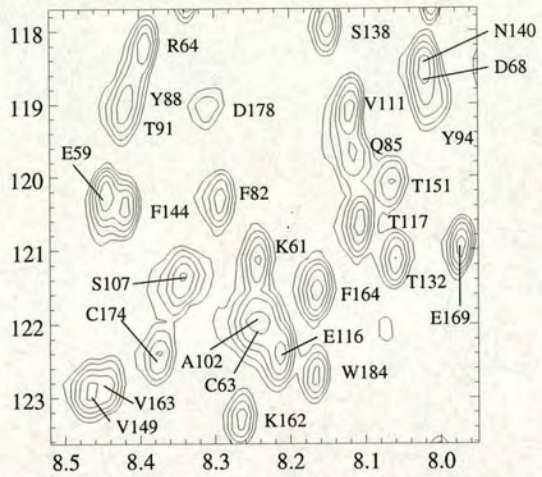


Figure 5.3: Central region of assigned $^{15}N,^{1}H$-HSQC of CR1~2-3.

The residues of the native sequence that did not appear to have crosspeaks in the $^{15}N,^{1}H$-HSQC were as follows: in module 2 - H75 and G108; in module 3 - T139, H145, S158, G159, N177, D179 and V181.

Of the non-native N-terminal residues, E59 and A60 were assigned although E57 and A58 remained unassigned. The presumably mobile nature of the N-terminus was likely responsible for the absence of E57 and A58, and indeed in the analogous module 16 both of these residues remained unassigned.

Regarding the two histidine residues (H75 and H145), the *P. pastoris* strain used to express the CR1~2-3 proteins was auxotrophic for histidine. Therefore, the labelled media on which they were fed contained unlabelled histidine which was incorporated into the constructs. Hence neither of the histidine residues were carbon or nitrogen labelled, and they did not appear in the $^{15}N,^{1}H$-HSQC spectra. However, their protons would be expected to be present in homonuclear spectra and NOESY spectra.

The only non-histidine residue in module 2 with no $^{15}N,^{1}H$-HSQC crosspeak is G108. In CR1~16 the analogous residue G1008 was missing, as was its neighbour N1009. In modules 16-17, on the other hand, the amide group of G1008 could be assigned,

though its other spins could not. While N1009 in modules 16-17 could not be assigned in the $^{15}N,^{1}H$-HSQC, D109, in 2-3, has a clearly visible crosspeak. The putative loop containing G108 and D109 is predicted to be close to the N-terminus of module 2, thus disallowing the suggestion that the presence of module 3 could provide stability. Therefore the absence of a crosspeak for G108 and the presence of a strong $^{15}N,^{1}H$-HSQC crosspeak for D109 suggests that the dynamics of modules 16 and 2 exhibit differences in this region.

Module 3 has 96% homology with module 17 - the three residues that are different are T132/A1032, P156/L1056 and G159/R1059. L1056 and R1059 are found in the C-terminal, DE loop of module 17, and A1032 is in the BC loop. Despite this similarity in the identity of modules 3 and 17, T139 was absent from the $^{15}N,^{1}H$-HSQC of 2-3 while T1039 was assigned in the $^{15}N,^{1}H$-HSQC of 16-17. T1039 is located in the hypervariable loop of module 17. It is possible, that in module 3 the nearby CD loop is altered in structure and/or dynamics (relative to the CD loop of 17) by the presence of module 2 (instead of module 16). This could lead to changes in dynamic behaviour within T139 and account for its absence from the $^{15}N,^{1}H$-HSQC of CR1~2-3.

"Missing" residues S158 and G159 are predicted to appear at the very tip of the putative DE loop and their crosspeaks could be expected to be absent from the $^{15}N,^{1}H$-HSQC because this region is likely to be mobile. On the other hand, both S1058 and R1059 were assigned in the $^{15}N,^{1}H$-HSQC of modules 16-17. The bulkier Arg residue, in comparison with the Gly in 2-3, could reduce the motion that this loop undergoes. Similarly, residues N177, D179 and V181 are all predicted to be in the FG loop that lies between modules 2 and 3, suggesting this loop is highly mobile. In support of such a hypothesis, the $^{15}N,^{1}H$-HSQC peaks for D178 and Q180 were very weak in intensity.

Many of these nine residues that could not be assigned in the $^{15}N,^{1}H$-HSQC also proved difficult in terms of assignment of the rest of their backbone atoms. In total, nine residues were missing their C_α assignments: P66, H75, P128, H145, S158, N177, D178, D179 and Q180. Of these residues S158, N177, D179 and Q180 were also missing amide nitrogen and proton assignments. Prolines 66 and 128 were both followed

in the sequence by another proline residue, increasing the difficult of completing their assignment. Residue D178 was very weakly present in the ^{15}N,^{1}H-HSQC. This meant that four residues (N177, D178, D179 and Q180) in the predicted FG loop in module 3 had C_αs that were not assigned. The histidines were unlabelled and therefore could not be assigned by any of the heteronuclear experiments.

Carbonyl ^{13}C assignments were missing for eleven residues. These were the nine above with no C_α assignments, plus P67 and V181. Table 5.1 summarises the proportion of backbone atoms assigned.

Backbone atom	Number assigned	Percentage
Amide N	111/119	93.2%
Amide H	111/119	93.2%
C_α	123/130	94.6%
Carbonyl C	121/130	93.1%

Table 5.1: Number of backbone atoms successfully assigned in CR1~2-3. These statistics exclude the non-native residues E57-A60 and the unlabelled His residues, H75 and H145, from consideration.

Thus a large proportion of the missing assignments were in the stretch N177-Q180 located in the FG loop region of module 3 that lies close to its junction with module 2. Excluding these residues from consideration raises the percentages assigned for amide nitrogen, amide proton, C_α and Carbonyl C to 95.6%, 95.6%, 97.6% and 96.8% respectively.

5.2 Sidechain assignment

5.2.1 ^{15}N spectra

While the ^{15}N HSQC-NOESY provided high quality data (an example of which is shown in Figure 5.4) that could be used subsequently in a structure calculation, the ^{15}N HSQC-TOCSY was not adequate. After a range of mixing times from 20-100 ms had been tried, 47 ms was selected as giving the strongest signal. However, the spectrum collected was still of poor quality. The H_αs of many residue strips were

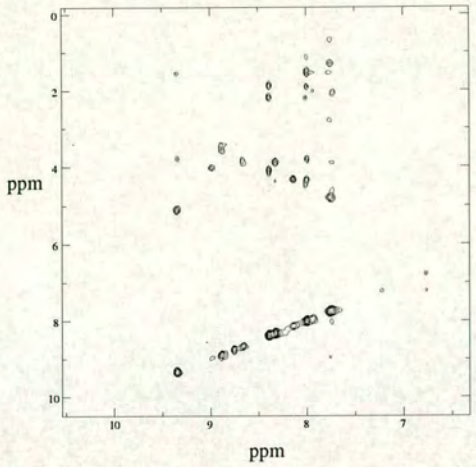


Figure 5.4: A plane from the ^{15}N HSQC-NOESY. This plane (^{15}N shift 122.0 ppm) shows good dispersion of proton resonances and was typical of the spectrum.

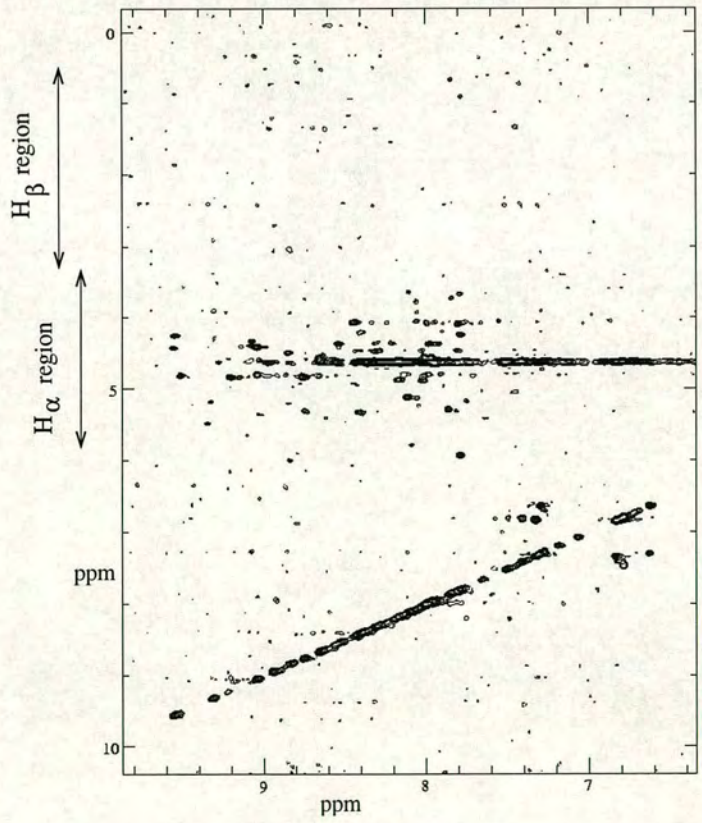


Figure 5.5: All ^{15}N HSQC-TOCSY planes overlaid. The absence of peaks in the H_β region highlights the ineffective transfer of magnetisation during this TOCSY experiment.

clear, but only in some cases were the H_β protons also visible. In practically none of the residues were peaks seen from protons beyond H_βs. The slow tumbling time of the module pair was most likely responsible. This would cause the signal to relax back to undetectable levels before any recording was possible. Similar problems were encountered with CR1˜15-16 and 16-17. Figure 5.5 shows an overlay of all the planes of the ^{15}N HSQC-TOCSY illustrating the absence of sidechain protons.

5.2.2 $^{13}C^{15}N$ spectra

Due to the poor quality of the ^{15}N spectra, ^{13}C spectra were required to complete the sidechain assignment. The majority of the sidechains assignments were achieved using the H(C)(CO)NH-TOCSY and (H)C(CO)NH-TOCSY, which allowed assignment of the chemical shifts of the preceding residue's ($i-1$, where i refers to the residue whose N−H chemical shifts define the strip) proton and carbon resonances, respectively. The HCCH-TOCSY was crucial in confirming and completing the sidechain assignments. These spectra are described in more detail in methods section 3.3.8. Figures 5.6 and 5.7 show examples from the H(C)(CO)NH-TOCSY, (H)C(CO)NH-TOCSY and HCCH-TOCSY.

5.2.3 Extent of sidechain assignment

Twenty-three of the 132 native residues had one or more sidechain nuclei that remained unassigned (i.e. excluding amide nitrogens and protons, C_α and carbonyl carbons).

Seven residues had totally unassigned sidechains: P66, H75, P128, H145, S158, D178 and D179, as did non-native E57 . While the histidine residues were unlabelled, assigning their protons using homonuclear spectra was a possibility. However, without a single shift from either residue with which to start from, any crosspeaks which appeared to belong to a histidine ring could not be attributed with certainty to a specific ring. Due to this the histidine ring protons remained unassigned. The two proline residues were both followed sequentially by another proline and consequently the majority of the 3D backbone spectra were of little use in obtaining any chemical shift information for them. The complete absence of the three residues in the predicted loop regions

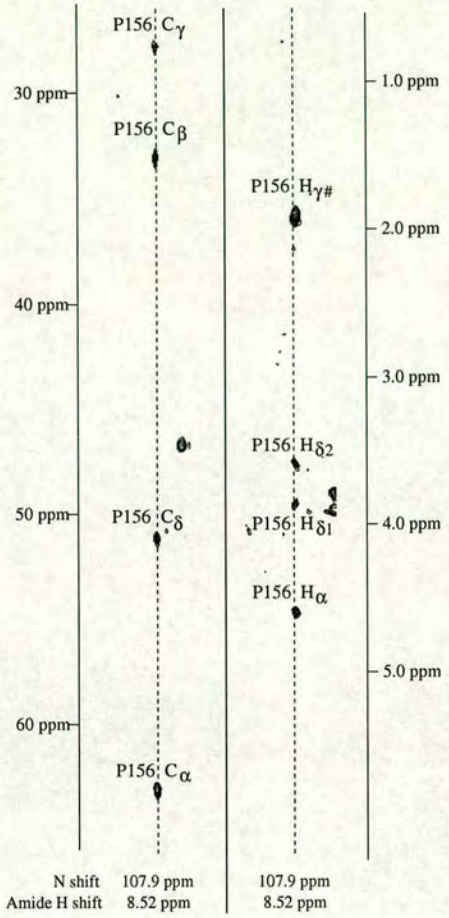


Figure 5.6: (H)C(CO)NH-TOCSY and H(C)(CO)NH-TOCSY strips for residue G157. The (H)C(CO)NH-TOCSY strip (left) shows the carbon atoms in the residue preceding G157 and the variation in shift suggests which is which (and later confirmed by the HCCH-TOCSY). The H(C)(CO)NH-TOCSY strip (right) shows the protons in the preceding residue. In this case the H_βs are not visible.

(S158, D178, D179) is not entirely unexpected as they could be involved in exchange motion.

Only four out of the eleven proline residues were fully assigned (P129, P156, P170 and P187). As mentioned above, P66 and P128 were completely unassigned. P67 was missing the C_γ, H_γs and carbonyl carbon. P69 was missing the C_δ and the H_δs. The only assigned atoms in P118 were the C_α, H_α, C_β and carbonyl carbon. P124 was missing the C_γ, H_γs and the H_δs. P189 was missing the H_βs, C_γ and the H_γs. In total, 49 of the 132 proline atoms (37%) could not be assigned. As the NMR sample used was dilute, the H(C)(CO)NH-TOCSY and (H)C(CO)NH-TOCSY spectra did not

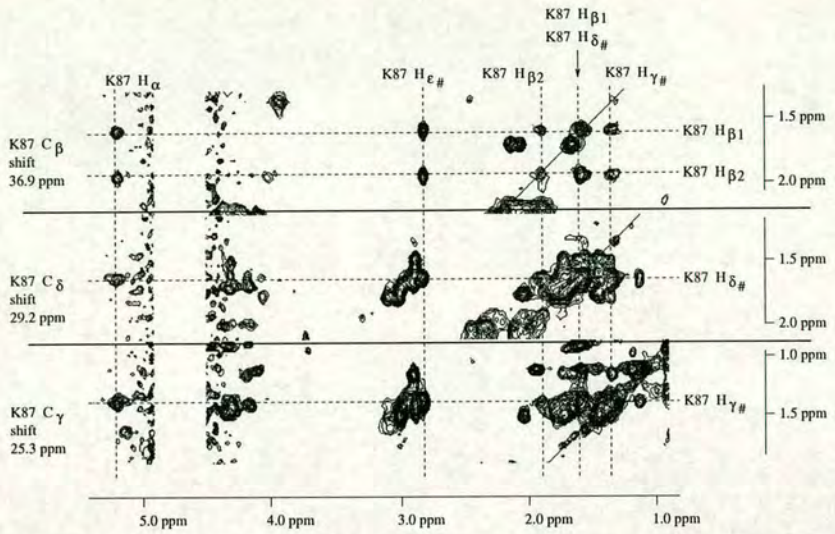


Figure 5.7: HCCH-TOCSY spectra for residue K87. This Figure shows planes at three different carbon shifts, corresponding to the C_β, C_γ and C_δ of K87. The protons attached to each carbon produce a diagonal peak at the requisite carbon shift. From this, crosspeaks to the other protons within the residue are visible.

always contain the full range of information from the C_γ, H_γs, C_δ and H_δs of prolines (as shown in Figure 5.6). Also, similarity of H_β and H_γ shifts led to many proline HCCH-TOCSY crosspeaks being overlapped with the diagonal and difficult to resolve with confidence.

Both Trp residues were fully assigned. Several aromatic residues were missing some ring proton and carbon shifts. The entire phenyl ring of F82 remained unassigned. The C_δ shifts for Y88 were missing. The C_ζ shift for F136 was missing, as were both the C_ζ and H_ζ shift for F144. Only the H_δs were assigned for F164 - a NOESY crosspeak from the H_βs made this possible. The rest of the ring could not be assigned with confidence. In the above aromatic cases, overlap of crosspeaks was responsible for the absence of the majority of the missing assignments. A plateau of signal intensity appeared in the aromatic random coil shift region in the aromatic spectra obscuring any resonances at those shifts. Figure 5.8 shows a section of overlap in the (HB)CB(CGCD)HD spectrum.

As commented upon previously, residues N177-Q180 account for a large proportion of the missing chemical shifts. Only the sidechain amide group nitrogen and attached

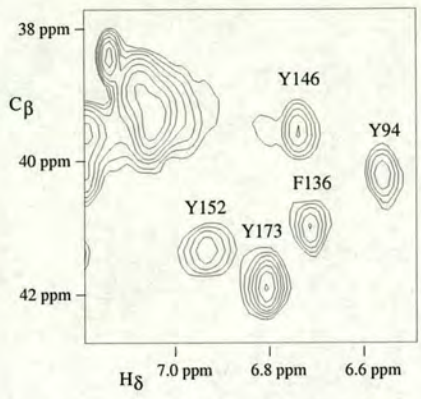


Figure 5.8: Overlap in the (HB)CB(CGCD)HD spectrum. In the top left of the figure there is a wide plateau of signal intensity. Any signals at these shifts are obscured. The nearby aromatic resonances with distinct chemical shifts (labelled on the figure) were straightforward to assign.

H_δs were assigned in N177. D178 and D179 were completely unassigned, as mentioned above. In terms of sidechain, Q180 had only its sidechain nitrogen and attached H_δs assigned. In the cases where these residues produced crosspeaks in any of the spectra they were of low intensity and in some cases hard to distinguish from the noise. A more complete assignment of equivalent residues was made in modules 16-17. Therefore there must be distinct differences between the dynamic motion of the FG loops in modules 3 and 17. Differences between the residues within the C-terminal portions of modules 2 and 16, for instance in the DE loop of 2/16, could account for this. A comparison of the isotropic dynamics of module pairs 2-3 and 16-17 can be found in section 6.2.

The C_ε and H_εs of M73 could not be assigned. In module 16, the equivalent protons were assigned, but in modules 16-17 they were not. The H_ε protons would normally be assigned in the homonuclear spectra or ^{15}N HSQC-TOCSY. The difficulty in assigning these protons in double module constructs comes from homonuclear spectra being heavily overlapped and the ^{15}N HSQC-TOCSY containing a reduced amount of data as mentioned. The C_{γ_1} of I105 could not be assigned due to overlap in the HCCH-TOCSY.

All sidechain amide group nitrogens and attached H_δs for all Asn and Gln residue were assigned. The sidechain amine groups for the lysine and arginine residues within

the construct could not be assigned. Neither could the protons in the hydroxyl groups of the serine and threonine residues. This is not unexpected as these protons are often exposed to the solvent and therefore labile and exchangeable, and rarely detected by NMR. Hydroxyl protons contained within hydrogen bonds are more stable, but can only be assigned through the use of NOESY spectra. None of these labile sidechain spins have been included in calculating the extent of the sidechain assignment.

A full chemical shift list for modules 2-3 can be found in Appendix B.

5.3 Modelling modules 2-3

Following the assignment of the backbone and sidechains of CR1˜2-3 the next logical step would be calculation of the three-dimensional solution structure. However, due to the high sequence similarity between modules 2-3 and modules 16-17, the level of novel data available from such a task would be limited. A homology based model would be likely, particularly in the case of module 3, to provide a representation of each module suitably accurate for some further studies. Studying the relaxation data was considered a higher priority than the structure calculation as there was no Modelfree-based dynamics data available for either site 1 or site 2 modules, apart from the module 16 analysis that was part of this project. It was therefore decided on the grounds of both novelty of results, and also due to time constraints, to pursue the dynamics of modules 2-3 and model the 2-3 structure by homology.

5.3.1 Alignment

Due to the homology between CR1˜2-3 and CR1˜16-17, comparative modelling using Modeller [124, 123] was a possible route to producing a good estimation of the structure of modules 2-3. The lowest five NOE energy structures of modules 16-17 were used as a template for modelling modules 2-3. The alignment, usually one of the most difficult parts of homology modelling, was simple in this case due to similarity of the sequences and an absence of insertions or deletions. The Trp residues and all eight Cys residues are in the same position in the sequences when comparing modules 2-3 with 16-17. The full alignment is shown in Figure 5.9.

```
          63                              90              104         *     120
CR1_2  EAKSCRNPPDPVNGMVHVIKGIQFGSQIKYSCTKGYRLIGSSSATCIISGDTVIWDTETPICDR
CR1_16      KT            TD  V  R NY   T H    H   E  L  N AH SNKP   Q
          963                             990             1004        *     1020

         125                             154                 174           *     191
CR1_3  IPCGLPPTITNGDFISTNRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSNDDQVGIWSGPAPQCI
CR1_17         A                          L  R
         1025                            1054                1074          *     1091
```

Figure 5.9: Sequence alignment of CR1, modules 2-3 and 16-17. Only residues which differ between modules 16-17 and 2-3 are listed for 16-17. Cysteine residues are numbered. Tryptophan residues are marked with a star.

Residues E59-I192 were modelled on residues E959-I1092. The N-terminal, non-native, EA residues were excluded as they were unassigned and unstructured in both 2-3 and 16-17.

5.3.2 Models

Ten models were outputted and the structures of the five with lowest overall energies were selected as most representative. In Figure 5.10 there is a representation of the lowest energy model along with the five template 16-17 structures.

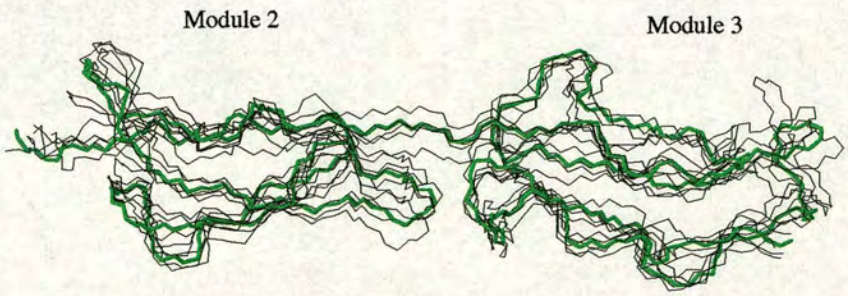


Figure 5.10: Model of the structure of CR1˜2-3. Modeller-based model of CR1˜2-3 (shown in green) overlaid with the five lowest energy structures of modules 16-17 (shown in black)

The five modelled structures with lowest energy were analysed using PROCHECK [77] to determine their suitability as representations of real proteins. The summary of the

Ramachandran plot (not shown) is tabulated below.

Residue type	Number	Percentage
Residues in most favoured regions	433	81.7%
Residues in additional allowed regions	76	14.3%
Residues in generously allowed regions	19	3.6%
Residues in disallowed regions	2	0.4%

Table 5.2: Ramachandran analysis of the top five models of CR1~2-3.

This good distribution (i.e. over 95% of residues in the favoured or generously allowed regions) shows that the set of structures have reasonable stereochemistry.

Within this set of structures, PROCHECK identified eleven potential β-strands within the module pair, corresponding to those found in modules 16-17, as well as one further strand consisting of F136-S138. Table 5.3 shows the residues comprising each identified strand.

Modelled module 2 strands	Residues
B	M73-H75
D	G85-S89
E	R95-I97
F	S101-I106
G	V111-W113
H	I119-D121
Modelled module 3 strands	**Residues**
B*	F136-S138
D	V149-Y152
E	E165-V167
F	S171-Y173
H	Q190-I192

Table 5.3: β-strands predicted on the model of CR1~2-3. Strands labelled using CCP module fold convention, as detailed in section 1.3.2. Strand B in module 3 is starred as it is the only strand in the models not present in the module 16-17 template structures.

Module 3 differs from module 2 in that residues equivalent to strand G are not recognised as having the correct ϕ, ψ angles to be consistent with β-strand structure. Consequently the module 3 strand F, which in module 2 hydrogen bonds to strand G, is shorter.

It is not unexpected that the strands are very similar to those found in the CR1~16-17 structures because Modeller relies heavily on the template files provided. That an extra strand (F136-S138) that was not identified in CR1~16-17 is present in the modelled structures, strand B in module 3, is not unusual. This region is frequently part of a strand in CCP modules (as in module 16). Whether or not this reflects a real difference between modules 3 and 17 is unknown. None of the three amino acid changes between modules 3 and 17 appear to be close to residues F136-S138.

One major different between the models of 2-3 and the structures of 16-17 relates to the linker region. In the models of 2-3, the tyrosine rings of residues Y94 and Y146 appear to be reasonably close in space, as shown in Figure 5.11. The rings are eclipsing each other with the hydroxyl groups arranged "head to tail".

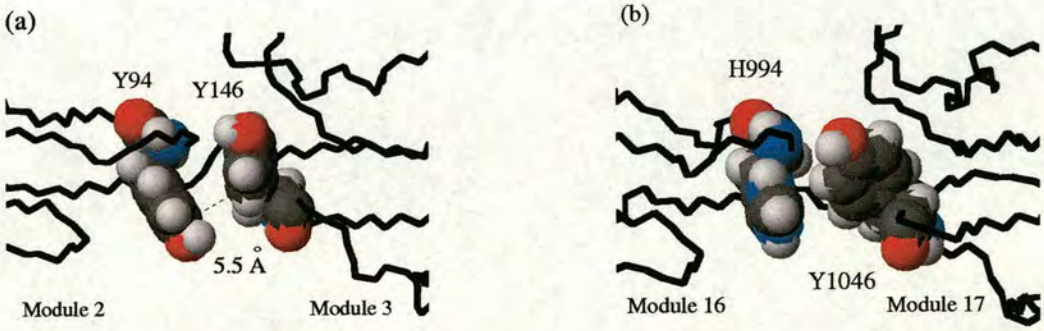


Figure 5.11: Possible Tyr ring stacking at the module 2-3 interface. Figure (a) shows possible Tyr ring stacking at the module 2-3 interface. The rings are parallel and around 5.5 Å apart. Figure (b) shows the His and Tyr rings at the module 16-17 interface. The two aromatic rings are perpendicular, ruling out the possibility of ring stacking.

In the case of modules 16-17, Y94 is replaced by a histidine residue. The rings are perpendicular to one another and could not be involved in ring stacking. Normally for π system interaction, aromatic rings should be around 3.5-4.0 Å apart [146]. In the

module 2-3 models the rings are around 5.5 Å apart which would preclude π-stacking. However, it is possible that the model is not showing a real representation of the interface as it is biased by modules 16-17. The stacking of these two Tyr rings, and the creation of a hydrophobic region close to or including parts of the linker, remains a possibility. Y94 is one of the residues shown by mutagenesis to be important in the functions of site 1.

These models are suitable for the interpretation of possible structural differences and also for discussing those NOEs which have had their contributing resonances assigned. Moreover, they are important for analysing the relaxation data and Modelfree results which form the next stage of the project.

5.4 Relaxation data

The relaxation data was acquired as described in sections 3.5.3 and 3.5.5. The way in which protein motions can be interpreted on the basis of raw relaxation data was described in section 1.3.4.

5.4.1 Omissions

Various factors contributed to reducing the number of residues in the 2-3 construct suitable for ^{15}N backbone relaxation data analysis. Obviously, the proline residues have no amide protons and therefore could not be included. Several residues (H75, G108, T139, H145, S158, G159, N177, D179 and V181) had not been assigned in the ^{15}N,^{1}H-HSQC and so could not be used. Similarly, though K61 and G79 had weak ^{15}N,^{1}H-HSQC crosspeaks, they did not have detectable peaks in the relaxation spectra and could not be analysed either. K61 and G79 are located at the N-terminus and in a loop, respectively, of the model structure and their absence from the relaxation spectra is probably due to intermediate exchange of their amide protons. There were also problems specifically relating to the heteronuclear NOE (hetNOE) data. Three residues, V74, R141 and G160, had their amide nitrogen and proton assigned (using the 3D spectra) but had no visible ^{15}N,^{1}H-HSQC or hetNOE peaks. These residues therefore could not be included. A further three residues, I137, T175 and S185, had their

hetNOE crosspeaks partially obscured by artefacts in the spectra and were deemed unreliable with regards to integration.

The following pairs or groups of residues could not be analysed as they were too overlapped in the relaxation spectra to be reliably integrated: E59 and F144; C63 and A102; D68 and N140; N71, I112 and C154; K78 and D114: I86 and S148; I97 and V167; S99, S100 and G182; I105 and Y152; T130 and E142; F136 and Q180; V149 and V163; V150 and I192. Table 5.4 shows the numbers of residues in each category. (Non-native residue A60 was also suitable for analysis.)

Residue category	**Number of residues**
Proline residues	11
Unassigned in ^{15}N,^{1}H-HSQC	9
Absent from NOE spectra	5
Obscured by artefacts in NOE spectra	3
Overlapped	27
Suitable for analysis	77
Total	132

Table 5.4: Residues available for relaxation data analysis in CR1˜2-3.

5.4.2 T_1 relaxation data

Figure 5.12 shows the T_1 relaxation data for modules 2-3. Excluding non-native signal sequence residue A60, the mean amide ^{15}N T_1 was 799 $\pm$ 132 ms. The most obvious outliers were G109, N143 and D178, all of which had a T_1 of at least two standard deviations above the average. Based on the model of modules 2-3, all of these residues were located in potentially flexible positions in or near loops. No residues displayed a notably low T_1 value. The linker region showed consistent T_1 values close to the average.

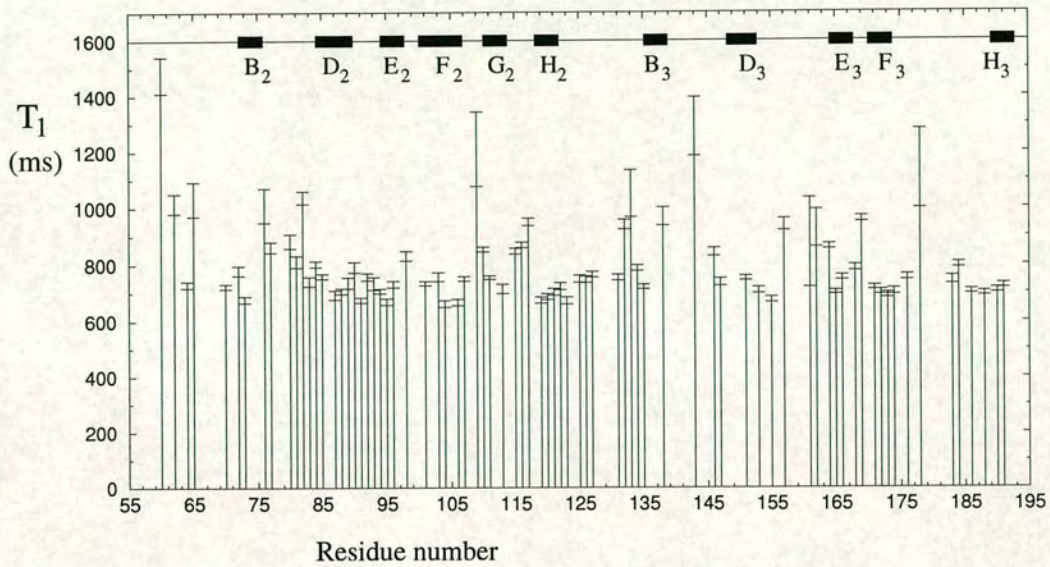


Figure 5.12: T_1 values for CR1~2-3. Error bars are shown for each relaxation time. The predicted β-strands of the 2-3 model are shown as black rectangles and labelled using the CCP module convention (section 1.3.2).

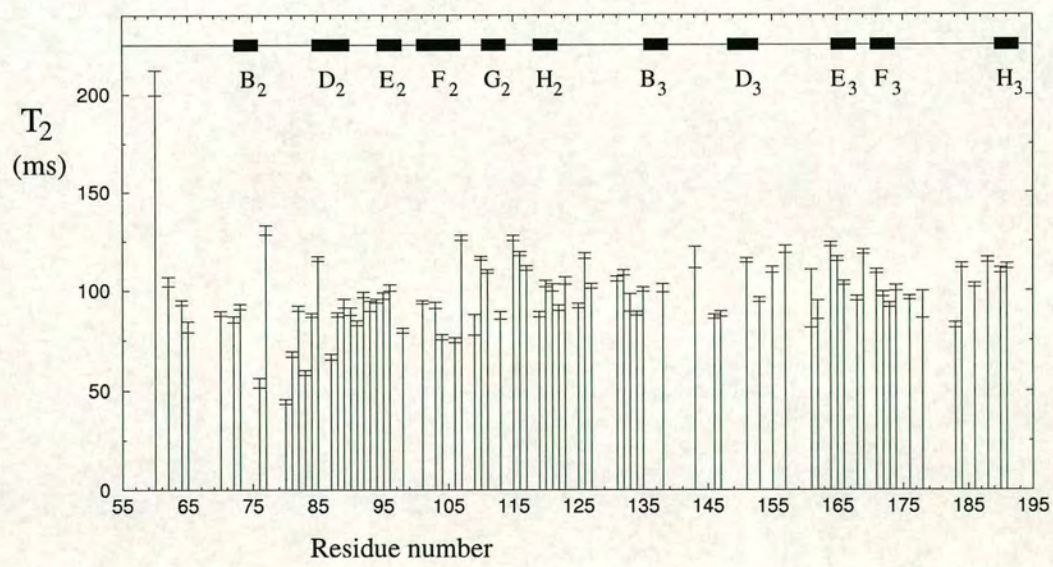


Figure 5.13: T_2 values for CR1~2-3. Error bars are shown for each relaxation time. The predicted β-strands of the 2-3 model are shown as rectangles and labelled using the CCP module convention.

5.4.3 T_2 relaxation data

Figure 5.13 shows the T_2 relaxation data for modules 2-3. The mean amide ^{15}N T_2 value for the module pair (excluding signal sequence residue A60) was 98 ± 17 ms. Obvious outliers (T_2 less than two standard deviations below the mean) included V76, I80 and G83. These three residues, with their low T_2's, may be undergoing chemical exchange (μs-ms motion). Thirteen residues, I77, S107, T110, T115, E116, G126, N143, T151, G157, F164, E165, E169 and A188 had a T_2 greater than one standard deviation above average. Apart from residues N143, G157 and E169, none of these above residues had a T_1 greater than one standard deviation above the mean. Therefore diffusional anisotropy could be contributing to the relaxation within these residues as it could increase T_2 while reducing T_1. Again, the linker region showed consistent T_2's that did not deviate far from the mean.

5.4.4 Heteronuclear NOE data

Figure 5.14 shows the ^{15}N heteronuclear NOE data for modules 2-3. The most N-terminal residue for which data was available (non-native A60) was the only residue to have a negative NOE. Excluding this residue, the range was 0.30−0.92 and the mean was 0.69 ± 0.14. This indicates a wide range of fast timescale flexibilities throughout the module pair. Notable areas of low NOE include residues G157-F164, where there is an obvious dip in NOEs - this corresponds to the DE loop in module 17. The linker region showed, on average, slightly higher than average heteronuclear NOEs - the mean for the three linker residues with available data being 0.74 ± 0.11. (Proline 124 had no relaxation data available.) Of the linker residues, only R122 showed a lower than average value of 0.60. Including the residues C120 and C125 with the linker gave the same mean but with a higher standard deviation: 0.74 ± 0.15.

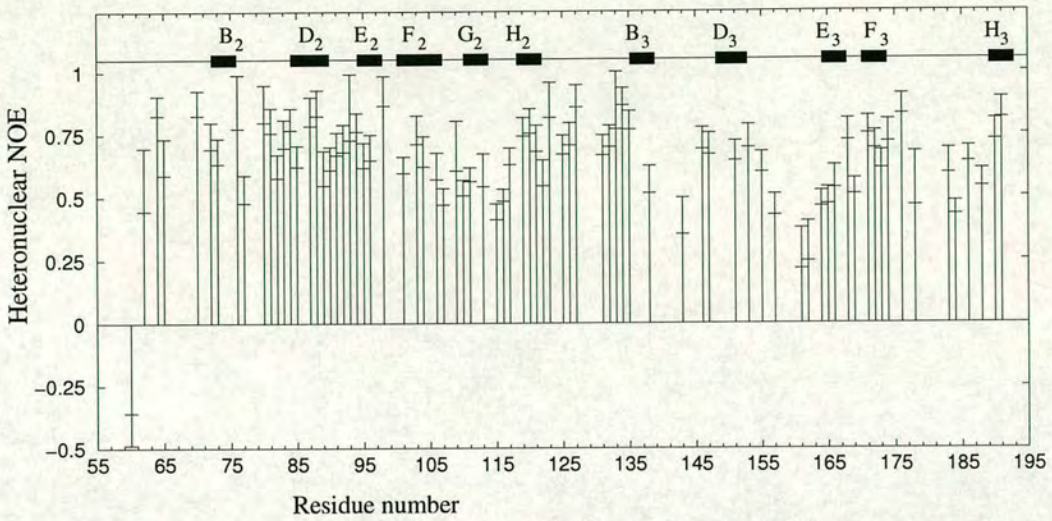


Figure 5.14: Heteronuclear NOE values for CR1˜2-3. Error bars are shown for each NOE value. The predicted β-strands of the 2-3 model are shown as rectangles and labelled using the CCP module convention.

5.4.5 Tryptophan sidechains

The relaxation data of the tryptophan $N_{\varepsilon_1}/H_{\epsilon_1}$ bond were also collected for both W113 and W184, as shown in Table 5.5. All six pieces of relaxation data for these sidechains fell within one standard deviation of the mean.

Trp sidechain	T_1 (ms)	T_2 (ms)	NOE
W113	867 ± 15	110 ± 2	0.64 ± 0.05
W184	857 ± 12	106 ± 1	0.66 ± 0.03

Table 5.5: Tryptophan sidechain ^{15}N relaxation data in CR1˜2-3.

5.5 Isotropic dynamics

5.5.1 Isotropic Modelfree results

Nine of the 77 residues for which data were available could not be fitted by Modelfree. The majority of these are in module 3. They are C120, I123, G126, G134, D135, F164, S171, Q190 and C191. The remaining 68 residues were fitted by Modelfree. Table 5.6

shows the number of residues fitted to each model.

Fitted model	1	2	3	4	5
No. of residues in module 2	14	3	9	4	9
No. of residues in module 3	8	4	1	7	8
Total	22	7	10	11	17

Table 5.6: Number of CR1~2-3 residues fitted to each Modelfree model. Poorly fitted C104 is not included.

One of the residues, C104, although fitted, was not included in any further analysis and is not counted in Table 5.6. C104 was fitted to the simplest case, model 1, with an S^2 value of 1.00 ± 0.01. Such a value would imply that its amide bond is entirely fixed in one position exhibiting no motion of any kind. Therefore this value cannot accurately represent the mobility of this bond. Even though C104 forms a disulphide bond, increasing stability in its vicinity, this would not account for complete rigidity. Originally, the SSE of C104 failed to be low enough to be fitted to model 1. It was during the re-evaluation of model 1 fitting (see section 3.5.9) that C104 fell within the acceptable bounds and was fitted to model 1. However, the values calculated for the relaxation data from the fitted parameters were not particularly accurate. (R_1 of 1.51 s^{-1} was fitted with 1.57 s^{-1} and R_2 of 13.06 s^{-1} was fitted with 12.08 s^{-1}.) It seems likely, therefore, that residue C104 is indeed quite rigid, but in this case Modelfree has overestimated the S^2 value.

5.5.2 S^2 values

Figure 5.15 shows the S^2 values fitted by Modelfree to CR1~2-3. The range of S^2 values was 0.45−0.96. Mean S^2 for all native residues was 0.76 ± 0.15. Using the β-strands predicted from the Modeller-based model structure, it was possible to consider the S^2 values in terms of predicted secondary structure (Table 5.7).

The mean S^2 for the β-strands in module 2 was 0.86 ± 0.08, whereas it was 0.78 ± 0.09 in module 3, although the sample used for module 3 was small. The mean for

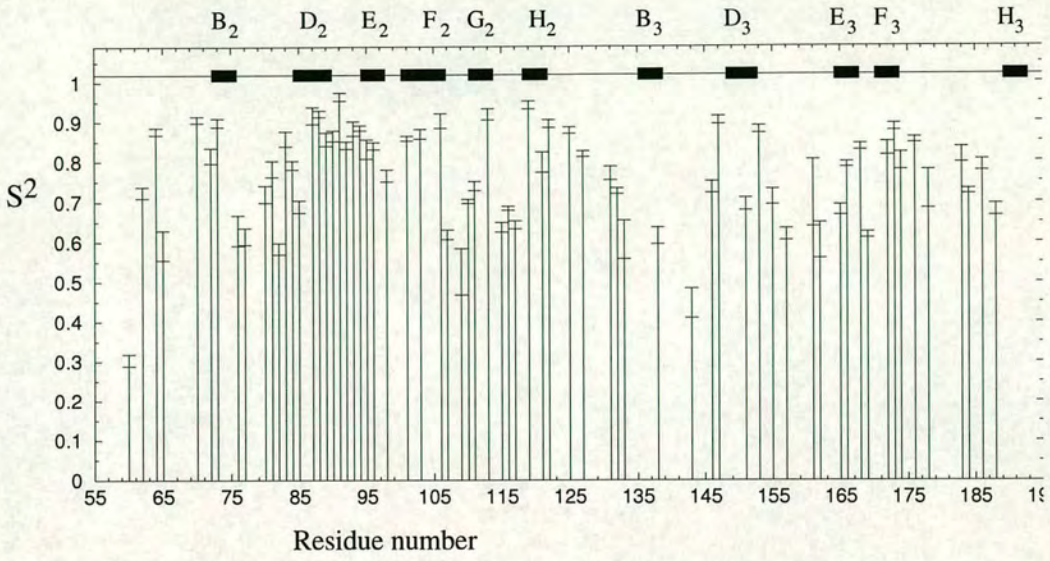


Figure 5.15: S^2 values of CR1˜2-3. Error bars are shown for each S^2 value. The predicted β-strands of the 2-3 model are shown as rectangles and labelled using the CCP module convention.

Modelled module 2 strand	**No. of residues**	**Mean S^2 ± sd**
B	1	0.90
D	4	0.85 ± 0.11
E	2	0.83, 0.84
F	3	0.88 ± 0.02
G	2	0.74, 0.92
H	2	0.94, 0.80
Modelled module 3 strand	**No. of residues**	**Mean S^2 ± sd**
B	0	-
D	1	0.70
E	2	0.68, 0.79
F	2	0.83, 0.89
H	0	-

Table 5.7: S^2 values of modelled β-strands in models of CR1˜2-3. Where only two or fewer residues within a strand has an associated S^2 value, a mean is not taken and instead the S^2 of the individual residue(s) are shown.

all residues in strands is 0.84 ± 0.08, whereas the mean for all residues not found in strands is significantly lower - 0.74 ± 0.12 (excluding non-native A60). Therefore, as expected, the residues making up the β-strands were significantly more rigid.

The overall mean S^2 for module 2 residues was 0.78 ± 0.12. The overall mean S^2 for module 3 residues was 0.74 ± 0.11. A statistical Student's t test showed that these results were significantly different only to a 90% confidence limit, showing neither module can be said to be more flexible than the other.

However, one difference between the two modules lies in the number of residues for which there are available data. In terms of S^2 fitting, 39/67 were found in module 2, compared to only 28/67 in module 3. This difference is accounted for by the slightly larger number of overlapped residues in module 3 as well as the fact that seven out of eight of the residues that could not be fitted by Modelfree were found in module 3. This may point to the fact that module 3 is undergoing more motion on a ns-μs timescale which Modelfree is not able to deal with.

Several residues had an S^2 much lower than the average. Although data is sparse from residues S138-H145, the two residues that have been fitted show high flexibility (S138, 0.62 ± 0.02; N143 0.44 ± 0.04). Residue T139 was also completely absent from the ^{15}N,^{1}H-HSQC, suggesting that exchange motion is occurring. From the modelled structure, this region (the hypervariable loop and CD loop of module 3) is towards the 2-3 intermodular junction, so it seems that flexibility here is unlikely to result artefactually from the absence of modules 1 or 4 from this construct. The S^2 values therefore indicate that the section of module 3 which would comprise the CD loop and would normally comprise β-strand C (if it were present) are highly mobile. As with the lone module 16, perhaps the motion occurring is preventing the formation of the AC β-sheet in module 3.

Looking at the linker region itself (and including the two Cys residues), only three of the six residues from C120-C125 inclusive were analysed by Modelfree. Residues D121, R122 and C125 gave S^2 values of 0.80 ± 0.03, 0.90 ± 0.01 and 0.88 ± 0.01. All

three of these S^2 values are above average for the module pair, and the values for R122 and C125 are higher than 80% of other residues. Indeed, these results show that the residues in the linker have fast timescale rigidities comparable with those found within the most rigid β-strands present in this construct. This high level of rigidity within some residues in the linker region indicates that the range of intermodular motion available on this timescale may be limited. Residues R122 and C125 were both fitted by the simplest case, model 1. D121 was fitted by model 5, indicating motion on two timescales. The overall S^2 was contributed to by S_f^2 of 0.88 $\pm$ 0.02 and an S_s^2 of 0.91 $\pm$ 0.02. From this, both the fast and slower ps-ns timescale motions of the residue have low amplitudes. Of the other three residues in the linker region, one was a proline and the other two were analysed but could not be fitted by Modelfree. This could indicate that they were undergoing motion on the ns-μs timescale which cannot be interpreted by Modelfree.

5.5.3 S^2 values mapped on to the modelled structure of CR1, 2-3

Using the previously generated model of the structure of module pair 2-3, the S^2 values were plotted as colours to show the pattern of overall flexibility. While the model will probably differ in intermodular orientation from the real structure, the individual module structures are likely to be fairly accurate and suitable for this purpose.

From Figure 5.16 it can be seen that module 2 has a similar distribution of S^2 values to that of module 16 (see section 4.5.2 and Figure 4.21 therein). The face of module 2 that would contain β-strands B_2, G_2 and H_2 contains the majority of residues which show very high or moderate flexibility (coloured red or orange, respectively). The other face of the module shows markedly higher S^2 values. For example the stretch from G83-I106 (which comprises strands D_2, E_2 and F_2) contains only a single residue coloured orange (Q85) and none coloured red. While this area is therefore relatively rigid and a distinction exists between the two faces, an even clearer distinction is visible by dividing the structure of module 2 into N- and C-terminal halves. Although only a few residues were analysed using Modelfree in the AB_2 loop region, those that were (V70, G72-M73) also show a low degree of flexibility. (It should be noted however,

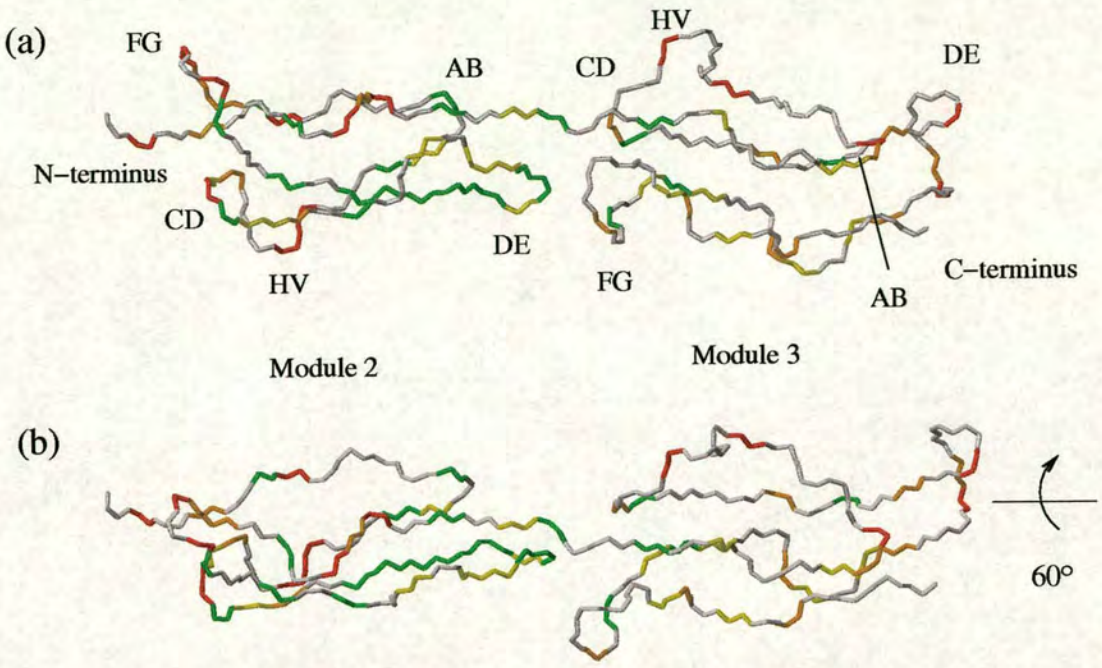


Figure 5.16: Distribution of S^2 values plotted on CR1~2-3 as colours. Red depicts areas of high flexibility ($S^2 < 0.65$); orange corresponds to areas of moderate flexibility ($0.65 \leq S^2 < 0.75$); yellow shows moderately rigid regions ($0.75 \leq S^2 < 0.85$); green shows regions of highest rigidity ($S^2 \geq 0.85$); white depicts residues for which no S^2 value could be determined. Two figures of the structure, rotated by 60°, are shown. Loop regions are labelled.

that residue M73 is part of strand B.) The AB_2 region and the D_2, E_2 and F_2 strand region are both located in the C-terminal half of module 2. Fitted residues preceding and within the linker (I119, D121-R122) continued the trend of high S^2 values in the entirity of the C-terminal half of module 2. Even the residues found in the tip of the DE_2 loop were not highly flexible on the fast timescale.

The reduced amount of S^2 data for module 3 should be noted as the results here cover less than half of the residues present in module 3. However, the data present show a particular pattern. All the residues coloured red are confined to the face of the module containing the hypervariable loop, although some orange residues do appear on the other face.

In terms of the overall pattern, it appears as if the module pair is most rigid towards the linker. It is possible that some flexible residues towards the N-terminus of module 2 and the C-terminus of module 3 would be less flexible in the intact protein due to contacts with modules 1 and 4. It is interesting that the least flexible region (on the fast timescale) within the construct is around the 2-3 linker as is the case here, because the 16-17 junction was reported to display flexibility (although not necessarily on the fast timescale).

5.5.4 Mutagenesis data

As mentioned in section 1.4.5, three residues within module 2 have been implicated in providing the C4b-binding and decay accelerating activity of functional site 1: R64, N65 and Y94. The three mentioned residues are all located on a single face of the model of module 2.

Residues R64 and N65 are located towards the N-terminus of the module whereas Y94 is located towards the C-terminus, in the EF loop. While the backbone atoms of this residue are actually on the opposing face to residues R64 and T65, the sidechain of Y94, on the model, is angled into the junction region. This places the sidechain against the linker and it becomes exposed on the flexible face. From the model, the Tyr ring is in close contact with linker residues D121 and R122. This close interaction could be stabilising the junction area, preventing motion and therefore contributing to the fast and slow timescale (described in section 5.5.6) spatial restriction of the 2-3 junction, as measured by Modelfree. Residue Y94, which is possibly involved in aromatic ring stacking (see Figure 5.11 in section 5.3.2) was fitted to a high (0.89) S^2 value. It is possible that the stability increase gained from this intermodular interaction plays a part in the reduced flexibility which might be located in this junction, compared with other CCP module linkers.

5.5.5 S_s^2 and S_f^2 values

Seventeen residues were fitted by model 5 and therefore required two separate S^2 values to be fitted. Eight of these residues (I77, Q85, R95, S107, T110-V111 and T115-E116) were located in module 2. A further eight (I131, T151, N155, E165, I172, C174, G186 and A188) were found in module 3. In the linker a single residue, D121, was fitted to two timescales.

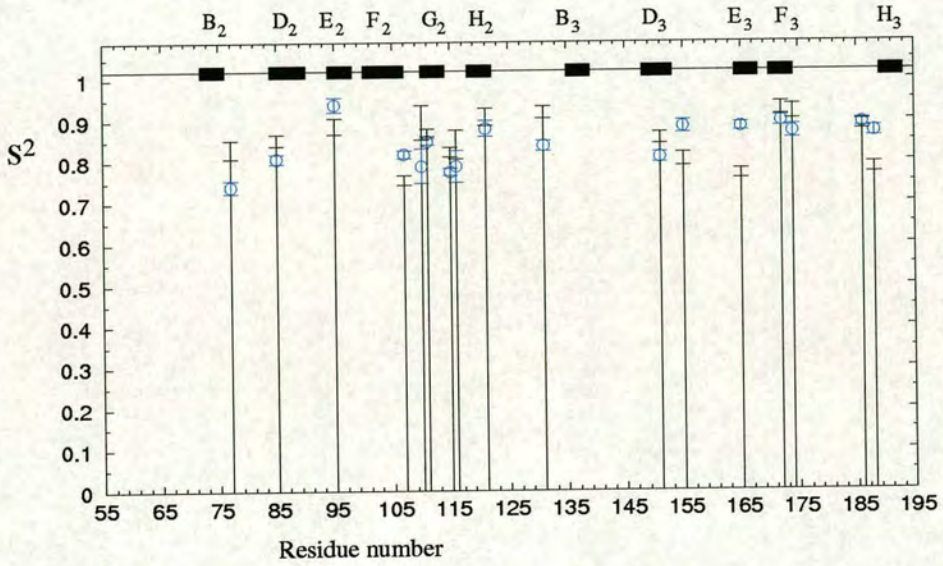


Figure 5.17: S_s^2 and S_f^2 values of CR1~2-3. S_s^2 values are represented by columns. S_f^2 values are represented by circles. Error bars are shown on each column and on each circle for the respective S_s^2 and S_f^2 values. The predicted β-strands of the 2-3 model are shown as rectangles and labelled using the CCP module convention.

The range of S_s^2 was 0.61−0.83 and the range of S_f^2 was 0.74−0.94. In most cases, the S_s^2 and S_f^2 values were similar. The only exceptions were residues T110 and E165 which had a difference of 0.1 or more. In T110, the S_s^2 was the lower value while in E165 the S_f^2 was lower. This similarity between S_s^2 and S_f^2 values is to be expected as none of these residues have a low overall S^2 and so high flexibility within either S_s^2 or S_f^2 (which multiply together to give S^2) was not possible.

5.5.6 Chemical exchange

A total of 21 residues required a chemical exchange term to be fitted and of these 11 also required a correlation time. (Non-native A60 also required both a chemical exchange term and a correlation time.) The fitted values ranged from 0.83–13.79 s^{-1} and are shown in Figure 5.18. Figure 5.19 shows the distribution of the residues fitted for chemical exchange on the model of the modules 2-3 structure.

The majority of the residues requiring the R_{ex} term were in module 2 (thirteen of the 21) and in particular every residue in the stretch from I80-S84, corresponding to the CD loop, had an R_{ex} term. The eight residues fitted in module 3 showed no overall pattern, appearing singly on all sides of the module. One of the eight was located within a β-strand (S138) and four more were in loop regions - N133 in the AB_3 loop, Y146 in the CD_3 loop and G157 and K162 in the DE_3 loop region. Residue N143 is located close to the CD_3 loop.

The module 2 residues undergoing chemical exchange appear, as they did in CR1˜16, mostly on one face. (See section 4.5.4 for a complete comparison.) Residues N65, G72, V76, I80-S84[2], I106, D109 and T117 again appear on what is analogous to the "front", binding face of site 2 on module 16. Only residues K87 and G98 have a chemical exchange term but are located on the "back" face of the module.

None of the residues within the linker required an R_{ex} term to be successfully fitted.

5.5.7 Correlation times

Eighteen residues required a correlation time to be fitted (as shown in Figure 5.20). Eleven of these also required a chemical exchange term. In Figure 5.20, the residue I131 has been excluded to better show the data. I131 had a fitted t_e of 771 ± 250 ps. This value, while higher than that of the others, is still within the region that Modelfree can fit with certainty. The range of t_e's fitted was 17.2 ± 7.5 to 771 ± 250

---

[2] Residues G83 and S84 do not appear square on the "front" face but as the CD loop in which they are located in (along with R_{ex} residues I80-F82) could be considered part of either face, they are included in the residues exposed on the binding face.

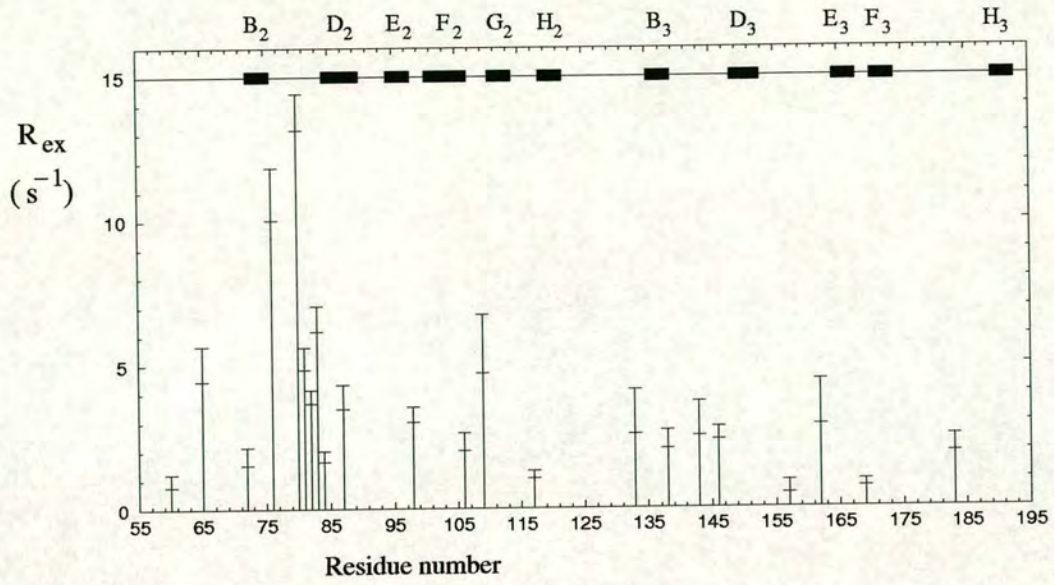


Figure 5.18: R_{ex} values of CR1˜2-3. Error bars are shown for each chemical exchange term. The predicted β-strands of the 2-3 model are shown as rectangles and labelled using the CCP module convention.

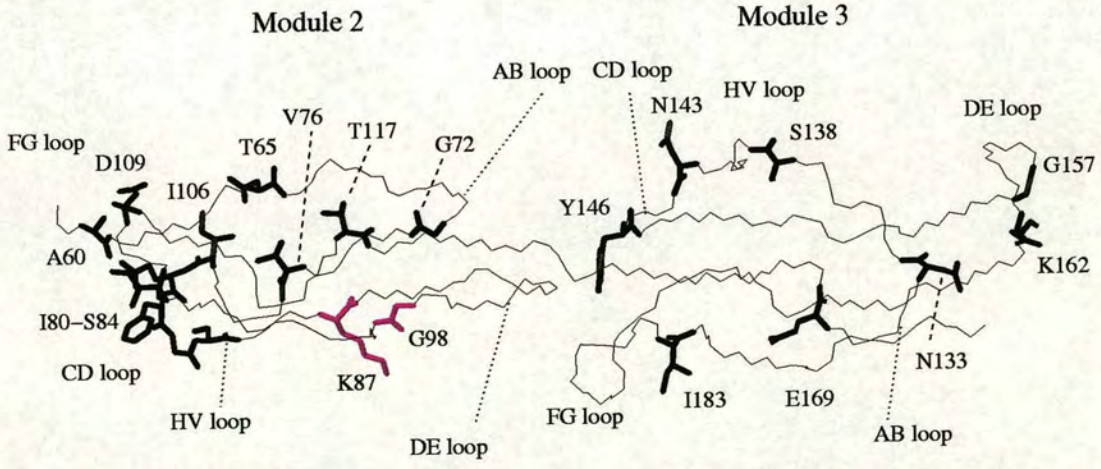


Figure 5.19: Residues with R_{ex} values mapped on the model of CR1˜2-3. Residues on the "back" face of module 2 are shown in magenta. Residues G83 and S84 do not appear square on the "front" face but as the CD loop in which they are located (along with R_{ex} residues I80-F82) could be considered part of either face, they are included in the residues exposed on the binding face. Loop regions are labelled using the CCP module convention.

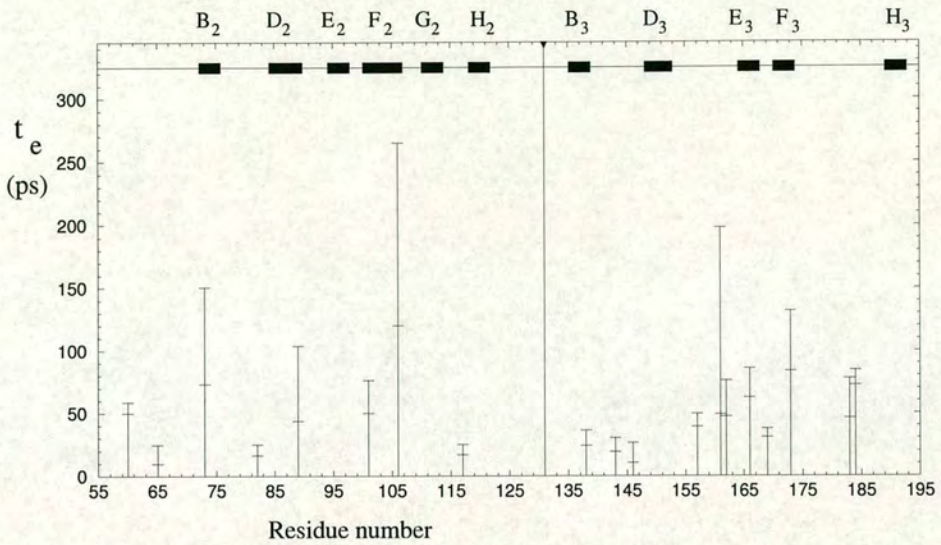


Figure 5.20: t_e values of CR1˜2-3. Error bars are shown for each correlation time. The predicted β-strands of the 2-3 model are shown as rectangles and labelled using the CCP module convention. I131 (t_e = 771 ± 250 ps) is shown as a vertical line.

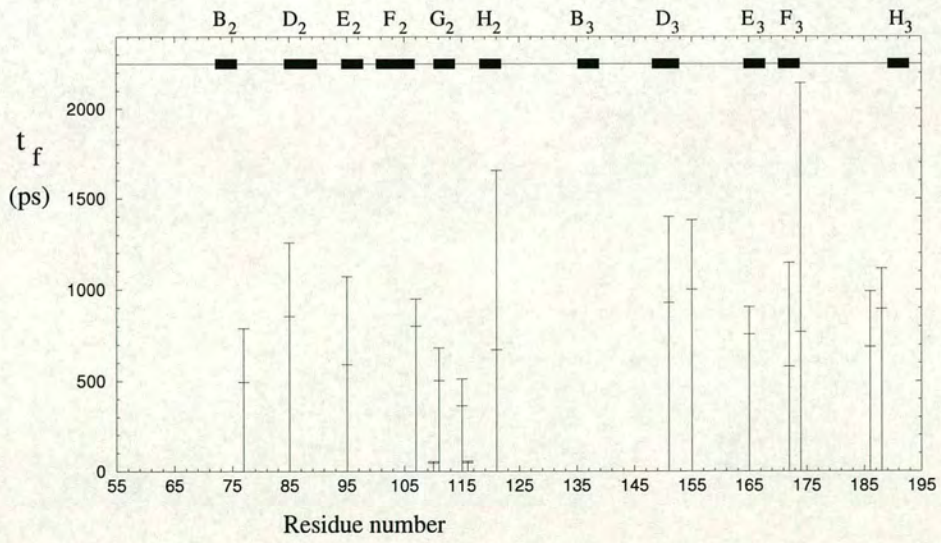


Figure 5.21: t_f values of CR1˜2-3. Error bars are shown for each correlation time. The predicted β-strands of the 2-3 model are shown as black rectangles and labelled using the CCP module convention.

ps.

Furthermore, a further seventeen residues were fitted to model 5 motion and therefore had a correlation time associated with S_s^2, as shown in Figure 5.21. The range of these t_f correlation times was 48.6 ± 5.0 to 1457 ± to 686 ps.

5.5.8 Tryptophan sidechains

Both tryptophan $N_{\varepsilon_1}/H_{\varepsilon_1}$ sidechain bonds were fitted by Modelfree, and both were fitted to the simplest case, model 1. The W113 sidechain had an S^2 value of 0.75 ± 0.01, whereas the W184 sidechain was fitted to an S^2 of 0.76 ± 0.01. As the sidechains of the invariant Trp residues are always buried in CCP modules, consistency between these two results is also expected.

5.6 Anisotropic dynamics

5.6.1 Anisotropic fitting

Because the anisotropic fitting by Modelfree requires knowledge of the angle between amide bond vectors and the diffusion tensor for the molecule (as described in 3.5.10), use of the homology model of CR1~2-3 would not be legitimate. This is because a small change in protein structure can produce a large change in the orientation of N−H bonds. However, the model was tested, and in each case the fitting of the Modelfree parameters was improved compared with the isotropic version. (F tests confirmed the improvement was real.) The ratio of $D_{||}/D_{\perp}$ was determined to range from 1.52−1.96, suggesting that the ratio of length to breadth for the double module construct was less than 2:1. The models of modules 2-3, however, showed an extended structure with a larger ratio. Due to this discrepancy and the fact that there was no certainty in accurate of the models, the Modelfree axially anisotropic fitting was not completed because the results generated could not be robustly defended. When a reliable, NOE-derived structure of CR1~2-3 is available, the fitting can be re-attempted.

5.6.2 Spurious R_{ex} fittings

There are possible consequences to using an isotropic diffusion Modelfree fitting of modules 2-3, when modules 2-3 are unlikely to be isotropic in shape. Analysis of relaxation data to which diffusional anisotropy is contributing can create spurious chemical exchange fittings, as T_1 can be raised while T_2 is lowered. In the Modelfree fitting of lone module 16, this is likely to have a limited effect as lone CCP modules have limited anisotropy of shape. Indeed, in the case of MCP~1, there was good correlation between the isotropic and the axially symmetric diffusion Modelfree fittings [106, 15]. In CR1~2-3, the effects could be more pronounced. However, the majority of the R_{ex} residues in modules 2-3 are located in module 2^3, and modules 2^3 and 16 show some consistency in the residues fitted with an R_{ex} term[3]. This suggests that the 2-3 fitting is representative of reality in modules 2-3. The larger R_{ex} values fitted in module 2^3 residues compared with those in module 16 may, however, have a contribution from diffusional anisotropy.

5.7 Linker NOEs

Identifying NOEs from and to the linker, or between modules, is a useful way of determining sustained contact between separate regions of the fragment. The Modeller-based model of CR1~2-3 was used to examine the NOEs mentioned below and calculate distances, although it cannot be assumed to be representative of the real structure of modules 2-3. In the case of the ^{15}N HSQC-NOESY, the majority of the NOEs detected from the linker residues' amide bonds were either intra-residue or sequential. There were three strong NOEs that did not fit this category and one amide-amide NOE.

In the ^{15}N HSQC-NOESY strip of D121 there was an NOE from the amide of D121 to presumably another amide, as it occurred at 9.52 ppm. This shift closely matched that of the amide proton of R95 and from the 2-3 model, this residue is sufficiently close to the linker to be involved in this interaction. It seems likely that linker residue

[3] Out of a possible 11, five of the R_{ex} residues in module 16 had the equivalent residue in module 2^3 fitted to R_{ex}. In several cases, module 2^3 R_{ex} residues had a neighbouring residue whose equivalent in module 16 had been fitted (e.g. A60/K961, G98/H999, I106/S1007).

D121 is in close contact with module 2 DE loop residue R95.

In the ^{15}N HSQC-NOESY strip of R122, there was a strong NOE crosspeak at 6.60 ppm. An attempt to assign this NOE was made by comparing the shifts of all assigned protons in the modules 2-3 construct. Only three possibilities were found: the H_δs and H_εs of Y94 and the H_δs of Y88. Y146 was close in space to R122 but its ring protons were at the wrong chemical shifts (6.75 and 6.82 ppm). F144 was just close enough to perhaps warrant an NOE, but again its shifts were incompatible (7.21 and 7.18 ppm). All other aromatic residues were too distant in the modelled 2-3 structures to be interacting with R122, including Y88. Of the Y94 protons, the modelled structure showed that the H_δ protons were 1 Å closer to the R122 amide group than the H_ε protons. This suggested that the Y94 H_δs were causing the NOE crosspeak, but alone is not conclusive. Either way, there appeared to be a strong NOE between the linker residue R122 and residue Y94 in the DE loop of module 2.

One further linker NOE was from the I123 N−H and occurred at 4.81 ppm. The closest shift matches were the H_αs of both Y94 and R95 (which have chemical shifts 4.82 and 4.81 respectively). Both these residues are again situated near the linker and using the modules 2-3 structure, the H_α of Y94 was shown to be around 5 Å from the amide proton of I123. The H_α of R95 was about 7.5 Å from the amide proton of I123, which implies that Y94 is the interaction partner.

The final linker NOE of interest in the ^{15}N HSQC-NOESY was from the amide group of C125. While this residue forms a disulphide bridge with C174 and is not part of the linker, this unidentified NOE still has ramifications for junction interactions. The crosspeak occurred at 5.22 ppm and did not have any obvious partners. Though the H_αs of K87, D121 and C154 were all at approximately the right shift (5.19, 5.19 and 5.22 respectively), all of these possibilities were too distant in the 2-3 models to explain such an intense NOE. The shift implied an H_α and residue H145 was positioned extremely close to C125 in the model structures. Furthermore, there was a strong crosspeak in ^{15}N HSQC-NOESY strip for residue Y146 at the same shift, suggesting that the interaction partner was the previously unassigned H_α of unlabelled H145.

No intermodular NOEs were identified in the ^{15}N HSQC-NOESY. The process of assigning of a ^{13}C-edited HSQC-NOESY, as part of the structure calculation, would make possible the identification of any intermodular NOEs, as well as a greater number of ^{15}N NOEs.

5.8 Conclusions

The assignment of the ^{13}C, ^{15}N and ^{1}H resonances of CR1~2-3 has allowed the analysis of the dynamics using Modelfree. Together with a modelled structure, using modules 16-17 as a template, interpretation of the dynamics data has been possible.

In the case of module 2, the majority of the chemical exchange and fast timescale motion are occurring on the "front" or binding face of the module. Of the residues important in C4b-binding in module 2 - R64, N65 and Y94 - the latter two are undergoing slow R_{ex} motion. It could be the case, as was suggested previously for module 16 in site 2, that slow motions allow the specific contacts of hydrophobic residues on both ligand and receptor to consolidate binding. Chemical exchange motion is present in the N-terminal CD_2 and FG_2 loops of module 2. Whether this motion is independent of context and would be present if the construct contained module 1 is unknown.

On the back face of module 2, the DE loop and and the E and F β-strands make up a particularly rigid region. Despite the presence of a loop in this region, the only residue with an even slightly lowered S^2 is G98, which has a value of 0.77. Residues K87 and G98 are the only residues in this stretch which are undergoing chemical exchange and as they have similar R_{ex} values (3.89 $\pm$ 0.41 and 3.28 $\pm$ 0.26 s^{-1}, respectively) there may be a concerted hinge motion of the internally rigid loop between them.

While the F_3 strand in module 3 is rigid, the FG loop section of module 3 (which is close in the Modeller models to the module 2 DE_2 loop) is flexible. Three of the residues (N177, D179 and V181) could not even be assigned in the ^{15}N,^{1}H-HSQC, and a further two (D178 and Q180) had extremely weak ^{15}N,^{1}H-HSQC peaks and no data

in most 3D experiments. The flexibility of the FG_3 loop could contribute to function.

The hypervariable loop section of module 3 contains residues which are flexible on the fast timescale. The large DE_3 loop also has lowered S^2 values and appears flexible on the fast timescale.

The linker region itself appears to have a much higher rigidity than expected. Both the residues for which S^2 values were available showed low fast timescale flexibility (D121 0.80, R122 0.90) and none of the residues displayed slow timescale chemical exchange motion. As residue I123 could not be fitted, intermediate μs-ms motion may be occurring around this residue.

Chapter 6

Analysis and discussion

Prior to the work carried out on CR1 during this project, only five CCP modules had been analysed using Modelfree. These are VCP modules 2, 3 and 4 (as 2-3 and 3-4 module pairs) [15], isolated MCP module 1 [106] and isolated $GABA_B$ receptor module 2 (personal communication, Dr. Stan Blein, University of Edinburgh). The completion of the Modelfree analysis of CR1~16 and CR1~2-3 therefore brings the total number of CCP modules studied to eight. The CR1~2-3 pair provides the third Modelfree analysis of an intermodular junction for this module type.

6.1 Comparison of CCP module pairs CR1~2-3 and VCP~2-3

As both the CR1~2-3 module pair and the VCP~2-3 module pair represent the second and third CCP modules in a functional site, a comparison is of interest. Modules 2 and 3 of CR1 share 44% and 30% sequence identity, respectively, with modules 2 and 3 of VCP. Note that module 1 of CR1 shares only 14% identity with VCP~1. The alignment shown in Figure 6.1 was used to identify equivalent residues for comparison [59].

VCP is a four CCP module protein expressed by the *Vaccinia* virus to inhibit complement within the host that it has invaded. It has similar functions to CR1, Like CR1, VCP can bind C3b and C4b and has cofactor activity for the factor I mediated cleavage of both these ligands [118, 66]. Unlike CR1, VCP also binds heparin and has been re-

ported to use a positively charged Lys/Arg-X-Lys/Arg amino acid motif to do so [129].

The VCP construct was cloned and purified by Dr. Colin Henderson (University of Edinburgh). A 2.5 mM solution of VCP˜2-3 in 10 mM sodium phosphate buffer at pH 6.0 was used for NMR spectroscopy. The structure calculation based on NMR restraints was completed by Dr. Henderson [41]. NMR spectra were acquired and the relaxation data analysis was completed by Dr. Krystyna Bromek-Burnside (University of Edinburgh) [15]. Subsequently, the crystal structure of intact VCP was solved by X-ray diffraction [97].

The NMR-derived structure of VCP˜2-3 [41] and the modelled structure of CR1˜2-3 are compared in Figure 6.2. The structures are similar, particularly in the case of the module 2s, which have a backbone atom RMSD of < 1 Å (for residues between the first and last cysteine, inclusive). This is to be expected given the high level of sequence similarity. Module 3 of each protein shares less sequence identity with its counterpart and much less structural similarity (> 3 Å backbone RMSD for residues between the first and last cysteine, inclusive). For reasons discussed previously, the intermodular orientation within the CR1˜2-3 model cannot be regarded as an accurate representation. The intermodular orientation of VCP˜2-3 in solution is not known either, in this case due to the lack of observable intermodular NOEs. The crystal structure of VCP [97], on the other hand (not shown), is consistent with a fixed orientation at the 2-3 junction.

```
            63                              90              104       +      120
CR1_2  EAEAKSCRNPPDPVNGMVHVIKGIQFGSQIKYSCTKGYRLIGSSSATCIISGD TVIWDTETPICDR
VCP_2  EFIKRRCPSPRDIDNGQLD IGGVDFGSSITYSCNSGYHLIGESKSYCELGSTGSMVWNPEAPICES
             9                              35              49       +      66

          125                             154                 174       +      191
CR1_3  IPCGLPPTITNGDFISTNRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSNDDQVGIWSGPAPQCI
VCP_3  VKCQSPPSISNGRHN GYEDFYTDGSVVTYSCNSG       YSLIGNSGVLCSGG      EWS DPPTCQIV
          71                              99                  114       +      125
```

Figure 6.1: Sequence alignment of CR1~2-3 and VCP~2-3. Cysteine residues are numbered. Invariant tryptophan residues are marked with a plus. Gaps show insertions/deletions. VCP residues are numbered from the start of module 2, as in [40].

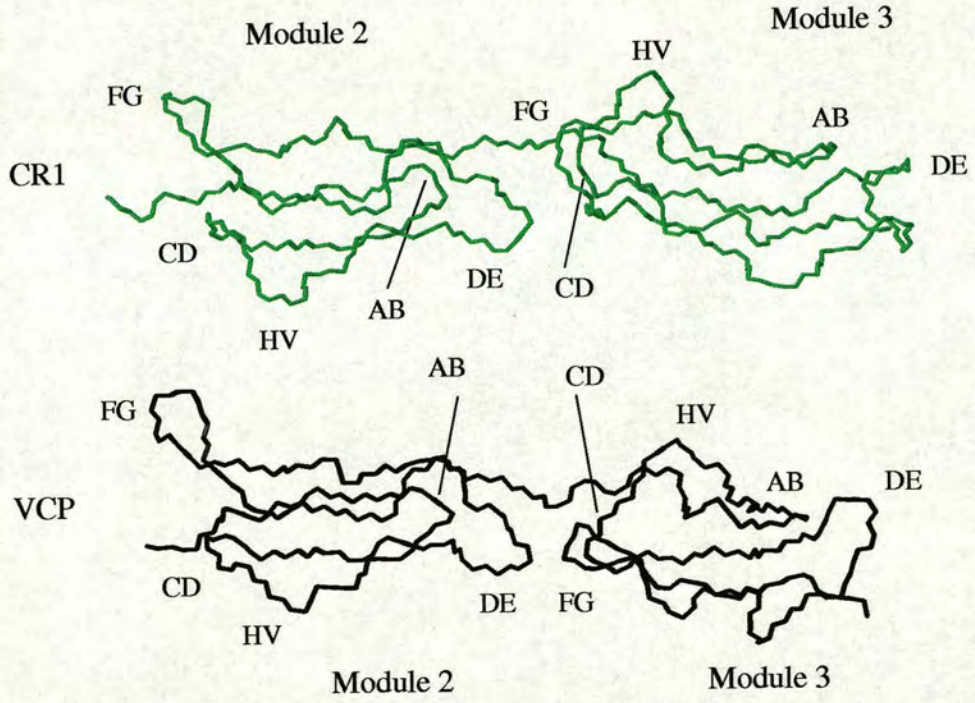


Figure 6.2: Model of CR1~2-3 (shown in green) and structure of VCP~2-3 (shown in black). The structures of module 2 have been aligned to show the similarity between them. In the case of the module 3s, there is much less structural similarity. Loops are labelled using the CCP module convention (section 1.3.2). HV is the hypervariable loop.

6.1.1 Comparison of S^2 values of CR1˜2-3 and VCP˜2-3

Figure 6.3 shows a plot comparing the S^2 values for these two module pairs. While CR1˜2^3 and CR1˜^{2}3 did not have statistically different mean S^2 values (0.78 ± 0.14 and 0.74 ± 0.11, respectively, see section 5.5.2), the two modules of VCP˜2-3 did have significantly different mean S^2 values - 0.73 ± 0.08 for VCP˜2^3 and 0.65 ± 0.09 for VCP˜^{2}3.

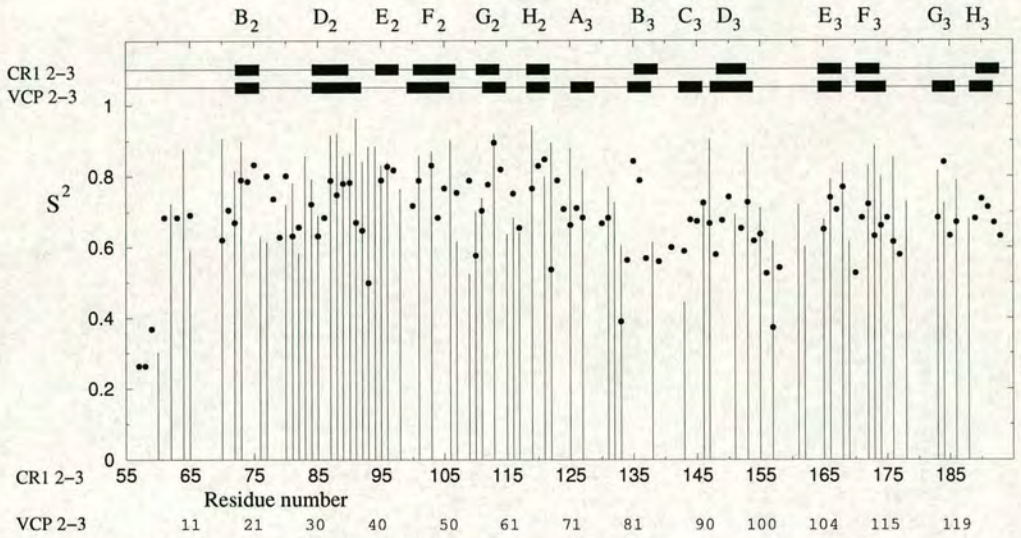


Figure 6.3: S^2 values for CR1, modules 2-3 (shown as columns) and VCP, modules 2-3 (shown as black circles) with aligned sequences. VCP residue numbers can be determined from the above key that uses the residue numbering system for VCP taken from [40]. β-strands are shown as black rectangles for CR1˜2-3 (top row) and VCP˜2-3 (bottom row). Strands are labelled using the CCP module convention.

Distinct differences between CR1˜2^3 and VCP ˜2^3 are mainly restricted to loop regions. The region of CR1˜2^3 corresponding to strands D_2 and E_2 and the DE_2 loop residues shows a consistently high S^2 value (0.83−0.96). It is conceivable that the presence of CR1˜^{2}3 provides some stability on the fast timescale in the DE_2 loop (Figure 6.2). In the case of VCP˜2^3, however, three of the residues in the DE region show lowered S^2 (N36 0.67, S37 0.65, G38 0.50) and a further residue (Y39) could not be fitted. The VCP˜2^3 D_2 strand also shows lower S^2 values than its CR1 counterpart. So despite the presence of VCP module 3, this region of module 2^3 retains a high degree of fast timescale flexibility.

In the case of the FG_2 loop in each module 2, CR1~2^3 shows a greater degree of flexibility compared with VCP~2^3. CR1 residues S107 and D109 have S^2 values of 0.62 and 0.53, whereas corresponding VCP residues G52 and T54 have S^2 values of 0.75 and 0.79. Residues G108 and S53 in CR1 and VCP respectively were not assigned, probably due to exchange with the solvent or broadening due to motion. The FG_2 loop is in the N-terminal half of each module and has the potential to be interacting with module 1 in the intact proteins.

In the case of CR1~2^3, the C-terminal half of the structure (containing the more rigid D_2, E_2 and F_2 region, as shown in section 5.5.2) is the less flexible compared to the N-terminal half. However, in VCP~2^3, with the flexible AB_2 loop and DE_2 region, both halves of the structure contain flexible loops.

Comparing limited data available for the linker regions highlights another possible difference. The 2-3 linker contains fewer "bulky" residues in VCP than in CR1 - Glu-Ser-Val-Lys in VCP~2-3 compared with Asp-Arg-Ile-Pro in CR1~2-3. S^2 values for the available CR1~2-3 linker residues are 0.80 (Asp) and 0.90 (Arg). (Residue I123 could not be fitted[1] and there are obviously no ^{15}N relaxation data for residue P124). In VCP~2-3, the S^2 values are 0.85 (Glu), 0.54 (Ser), 0.79 (Val) and 0.71 (Lys). The VCP values are therefore somewhat lower and the smallest residue (S68 0.54, c.f. R122 0.90) has a highly flexible nature on a fast timescale.

With regard to module 3 of each pair, differences are more difficult to identify. This is due to the reduced amount of data available both for CR1~^{2}3 and VCP~^{2}3.

It may be observed, overall, that in both modules the β-strands in CR1~2-3 are significantly more rigid than those in VCP~2-3. In CR1~2-3 the mean S^2 of residues predicted to be in strands in the model was 0.84 ± 0.08, whereas in VCP~2-3 the mean for strand residues was 0.73 ± 0.08.

---

[1] This could imply either that intermediate timescale motion which Modelfree cannot account for is occurring or that the anisotropy of diffusion is affecting this residue and preventing a fit.

6.1.2 Comparison of chemical exchange in CR1~2-3 and VCP~2-3

According to the analyses[2], VCP~2-3 contained a higher number of residues that appear to be in slow timescale motion: 31 compared with the 21 in CR1~2-3. However, data for 11 of the 31 VCP residues were fitted to small R_{ex} values which are lower in magnitude than the associated error. These residues therefore cannot be said, with certainty, to be undergoing chemical exchange and will be excluded from the comparison below. In both pairs, module 2^3 contained more R_{ex}-fitted residues than module ^{23}. In VCP there were thirteen in module 2^3 compared with seven in module ^{23}. For CR1, there were thirteen in module 2^3 and only eight in module ^{23}. The differences in number of chemically exchanging residues between modules 2^3 and ^{23} in both VCP and CR1 can mainly be accounted for by the regions encompassing the hypervariable and the CD loops (see Figure 6.2 for position of loops within the 3D-structure). In both module 2s, these adjoining regions each have a high number of residues fitted with an R_{ex} term. In VCP~2^3, there are five residues in this area (I22, G24, V25, F27 and S29) that required R_{ex} terms, while the number in CR1~2^3 was six residues (V76, I80, Q81, F82, G83 and S84).

Overall, VCP~2-3 contains a higher proportion of R_{ex} residues within its β-strands than does CR1~2-3. In VCP~2^3, six of the 13 are located within strands. In VCP~^{23}, six of the seven are within strands. This contrasts with CR1, which has only two out of 13 R_{ex} residues within strands in module 2^3 and only one of eight in module ^{23}. Therefore β-strands in VCP~2-3, as well as being more flexible on a fast timescale than those in CR1~2-3, are also undergoing slow timescale motion that is not occurring in the β-strands of CR1.

As previously stated (section 5.5.6), CR1~2^3 has all but two (K87 and G98) of its residues that undergo chemical exchange confined to the more flexible "binding face". The distribution of chemical exchange in VCP~2^3 follows roughly the same pattern. The majority of R_{ex} residues are located in strands on the face equivalent to the binding face on CR1~2^3. There are three exceptions - Y33, L41 and I42.

---

[2] In the case of VCP, chemical exchange was measured directly using NMR spectroscopy [15]. The chemical exchange terms were supplied to the subsequent Modelfree fitting.

In both module 3s, the D_3 strand represents a region where there is a markedly different distribution of residues undergoing chemical exchange. CR1~2~3 has a single R_{ex} residue in this region - S138. In VCP~2~3 there are three residues - G92, V95 and Y97. Other than this strand, and the earlier observation that in general residues undergoing chemical exchange are more common within strands in VCP, there are no further obvious differences between R_{ex} distributions within the two versions of module 3. In each case, a small number of R_{ex} residues is distributed over both faces and found within both termini of the module.

As was seen in the comparison of fast timescale motion, the VCP~2-3 linker exhibits a difference in μs-ms timescale motion compared to the CR1~2-3 linker. Within CR1~2-3, D121 and R122 were not fitted with an R_{ex} term. I123, which could not be fitted by Modelfree, had an elevated T_2 value and so is unlikely to be undergoing chemical exchange motion. ^{15}N relaxation data cannot, of course, be obtained for P124. In VCP~2-3, two linker residues, E67 and K70, were fitted with chemical exchange terms. This indicates that as well as being flexible on the fast ps-ns timescale, motion at the VCP~2-3 linker is also occurring on the slow μs-ms timescale. While the 2-3 linker in CR1 is apparently flexible on neither timescale, in VCP it is flexible on both.

6.1.3 Summary of CR1~2-3 and VCP~2-3 comparison

A similarity between the dynamics of VCP and CR1 site 2 (as opposed to site 1), which may point to a common feature underlying their similar functions, is the apparent flexibility in the vicinity of both the VCP~2-3 and the CR1~16-17 junctions. A comparison of sites 1 and 2 in CR1 is described in section 6.2 below.

Perhaps the most interesting difference between the CR1 and VCP 2-3 module pairs lies within the linker itself. The VCP~2-3 linker shows evidence of a reasonably high flexibility on the fast timescale as well as motion occurring on the slow timescale. The CR1~2-3 linker appears to show neither. Another difference is that while the β-strands are the most rigid sections of VCP on a fast timescale (as in CR1), they are

also undergoing slow timescale motion, again in contrast to CR1 site 1.

6.2 Comparison of CR1 module pairs 2-3 and 16-17

Relaxation data is now available for the second and third modules of both CR1 functional sites 1 and 2 (copy 2). Therefore a comparison of their dynamics is possible. The Modelfree analysis of the module pair 2-3 was described in the previous section, 5.5.1. In addition to the studies of single module 16 described in section 4.5.1, work has also been done in Edinburgh on the dynamics of module pair 16-17 [128]. In the 16-17 work, a construct consisting of native residues K961-I1092 was made by collaborators Dr. Malgorzata Krych and Prof. John Atkinson (University of Washington Medical School, St. Louis), using the methods described in section 3.1.1. ^{15}N relaxation data (T_1, T_2 and heteronuclear NOE) were collected on a 1 mM sample on a 600 MHz Varian INOVA NMR spectrometer. The data acquisition and handling were completed by Dr. Krystyna Bromek-Burnside (University of Edinburgh). Although the Modelfree fitting of the ^{15}N relaxation data of modules 16-17 could not be completed (see section 1.4.3), the raw relaxation data can still be used for comparison. The raw relaxation data from the experiments using the single module 16 construct will not be included in this comparison, although the results of the Modelfree analysis for this module will be. Relaxation data for single module 16 and the 16-17 pair have already been compared in section 4.4.4.

Knowledge of differences in dynamics can contribute to further understanding of the factors that lead to proteins with high sequence and structural similarity, as in this case, having differing binding functions. As both double module pairs contain an intact intermodular junction, the comparison of the linker residues and nearby loops is of especial interest.

6.2.1 T_1 and T_2 values of CR1~2-3 and 16-17

Figure 6.4 shows the T_1 and T_2 data on a residue-by-residue basis for both the 2-3 and 16-17 module pairs. Table 6.1 summarises the mean values of T_1 and T_2 for each module pair, and for the constituent modules. Figure 6.5 shows a plot of the

CR1~2-3 vs. CR1~16-17 T_1 and T_2 values for equivalent residues. The residues used for the comparison were only those from the native sequence, i.e. signal peptide sections E57-A60 and E957-A960 were excluded. For the calculation of mean values for the constituent modules, both the signal peptide sequences and the linker residues, D121-P124 or Q1021-P1024, were excluded. Thus, 77 residues were included in the calculation for modules 2-3, and 107 residues for 16-17. For modules 2^3, ^{2}3, 16^{17}, and ^{16}17, 39, 35, 51 and 53 residues, respectively, were included.

Site 1 modules		**Site 2 modules**	
Module(s)	**Mean T_1 (ms)**	**Module(s)**	**Mean T_1 (ms)**
CR1~2-3	799 ± 132	CR1~16-17	709 ± 63
CR1~2^3	795 ± 126	CR1~16^{17}	700 ± 41
CR1~^{2}3	811 ± 141	CR1~^{16}17	717 ± 80
Module(s)	**Mean T_2 (ms)**	**Module(s)**	**Mean T_2 (ms)**
CR1~2-3	98 ± 17	CR1~16-17	84 ± 14
CR1~2^3	92 ± 19	CR1~16^{17}	85 ± 15
CR1~^{2}3	103 ± 11	CR1~^{16}17	83 ± 14

Table 6.1: Comparison of the mean T_1 and T_2 for CCP modules present in module pairs CR1~2-3 and CR1~16-17.

Looking at Table 6.1 and Figures 6.4 and 6.5, the mean T_1 and T_2 values suggest that there are significant differences between the ^{15}N relaxation properties of the two module pairs. The difference between the pairs could be due to an overall difference in macromolecular motion. Variation in junction flexibility could be responsible for this.

Modules 2-3 have significantly higher T_1 values than modules 16-17, whether considered as a module pair or at the level of the constituent modules. The site 1 modules also have a wider range of T_1 values; on the other hand module 16^{17} shows the most limited range of T_1 values out of the four modules. This is consistent with the relatively uniform relaxation data, and Modelfree analysis, for the single CCP 16 construct, as detailed in section 4.5.1. Figure 6.5 shows that while the T_1 values of CR1~2-3 are generally higher than the T_1 values of CR1~16-17 (as was evident in Table 6.1), there is in fact little correlation between the values for equivalent residues. Neither module

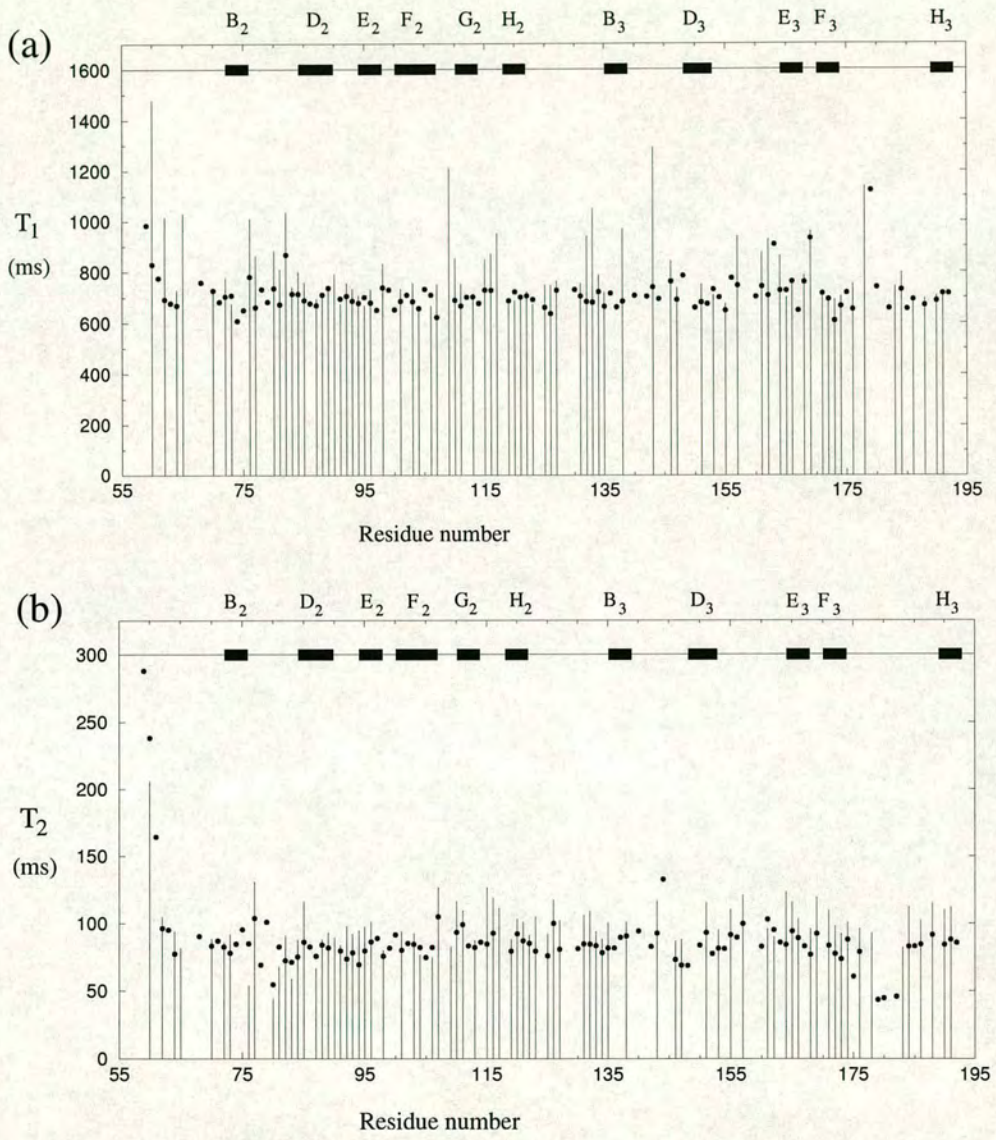


Figure 6.4: T_1 (Figure a) and T_2 (Figure b) data for CR1~2-3 (shown as black columns) and CR1~16-17 (shown as black circles). β-strands corresponding to the models of modules 2-3 are shown as black rectangles. Strand B_3 in module $^{2}3$ does not occur in module $^{16}17$. Residue numbers for modules 16-17 can be obtained by adding 900 to the stated values.

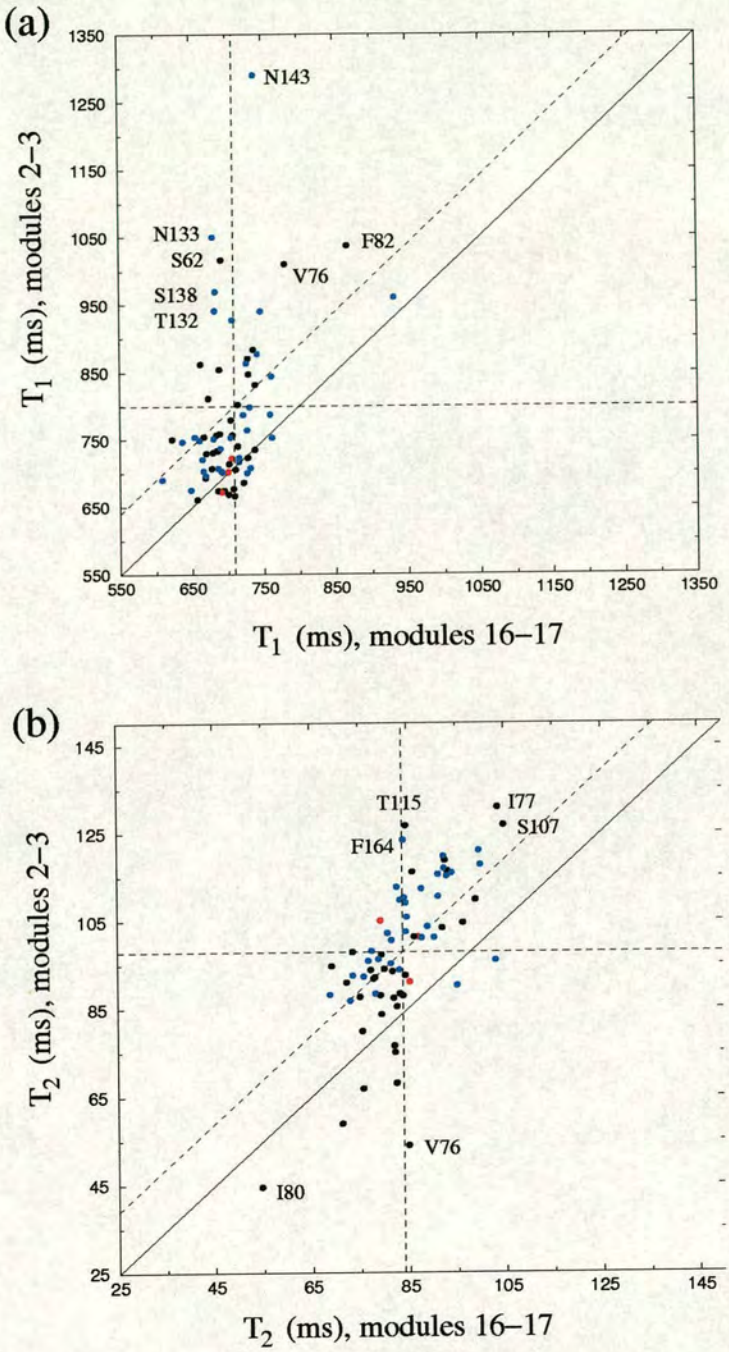


Figure 6.5: T_1 and T_2 times for equivalent residues in CR1 pairs 2-3 and 16-17 plotted against each other. Black points show module 2/16 residues. Blue points show module 3/17 residues. Red points show linker residues (D121/Q1021-P124/P1024). The orthogonal dashed lines represent the mean relaxation times of each module pair, with the dashed diagonal showing the line of perfect correlation allowing for the differences in mean values. Selected outliers (lying far from the dashed diagonal) are labelled, using the 2-3 residue numbering. There is little correlation between the T_1 relaxation times of CR1 pairs 2-3 and 16-17. There is a systematic shift to higher T_2 times for module $^{2}3$ residues compared with module $^{16}17$ residues. Overall, there is a stronger correlation between the 2-3/16-17 T_2 times than the T_1 times.

2/16 residues (black points) or module 3/17 residues (blue points) show a strong correlation - the respective coefficients are 0.55 ± 0.35 and 0.37 ± 0.40. Data points well away from the diagonal represent pairs of residues with distinctly different T_1 values in the two fragments. These are identified on the plot.

Focussing on T_2 values in Table 6.1, module $^{2}3$ has a significantly higher mean T_2 value than module $^{16}17$, although module 2^3 values are not significantly higher, on average, than module 16^{17} values. As it is modules $^{2}3$ and $^{16}17$ that are so similar in sequence (96% identical residues), this result is unexpected. From the T_2 plot in Figure 6.5, the correlation between module 2/16 T_2 values (black points) is stronger than that for module 3/17 residues (blue points) - 0.71 ± 0.25 compared to 0.60 ± 0.19, respectively. However, there is a systematic shift to higher T_2 values in module $^{2}3$ compared with module $^{16}17$. While the correlation coefficient may be higher for modules 2/16 compared with modules 3/17, there is a much lower standard deviation in the 3/17 T_2 plot, compared with the 2/16 T_2 plot - 10.7 and 19.9 respectively. Therefore, the module 3/17 data points show less dispersion than those for module 2/16. This is evident from the fewer outliers (from the dashed diagonal) in the module 3/17 data set.

Overall, the T_2 values show better correlation between the two fragments than the T_1 values. An analysis of the T_1 and T_2 values in order to identify and compare possible chemical exchange within the two pairs is detailed in section 6.2.5 below.

T_1 and T_2 value interpretation in terms of backbone dynamics is not as straightforward as with heteronuclear NOEs, which are described next.

6.2.2 Heteronuclear NOE values of CR1~2-3 and 16-17

Figure 6.6 shows the heteronuclear NOE (hetNOE) data for both module pairs. Figure 6.7 shows a plot of the CR1~2-3 hetNOE values against the CR1~16-17 hetNOE data for equivalent residues. There is a stronger correlation between the hetNOE values of modules 3/17 (coefficient 0.63 ± 0.26) than modules 2/16 (coefficient 0.48 ± 0.28), as would be expected due to the similarity of modules 3 and 17. These correlations are

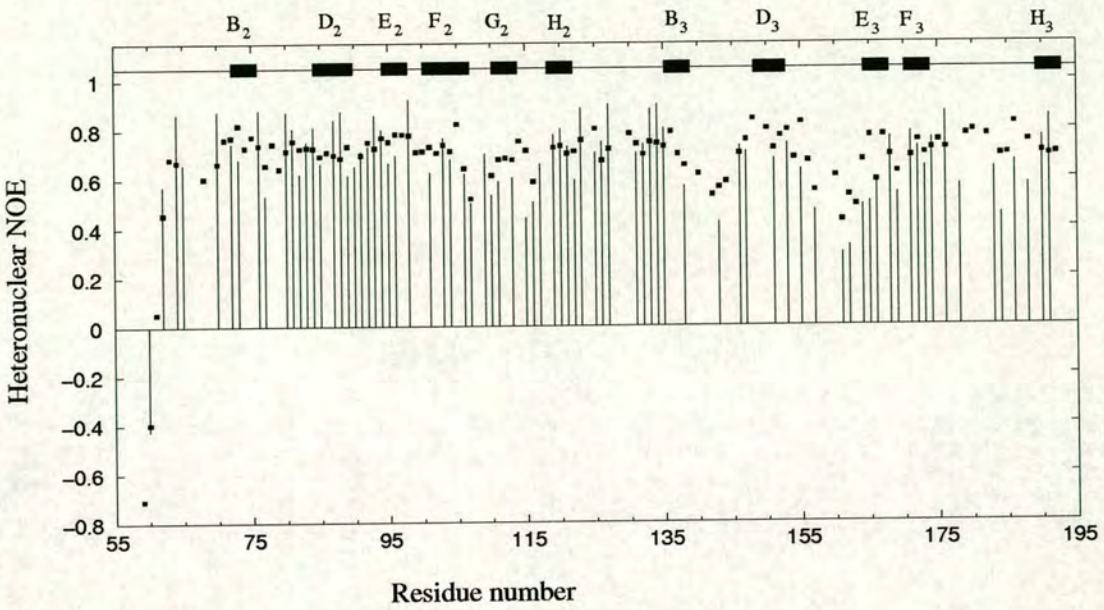


Figure 6.6: Heteronuclear NOE relaxation data for CR1˜2-3 (shown as black columns) and module pair CR1˜16-17 (shown as black squares). β-strands corresponding to the models of modules 2-3 are shown as black rectangles. Strand B_3 does not occur in module $^{16}17$. Residue numbers for modules 16-17 can be obtained by adding 900 to the stated values.

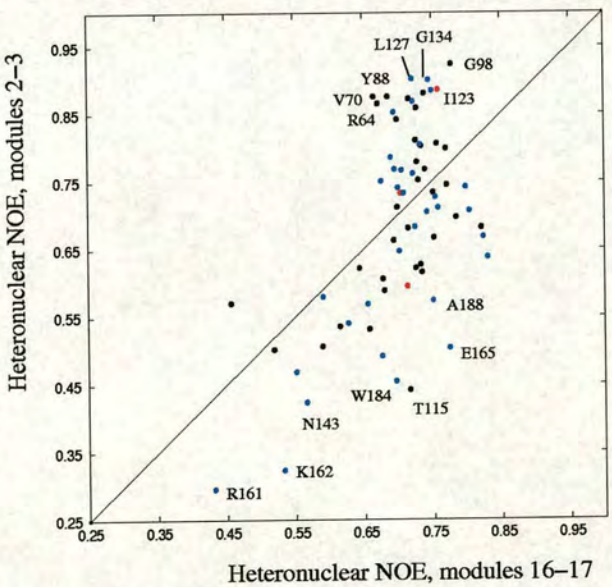


Figure 6.7: Heteronuclear NOEs of equivalent residues within CR1˜2-3 and 16-17 plotted against each other. Black points show module 2/16 residues. Blue points show module 3/17 residues. Red points show linker residues (D121/Q1021-P124/P1024). Selected outliers are labelled, using the 2-3 residue numbering. CR1˜2-3 has a larger number of residues with very high heteronuclear NOE, although the mean values for the two pairs are not significantly different.

Site 1 modules		Site 2 modules	
Module(s)	**Mean NOE**	**Module(s)**	**Mean NOE**
CR1~2-3	0.69 ± 0.14	CR1~16-17	0.70 ± 0.10
CR1~2^3	0.71 ± 0.12	CR1~16^{17}	0.70 ± 0.11
CR1~^{2}3	0.66 ± 0.16	CR1~^{16}17	0.70 ± 0.09

Table 6.2: Comparison of the mean heteronuclear NOE values for CCP modules present in double module pairs 2-3 and 16-17. For comparison, the mean hetNOE value for the lone module 16 construct was 0.66 ± 0.10.

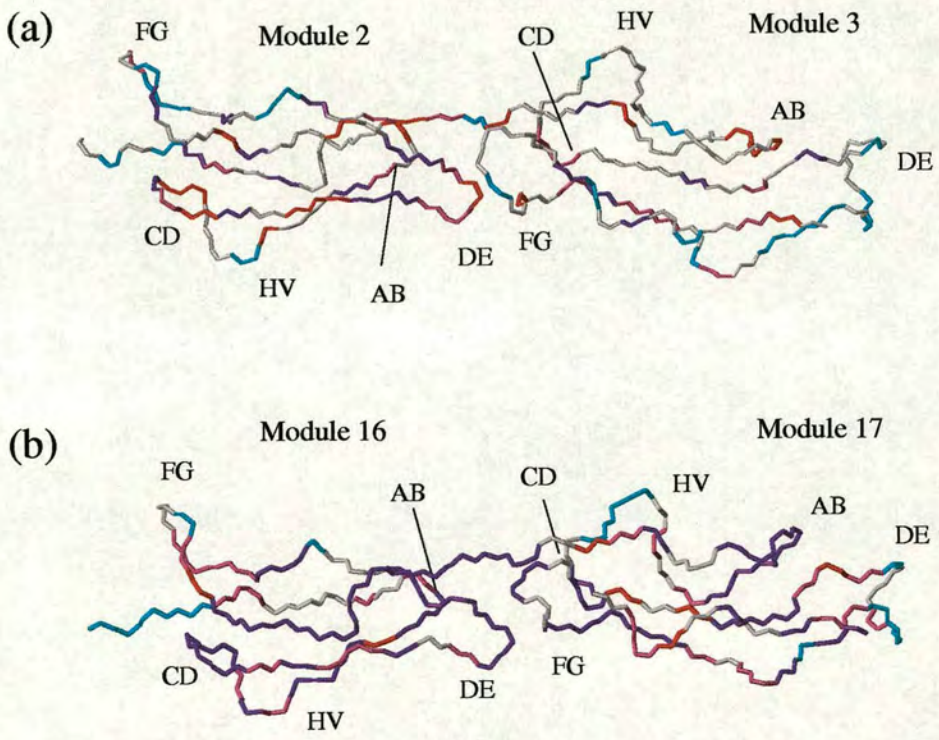


Figure 6.8: Heteronuclear NOE relaxation data for CR1 module pairs 2-3 and 16-17 mapped on the structures as colours. Red depicts low hetNOE (hetNOE < 0.6); pink depicts moderately low hetNOE (0.6 ≤ hetNOE < 0.7); purple depicts moderately high hetNOE (0.7 ≤ hetNOE < 0.8); cyan depicts high hetNOE (hetNOE ≥ 0.8); white depicts residues for which no hetNOE is available. Loops regions are labelled using the CCP module convention. HV is the hypervariable loop.

both poorer than would be anticipated. The figure also shows that the 2-3 pair have a higher proportion of residues with highly elevated hetNOE located in both modules 2 and 3 compared with modules 16 and 17. As with the relaxation times, the hetNOE values indicates significant dynamic differences between modules 2-3 and 16-17, despite the very high sequence similarity.

Table 6.2 shows the mean hetNOE values for each module pair and each constituent CCP module. None of the mean hetNOE values for the constituent modules are significantly different. Figure 6.8 shows the hetNOE values mapped on to the CR1˜2-3 model, and the CR1˜16-17 structure, in colour. One distinct difference between modules 2-3 and 16-17 evident in Figure 6.8 is the FG loop region in modules 2^3 and 16^{17}. The F_{16} and G_{16} strands in module 16 show relatively low hetNOE while the equivalent residues in module 2^3 do not.

All four modules have similar mean hetNOE values. There is a difference in the ranges of values measured. Modules 2^3 and $^2 3$ had ranges of 0.44−0.92 and 0.30−0.90 respectively (excluding signal peptide residue A60, which had an NOE value of −0.42). Modules 16^{17} and $^{16}17$ had ranges of 0.46−0.82 and 0.43−0.84 respectively. (Again excluding residues E959 and A960 which were −0.71 and −0.40, but also K961 (0.05) as there was no data available for K61.) The larger range of hetNOE in 2-3 compared with 16-17 could point to a wider range of flexibilities within the site 1 pair.

The comparable residues (i.e. those for which there are relaxation data from both the 2-3 and 16-17 pairs) which differ most in hetNOE between modules 2^3 and 16^{17} (see Figure 6.7) are R64/K964, V70/V970, V76/V976, I80/I980, K87/T987, Y88/Y988, G98/G998 and T115/T1015. In all cases, except T115/T1015, the NOE in module 2^3 is greater then that in module 16^{17}. The residues which differ most between modules $^2 3$ and $^{16}17$ are L127/L1027, G134/G1034, N155/N1055, K162/K1062, F164/F1064, E165/E1065, W184/W1084, I186/I1086, A188/A1088 and C191/C1091. In all but L127/L1027, G134/G1034 and I191/I1091 the NOE is higher in module $^{16}17$. In the linker, residues R122/R1022 and I123/I1023 differ in their hetNOE values. In the case of R122/R1022, the site 2 residue (R1022) has the higher value. For I123/I1023, the

site 1 residue (I123) has the higher value. The positions of these residues are shown on the structure of modules 16-17 in Figure 6.9. Some of them are also labelled in Figure 6.7. As the mean hetNOE values are similar for all four modules, these residues which differ between sites 1 and 2 are likely to be undergoing significantly different motions.

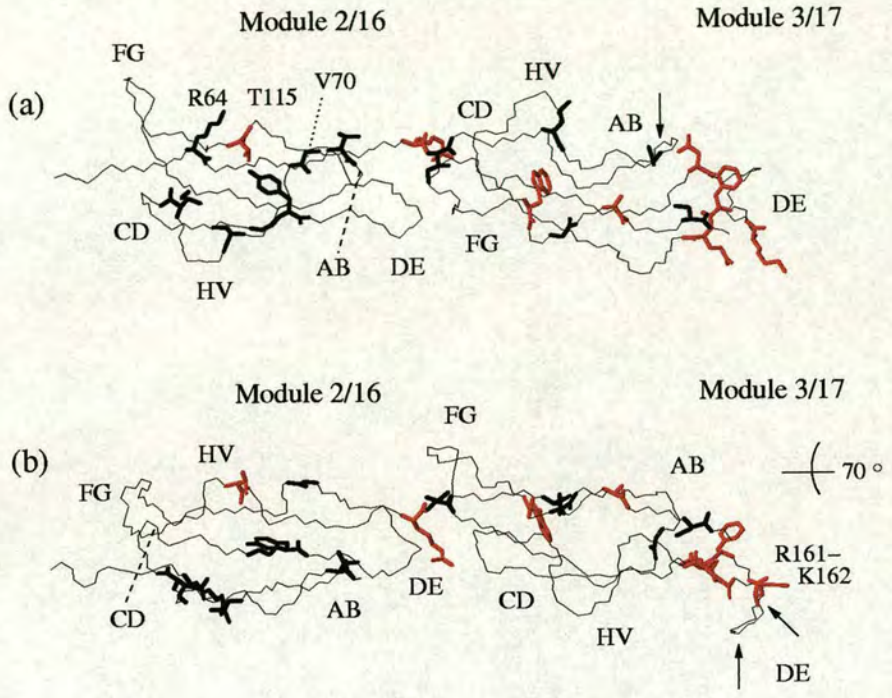


Figure 6.9: Residues which differ in heteronuclear NOE between modules 2-3 and 16-17 by 0.1 or more. The backbone and sidechain heavy atoms of residues with higher hetNOE in modules 2-3 (shown in black) and modules 16-17 are highlighted (shown in red). The three residues which differ between modules 3 and 17 are indicated by arrows. Loops are labelled using the CCP module convention. HV is the hypervariable loop. The most extreme outliers are labelled for orientation.

Comparing modules 2^3 and 16^{17}, several residues have higher hetNOE in module 2^3 on the non-binding face, implying that this region may be more rigid than in module 16. From Figure 6.9(b), it can be seen that the residues which vary the most between modules 3 and 17 are mostly localised on the "back" of the module, i.e. the side which does not contain the residues implicated in site 2 function.

The intermodular junction and surrounding loops contained just two residues which varied in hetNOE between CR1~2-3 and CR1~16-17 both of which are in the linker

itself: R122/R1022 and I123/I1023. While R122 showed a slightly depressed NOE (0.60), it was fitted to a very high S^2 by Modelfree (0.90). R1022 was also fitted to a high S^2 (0.87) in lone module 16, despite it being on the C-terminus following the final cysteine residue. It appears as though both R122 and R1022 are rigid in terms of the fast timescale.

I1023 shows a hetNOE value similar to the other residues in the 16-17 linker. I123 shows a highly elevated hetNOE and this pair is one of the principal outliers in Figure 6.7. While the is no Modelfree analysis for residue I123, from the hetNOE data alone it appears as if this residue is reasonably rigid on the fast ps-ns timescale.

6.2.3 Comparison of models fitted by Modelfree for residues in CR1 module 2^3 and module 16

A breakdown of the results of the fitting by Modelfree for comparable modules is shown in Table 6.3. From the distribution of fitted models it can be seen that module 2^3 is more varied and diverse in terms of internal motions than module 16, which had the majority of its residues fitted by the simplest case of S^2 alone.

Fitted model	**Residues in module 2^3**	**Residues in module 16**
1 (S^2 only)	13	34
2 (S^2, t_e)	3	1
3 (S^2, R_{ex})	9	7
4 (S^2, t_e, R_{ex})	5	4
5 (S_s^2, S_f^2, t_f)	9	0

Table 6.3: Comparison of model selection by Modelfree for module 2^3 and module 16

Of the 37 residues for which there was relaxation data available for both modules 2^3 and 16, 29 were fitted by Modelfree in both modules. The sidechain data for W113/W1013 was also successfully fitted in both modules. These residues are listed in Table 6.4 along with the values fitted to each parameter.

Eighteen of the 29 residue-pairs were of identical amino acid type. Of these, eight were

Residue	Model	S^2	t_e (ps)	R_{ex} (s^{-1})
S62/S962	1/1	0.72/0.72		
R64/K964	1/1	0.88/0.83		
N65/T965	4/4	0.59/0.72	17/54	5.1/2.3
V70/V970	1/1	0.91/0.78		
G72/G972	3/1	0.82/0.85		1.8/-
V76/V976	3/1	0.63/0.79		11.0/-
I80/I980	3/3	0.72/0.82		13.8/4.3
Q81/Q981	3/1	0.78/0.86		5.2/-
F82/V982	4/3	0.58/0.78	20/-	3.9/1.1
G83/G983	3/3	0.86/0.86		6.6/0.5
S84/S984	3/1	0.79/0.86		1.8/-
Q85/R985	5/1	0.69*/0.82	1055/-	
K87/T987	3/3	0.92/0.89		3.9/0.7
Y88/Y988	1/1	0.92/0.82		
K92/T992	1/1	0.84/0.82		
Y94/H994	1/1	0.89/0.83		
R95/R995	5/1	0.83*/0.82	833/-	
L96/L996	1/1	0.84/0.82		
G98/G998	3/1	0.77/0.81		3.3/-
S101/S1001	2/1	0.86/0.81	63/-	
T103/E1003	1/1	0.87/0.85		
I106/L1006	4/1	0.90/0.80	193/-	2.3/-
S107/S1007	5/4	0.62*/0.75	876/105	0.8/-
T110/T1010	5/1	0.70*/0.77	48/-	
W113s/W1013s	1/1	0.75/0.82		
E116/K1016	5/1	0.68*/0.79	51/-	
I119/I1019	1/1	0.94/0.86		
D121/Q1021	5/1	0.80*/0.80	1165/-	
R122/R1022	1/3	0.90/0.87		-/1.1

Table 6.4: Comparable residues in module 2^3 and in module 16. Modelfree model fittings and parameter values are given. S^2 marked with * show the composite $S_s^2 \times S_f^2$.

fitted by Modelfree to the same model. The other ten required a different model, and in all cases bar R122/R1022 the module 2^3 residue required the more complex model to fit the data. Again, this highlights the greater variety of dynamics within the residues of module 2^3. For the residues which were non-identical between the two module, six were fitted using the same model. Again, in every case where the fitting was not the same, the module 2^3 residue required the more complex model.

6.2.4 Comparison of S^2 values between CR1 module 2^3 and module 16

The data comparing the S^2 values for the two analogous modules is shown in Figure 6.10. The mean S^2 values for each module are listed in Table 6.5. From the S^2 data it can be seen that CR1~2^3 has a much larger range of flexibility in its residues than CR1~16. The mean S^2 for all three categories in Table 6.5 are not significantly different between modules 2^3 and 16. Figures 4.21 and 5.15 in sections 5.5.2 and 4.5.3 showed the S^2 values of modules 16 and 2^3, respectively, mapped on to the structures.

Notable differences (with an S^2 difference of 0.1 or more) include residues N65/T965*, V70/V970, V76/V976*, I80/I980*, F82/V982*, Q85/R985*, Y88/Y988, I106/L1006, S107/S1007* and E116/K1016. Residue pairs marked with asterisks have the higher S^2 value in module 16, otherwise they have the higher value in module 2^3. These residues are shown highlighted on the structure of modules 16-17 in Figure 6.11.

In the case of V70/V970, this residue position is located in the AB loop. However, this loop has no obvious contact with any part of the next module. The presence of the following module is therefore unlikely to be responsible for restricting the space available in which V70/V970 can move. This residue position is located close to the linker, however, so it may be the case that the structuring of the linker region plays a part in determining the motion which residues V70/V970 can undergo. As the 16-17 linker is believed to be flexible [128], the observation that V70 has a higher S^2 suggests that the area around the 2-3 junction could be more rigid. However, without an S^2 value for V970 in a construct of modules 16-17, this cannot be fully confirmed.

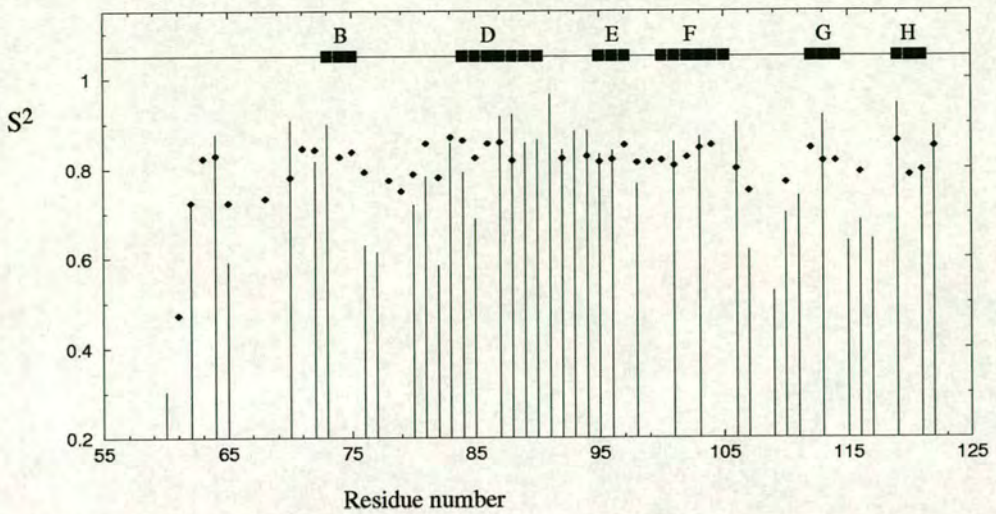


Figure 6.10: S^2 values of CR1 modules 2^3 and 16. Module 2^3 shown as black columns, module 16 shown as diamonds. Residue numbers for module 16 residues can be obtained by adding 900 to the stated values. β-strands are shown as black rectangles and labelled according to CCP module convention.

Attribute	**Module 2^3**	**Module 16**
Mean S^2	0.78 ± 0.14	0.81 ± 0.06
Mean β-strand residue S^2	0.86 ± 0.08	0.83 ± 0.02
Mean non-β-strand residue S^2	0.73 ± 0.13	0.78 ± 0.08

Table 6.5: Comparison of S^2 data for module 16 and module 2^3

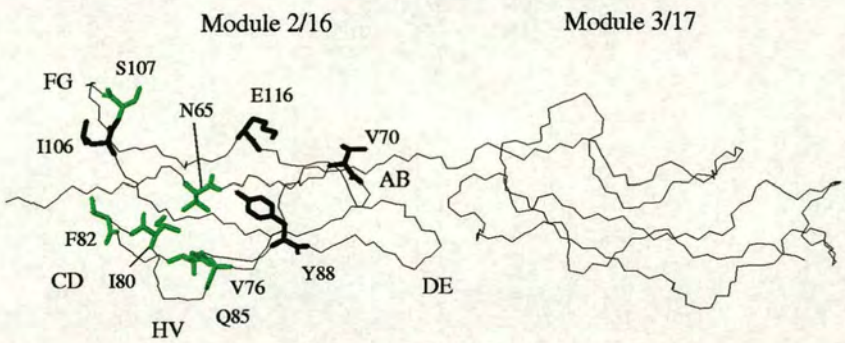


Figure 6.11: Residues which differ in S^2 between CR1 modules 2-3 and 16-17 by 0.1 or more. The backbone and sidechain heavy atoms of residues with higher S^2 in modules 2-3 (shown in black) or modules 16-17 are highlighted (shown in green). Loops are labelled using the CCP module convention. HV is the hypervariable loop.

For residues Y88/Y988, the module 2^3 residue again has the higher S^2. These residues are found within β-strand D in the model/structure (remote from neighbouring modules) and would be expected to be relatively rigid. When comparing the "back" face of the modules, module 2^3 is more rigid than module 16. In support of this statement, residues Y88/Y988 display evidence of this, as do all residues in this area for which there are data. The pairs K87/T987, S101/S1001 and I106/L1006 all have higher S^2 values in module 2^3. This shows that while both modules contain one flexible face and one rigid face, in the case of module 2^3 the results are exaggerated: the flexible face being more flexible and the rigid face being more rigid than the analogous regions in module 16.

Residues S107/S1007 are located on the FG loops. Both have S^2 values well below the mean for the module in which they are found, though S107 in module 2^3 has the lower value. Cases like S107/S1007, where module 16 has a higher S^2 than module 2^3, are often residues that are amongst the most flexible within their respective modules (e.g. V76/V976, F82/V982).

6.2.5 Comparison of chemical exchange results in CR1~2-3 and 16-17

Various criteria for deciding which residues in these two CCP module pairs are undergoing chemical exchange motion have been described so far. For modules 2-3, the Modelfree analysis is definitive and more reliable than any other method. For modules 16 and 17, there is the Modelfree analysis of the single module 16, which may not be entirely representative of the module when it is part of a larger fragment. There is also the application of the qualitative *Barbato et al.* criteria to modules 16-17 which was completed by Dr. Krystyna Bromek-Burnside (University of Edinburgh). The conclusions of these methods are summarised in Table 6.6, which firstly gives the comparison of modules 2 and 16 and then the comparison of modules 3 and 17. Figure 6.12 shows the residues apparently involved in chemical exchange on the model of modules 2-3 and the structure of modules 16-17.

Modules 2-3 show markedly less chemical exchange motion than modules 16-17. In

Module 2^3 (Modelfree)	**Module 16** (Modelfree)	**Module 16^{17}** (Barbato)
No fit	K961 (0.41)	
		K964
N65 (5.05)	T965 (2.31)	
G72 (1.83)		
		M973
V76 (11.0)		
No fit	T978 (1.97)	T978
I80 (13.8)	I980 (4.32)	I980
Q81 (5.23)		
F82 (3.90)	V982 (1.07)	
G83 (6.61)	G983 (0.47)	G983
S84 (1.83)		S984
No fit	I986 (0.66)	
K87 (3.89)	T987 (1.80)	T987
		T991
		T992
		G993
		H994
		R995
G98 (3.28)		
No fit	H999 (0.63)	
		S1001
I106 (2.33)		
	S1007 (0.81)	
D109 (5.72)	No fit	
T117 (1.17)	No fit	
		I1019
	R1022 (1.10)	
No fit		I1023

Module ^{2}3 (Modelfree)	**Module ^{16}17** (Barbato)
N133 (3.33)	
S138 (2.38)	
N143 (3.09)	
Y146 (2.59)	
	G1047
	Y1052
G157 (0.70)	
K162 (3.65)	
E169 (0.83)	
	I1072
	Y1073
	T1075
	S1076
	Q1080
	G1082
I183 (2.20)	

Table 6.6: Residues undergoing chemical exchange, as determined by Modelfree analysis (for modules 2-3 and 16) and *Barbato et al.* analysis of relaxation data (module 16-17). Note, residues in the N-terminal EAEA secretion signal peptide have not been included, even if they were fitted with an R_{ex} term. R_{ex} values fitted by Modelfree are shown in brackets, where appropriate.

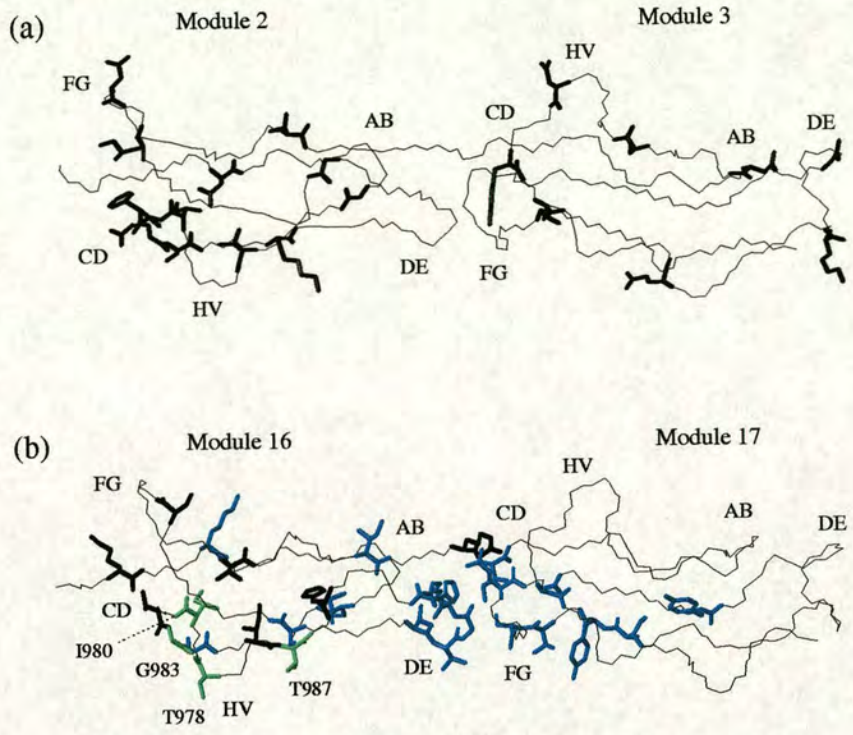


Figure 6.12: Residues apparently undergoing chemical exchange in modules 2-3 and 16-17. (a) R_{ex} residues identified by Modelfree in CR1~2-3 are shown on the model of modules 2-3 in black. (b) R_{ex} residues identified by Modelfree in lone module 16 are shown on the structure of modules 16-17 in black. R_{ex} residues identified by *Barbato et al.* method in CR1~16-17 are shown on the structure of modules 16-17 in pale blue. The residues in CR1~16-17 identified by both Modelfree and *Barbato et al.* are coloured pale green. These residues are labelled for orientation. Loops are labelled using the CCP module convention. HV is the hypervariable loop.

particular, the 16-17 junction appears to have a high number of residues undergoing slow timescale motion. Chemical exchange motion is almost entirely absent from the 2-3 junction.

Excluding residue A60, there are thirteen equivalent residues which were fitted by Modelfree to the same model in both modules 2^3 and 16. Of these, five required an R_{ex} term: N65/T965, I80/I980, F82/V982, G83/G983 and K87/T987. In every case the module 2^3 residue required a larger chemical exchange value to be fitted.

In module 2^3, residues V76, Q81, S84, G98 and I106 are involved in chemical exchange motion while the analogous residues in module 16 appear not to be. Conversely,

residues S1007 and R1022 are involved in chemical exchange in module 16, but their equivalents in module 2^3 are not. In the case of R1022, the difference may be related to the absence of module 17 in the construct. Residues I1019 and I1023 were not fitted to a R_{ex} term by Modelfree in the lone 16 case, but the *Barbato et al.* method highlights them as potential chemical exchange candidates in the 16-17 construct. One residue (I1023) is located within the linker and the other (I1019) is found very close to it.

The qualitative method provides further information. This method indicates that the DE_{16} loop region of module 16^{17} is also undergoing chemical exchange (residues T991-R995). This was not detected by Modelfree in the single module, but as this loop is at the 16-17 interface, it is likely that the double module data is more representative, and some or all of these residues are likely involved in chemical exchange.

In module $^{16}17$, the *Barbato et al.* method highlights strand F_{17} and the FG_{17} loop as being involved in chemical exchange. Such an observation was not made in module $^{2}3$, although several FG_3 loop residues were absent from the ^{15}N,^{1}H-HSQC so there appears to be some slow motion in this loop in both module pairs. However, the F_{17} strand contains chemically exchanging residues - I1072, Y1073 - (according to the *Barbato et al.* criteria), while the F_3 does not.

In module $^{2}3$, there are several residues in the hypervariable loop and CD_3 loop region undergoing chemical exchange, all with similar values. It is possible that concerted bending of the hypervariable loop and what would be the C strand is occurring, with two or more of residues N133, S138, N143 and Y146 acting as hinges. This mobility may explain the failure of residues in this area to conform to the definition of a β-strand (these residues form strand C in many other CCP modules). The absence of several $^{2}3$ FG_3 loop residues from the ^{15}N,^{1}H-HSQC suggests that this loop is also undergoing slow motion.

In summary, modules 16-17 show a far greater number of residues at the junction likely to be undergoing slow timescale motion, including residues in both modules 16^{17} and $^{16}17$ and perhaps the linker. In contrast, in modules 2-3 there are far fewer residues

undergoing chemical exchange at the junction, and apparently none in the 2-3 linker. The only evidence for slow timescale motion in the 2-3 junction region is in the loops of module $^{2}3$.

6.2.6 Linker residues in CR1˜2-3 and 16-17

The relaxation data and Modelfree results, where available, are collated for the linker region residues in Figure 6.13.

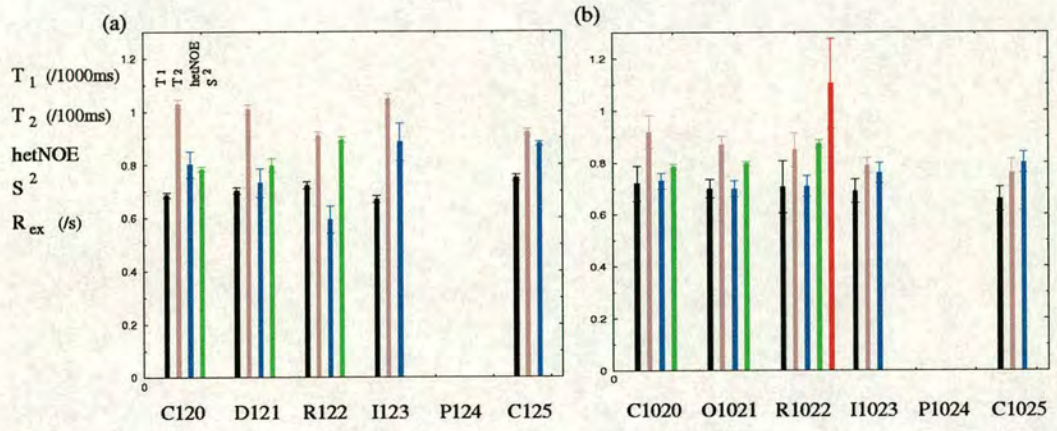


Figure 6.13: Relaxation data and Modelfree results for comparable residues in the linker regions of modules 2-3 and 16-17. T_1 shown as black columns, T_2 shown as brown columns, heteronuclear NOE shown as blue columns, S^2 shown as green columns and R_{ex} shown as a red column. Errors bars are shown for each value. Module 16 S^2 and R_{ex} data are from the Modelfree analysis of lone module 16.

Both linker regions show relatively high S^2 values, although the data set is complete for neither. One difference is the fitting of a chemical exchange term to residue R1022. This could indicate a contribution to linker flexibility in modules 16-17 from a slow timescale hinging motion around residue R1022. In residue R122, despite the slightly lower than average T_2 value, Modelfree detected no evidence of this type of motion. While Modelfree fitting was undertaken on module 16 as a lone module, the relaxation data for R1022 in 16 as a single module, and as part of 16-17, are similar.[3] This R_{ex} motion cannot be ruled out in the case of the 16-17 construct. The chemical exchange

[3] The T_1 and T_2 data for R1022 in 16 and 16^{17} cannot be directly compared due to the differences in macromolecular size, however the T_2 is slightly lowered from the average in both cases. T_1 is close to the mean in both cases.

motion present in two loops at the 16-17 junction is consistent with this. The C-terminus in the lone module 16 is structured - residues I1019-Q1021 comprise β-strand H - showing that the data obtained for the lone module may well be representative of the real structure.

Residue D121 was fitted to model 5 with a correlation time of 1.164 ns. However, the S^2 associated with the ns and ps motions were 0.80 and 0.88, indicating that there is limited flexibility on the fast timescale. There was no flexibility on the fast ps-ns timescale measured for either linker.

6.2.7 Summary of differences in dynamics in functional sites 1 and 2

CR1 modules 2 and 16

Considering CR1 modules 2^3 and 16 first, it is clear that while a high percentage of identical residues between CCP modules gives rise to similar 3D-structures, the same cannot be said of dynamics. At both the level of individual residues and of sub-structures within the modules, similar sequences exhibit dynamical differences. Non-identical residue types showed marked changes in dynamics in many cases (examples in CR1 modules 2 and 16 are N65/T965, F82/V982, Q85/R985, E116/K1016)[4]. Even in cases of identical residue type, however, different Modelfree models were fitted in some cases (V76/V976, S84/S984, G98/G998, S101/S1001, L106/L1006, S107/S1007, T110/T1010). In instances where the same dynamical model was required for fitting conserved residues, notably different levels of flexibility were sometimes observed (I80/I980, Y88/Y988).

Other differences between modules 2^3 and $16/16^{17}$ include the wider range of dynamical motion within module 2^3. Module 2^3 has a larger range of hetNOE than 16^{17} and a larger range of S^2 values than lone module 16. Both modules 2^3 and 16 have a more flexible face which contains the majority of the residues implicated in binding and almost all the residues showing chemical exchange. In module $16/16^{17}$, residues

[4] While the examples given include results from the Modelfree analysis of lone module 16, residues listed are not found near the 16-17 junction and so should be less affected by the absence of module 17.

K964, N1009 and K1016 are clearly located on one face. In module 2^3, residues R64 and T65 are located on this same face and while residue Y94 is located towards the back of the module, its sidechain ring is at the junction and exposed in the active face. However, in the case of module 2^3 the active site face is more flexible and the scaffold-like face is more rigid than their module 16 counterparts. One difference in the chemical exchange motion between the modules is that the DE_{16} loop in module 16^{17} (residues T991-R995) appears to be undergoing slow timescale motion (from the 16-17 relaxation data) whereas the same cannot be said for module 2^3. In module 2^3, the DE_2 and AB_2 loops (both of which are at the 2-3 junction) are also rigid on the fast timescale.

CR1 modules 3 and 17

Considering modules $^{2}3$ and $^{16}17$: without a Modelfree analysis for both modules it is more difficult to make assured comparisons. However, given the identity between these two modules, the apparent differences based on relaxation data are unexpected. The relaxation data alone are sufficiently different to show that these two modules have significantly different dynamics - the T_1, T_2 and hetNOE data for modules $^{2}3$ and $^{16}17$ are plainly different (see Figures 6.5 and 6.7). The FG loop of both modules $^{2}3$ and $^{16}17$ appears to be undergoing exchange motion, but in the case of module $^{16}17$ the motion appears to continue into the F_{17} and G_{17} β-strands. This region is identical between modules, but adjacent to the junction so variations in the preceding module could account for the dynamic differences. Similarly, the hypervariable loops of module $^{2}3$ seems flexible on a fast timescale and may also be undergoing chemical exchange, although there is no evidence of this in module $^{16}17$.

CR1 2-3 and 16-17 linker regions

Comparing the linker region of these double modules, three out of of four of the residues are conserved and D121/Q1021 represents the only difference. Despite this, there are notable differences in the relaxation data when comparing the two linkers, even for the conserved residues. Neither linker shows evidence of being flexible on the fast, ps-ns timescale. The module 16-17 linker might be undergoing slow timescale (μs-ms)

motion, while the 2-3 linker does not have apparent flexibility on either fast or slow timescales. It is still possible that both linkers are flexible on the intermediate, ns-μs, timescale because Modelfree cannot measure this.

CR1 2-3 and 16-17 loop regions

The loop regions in modules 2-3 and 15-16 are shown in Figure 6.14, with the likely motions mapped on to the structures as colours. The 2-3 junction contains loop regions which are rigid on the fast timescale and also evidence of slow motion only in residues distant from the junction which may have hinging actions. From the dynamics data, it appears as though there is a reduced level of flexibility at the 2-3 interface, in comparison with the 16-17 interface.

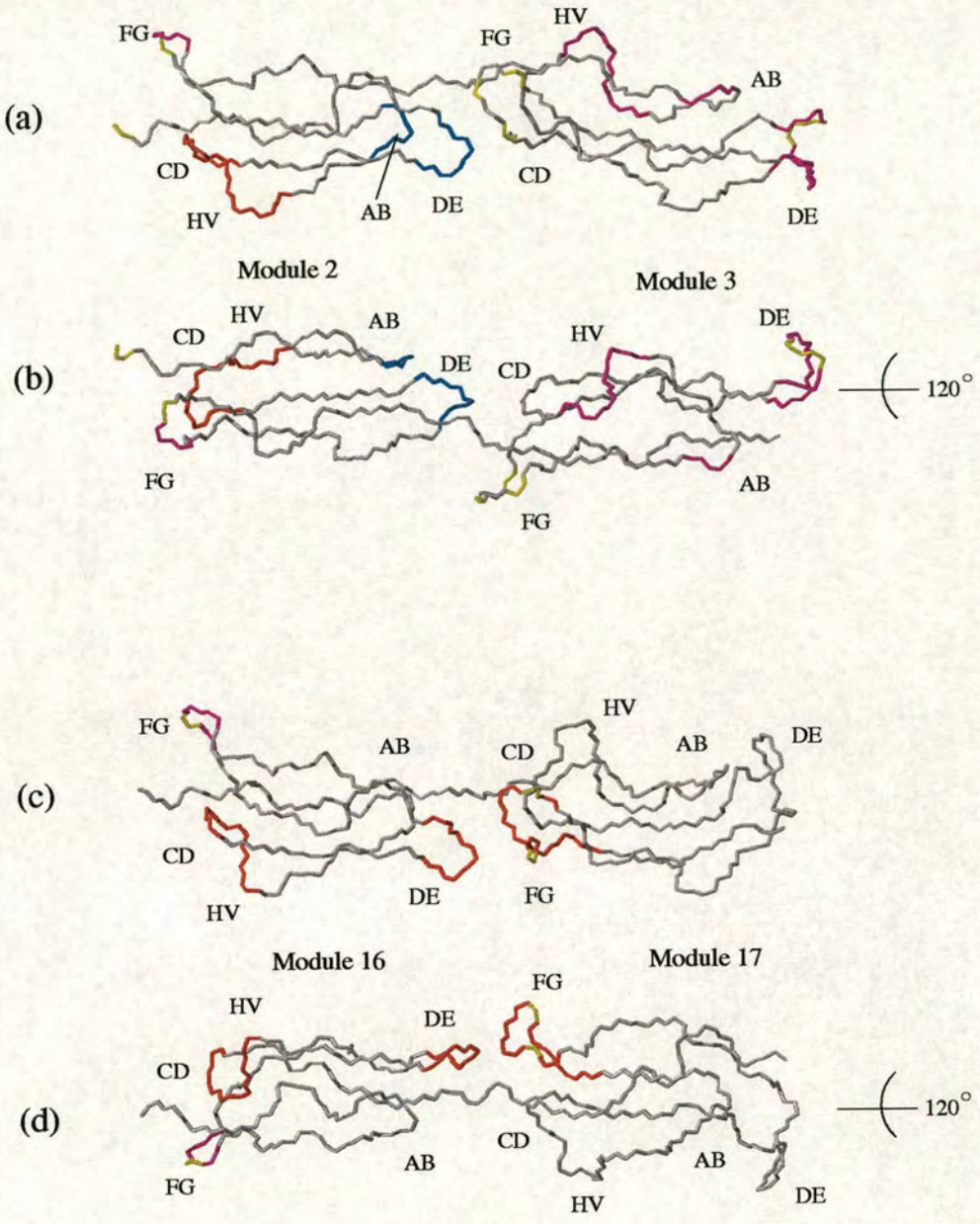


Figure 6.14: Slow and fast timescale motion of loops within modules 2-3 and 16-17. Loop regions likely to be undergoing chemical exchange are shown in red. (The DE_2 and CD_3 loops contain no residues likely undergoing chemical exchange but there may be hinging motion of the whole loop regions occurring in residues far from the junction.) Loops which are flexible on the fast timescale (from Modelfree analysis in the case of 2-3 and heteronuclear NOE data in the case of 16-17) are shown in magenta. Loops which are rigid on a fast timescale for modules 2-3 are shown in blue. Residues unassigned in the $^{15}N,^{1}H$-HSQC (and therefore likely to be exchanging) are shown in yellow. Though coloured magenta for fast timescale flexibility, the FG_2 loop may also be undergoing chemical exchange. Loops are labelled using the CCP module convention. HV is the hypervariable loop.

6.3 Implications of dynamics to binding functions of CR1

6.3.1 Binding function in CR1, site 2

In module 16, the three residues most strongly implicated in binding are located on the more flexible face of the module. Similarly, most key binding residues in module 15 are located on one face. Module 15 and 16 together present a contiguous binding face that extends over both modules. Such a contiguous binding face does not, however, extend into module 17 of the solved 3D-structure. In module 17, important binding residues are also localised on one face of the module - the face consisting of the hypervariable loop, the CD loop and the D-strand - but in the closest-to-the-mean NMR-derived structure, this face is twisted approximately 60° from the 15-16 binding face. It has been postulated that a change of the twist angle between modules 16 and 17 will present a contiguous binding face for C3b/C4b [89]. From the available structural evidence it seems that the 16-17 linker has the requisite flexibility to accomplish this. Figure 6.15 shows the residues important for binding in site 2 mapped on to the structure.

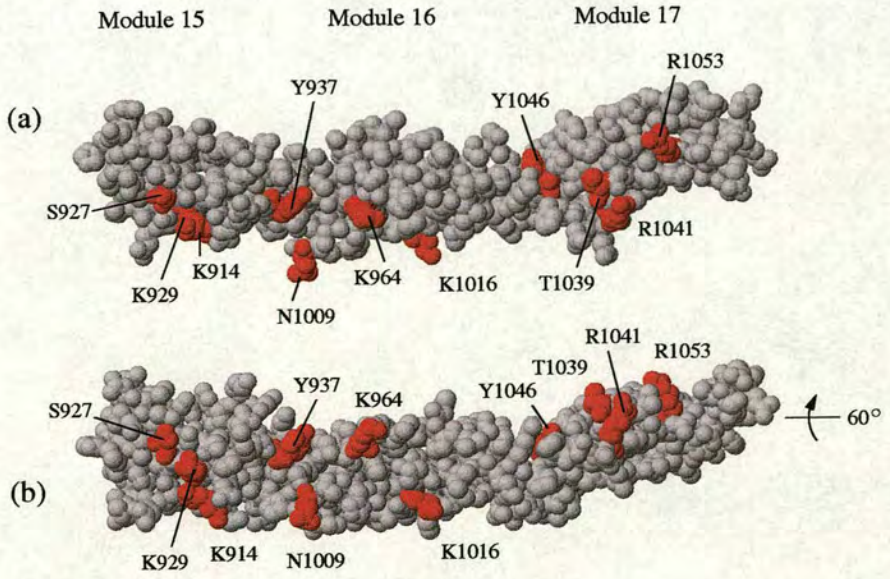


Figure 6.15: The structure of CR1 site 2 with residues important for the binding functions of site 2 shown in red.

6.3.2 Binding function in CR1, site 1

From the knowledge of the binding face of site 2, the available mutagenesis data, and the current work, it is possible to infer the binding face of site 1. Modules 1 and 15 share only 56% identity. An alignment of their sequences is shown in Figure 6.16. Comparing modules 1 and 15, none of the four important binding residues listed in Table 1.5 are conserved in module 1. Lys residues K914 and K929 are mutated to residues without positive charge, T14 and N29. In the case of S927, the residue type is altered to provide a more bulky sidechain - Y27. A converse swap is made in the case of Y937/S37. Mutation of residue G35 is known to have a large effect on the C4b-binding function of site 1. Using the 15-16 structures as a template, in the structure of 1-2, G35 would be tucked into the linker region and not exposed. It is possible, therefore, that the mutation G35E - which disrupted C4b-binding and decay accelerating activity functions - had a major effect on the module 1-2 linker which in turn affected inter-modular orientation. If alkyl and aromatic sidechains form an organised hydrophobic region at the 1-2 junction (P34, G35 and Y36 could all participate), the addition of a Glu residue could disrupt the packing involved.

```
          2                                  32            45    +       58
CR1_1   WGQCNAPEWLPFARPTNLTDEFEFPIGTYLNYECRPGYSGRPFSIICLKNSVWTGAKDRCRR
CR1_15  L H Q  DHFL  KLKTQTNASD     S K     E Y      T  D L  SSP  V K
                     *              * *       *
```

Figure 6.16: Sequence alignment of CR1 modules 1 and 15. Only residues which differ in sequence identity are shown for module 15. Cysteine residues in module 1 are numbered. Residues implicated in site 2 function are shown by an asterisk beneath the module 15 sequence.

For the purposes of comparing the structures of both functional sites in their entirety, a Modeller-based model of the entire functional site 1 was generated using the previously described method (section 3.6 and 5.3) and the 15-16 and 16-17 double module pairs as templates. Due to the relatively low sequence identity between modules 1 and 15, the 3D-structural model of this portion is unlikely to be completely accurate. It is not described in detail here and is used only as a guide to identify probable locations (e.g.

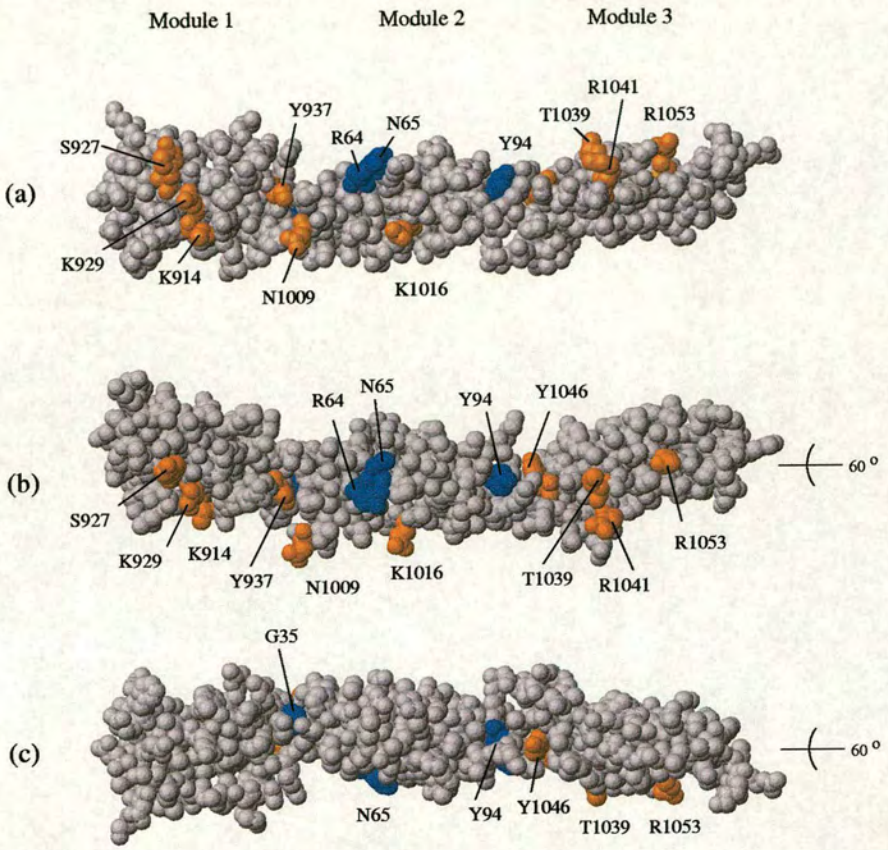


Figure 6.17: Space filling representation of the modelled structure of site 1 with residues indicated by mutagenesis to be important for its binding function shown in blue. Residues equivalent to those important in the binding function of site 2 are shown in orange. Figure (c) shows the probable "back" of the site and the reduced number of key residues.

in loops, turns) of key residues. The intermodular junctions in this model cannot be taken as representative of the actual junctions in site 1. Figure 6.17 shows the residues important for binding in site 1 mapped on to the model surface in blue. Residues equivalent to the key binding residues of site 2 are mapped on to the equivalent site 1 residues and shown in orange. Residue R64 is coloured blue along with the other residues important for site 1 binding, despite K964 also being important in site 2 binding.

Site 1 requires module 3 for its binding function. In the case of site 2 at least, specific residues in the third module have been identified as being important for binding. At present, it is unknown if the equivalent residues - T139, R141, Y146 and R153 - are important to the C4b-binding of site 1. However, two of the three differences between modules 3 and 17 occur in the same face as the crucial binding residues in module 17. These are P156/L1056 and G159/R1059. Neither L1056 or R1059 has been investigated in site 2 as functionally important residues, and mutagenesis data on these residues would be extremely useful. The change of L1056 residue to P156 could have ramifications for altering the motions available to the DE_3 loop in module 3. As Arg residues have previously been shown to be crucial to binding in CR1 (R64, R1041, R1053) R1059 is a prime candidate for mutagenesis. The third point change, T132/A1032 occurs in the vicinity of the C-terminus of the module. The hydroxyl group could reduce hydrophobic packing around the 3-4 linker region and perhaps alter this junction's properties.

Potential binding mechanisms

CR1 site 1 binds C4b (and C3b very weakly), whereas site 2 binds C4b with similar affinity and C3b with greater affinity. It is possible that electrostatics play some part in determining the functional differences between the sites. In module 2, there are three more negatively charged (Asp, Glu) residues compared with module 16. Several differences between modules 2 and 16 either remove positive charge or introduce negative charge (e.g. D109/N1009, D114/S1014, E116/K1016) on the proposed binding face of module 2. While the 3D solution structures of the C4b and C3b ligands are not known, one area in C3b has been identified as important for binding to CR1 [107].

This is the segment comprising residues C727-F767 and it contains five positive and seven negative residues. In contrast, the same segment in C4b contains six positive and only five negative residues. From this, C3b may possibly possess a more negative binding area than C4b, while site 2 has a more positive C3b/C4b-binding face than site 1. This could go some way to explaining the preference of site 2 to bind C3b over C4b. However, it does not explain why site 1 binds C3b only very weakly.

Regarding the chemical exchange motion of residues on the binding face of modules 2 and 16, it appears as though this slow timescale flexibility in the backbone of key residues could be involved in binding function in both sites. In one possible mechanism for specific interactions, positive residues across the binding face could attract the ligand to the correct face through electrostatic interactions. This face would then possess the flexibility to subsequently "fit" correctly around the ligand. The μs-ms timescale motion across the binding face could provide sufficient time for the correct hydrophobic residues on both receptor and ligand to organise themselves so as to maximise van der Waals contacts. These would provide increased specificity and an enthalpic gain to offset the entropic cost of rigidification.

The differences in the affinities of site 1 and site 2 for their binding partners could be contributed to by the differences in distribution of residues undergoing chemical exchange. In site 2 there exists the possibility that a slow timescale hinging motion is occurring at the 16-17 junction, as suggested by the apparent μs-ms motions in the module 16 DE_{16} loop, the module 17 FG_{17} loop and possibly the 16-17 linker residues. Therefore it is conceivable that this junction flexibility allows site 2 to take on conformations required for C3b-binding, including aligning the binding faces across all three modules. However, in site 1, there are many fewer residues undergoing chemical exchange at the 2-3 junction. Unlike the 16-17 junction (which has chemical exchange in the loops of both modules 16 and 17 at the junction), the only evidence of slow motion at the 2-3 junction is in the module 3 loops. This apparent lack of flexibility could prevent site 1 from assuming the correct orientation to bind C3b. Similarly, this potential difference could contribute to the fact that site 1 is responsible for decay accelerating activity while site 2 governs cofactor activity.

If the Modeller-based model of CR1~2-3 were correct in terms of intermodular orientation, then the key module 2 residues (R64, N65, Y94) are well aligned with the module 3 face that is equivalent to the proposed site 2 binding face on module 17. (Figure 6.17 shows that the blue coloured residues in the central module line up well with the orange coloured ones on module 3/17.) This is based on the assumption that the same residues in module 3/17 are involved in binding in both sites. It may be that the 15-16 junction is somewhat rigid to maintain its contiguous binding face, and the 2-3 junction may be similar. Figure 6.18 shows the possible differences between the binding sites.

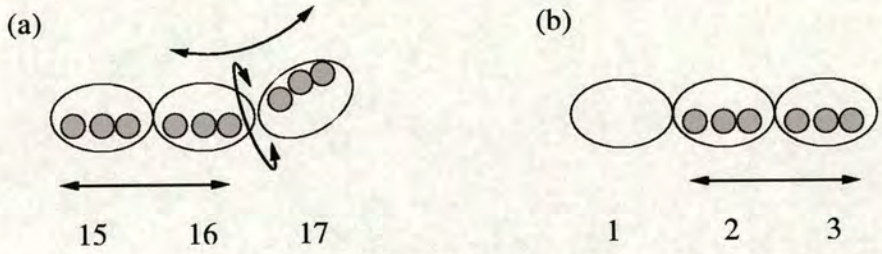


Figure 6.18: Possible differences between the flexibility of the junctions in CR1 sites 2 (figure a) and 1 (figure b). The binding regions are shown as grey circles. (a) The 16-17 junction is believed to be flexible, while the 15-16 shows more intermodular contacts. The binding face is contiguous across modules 15 and 16. A relatively rigid junction could maintain this alignment. The module 17 binding face is not contiguous with that of 15 and 16. A flexible 16-17 junction could allow all three binding regions to align. (b) In modules 2 and 3, the binding faces may be contiguous. A rigid 2-3 linker could maintain this face, in a similar way to the 15-16 junction. Modules 1 and 15 are so different that neither it's binding face or the 1-2 junction cannot be speculated on.

One further point relates to junction melting studies. Modules 15-17 seem to have melting transitions for junctions that are distinct from the modules melting, which imply that the junctions are less stable than the body of the modules. The 16-17 junction appears the less stable of the two junctions [62, 60]. In modules 1-3, there seems to be only one overall melting transition, implying that the whole site behaves like one protein domain, with consistent stability along its length without clearly defined junctions [24]. The wider ranges of the relaxation data and Modelfree parameter fittings for the residues within the bodies of modules 2 and 3 compared with 16 and 17 supports this

hypothesis. This suggestion is also consistent with the reduced flexibility seen around the 2-3 linker and would presumably require similar rigidity at the module 1-2 junction.

While rigid, rod-like structures are presented by several of the CCP double module structures solved using X-ray crystallography [97, 23, 87], the NMR-derived structures of CCP module pairs have mostly suggested limited intermodular contacts or a variety of intermodular orientations [41, 128, 135]. VCP˜3-4 is the only example of a CCP module pair in which the NMR-derived data suggests junction rigidity. The solution structure of VCP˜3-4 shows a reduced RMSD compared with VCP˜2-3 - backbone RMSD over both modules of 0.73 Å in 3-4 compared with 1.41 Å in 2-3 [40, 142]. Unfolding of VCP 2^3 was also shown to have little effect on the chemical shift dispersion of module ^{2}3, suggesting a lack of intermodular contact at the 2-3 interface [61]. The relaxation data analysis of VCP˜2-3 (as described previously in section 6.1) suggested both slow and fast timescale flexibility within the linker and junction regions. The analysis of VCP˜3-4, however, showed flexibility on neither the fast or slow timescales [15]. While CR1˜2-3 may be homologous to VCP˜2-3, it appears if the dynamics within the CR1˜2-3 junction are more similar to those of VCP˜3-4. CR1˜16-17 and VCP˜2-3 appear to have junction flexibility, and form part of functional sites which can bind C3b. CR1˜2-3 appears to have less junction flexibility, and forms part of a functional site which cannot bind C3b.

Rigidity at the junctions within other modular proteins has been previously characterised with NMR spectroscopy. In the NMR-derived structure of two pairs (modules 12-13 and 32-33) of calcium binding epidermal growth factor (cbEGF) domains found within fibrillin-1, the relative orientation of the module pairs is fixed [31, 141]. Relaxation data analysis of these pairs has confirmed that the least flexible portions of these fragments are the junctions [141, 126]. In the case of cbEGF domains, a calcium atom is bound within the junction and hydrophobic packing (including the involvement of a Tyr residue) is involved in providing the junction stability. However, this family of proteins are involved in forming scaffold structures within the extracellular matrix and have only a single residue in the intermodular linkers, in contrast with RCA proteins [22].

6.3.3 Intermediate timescale motions

In terms of flexibility, it may not be that the 2-3 junction is inflexible on all timescales. The 2-3 junction may be undergoing intermediate timescale motion that Modelfree has not measured. It is possible that the apparent flexibility in the 16-17 junction also comes from motion on this timescale. However, if the 16-17 junction is moving on a slow timescale, and the 2-3 junction on an unmeasured intermediate timescale, this could still differentiate the sites in terms of dynamics.

The deficiencies in Modelfree regarding intermediate timescale motion are soon to be circumvented. The slowly relaxing local structure (SRLS) model has been introduced in the recent past precisely to probe the motions that Modelfree cannot [134]. It is based on Modelfree (in fact SRLS can be thought of as a generalised version of Modelfree) but does not not make the assumption that the global and local motions are not coupled. Using this mode-coupling approach, study of protein ^{15}N backbone dynamics has begun [125].

6.4 Conclusions

The work completed within this project and the main conclusions are detailed below.

1. The assignment of ^{15}N and ^{1}H resonances of the backbone and sidechains of CR1~16 was accomplished to near-completion.

2. The three-dimensional solution structure of CR1~16, as a lone module, was determined using NMR-derived NOE restraints. The scaffold of the structure was shown to be independent of context when compared to larger fragments containing module 16. The positions of loops and turns was context dependent.

3. The ^{15}N isotropic dynamics of CR1~16 were analysed using Modelfree. The module was shown to have relatively reduced variation in dynamics when compared to other CCP modules. The face of the module containing the key binding residues was shown to be the more flexible on the ps-ns timescale than the opposite face. This binding face also contained the majority of the residues undergoing slow

timescale motions which could have implications for function.

4. The ^{15}N dynamics of CR1~16 taking into account anisotropy of diffusion could not be successfully analysed. This was due to the lack of precision within the solution structures of CR1~16.

5. The assignment of ^{13}C, ^{15}N and ^{1}H resonances of CR1~2-3 was accomplished to near-completion. This assignment can, in future, be used to determine the solution structure of this module pair.

6. The isotropic dynamics of CR1~2-3 were analysed using Modelfree. Module 2 was shown to have the same overall pattern of rigidity as module 16, though the rigid regions were more rigid and the flexible regions more flexible. As with module 16, slow timescale motions were detected on the proposed binding face of module 2. The module pair was shown to be rigid around the linker region in terms of both fast and slow motions.

7. The structure of 2-3 was modelled using Modeller, taking advantage of the high sequence identity of CR1~16-17, which served as a structural template. Comparisons of the dynamics of CR1~2-3 were made with those of VCP~2-3 and CR1~16-17. The VCP linker appears flexible on both fast and slow timescales, in contrast to CR1~2-3. The majority of chemical exchange occurs within strands in VCP, unlike CR1. Both the CR1~2-3 and CR1~16-17 linkers lack residues with obvious backbone motion on the fast timescale. The CR1~16-17 junction and possibly the linker appear to be involved in slow timescale motion, unlike the majority of the CR1~2-3 junction. Thus dynamics of CCP modules can vary greatly despite high sequence similarity between homologous proteins.

There are several steps which could now be taken to further our understanding of the sequence-structure-function relationship within CR1. Primarily, the solution structure of modules 2-3 must be completed using the above mentioned assignment data. This would allow the anisotropic dynamics of modules 2-3 to be analysed. The assignment and structure determination of modules 1-2, providing a construct of suitable quality can be cloned, would complete the structure of the entirety of functional site 1. Anisotropic dynamics would complete the picture in terms of structure and flexibility.

The application of mode-coupling approaches could conclusively identify whether the 2-3 junction was rigid on all timescales. From there, more site-directed mutagenesis would be possible to further elucidate function, including mutations to increase, reduce or remove junction flexibility.

Bibliography

[1] www.wellcome.ac.uk/en/malaria/.

[2] M. Assfalg, L. Banci, I. Bertini, S. Ciofi-Baffoni, and P.D. Barker. ^{15}N backbone dynamics of ferricytochrome b_{562}: comparison with the reduced protein and the R98C variant. *Biochemistry*, 40:12761–12771, 2001.

[3] W.P. Aue, E. Bartholdi, and R.R. Ernst. Two-dimensional spectroscopy. Application to nuclear magnetic resonance. *J. Chem. Phys.*, 64:2229–2246, 1976.

[4] G. Barbato, M. Ikura, L.E. Kay, R.W. Pastor, and A. Bax. Backbone dynamics of calmodulin studied by ^{15}N relaxation using inverse detected 2-dimensional NMR-spectroscopy – the central helix is flexible. *Biochemistry*, 31:5269–5278, 1992.

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

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

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

[8] A. Bax and R. Freeman. Investigation of complex networks of spin-spin coupling by two-dimensional networks. *J. Magn. Res.*, 44:542–561, 1981.

[9] D. Bentrop, I. Bertini, R. Iacoviello, C. Luchinat, Y. Niikura, M. Piccioli, C . Presenti, and A. Rosato. Structural and dynamical properties of a partially unfolded Fe_4S_4 protein: role of the cofactor in protein folding. *Biochemistry*, 38:4669–4680, 1999.

[10] A.M. Blom, J. Webb, B.O. Villoutreix, and B. Dahlback. A cluster of positively charged amino acids in the C4BP α-chain is crucial for C4b binding and factor I cofactor function. *J. Biol. Chem.*, 274:19237–19245, 1999.

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

[12] B. Bouma, P.G. De Groot, J.M. Van Den Elsen, R.B.G. Ravelli, A. Schouten, M.J.A. Simmelink, R.H.W.M. Derksen, J. Kroon, and P. Gros. Adhesion mechanisms of human β2-glycoprotein I to phospholipids based on its crystal structure. *EMBO J.*, pages 5166–5174, 1999.

[13] W. Braun, C. Bosch, L. Brown, N. GO, and K. Wütrich. Combined use of proton-proton overhauser enhancements and a distance geometry algorithm for determination of polypeptide conformations. Application to micelle-bound glucagon. *Biochim. Biophys. Acta.*, 667:377–396, 1981.

[14] L. Braunschweiler and R.R. Ernst. Coherence transfer by isotropic mixing: application to proton correlation spectroscopy. *J. Magn. Res.*, 53:521–528, 1983.

[15] K. Bromek-Burnside. *Nuclear magnetic resonance methods for obtaining structural and dynamical information for biomolecules of modular composition.* PhD thesis, The University of Edinburgh, 2002.

[16] A.T. Brünger. http://cns.csb.yale.edu/v1.0.

[17] A.T. Brünger. *X-PLOR: Version 3.1 A system for X-ray crystallography and NMR.* Yale University Press, 1992.

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

[19] M. Budayova-Spano, W. Grabarse, N.M. Thielens, H. Hillen, M. Lacroix, M. Schmidt, J.C. Fontacilla-Camps, G.J. Arlaud, and C. Gaboriaud. Monomeric structures of the zymogen and active catalytic domain of complement protease C1r: further insights into the C1 activation mechanism. *Structure*, 10:1509–1519, 2002.

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

[21] L.J. Patrick C. Roger. FAST-Modelfree: a program for rapid automated analysis of solution NRM spin-relaxation data. *J. Biomol. NMR.*, 26:203–213, 2003.

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

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

[24] N.S. Clark, I. Dodd, D.E. Mossakowska, R.A.G. Smith, and M.G. Gore. Folding and conformational studies on SCR1-3 domains of human complement receptor 1. *Protein Eng.*, 9:877–884, 1996.

[25] M.G. Clore, A. Szabo, A. Bax, L.E. Kay, P.C. Driscoll, and A.M. Gronenborn. Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *J. Am. Chem. Soc.*, 112:4989–4991, 1990.

[26] I.A. Cockburn, M.J. Mackinnon, A. O'Donnell, S. J. Allen, J.M. Moulds, M. Baisor, M. Bockarie, J.C. Reeder, and A.J. Rowe. A human complement receptor 1 polymorphism that reduces *Plasmodium falciparum* from rosetting confers protection against severe malaria. *Proc. Nat. Acad. Sci.*, 101:272–277, 2004.

[27] Protein Databank. http://www.rcsb.org/pdb/.

[28] S. Deep, K.P. Walker, Z. Shu, and A.P. Hinck. Solution structure and backbone dynamics of the TGFβ type II receptor extracellular domain. *Biochemistry*, 42:10126–10139, 2003.

[29] A. E. Derome. *Modern NMR techniques for chemistry reseacrh.* Pergamon Press, 1993.

[30] I. Dodd, D.E. Mossakowska, P. Camilleri, M. Haran, P. Hensley, E.I. Lawlor, D.I. McBay, W. Pindar, and R.A.G. Smith. Overexpression in *Escherichia coli*, folding, purification, and characterization of the first three short consensus repeat modules of human complement receptor type 1. *Protein Express. Purif.*, 6:727–736, 1995.

[31] A.K. Downing, V. Knott, J.M. Werner, C.M. Cardy, I.D. Campbell, and P.A. Handford. Solution structure of a pair of calcium-binding epidermal growth factor-like domains: implications for the Marfan syndrome and other genetic disorders. *Cell*, 85:597–605, 1996.

[32] B.M. Duggan, H.J. Dyson, and P.E. Wright. Inherent flexibility in a potent inhibitor of blood coagulation, recombinant nematode anticoagulant protein c2. *Eur. J. Biochem.*, 265:539–548, 1999.

[33] V.A. Feher and J. Cavanagh. Millisecond-timescale motions contribute to the function of the bacterial response regulator protein Spo0F. *Nature*, 400:289–293, 1999.

[34] R. Freeman. *A handbook of nuclear magnetic resonance.* Longman Scientific and Technical, 1988.

[35] R. Freeman. *Spin choreography.* Spektrum, 1997.

[36] C. Gaboriaud, V. Rossi, I. Bally, G.J. Arlaud, and J.C. Fontecilla-Camps. Crystal structure of the catalytic domain of human complement C1s: a serine protease with a handle. *EMBO J.*, 19:1755–1765, 2000.

[37] I.J Griswold and F.W. Dahlquist. Bigger is better: megadalton protein NMR in solution. *Nat. Struc. Biol.*, pages 567–568, 2002.

[38] I.J Griswold and F.W. Dahlquist. The dynamic behaviour of CheW from *Thermotoga maritima* in solution, as determined by nuclear magnetic resonance: implications for potential protein-protein interaction sites. *Biophys. Chem.*, 101:359–373, 2002.

[39] N. Guex. http://au.expasy.org/spdbv/.

[40] C.E. Henderson. *Solution structure of the central CCP module pair of a poxvirus complement control protein.* PhD thesis, The University of Edinburgh, 2001.

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

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

[43] S.W. Homans. *A dictionary of concepts in NMR.* Oxford Science Publications, 1993.

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

[45] D.E. Hourcade, D.R. Miesner, C. Bee, W. Zeldes, and J.P. Atkinson. Duplication and divergence of the amino-terminal coding region of the complement receptor 1 (CR1) gene. An example of concerted (horizontal) evolution within a gene. *J. Biol. Chem.*, 265:974–980, 1990.

[46] D.E. Hourcade, L. Mitchell, L.A. Kuttner-Kondo, and J.P. Atkinson. Decay-accelerating factor (DAF), complement receptor 1 (CR1) and factor H dissociate the complement AP C3 convertase (C3bBb) via sites on the type A domain of Bb*. *J. Biol. Chem.*, 277:1107–1112, 2002.

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

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

[49] M. Ikura, L.E. Kay, and A. Bax. A novel approach for sequential assignment of ^{1}H, ^{13}C and ^{15}N spectra of larger proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Applicaiton to calmodulin. *Biochemistry*, 29:4659–4667, 1990.

[50] N. Ishii, P. Wang, H. Kato, and M. Mizuno. Solubilized complement receptor type 1 in crystal form. *Eur. Pat. Appl.*, 2003.

[51] O. Jardetzky and J-F. Lefèvre. Protein dynamics. *FEBS Letters*, 338:246–250, 1994.

[52] B. Jimenez, M. Piccioli, J-M. Moratal, and A. Donaire. Backbone dynamics of rusticyanin: the high hydrophobicity and rigidity of this blue copper protein is responsible for its thermodynamic properties. *Biochemistry*, 42:10396–10405, 2003.

[53] W. Kabsch and C. Sander. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22:2577–2637, 1983.

[54] L.E. Kay. Protein dynamics from NMR. *Nat. Struc. Biol.*, 5:513–517, 1998.

[55] L.E. Kay, M. Ikura, R. Tschudin, and A. Bax. Three-dimensional triple-resonance spectroscopy of isotopically enriched proteins. *J. Magn. Reson.*, 89:496–514, 1990.

[56] L.E. Kay, L.K. Nicholson, F. Delaglio, A. Bax, and D.A. Torchia. Pulse sequences for removal of the effects of cross-correlation between dipolar and chemica l-shift anisotropy relaxation mechanism on the measurement of heteronuclear T_1 and T_2 values in pro teins. *J. Magn. Reson.*, 97:359–375, 1992.

[57] L.E. Kay, D.A. Torchia, and A. Bax. Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spe ctroscopy: application to staphylococcal nuclease. *Biochemistry*, 28:8972–8979, 1989.

[58] L.E. Kay, G.Y. Xu, D.R. Singer, A.U .and Muhandiram, and J.D. Forman-Kay. A gradient-enhanced HCCH-TOCSY experiment for recording side-chain ^1H and ^{13}C correlations in H_2O samples of proteins. *J. Magn.Reson. B*, 101:333–337, 1993.

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

[60] M.D. Kirkitadze, D.T.F. Dryden, S.M. Kelly, N.C. Price, X. Wang, M. Krych, J.P. Atkinson, and P.N. Barlow. Co-operativity between modules within a c3b-binding site of complement receptor type 1. *FEBS Letters*, 459:133–138, 1999.

[61] M.D. Kirkitadze, C.E. Henderson, N.C. Price, S.M. Kelly, N.P. Mullin, J.P. Parkinson, D.T.F. Dryden, and P.N. Barlow. Central modules of the vaccinia virus complement control protein module are not in extensive contact. *Biochem. J.*, 344:167–175, 1999.

[62] M.D. Kirkitadze, M. Krych, D. Uhrín, D.T.F. Dryden, B.O. Smith, A. Cooper, Hauhart R. Wang, X., J.P. Atkinson, and P.N. Barlow. Independently melting modules and highly structured intermodular junctions within complement receptor type 1. *Biochemistry*, 38:7019–7031, 1999.

[63] L. B. Klickstein, T. J. Bartow, V. Miletic, L. D. Rabson, J. A. Smith, and D. T. Fearon. Identification of distinct C3b and C4b recognition sites in the human C3b/C4b receptor (CR1, CD35) by deletion mutagenesis. *J. Experimental Medicine*, 168:1699–1717, 1988.

[64] L.B. Klickstein, W.W. Wong, J.A. Smith, J.H. Weis, J.G. Wilson, and D.T. Fearon. Human C3b/C4b receptor (CR1). Demonstration of long homologous repeating domains that are composed of the short consensus repeats characteristics of C3/C4 binding proteins. *J. Exp. Med.*, 165:1095–1112, 1987.

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

[66] G.J. Kotwal, S.N. Isaacs, R. McKenzie, M.M. Frank, and B. Moss. Inhibition of the complement cascade by the major secretory protein of vaccinia virus. *Science*, 250:827–830, 1990.

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

[68] M. Krych, L. Clemenza, D. Howdeshell, R.E. Hauhart, D.E. Hourcade, and J.P. Atkinson. Analysis of the functional domains of complement receptor type 1 (C3b/C4b receptor; CD35) by substitution mutagenesis. *J. Biol. Chem.*, 269:13273–13278, 1994.

[69] M. Krych, R.E. Hauhart, and J.P. Atkinson. Structure-function analysis of the active sites of complement receptor type 1. *J. Biol. Chem.*, 273:8623–8629, 1998.

[70] M. Krych, D.E. Hourcade, and J.P. Atkinson. Sites within the complement C3b/C4b receptor important for the specificity of ligand binding. *Proc. Natl. Acad. Sci.*, 88:4353–4357, 1991.

[71] M. Krych-Goldberg and J.P. Atkinson. Structure-function relationship of complement receptor type 1 . *Immunol. Rev.*, 180:112–122, 2001.

[72] M. Krych-Goldberg, R.E. Hauhart, V. Bala Subramanian, B.M. Yurcisin, D.L. Crymmins, D.E. Hourcade, and J.P. Atkinson. Decay accelerating activity of complement receptor type 1 (CD35). *J. Biol. Chem.*, 274:31160–31168, 1999.

[73] M. Krych-Goldberg, J.M. Moulds, and J.P. Atkinson. Human complement receptor type 1 (CR1) binds to major malarial adhesin. *Trends in Molec. Medicine*, 8:531–537, 2002.

[74] L. Kuttner-Kondo, M.E. Medof, W. Brodbeck, and M. Shoham. Molecular modeling and mechanism of action of human decay-accelerating factor. *Protein Eng.*, pages 1143–1149, 1996.

[75] L. Kuttner-Kondo, L. Mitchell, D.E. Hourcade, and M.E. Medof. Characterization of the active sites in decay-accelerating factor. *J. Immunol.*, pages 2164–2171, 2001.

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

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

[78] E.D. Laue, J. Skilling, J. Staunton, S. Sibisi, and R.G. Brereton. Maximum entropy method in nuclear magnetic resonance. *J. Magn. Res.*, 62:432–452, 1985.

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

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

[81] J.S. Li, S.P. Sander, A.E. Perry, S.S. Stinnett, J. Jaggers, P. Bokesh, L. Reynolds, R. Nassar, and P.A.W. Anderson. Pharmacokinetics and safety of TP10, soluble complement receptor 1, in infants undergoing cardiopulmonary bypass. *Am. Heart. J.*, 147:173–180, 2004.

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

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

[84] G. Lipari and A. Szabo. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and rage of validity. *J. Am. Chem. Soc.*, 104:4546–4559, 1982.

[85] G. Lipari and A. Szabo. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J. Am. Chem. Soc.*, 104:4559–4570, 1982.

[86] M.K. Liszewski, M. Leung, W. Cui, V.B. Subramanian, J. Parkinson, P.N. Barlow, M. Manchester, and J.P. Atkinson. Dissecting sites important for complement regulatory activity in membrane cofactor protein (MCP; CD46). *J. Biol. Chem.*, pages 37692–37701, 2000.

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

[88] S. Macura and R.R. Ernst. Elucidation of cross relaxation in liquids by two-dimensional NMR spectroscopy. *Mol. Phys.*, 41:95–117, 1980.

[89] R.L. Mallin. *A structural study of the C3b-binding site of complement receptor type I (CD 35).* PhD thesis, The University of Edinburgh, 2003.

[90] A.M. Mandel, M. Akke, and A.G. Palmer. Backbone dynamics of *Escherichia coli* ribonuclease HI: correlations with structure and function in an active enzyme. *J. Mol. Biology*, 246:144–163, 1995.

[91] J. McMurry. *Organic chemistry.* Brooks/Cole, 1995.

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

[93] S. Mori, C. Abeygunawardana, M.O. Johnson, and P.C.M. van Zijl. Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation. *J. Magn. Reson. B*, 108:94–98, 1995.

[94] G.A. Morris and R. Freeman. Enhancement of nuclear magnetic resonance signals by polarization transfer. *J. Am. Chem. Soc.*, 101:760–762, 1979.

[95] J.M. Moulds, L. Kassambara, J.J. Middleton, M. Baby, I. Sagara, A. Guindo, Coulibaly S., Yalcouye D., Diallo D.A., Miller L., and Doumbo O. Identification of complement receptor 1 (CR1) polymorphisms in west Africa. *Genes Immun.*, 5:325–329, 2000.

[96] D.R. Muhandiram and L.E. Kay. Gradient enhanced triple resonance three-dimensional NMR experiments with improved sensitivity. *J. Magn. Reson. B*, 103:203–216, 1994.

[97] K.H.M. Murthy, S.A. Smith, V.K. Ganesh, K.W. Judge, N. Mullin, P. N. Barlow, Ogata C. M., and Kotwal G.J. Crystal structure of a complement control protein that regulates both pathways of complement activation and binds heparan sulfate proteoglycans. *Cell*, 104:301–311, 2001.

[98] D. Neuhaus and M. Williamson. *The nuclear Overhauser effect in structural and conformational analysis.* VCH, 1989.

[99] M. Nilges. Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities. *J. Mol. Biol.*, 245, 1995.

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

[101] M. Nilges, G. M. Clore, and A. M. Gronenborn. Determination of three-dimensional structures of proteins from interproton distance data by dynamical simulated annealing from a random array of atoms: circumventing problems associated with folding. *FEBS Letters*, 239:129–136, 1988.

[102] M. Nilges, M.J. Macias, S.I. O'Donoghue, and H. Oschkinat. Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from β-spectrin. *J. Mol. Biol.*, 269:408–422, 1997.

[103] M. Nilges and S. I O'Donoghue. Ambiguous NOEs and automated NOE assignment. *Progr. NMR Spectr.*, 32:107–139, 1998.

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

[105] J. O'Leary. *Chemical synthesis, recombinant expression and structural characterisation of complement protein modules.* PhD thesis, The University of Edinburgh, 2001.

[106] J. O'Leary, K. Bromek, G.M. Black, S. Uhrínová, C. Schmitz, X. Wang, M. Krych, J.P. Atkinson, D. Uhrín, and P.N. Barlow. Backbone dynamics of complement control protein (CCP) modules reveals mobility in the binding surfaces. *Prot. Sci.*, 13, 2004.

[107] A.E. Oran and D.E. Isenman. Identification of residues within the 727-767 segment of human complement component C3 important for its interaction with factor H and with complement receptor 1 (CR1, CD35). *J. Biol. Chem.*, 274(8):5120–5130, 1999.

[108] A.G. Palmer. http://cpmcnet.columbia.edu/dept/gsas/biochem/labs/palmer/.

[109] A.G. Palmer, M. Rance, and P.E. Wright. Intramolecular motions of a zinc finger DNA-binding domain from xfin characterised by proton-detected natural abundance ^{13}C heteronuclear NMR spectroscopy. *J. Am. Chem. Soc.*, 113:4371–4380, 1991.

[110] S.M. Pascal, D.R. Muhandiram, T. Yamasaki, J.D. Porman-Kay, and L.E. Kay. Simultaneous acquisition of ^{15}N-edited and ^{13}C-edited NOE spectra of proteins dissolved in H_2O. *J. Magn.Reson. B*, 103:197–201, 1994.

[111] K. Pervushin, R. Riek, G. Wider, and K. Wüthrich. Attenuated T-2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *P. Natl. Acad. Sci. USA*, 94:12366–12371, 1997.

[112] K. Pervushin, R. Riek, G. Wider, and K. Wüthrich. Transverse relaxation-optimized spectroscopy (TROSY) for NMR studies of aromatic spin systems in ^{13}C-labelled proteins. *J. Am. Chem. Soc.*, 120:6394–6400, 1998.

[113] G.N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan. Stereochemistry of polypeptide and polysaccharide chain conformations. *Aspects Protein Struc.*, pages 121–134, 1963.

[114] S. Ranganathan, D.A. Male, R.J. Ormsby, E. Giannakis, and D.L. Gordon. Pinpointing the putative heparin/sialic acid-binding residues in the 'sushi' domain 7 of factor H: a molecular modeling study. *Pac. Symp. Biocomput.*, pages 155–167, 2000.

[115] H. Rattle. *An NMR primer for life scientists.* Partnership Press, 1993.

[116] D.G. Reid. *Methods in molecular biology. Volume 60. Protein NMR techniques.* Humana Press, 1997.

[117] G.C.K. Roberts. *NMR of macromolecules - a practical approach.* IRL Press, 1993.

[118] A.M. Rosengard, L.C. Alonso, L.C. Korb, W.M. Baldwin, F. Sanfilippo, L.A. Turka, and J.M. Ahearn. Functional characterization of soluble and membrane-bound forms of vaccinia virus complement control protein (VCP). *Mol. Immunol.*, 36:685–697, 1999.

[119] G.D. Ross, J.D. Lambris, J.A. Cain, and S.L. Newman. Generation of three different fragments of bound C3 with purified factor I or serum. *J. Immunol.*, 129:2051–2059, 1982.

[120] J.A. Rowe, J.M. Moulds, C.I. Newbold, and L.H. Miller. *P. falciparum* rosetting mediated by a parasite-variant erythrocyte membrane protein and complement-receptor 1. *Nature*, 388:292–295, 1997.

[121] J.A. Rowe, S.J. Rogerson, A. Raza, J.M. Moulds, D. Kazatchkine, K. Marsh, C. Newbold, J.P. Atkinson, and L.H. Miller. Mapping of the region of complement receptor (CR) 1 required for *Plasmodium falciparum* rosetting and demonstration of the importance of CR1 in rosetting in field isolates. *J. Immunol*, 165:6341–6346, 2000.

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

[123] A. Sali. http://salilab.org/modeller/.

[124] A. Sali and T.L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, 234:779–815, 1993.

[125] Y.E. Shapiro, E. Kahana, V. Tugarinov, Z. Liang, J.H. Freed, and E. Meirovitch. Domain flexibility in ligand-free and inhibitor-bound *Escherichia coli* adenylate kinase based on a mode-coupling analysis of ^{15}N spin relaxation. *Biochemistry*, 41:6271–6281, 2002.

[126] R.S. Smallridge, P. Whiteman, J.M. Werner, I.D. Campbell, P.A. Handford, , and A.K. Downing. Solution structure and dynamics of a calcium binding epidermal growth factor-like domain pair from the neonatal region of human fibrillin-1. *J. Biol. Chem.*, 278:12199–12206, 2003.

[127] B.O. Smith. Private communication.

[128] B.O. Smith, R.L. Mallin, M. Krych-Goldberg, X. Wang, R.E. Hauhart, K. Bromek, D. Uhrín, J.P. Atkinson, and P.N. Barlow. Structure of the C3b binding site of CR1 (CD35), the immune adherence receptor. *Cell.*, 108:1–20, 2002.

[129] S.A. Smith, N.P. Mullin, J. Parkinson, S.N. Shchelkunov, A.V. Totmentin, V.N. Loparev, R. Srisatjaluk, D.N. Reynolds, K.L. Keeling, D.E. Justus, P.N. Barlow, and G.J. Kotwal. Conserved surface-exposed K/R-X-K/R motifs and net positive charge on poxvirus complement control proteins serve as putative heparin binding sites and contribute to inhibition of molecular interactions with human endothelial cells: a novel mechanism for evasion of host defense. *J. Virol.*, 74:5659–5666, 2000.

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

[131] D. Soares. Private communication.

[132] V.B. Subramanian, L. Clemenza, M. Krych, and J.P.Atkinson. Substitution of two amino acids confers C3b binding to the C4b binding site of CR1 (CD35). *J. Immunol.*, 157:1242–1247, 1996.

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

[134] V. Tugarinov, Z. Liang, Y.E. Shapiro, J.H. Freed, and E. Meirovitch. A structural mode-coupling approach to ^{15}N NMR relaxation in proteins. *J. Am. Chem. Soc.*, 123:3055–3063, 2001.

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

[136] B.O. Villoutreix, A.M. Blom, J. Webb, and B. Dahlback. The complement regulator C4b-binding protein analyzed by molecular modeling, bioinformatics and computer-aided experimental design. *Immunopharm.*, pages 121–134, 1999.

[137] B.O. Villoutreix, Y. Hardig, A. Wallqvist, A. Covell, and P. Garcia de Frutos. Structural investigation of C4b-binding protein by molecular modeling: localization of putative binding sites. *Proteins*, pages 391–405, 1998.

[138] B.L. Volkman, D. Lipson, D.E. Wemmer, and D. Kern. Two-state allosteric behavior in a single-domain signaling protein. *Science*, 291:2429–2433, 2001.

[139] A.J. Wand. Dynamic activation of protein function: a view emerging from NMR spectroscopy. *Nat. Struc. Biol.*, 8:926–931, 2001.

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

[141] J.M. Werner, V. Knott, P.A. Handford, I.D. Campbell, and A.K. Downing. Backbone dynamics of a cbEGF domain pair in the presence of calcium. *J. Mol. Biol.*, 296:1065–1078, 2000.

[142] A.P. Wiles. *Structure of the C-terminal fragment of the secreted complement control protein from vaccinia virus.* PhD thesis, University of Oxford, 1996.

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

[144] P. Williams, Y. Chaudry, I.G. Goodfellow, J. Billington, R. Powell, O.B. Spiller, D.J. Evans, and S. Lea.

[145] W.W. Wong and S.A. Farrell. Proposed structure of the F' allotype of human CR1. Loss of a C3b binding site may be associated with altered function. *J. Immunol.*, 146:656–662, 1990.

[146] G. Wu, G. Wang, X. Fu, and L. Zhu. Synthesis, crystal structure, stacking effect and antibacterial studies of a nocel quaternary copper (II) complex with quinolone. *Molecules*, pages 287–296, 2003.

[147] K. Wüthrich. http://129.132.45.141/wuthrich/software/molmol/index.html.

[148] T. Yamazaki, J.D. Forman-Kay, and L.E. Kay. Two-dimensional NMR experiments for correlating ^{13}C-β and ^{1}H-δ/ϵ chemical-shifts of aromatic residues in ^{13}C-labelled proteins via scalar couplings. *J. Am. Chem. Soc.*, 115:11054–11055, 1993.

[149] K. Yazdanbaksh, S. Kang, D. Tamasauskas, D. Sung, and A. Scaradavou. Complement receptor 1 inhibitors for the prevention of immune-mediated red cell destruction: potential use in transfusion therapy. *Blood*, 101:5046–5052, 2003.

[150] S. Yoon and D.T. Fearon. Characterization of a soluble form of the C3b/C4b receptor (CR1) in human serum. *J. Immunol.*, 134:3332–3338, 1985.

[151] X. Yuan, J.M. Werner, J. Lack, V. Knott, P.A. Handford, I.D. Campbell, and A.K. Downing. Effects of the N2144S mutation on backbone dynamics of a TB-cbEGF domain pair from human fibrillin-1. *J. Mol. Biol.*, 316:113–125, 2002.

[152] J.L. Zimmerman, R.P. Dellinger, R.C. Straube, and J.L. Levin. Phase I trial of the recombinant soluble complement receptor type 1 in acute lung injury and acute respiratory distress syndrome. *Critical Care Medicine*, 28:3149–3154, 2000.

Appendix A

Chemical shift table for CR1~16

The chemical shifts assigned to the protons and nitrogen spins of CR1~16 are listed in the table below.

Residue	Atom	Shift	Atom	Shift
E957	HA	-	CA	-
E957	HB1	-	CB	-
E957	HB2	-	CB	-
E957	HG1	-	CG	-
E957	HG2	-	CG	-
A958	HA	-	CA	-
A958	HB1	-	CB	-
A958	HB2	-	CB	-
A958	HB3	-	CB	-
A958	HN	-	N	-
E959	HA	-	CA	-
E959	HB1	-	CB	-
E959	HB2	-	CB	-
E959	HG1	-	CG	-
E959	HG2	-	CG	-
E959	HN	-	N	-
A960	HA	4.30	CA	-
A960	HB1	1.30	CB	-
A960	HB2	1.30	CB	-
A960	HB3	1.30	CB	-
A960	HN	8.28	N	126.1
K961	HA	4.29	CA	-
K961	HB1	1.60	CB	-
K961	HB2	1.60	CB	-
K961	HD1	1.66	CD	-
K961	HD2	1.66	CD	-
K961	HE1	2.91	CE	-
K961	HE2	2.91	CE	-
K961	HG1	1.36	CG	-
K961	HG2	1.36	CG	-
K961	HN	8.27	N	121.3
K961	HZ1	-	NZ	-
K961	HZ2	-	NZ	-
K961	HZ3	-	NZ	-
S962	HA	4.62	CA	-
S962	HB1	3.69	CB	-
S962	HB2	3.56	CB	-
S962	HN	7.98	N	114.4
C963	HA	4.07	CA	-
C963	HB1	2.15	CB	-
C963	HB2	2.06	CB	-
C963	HN	8.07	N	120.8

Residue	Atom	Shift	Atom	Shift
K964	HA	4.30	CA	-
K964	HB1	1.88	CB	-
K964	HB2	1.80	CB	-
K964	HD1	1.72	CD	-
K964	HD2	1.72	CD	-
K964	HE1	3.03	CE	-
K964	HE2	3.03	CE	-
K964	HG1	1.52	CG	-
K964	HG2	1.52	CG	-
K964	HN	8.41	N	119.6
K964	HZ1	-	NZ	-
K964	HZ2	-	NZ	-
K964	HZ3	-	NZ	-
T965	HA	4.11	CA	-
T965	HB	3.93	CB	-
T965	HG21	1.15	CG2	-
T965	HG22	1.15	CG2	-
T965	HG23	1.15	CG2	-
T965	HN	8.40	N	121.3
P966	HA	4.67	CA	-
P966	HB1	1.66	CB	-
P966	HB2	1.42	CB	-
P966	HD1	3.56	CD	-
P966	HD2	3.56	CD	-
P966	HG1	1.76	CG	-
P966	HG2	1.35	CG	-
P967	HA	4.46	CA	-
P967	HB1	2.13	CB	-
P967	HB2	1.75	CB	-
P967	HD1	3.75	CD	-
P967	HD2	3.59	CD	-
P967	HG1	1.96	CG	-
P967	HG2	1.89	CG	-
D968	HA	4.63	CA	-
D968	HB1	2.72	CB	-
D968	HB2	2.28	CB	-
D968	HN	8.00	N	119.4
P969	HA	4.33	CA	-
P969	HB1	2.15	CB	-
P969	HB2	1.34	CB	-
P969	HD1	3.91	CD	-
P969	HD2	3.27	CD	-
P969	HG1	1.64	CG	-
P969	HG2	1.59	CG	-
V970	HA	3.64	CA	-
V970	HB	1.81	CB	-
V970	HG11	1.00	CG1	-
V970	HG12	1.00	CG1	-
V970	HG13	1.00	CG1	-
V970	HG21	0.91	CG2	-
V970	HG22	0.91	CG2	-
V970	HG23	0.91	CG2	-
V970	HN	8.30	N	125.1
N971	HA	4.04	CA	-
N971	HB1	2.72	CB	-
N971	HB2	1.54	CB	-
N971	HN	8.96	N	120.1
N971	HD21	7.45	ND2	-
N971	HD22	6.91	ND2	-
G972	HA1	4.45	CA	-
G972	HA2	4.22	CA	-
G972	HN	7.88	N	105.9
M973	HA	4.81	CA	-
M973	HB1	1.72	CB	-
M973	HB2	1.65	CB	-
M973	HE1	-	CE	-
M973	HE2	-	CE	-
M973	HE3	-	CE	-
M973	HG1	2.20	CG	-
M973	HG2	2.10	CG	-
M973	HN	8.69	N	116.4
V974	HA	4.45	CA	-
V974	HB	1.73	CB	-
V974	HG11	0.57	CG1	-
V974	HG12	0.57	CG1	-
V974	HG13	0.57	CG1	-
V974	HG21	0.22	CG2	-
V974	HG22	0.22	CG2	-
V974	HG23	0.22	CG2	-
V974	HN	8.49	N	122.5
H975	HA	4.61	CA	-
H975	HB1	3.14	CB	-
H975	HB2	2.76	CB	-
H975	HD2	6.63	CD2	-
H975	HE1	7.51	CE1	-
H975	HN	8.86	N	125.0
H975	HD1	-	ND1	-
H975	HE2	-	NE2	-

Residue	Atom	Shift	Atom	Shift
V976	HA	3.97	CA	-
V976	HB	1.92	CB	-
V976	HG11	0.73	CG1	-
V976	HG12	0.73	CG1	-
V976	HG13	0.73	CG1	-
V976	HG21	0.73	CG2	-
V976	HG22	0.73	CG2	-
V976	HG23	0.73	CG2	-
V976	HN	8.45	N	126.2
I977	HA	3.89	CA	-
I977	HB	1.96	CB	-
I977	HD11	0.73	CD1	-
I977	HD12	0.73	CD1	-
I977	HD13	0.73	CD1	-
I977	HG11	1.49	CG1	-
I977	HG12	1.17	CG1	-
I977	HG21	0.94	CG2	-
I977	HG22	0.94	CG2	-
I977	HG23	0.94	CG2	-
I977	HN	7.65	N	129.1
T978	HA	4.40	CA	-
T978	HB	4.22	CB	-
T978	HG21	1.17	CG2	-
T978	HG22	1.17	CG2	-
T978	HG23	1.17	CG2	-
T978	HN	9.29	N	113.1
D979	HA	4.59	CA	-
D979	HB1	3.03	CB	-
D979	HB2	2.70	CB	-
D979	HN	8.63	N	123.4
I980	HA	4.88	CA	-
I980	HB	2.34	CB	-
I980	HD11	0.90	CD1	-
I980	HD12	0.90	CD1	-
I980	HD13	0.90	CD1	-
I980	HG11	1.38	CG1	-
I980	HG12	1.08	CG1	-
I980	HG21	0.96	CG2	-
I980	HG22	0.96	CG2	-
I980	HG23	0.96	CG2	-
I980	HN	7.63	N	110.7
Q981	HA	4.40	CA	-
Q981	HB1	2.04	CB	-
Q981	HB2	1.94	CB	-
Q981	HG1	2.26	CG	-
Q981	HG2	2.17	CG	-
Q981	HN	7.87	N	118.7
Q981	HE21	7.49	NE2	-
Q981	HE22	6.63	NE2	-
V982	HA	3.09	CA	-
V982	HB	1.77	CB	-
V982	HG11	0.80	CG1	-
V982	HG12	0.80	CG1	-
V982	HG13	0.80	CG1	-
V982	HG21	0.68	CG2	-
V982	HG22	0.68	CG2	-
V982	HG23	0.68	CG2	-
V982	HN	7.74	N	119.9
G983	HA1	4.42	CA	-
G983	HA2	3.80	CA	-
G983	HN	9.18	N	117.7
S984	HA	4.58	CA	-
S984	HB1	4.08	CB	-
S984	HB2	4.08	CB	-
S984	HN	8.29	N	119.1
R985	HA	5.67	CA	-
R985	HB1	1.82	CB	-
R985	HB2	1.76	CB	-
R985	HD1	3.14	CD	-
R985	HD2	3.14	CD	-
R985	HG1	1.68	CG	-
R985	HG2	1.52	CG	-
R985	HN	8.53	N	120.9
R985	HE	-	NE	-
R985	HH11	-	NH1	-
R985	HH12	-	NH1	-
R985	HH21	-	NH2	-
R985	HH22	-	NH2	-
I986	HA	4.93	CA	-
I986	HB	1.30	CB	-
I986	HD11	-0.64	CD1	-
I986	HD12	-0.64	CD1	-
I986	HD13	-0.64	CD1	-
I986	HG11	0.57	CG1	-
I986	HG12	0.43	CG1	-
I986	HG21	0.61	CG2	-
I986	HG22	0.61	CG2	-
I986	HG23	0.61	CG2	-
I986	HN	8.80	N	117.2

Residue	Atom	Shift	Atom	Shift
T987	HA	5.17	CA	-
T987	HB	4.26	CB	-
T987	HG21	1.09	CG2	-
T987	HG22	1.09	CG2	-
T987	HG23	1.09	CG2	-
T987	HN	7.99	N	111.0
Y988	HA	5.40	CA	-
Y988	HB1	2.62	CB	-
Y988	HB2	2.38	CB	-
Y988	HD1	6.60	CD1	-
Y988	HD2	6.60	CD2	-
Y988	HE1	6.43	CE1	-
Y988	HE2	6.43	CE2	-
Y988	HN	8.47	N	118.6
S989	HA	4.64	CA	-
S989	HB1	3.95	CB	-
S989	HB2	3.83	CB	-
S989	HN	8.85	N	113.5
C990	HA	5.40	CA	-
C990	HB1	2.89	CB	-
C990	HB2	2.47	CB	-
C990	HN	8.67	N	116.4
T991	HA	4.21	CA	-
T991	HB	4.06	CB	-
T991	HG21	1.10	CG2	-
T991	HG22	1.10	CG2	-
T991	HG23	1.10	CG2	-
T991	HN	8.47	N	116.5
T992	HA	-	CA	-
T992	HB	4.05	CB	-
T992	HG21	1.25	CG2	-
T992	HG22	1.25	CG2	-
T992	HG23	1.25	CG2	-
T992	HN	8.10	N	117.4
G993	HA1	4.11	CA	-
G993	HA2	3.64	CA	-
G993	HN	-	N	-
H994	HA	5.10	CA	-
H994	HB1	3.14	CB	-
H994	HB2	2.49	CB	-
H994	HD2	6.50	CD2	-
H994	HE1	-	CE1	-
H994	HN	7.93	N	117.3
H994	HD1	-	ND1	-
H994	HE2	-	NE2	-
R995	HA	4.62	CA	-
R995	HB1	1.72	CB	-
R995	HB2	1.66	CB	-
R995	HD1	3.11	CD	-
R995	HD2	3.11	CD	-
R995	HG1	1.52	CG	-
R995	HG2	1.52	CG	-
R995	HN	9.41	N	122.4
R995	HE	7.81	NE	-
R995	HH11	-	NH1	-
R995	HH12	-	NH1	-
R995	HH21	-	NH2	-
R995	HH22	-	NH2	-
L996	HA	4.60	CA	-
L996	HB1	1.60	CB	-
L996	HB2	1.48	CB	-
L996	HD11	0.66	CD1	-
L996	HD12	0.66	CD1	-
L996	HD13	0.66	CD1	-
L996	HD21	0.42	CD2	-
L996	HD22	0.42	CD2	-
L996	HD23	0.42	CD2	-
L996	HG	1.15	CG	-
L996	HN	8.32	N	127.8
I997	HA	4.25	CA	-
I997	HB	1.87	CB	-
I997	HD11	0.66	CD1	-
I997	HD12	0.66	CD1	-
I997	HD13	0.66	CD1	-
I997	HG11	1.32	CG1	-
I997	HG12	1.15	CG1	-
I997	HG21	0.76	CG2	-
I997	HG22	0.76	CG2	-
I997	HG23	0.76	CG2	-
I997	HN	9.35	N	130.1
G998	HA1	4.45	CA	-
G998	HA2	3.48	CA	-
G998	HN	8.38	N	117.1
H999	HA	-	CA	-
H999	HB1	3.42	CB	-
H999	HB2	3.14	CB	-
H999	HD2	7.39	CD2	-
H999	HE1	8.29	CE1	-
H999	HN	8.79	N	120.9
H999	HD1	-	ND1	-
H999	HE2	-	NE2	-

Residue	Atom	Shift	Atom	Shift
S1000	HA	4.29	CA	-
S1000	HB1	3.90	CB	-
S1000	HB2	3.90	CB	-
S1000	HN	8.99	N	116.8
S1001	HA	5.30	CA	-
S1001	HB1	3.91	CB	-
S1001	HB2	3.87	CB	-
S1001	HN	7.74	N	115.2
A1002	HA	4.65	CA	-
A1002	HB1	1.55	CB	-
A1002	HB2	1.55	CB	-
A1002	HB3	1.55	CB	-
A1002	HN	8.32	N	121.6
E1003	HA	5.53	CA	-
E1003	HB1	1.86	CB	-
E1003	HB2	1.86	CB	-
E1003	HG1	2.06	CG	-
E1003	HG2	2.06	CG	-
E1003	HN	8.86	N	119.7
C1004	HA	4.03	CA	-
C1004	HB1	2.72	CB	-
C1004	HB2	1.66	CB	-
C1004	HN	8.65	N	126.2
I1005	HA	4.54	CA	-
I1005	HB	1.65	CB	-
I1005	HD11	0.66	CD1	-
I1005	HD12	0.66	CD1	-
I1005	HD13	0.66	CD1	-
I1005	HG11	0.98	CG1	-
I1005	HG12	0.84	CG1	-
I1005	HG21	0.76	CG2	-
I1005	HG22	0.76	CG2	-
I1005	HG23	0.76	CG2	-
I1005	HN	8.85	N	129.1
L1006	HA	4.61	CA	-
L1006	HB1	1.58	CB	-
L1006	HB2	1.58	CB	-
L1006	HD11	0.67	CD1	-
L1006	HD12	0.67	CD1	-
L1006	HD13	0.67	CD1	-
L1006	HD21	0.67	CD2	-
L1006	HD22	0.67	CD2	-
L1006	HD23	0.67	CD2	-
L1006	HG	1.30	CG	-
L1006	HN	8.24	N	122.5
S1007	HA	4.60	CA	-
S1007	HB1	3.72	CB	-
S1007	HB2	3.63	CB	-
S1007	HN	8.50	N	121.4
G1008	HA1	3.95	CA	-
G1008	HA2	3.61	CA	-
G1008	HN	-	N	-
N1009	HA	4.68	CA	-
N1009	HB1	2.88	CB	-
N1009	HB2	2.75	CB	-
N1009	HN	-	N	-
N1009	HD21	7.48	ND2	-
N1009	HD22	6.81	ND2	-
T1010	HA	4.39	CA	-
T1010	HB	4.05	CB	-
T1010	HG21	1.08	CG2	-
T1010	HG22	1.08	CG2	-
T1010	HG23	1.08	CG2	-
T1010	HN	7.79	N	114.9
A1011	HA	5.09	CA	-
A1011	HB1	0.80	CB	-
A1011	HB2	0.80	CB	-
A1011	HB3	0.80	CB	-
A1011	HN	8.40	N	129.4
H1012	HA	4.68	CA	-
H1012	HB1	3.15	CB	-
H1012	HB2	3.06	CB	-
H1012	HD2	6.99	CD2	-
H1012	HE1	8.32	CE1	-
H1012	HN	9.08	N	120.9
H1012	HD1	-	ND1	-
H1012	HE2	-	NE2	-
W1013	HA	4.97	CA	-
W1013	HB1	3.32	CB	-
W1013	HB2	3.15	CB	-
W1013	HD1	7.40	CD1	-
W1013	HE3	7.12	CE3	-
W1013	HH2	6.88	CH2	-
W1013	HZ2	7.17	CZ2	-
W1013	HZ3	6.91	CZ3	-
W1013	HN	8.44	N	122.6
W1013	HE1	10.38	NE1	128.2
S1014	HA	3.93	CA	-
S1014	HB1	4.18	CB	-
S1014	HB2	4.13	CB	-
S1014	HN	9.67	N	118.3

Residue	Atom	Shift	Atom	Shift
T1015	HA	4.64	CA	-
T1015	HB	4.19	CB	-
T1015	HG21	0.91	CG2	-
T1015	HG22	0.91	CG2	-
T1015	HG23	0.91	CG2	-
T1015	HN	7.22	N	112.0
K1016	HA	4.53	CA	-
K1016	HB1	1.81	CB	-
K1016	HB2	1.81	CB	-
K1016	HD1	1.72	CD	-
K1016	HD2	1.72	CD	-
K1016	HE1	-	CE	-
K1016	HE2	-	CE	-
K1016	HG1	1.54	CG	-
K1016	HG2	1.54	CG	-
K1016	HN	8.30	N	122.8
K1016	HZ1	-	NZ	-
K1016	HZ2	-	NZ	-
K1016	HZ3	-	NZ	-
P1017	HA	-	CA	-
P1017	HB1	-	CB	-
P1017	HB2	-	CB	-
P1017	HD1	3.80	CD	-
P1017	HD2	3.51	CD	-
P1017	HG1	1.96	CG	-
P1017	HG2	1.82	CG	-
P1018	HA	4.58	CA	-
P1018	HB1	-	CB	-
P1018	HB2	-	CB	-
P1018	HD1	-	CD	-
P1018	HD2	-	CD	-
P1018	HG1	-	CG	-
P1018	HG2	-	CG	-
I1019	HA	4.35	CA	-
I1019	HB	1.74	CB	-
I1019	HD11	-	CD1	-
I1019	HD12	-	CD1	-
I1019	HD13	-	CD1	-
I1019	HG11	1.36	CG1	-
I1019	HG12	1.11	CG1	-
I1019	HG21	0.82	CG2	-
I1019	HG22	0.82	CG2	-
I1019	HG23	0.82	CG2	-
I1019	HN	7.85	N	112.8
C1020	HA	5.32	CA	-
C1020	HB1	3.12	CB	-
C1020	HB2	2.33	CB	-
C1020	HN	8.85	N	122.9
Q1021	HA	4.63	CA	-
Q1021	HB1	1.99	CB	-
Q1021	HB2	1.83	CB	-
Q1021	HG1	2.29	CG	-
Q1021	HG2	2.29	CG	-
Q1021	HN	9.12	N	125.9
Q1021	HE21	7.42	NE2	-
Q1021	HE22	6.76	NE2	-
R1022	HA	3.71	CA	-
R1022	HB1	1.61	CB	-
R1022	HB2	1.61	CB	-
R1022	HD1	2.99	CD	-
R1022	HD2	2.99	CD	-
R1022	HG1	1.34	CG	-
R1022	HG2	1.23	CG	-
R1022	HN	9.12	N	129.7
R1022	HE	-	NE	-
R1022	HH11	-	NH1	-
R1022	HH12	-	NH1	-
R1022	HH21	-	NH2	-
R1022	HH22	-	NH2	-
I1023	HA	4.17	CA	-
I1023	HB	1.58	CB	-
I1023	HD11	0.67	CD1	-
I1023	HD12	0.67	CD1	-
I1023	HD13	0.67	CD1	-
I1023	HG11	1.37	CG1	-
I1023	HG12	0.82	CG1	-
I1023	HG21	0.87	CG2	-
I1023	HG22	0.87	CG2	-
I1023	HG23	0.87	CG2	-
I1023	HN	7.89	N	128.7
P1024	HA	-	CA	-
P1024	HB1	-	CB	-
P1024	HB2	-	CB	-
P1024	HD1	-	CD	-
P1024	HD2	-	CD	-
P1024	HG1	-	CG	-
P1024	HG2	-	CG	-

Table A.1: Chemical shifts for CR1~16 spins.

Appendix B

Chemical shift table for CR1~2-3

The chemical shifts assigned to the proton, nitrogen and carbon spins of CR1~2-3 are listed in the table below.

Residue	Atom	Shift	Atom	Shift
E57	HA	-	CA	-
E57	HB1	-	CB	-
E57	HB2	-	CB	-
E57	HG1	-	CG	-
E57	HG2	-	CG	-
A58	HA	4.25	CA	52.9
A58	HB1	1.31	CB	19.1
A58	HB2	1.31	CB	19.1
A58	HB3	1.31	CB	19.1
A58	HN	-	N	-
E59	HA	4.15	CA	56.9
E59	HB1	1.93	CB	30.2
E59	HB2	1.98	CB	30.2
E59	HG1	2.22	CG	36.5
E59	HG2	2.22	CG	36.5
E59	HN	8.44	N	120.2
A60	HA	4.32	CA	52.1
A60	HB1	1.36	CB	19.5
A60	HB2	1.36	CB	19.5
A60	HB3	1.36	CB	19.5
A60	HN	8.18	N	125.1
K61	HA	4.31	CA	56.2
K61	HB1	1.62	CB	34.0
K61	HB2	1.71	CB	34.0
K61	HD1	1.58	CD	29.3
K61	HD2	1.58	CD	29.3
K61	HE1	2.86	CE	42.4
K61	HE2	2.86	CE	42.4
K61	HG1	1.37	CG	25.4
K61	HG2	1.41	CG	25.4
K61	HN	8.24	N	121.2
K61	HZ1	-	NZ	-
K61	HZ2	-	NZ	-
K61	HZ3	-	NZ	-
S62	HA	4.73	CA	57.9
S62	HB1	3.59	CB	65.6
S62	HB2	3.73	CB	65.6
S62	HN	8.20	N	115.0
C63	HA	4.08	CA	60.9
C63	HB1	2.19	CB	41.0
C63	HB2	2.19	CB	41.0
C63	HN	8.24	N	121.9

Residue	Atom	Shift	Atom	Shift
R64	HA	4.20	CA	55.8
R64	HB1	1.91	CB	30.6
R64	HB2	1.91	CB	30.6
R64	HD1	3.28	CD	43.8
R64	HD2	3.28	CD	43.8
R64	HG1	1.71	CG	27.4
R64	HG2	1.71	CG	27.4
R64	HN	8.39	N	118.1
R64	HE	-	NE	-
R64	HH11	-	NH1	-
R64	HH12	-	NH1	-
R64	HH21	-	NH2	-
R64	HH22	-	NH2	-
N65	HA	4.69	CA	52.2
N65	HB1	2.68	CB	38.3
N65	HB2	2.68	CB	38.3
N65	HN	8.57	N	122.0
N65	HD21	7.50	ND2	112.6
N65	HD22	6.79	ND2	112.6
P66	HA	-	CA	-
P66	HB1	-	CB	-
P66	HB2	-	CB	-
P66	HD1	-	CD	-
P66	HD2	-	CD	-
P66	HG1	-	CG	-
P66	HG2	-	CG	-
P67	HA	4.48	CA	61.7
P67	HB1	2.18	CB	32.3
P67	HB2	1.79	CB	32.3
P67	HD1	3.69	CD	50.7
P67	HD2	3.81	CD	50.7
P67	HG1	-	CG	-
P67	HG2	-	CG	-
D68	HA	4.62	CA	53.0
D68	HB1	2.20	CB	40.6
D68	HB2	2.64	CB	40.6
D68	HN	8.02	N	118.4
P69	HA	4.25	CA	62.1
P69	HB1	1.27	CB	31.1
P69	HB2	2.08	CB	31.1
P69	HD1	-	CD	-
P69	HD2	-	CD	-
P69	HG1	1.47	CG	27.8
P69	HG2	1.71	CG	27.8
V70	HA	3.64	CA	64.7
V70	HB	1.85	CB	30.7
V70	HG11	0.95	CG1	20.9
V70	HG12	0.95	CG1	20.9
V70	HG13	0.95	CG1	20.9
V70	HG21	0.95	CG2	22.0
V70	HG22	0.95	CG2	22.0
V70	HG23	0.95	CG2	22.0
V70	HN	8.11	N	125.4
N71	HA	3.93	CA	55.8
N71	HB1	1.31	CB	36.5
N71	HB2	2.43	CB	36.5
N71	HN	8.85	N	119.3
N71	HD21	6.81	ND2	116.7
N71	HD22	7.38	ND2	116.7
G72	HA1	4.22	CA	47.0
G72	HA2	4.39	CA	47.0
G72	HN	7.64	N	104.6
M73	HA	4.97	CA	54.2110
M73	HB1	1.76	CB	30.2410
M73	HB2	1.76	CB	30.2410
M73	HE1	-	CE	-
M73	HE2	-	CE	-
M73	HE3	-	CE	-
M73	HG1	2.54	CG	51.7
M73	HG2	2.54	CG	51.7
M73	HN	8.81	N	116.2
V74	HA	4.65	CA	59.4
V74	HB	1.78	CB	35.0
V74	HG11	0.26	CG1	24.1
V74	HG12	0.26	CG1	24.1
V74	HG13	0.26	CG1	24.1
V74	HG21	0.68	CG2	18.7
V74	HG22	0.68	CG2	18.7
V74	HG23	0.68	CG2	18.7
V74	HN	8.83	N	122.8
H75	HA	-	CA	-
H75	HB1	-	CB	-
H75	HB2	-	CB	-
H75	HD2	-	CD2	-
H75	HE1	-	CE1	-
H75	HN	-	N	-
H75	HD1	-	ND1	-
H75	HE2	-	NE2	-

Residue	Atom	Shift	Atom	Shift
V76	HA	4.13	CA	62.4
V76	HB	2.04	CB	31.5
V76	HG11	0.77	CG1	20.8
V76	HG12	0.77	CG1	20.8
V76	HG13	0.77	CG1	20.8
V76	HG21	0.84	CG2	21.7
V76	HG22	0.84	CG2	21.7
V76	HG23	0.84	CG2	21.7
V76	HN	8.68	N	126.8
I77	HA	4.09	CA	62.9
I77	HB	1.92	CB	38.8
I77	HD11	0.78	CD1	12.8
I77	HD12	0.78	CD1	12.8
I77	HD13	0.78	CD1	12.8
I77	HG11	1.18	CG1	27.6
I77	HG12	1.44	CG1	27.6
I77	HG21	0.92	CG2	17.8
I77	HG22	0.92	CG2	17.8
I77	HG23	0.92	CG2	17.8
I77	HN	7.81	N	126.8
K78	HA	4.51	CA	55.9
K78	HB1	1.77	CB	33.1
K78	HB2	2.03	CB	33.1
K78	HD1	1.73	CD	29.3
K78	HD2	1.73	CD	29.3
K78	HE1	2.98	CE	42.4
K78	HE2	2.98	CE	42.4
K78	HG1	1.45	CG	25.2
K78	HG2	1.45	CG	25.2
K78	HN	9.27	N	123.1
K78	HZ1	-	NZ	-
K78	HZ2	-	NZ	-
K78	HZ3	-	NZ	-
G79	HA1	3.90	CA	44.8
G79	HA2	4.35	CA	44.8
G79	HN	7.82	N	108.7
I80	HA	4.89	CA	60.8
I80	HB	2.42	CB	38.6
I80	HD11	0.92	CD1	13.6
I80	HD12	0.92	CD1	13.6
I80	HD13	0.92	CD1	13.6
I80	HG11	1.05	CG1	25.6
I80	HG12	1.27	CG1	25.6
I80	HG21	1.00	CG2	18.7
I80	HG22	1.00	CG2	18.7
I80	HG23	1.00	CG2	18.7
I80	HN	8.02	N	110.0
Q81	HA	4.53	CA	55.8
Q81	HB1	2.03	CB	30.0
Q81	HB2	2.13	CB	30.0
Q81	HG1	2.38	CG	34.3
Q81	HG2	2.38	CG	34.3
Q81	HN	7.73	N	117.6
Q81	HE21	7.42	NE2	111.5
Q81	HE22	6.80	NE2	111.5
F82	HA	3.82	CA	59.6
F82	HB1	2.82	CB	39.4
F82	HB2	2.98	CB	39.4
F82	HD1	-	CD1	-
F82	HD2	-	CD2	-
F82	HE1	-	CE1	-
F82	HE2	-	CE2	-
F82	HZ	-	CZ	-
F82	HN	8.2925	N	120.3
G83	HA1	3.19	CA	44.6
G83	HA2	3.93	CA	44.6
G83	HN	8.65	N	117.2
S84	HA	4.37	CA	60.0
S84	HB1	3.89	CB	64.5
S84	HB2	3.89	CB	64.5
S84	HN	8.32	N	117.4
Q85	HA	5.80	CA	54.3
Q85	HB1	1.95	CB	32.6
Q85	HB2	2.08	CB	32.6
Q85	HG1	2.38	CG	34.8
Q85	HG2	2.38	CG	34.8
Q85	HN	8.10	N	119.6
Q85	HE21	7.28	NE2	111.4
Q85	HE22	6.71	NE2	111.4
I86	HA	4.80	CA	59.6
I86	HB	1.29	CB	41.2
I86	HD11	-0.63	CD1	12.6
I86	HD12	-0.63	CD1	12.6
I86	HD13	-0.63	CD1	12.6
I86	HG11	0.48	CG1	24.9
I86	HG12	0.51	CG1	24.9
I86	HG21	0.68	CG2	19.9
I86	HG22	0.68	CG2	19.9
I86	HG23	0.68	CG2	19.9
I86	HN	9.03	N	118.9

Residue	Atom	Shift	Atom	Shift
K87	HA	5.18	CA	53.8
K87	HB1	1.56	CB	36.6
K87	HB2	1.90	CB	36.6
K87	HD1	1.60	CD	29.2
K87	HD2	1.60	CD	29.2
K87	HE1	2.80	CE	42.2
K87	HE2	2.80	CE	42.2
K87	HG1	1.35	CG	25.3
K87	HG2	1.35	CG	25.3
K87	HN	7.76	N	117.9
K87	HZ1	-	NZ	-
K87	HZ2	-	NZ	-
K87	HZ3	-	NZ	-
Y88	HA	5.45	CA	57.0
Y88	HB1	2.42	CB	44.1
Y88	HB2	2.69	CB	44.1
Y88	HD1	6.63	CD1	-
Y88	HD2	6.63	CD2	-
Y88	HE1	6.43	CE1	117.5
Y88	HE2	6.43	CE2	117.5
Y88	HN	8.40	N	118.7
S89	HA	4.63	CA	57.7
S89	HB1	3.83	CB	65.8
S89	HB2	3.98	CB	65.8
S89	HN	8.95	N	113.4
C90	HA	5.32	CA	53.4
C90	HB1	2.48	CB	41.9
C90	HB2	2.97	CB	41.9
C90	HN	8.65	N	115.8
T91	HA	3.98	CA	63.1
T91	HB	3.80	CB	69.9
T91	HG21	1.19	CG2	22.0
T91	HG22	1.19	CG2	22.0
T91	HG23	1.19	CG2	22.0
T91	HN	8.42	N	119.2
K92	HA	4.05	CA	58.0
K92	HB1	1.72	CB	31.8
K92	HB2	1.76	CB	31.8
K92	HD1	1.74	CD	29.0
K92	HD2	1.74	CD	29.0
K92	HE1	3.00	CE	42.3
K92	HE2	3.09	CE	42.3
K92	HG1	1.34	CG	24.3
K92	HG2	1.43	CG	24.3
K92	HN	8.00	N	123.9
K92	HZ1	-	NZ	-
K92	HZ2	-	NZ	-
K92	HZ3	-	NZ	-
G93	HA1	1.82	CA	43.8
G93	HA2	3.25	CA	43.8
G93	HN	8.81	N	114.2
Y94	HA	4.80	CA	56.5
Y94	HB1	2.29	CB	40.4
Y94	HB2	3.20	CB	40.4
Y94	HD1	6.58	CD1	133.0
Y94	HD2	6.58	CD2	133.0
Y94	HE1	6.55	CE1	117.6
Y94	HE2	6.55	CE2	117.6
Y94	HN	7.99	N	118.9
R95	HA	4.80	CA	53.0
R95	HB1	1.69	CB	33.1
R95	HB2	1.85	CB	33.1
R95	HD1	3.18	CD	44.4
R95	HD2	3.27	CD	44.4
R95	HG1	1.63	CG	26.6
R95	HG2	1.63	CG	26.6
R95	HN	9.51	N	119.5
R95	HE	-	NE	-
R95	HH11	-	NH1	-
R95	HH12	-	NH1	-
R95	HH21	-	NH2	-
R95	HH22	-	NH2	-
L96	HA	4.64	CA	56.5
L96	HB1	1.48	CB	43.2
L96	HB2	1.61	CB	43.2
L96	HD11	0.48	CD1	25.6
L96	HD12	0.48	CD1	25.6
L96	HD13	0.48	CD1	25.6
L96	HD21	0.63	CD2	26.2
L96	HD22	0.63	CD2	26.2
L96	HD23	0.63	CD2	26.2
L96	HG	1.21	CG	28.3
L96	HN	8.31	N	126.2

Residue	Atom	Shift	Atom	Shift
I97	HA	4.26	CA	61.1
I97	HB	1.89	CB	37.7
I97	HD11	0.80	CD1	13.0
I97	HD12	0.80	CD1	13.0
I97	HD13	0.80	CD1	13.0
I97	HG11	1.11	CG1	27.1
I97	HG12	1.45	CG1	27.1
I97	HG21	0.83	CG2	17.0
I97	HG22	0.83	CG2	17.0
I97	HG23	0.83	CG2	17.0
I97	HN	9.55	N	130.2
G98	HA1	3.61	CA	43.1
G98	HA2	4.53	CA	43.1
G98	HN	8.36	N	116.3
S99	HA	4.58	CA	58.1
S99	HB1	3.85	CB	64.2
S99	HB2	4.15	CB	64.2
S99	HN	8.56	N	116.6
S100	HA	4.42	CA	59.2
S100	HB1	3.92	CB	64.2
S100	HB2	4.31	CB	64.2
S100	HN	8.54	N	116.5
S101	HA	5.29	CA	57.1
S101	HB1	3.81	CB	66.3
S101	HB2	3.89	CB	66.3
S101	HN	7.86	N	115.2
A102	HA	5.13	CA	51.6
A102	HB1	1.56	CB	24.4
A102	HB2	1.56	CB	24.4
A102	HB3	1.56	CB	24.4
A102	HN	8.22	N	121.9
T103	HA	5.50	CA	61.4
T103	HB	3.79	CB	71.2
T103	HG21	1.03	CG2	21.5
T103	HG22	1.03	CG2	21.5
T103	HG23	1.03	CG2	21.5
T103	HN	9.32	N	117.4
C104	HA	4.27	CA	55.5
C104	HB1	1.62	CB	39.3
C104	HB2	2.57	CB	39.3
C104	HN	8.30	N	127.4
I105	HA	4.83	CA	60.2
I105	HB	2.01	CB	42.1
I105	HD11	0.89	CD1	14.5
I105	HD12	0.89	CD1	14.5
I105	HD13	0.89	CD1	14.5
I105	HG11	1.04	CG1	-
I105	HG12	1.29	CG1	-
I105	HG21	0.92	CG2	18.3
I105	HG22	0.92	CG2	18.3
I105	HG23	0.92	CG2	18.3
I105	HN	9.20	N	126.4
I106	HA	4.38	CA	61.2
I106	HB	1.55	CB	39.4
I106	HD11	0.40	CD1	13.5
I106	HD12	0.40	CD1	13.5
I106	HD13	0.40	CD1	13.5
I106	HG11	1.36	CG1	28.9
I106	HG12	0.81	CG1	28.9
I106	HG21	0.79	CG2	17.7
I106	HG22	0.79	CG2	17.7
I106	HG23	0.79	CG2	17.7
I106	HN	7.93	N	118.4
S107	HA	4.48	CA	56.8
S107	HB1	3.42	CB	63.9
S107	HB2	3.50	CB	63.9
S107	HN	8.32	N	121.3
G108	HA1	3.60	CA	47.6
G108	HA2	3.96	CA	47.6
G108	HN	-	N	-
D109	HA	4.65	CA	54.0
D109	HB1	2.67	CB	41.1
D109	HB2	2.69	CB	41.1
D109	HN	8.64	N	126.1
T110	HA	4.65	CA	59.6
T110	HB	4.08	CB	71.1
T110	HG21	1.20	CG2	20.4
T110	HG22	1.20	CG2	20.4
T110	HG23	1.20	CG2	20.4
T110	HN	7.94	N	114.4
V111	HA	5.11	CA	59.7
V111	HB	1.91	CB	32.9
V111	HG11	0.57	CG1	20.4
V111	HG12	0.57	CG1	20.4
V111	HG13	0.57	CG1	20.4
V111	HG21	0.57	CG2	23.0
V111	HG22	0.57	CG2	23.0
V111	HG23	0.57	CG2	23.0
V111	HN	8.11	N	119.0

Residue	Atom	Shift	Atom	Shift
I112	HA	4.49	CA	59.6
I112	HB	1.93	CB	43.3
I112	HD11	0.82	CD1	13.7
I112	HD12	0.82	CD1	13.7
I112	HD13	0.82	CD1	13.7
I112	HG11	1.20	CG1	27.6
I112	HG12	1.44	CG1	27.6
I112	HG21	0.87	CG2	18.3
I112	HG22	0.87	CG2	18.3
I112	HG23	0.87	CG2	18.3
I112	HN	8.84	N	119.1
W113	HA	5.07	CA	56.5
W113	HB1	3.08	CB	30.4
W113	HB2	3.37	CB	30.4
W113	HD1	7.28	CD1	126.6
W113	HE3	7.12	CE3	120.7
W113	HH2	6.91	CH2	124.3
W113	HZ2	7.18	CZ2	113.4
W113	HZ3	6.92	CZ3	123.2
W113	HN	7.49	N	122.0
W113	HE1	10.48	NE1	127.9
D114	HA	4.16	CA	56.0
D114	HB1	2.84	CB	40.1
D114	HB2	3.09	CB	40.1
D114	HN	9.31	N	123.1
T115	HA	4.64	CA	60.3
T115	HB	4.27	CB	71.4
T115	HG21	1.37	CG2	21.4
T115	HG22	1.37	CG2	21.4
T115	HG23	1.37	CG2	21.4
T115	HN	7.51	N	112.4
E116	HA	4.37	CA	55.7
E116	HB1	1.88	CB	30.9
E116	HB2	2.04	CB	30.9
E116	HG1	2.32	CG	36.5
E116	HG2	2.32	CG	36.5
E116	HN	8.20	N	122.3
T117	HA	3.64	CA	61.8
T117	HB	3.74	CB	70.0
T117	HG21	1.10	CG2	21.3
T117	HG22	1.10	CG2	21.3
T117	HG23	1.10	CG2	21.3
T117	HN	8.09	N	120.6
P118	HA	4.53	CA	62.5
P118	HB1	-	CB	32.8
P118	HB2	-	CB	32.8
P118	HD1	-	CD	-
P118	HD2	-	CD	-
P118	HG1	-	CG	-
P118	HG2	-	CG	-
I119	HA	4.24	CA	59.1
I119	HB	1.73	CB	41.2
I119	HD11	0.81	CD1	13.2
I119	HD12	0.81	CD1	13.2
I119	HD13	0.81	CD1	13.2
I119	HG11	1.15	CG1	27.7
I119	HG12	1.43	CG1	27.7
I119	HG21	0.81	CG2	17.9
I119	HG22	0.81	CG2	17.9
I119	HG23	0.81	CG2	17.9
I119	HN	7.78	N	114.5
C120	HA	5.41	CA	53.9
C120	HB1	2.42	CB	41.2
C120	HB2	3.15	CB	41.2
C120	HN	8.61	N	123.1
D121	HA	5.18	CA	52.5
D121	HB1	2.42	CB	45.6
D121	HB2	2.63	CB	45.6
D121	HN	9.33	N	126.0
R122	HA	3.39	CA	57.5
R122	HB1	1.44	CB	30.8
R122	HB2	1.44	CB	30.8
R122	HD1	2.57	CD	43.5
R122	HD2	2.64	CD	43.5
R122	HG1	1.19	CG	27.5
R122	HG2	1.19	CG	27.5
R122	HN	8.52	N	127.2
R122	HE	-	NE	-
R122	HH11	-	NH1	-
R122	HH12	-	NH1	-
R122	HH21	-	NH2	-
R122	HH22	-	NH2	-
I123	HA	3.84	CA	60.9
I123	HB	1.37	CB	39.9
I123	HD11	0.39	CD1	13.3
I123	HD12	0.39	CD1	13.3
I123	HD13	0.39	CD1	13.3
I123	HG11	0.65	CG1	30.4
I123	HG12	1.33	CG1	30.4
I123	HG21	0.64	CG2	16.5
I123	HG22	0.64	CG2	16.5
I123	HG23	0.64	CG2	16.5
I123	HN	8.03	N	127.0

Residue	Atom	Shift	Atom	Shift
P124	HA	5.03	CA	62.2
P124	HB1	1.88	CB	32.7
P124	HB2	2.11	CB	32.7
P124	HD1	-	CD	51.0
P124	HD2	-	CD	51.0
P124	HG1	-	CG	-
P124	HG2	-	CG	-
C125	HA	4.47	CA	58.4
C125	HB1	1.89	CB	47.7
C125	HB2	2.89	CB	47.7
C125	HN	7.76	N	114.9
G126	HA1	3.80	CA	44.1
G126	HA2	4.32	CA	44.1
G126	HN	8.20	N	107.1
L127	HA	4.09	CA	54.3
L127	HB1	1.44	CB	40.6
L127	HB2	1.58	CB	40.6
L127	HD11	1.09	CD1	25.5
L127	HD12	1.09	CD1	25.5
L127	HD13	1.09	CD1	25.5
L127	HD21	1.19	CD2	23.9
L127	HD22	1.19	CD2	23.9
L127	HD23	1.19	CD2	23.9
L127	HG	1.92	CG	27.6
L127	HN	8.00	N	117.5
P128	HA	-	CA	-
P128	HB1	-	CB	-
P128	HB2	-	CB	-
P128	HD1	-	CD	-
P128	HD2	-	CD	-
P128	HG1	-	CG	-
P128	HG2	-	CG	-
P129	HA	4.19	CA	62.7
P129	HB1	1.60	CB	31.9
P129	HB2	2.18	CB	31.9
P129	HD1	2.83	CD	49.8
P129	HD2	2.83	CD	49.8
P129	HG1	1.79	CG	27.6
P129	HG2	1.79	CG	27.6
T130	HA	4.37	CA	62.2
T130	HB	4.12	CB	70.1
T130	HG21	1.22	CG2	22.1
T130	HG22	1.22	CG2	22.1
T130	HG23	1.22	CG2	22.1
T130	HN	7.98	N	115.9
I131	HA	4.60	CA	58.7
I131	HB	1.57	CB	41.2
I131	HD11	0.47	CD1	15.7
I131	HD12	0.47	CD1	15.7
I131	HD13	0.47	CD1	15.7
I131	HG11	0.75	CG1	26.6
I131	HG12	1.20	CG1	26.6
I131	HG21	0.73	CG2	16.6
I131	HG22	0.73	CG2	16.6
I131	HG23	0.73	CG2	16.6
I131	HN	8.65	N	119.7
T132	HA	4.04	CA	63.2
T132	HB	3.89	CB	68.8
T132	HG21	1.16	CG2	22.4
T132	HG22	1.16	CG2	22.4
T132	HG23	1.16	CG2	22.4
T132	HN	8.05	N	121.1
N133	HA	4.03	CA	55.1
N133	HB1	1.91	CB	36.1
N133	HB2	2.44	CB	36.1
N133	HN	8.99	N	117.9
N133	HD21	6.79	ND2	116.8
N133	HD22	7.26	ND2	116.8
G134	HA1	4.01	CA	46.3
G134	HA2	4.40	CA	46.3
G134	HN	7.06	N	102.5
D135	HA	4.83	CA	52.7
D135	HB1	2.50	CB	43.7
D135	HB2	2.69	CB	43.7
D135	HN	8.74	N	118.4
F136	HA	5.33	CA	55.0
F136	HB1	1.90	CB	41.3
F136	HB2	2.19	CB	41.3
F136	HD1	6.74	CD1	132.0
F136	HD2	6.74	CD2	132.0
F136	HE1	7.22	CE1	131.2
F136	HE2	7.22	CE2	131.2
F136	HZ	-	CZ	-
F136	HN	8.42	N	114.4

Residue	Atom	Shift	Atom	Shift
I137	HA	4.31	CA	60.1
I137	HB	1.76	CB	40.3
I137	HD11	0.74	CD1	13.5
I137	HD12	0.74	CD1	13.5
I137	HD13	0.74	CD1	13.5
I137	HG11	1.03	CG1	27.3
I137	HG12	1.38	CG1	27.3
I137	HG21	0.82	CG2	17.7
I137	HG22	0.82	CG2	17.7
I137	HG23	0.82	CG2	17.7
I137	HN	9.07	N	119.5
S138	HA	4.87	CA	57.3
S138	HB1	4.00	CB	64.9
S138	HB2	4.39	CB	64.9
S138	HN	8.14	N	117.9
T139	HA	4.15	CA	64.7
T139	HB	4.30	CB	68.9
T139	HG21	1.32	CG2	22.4
T139	HG22	1.32	CG2	22.4
T139	HG23	1.32	CG2	22.4
T139	HN	-	N	-
N140	HA	4.78	CA	52.9
N140	HB1	2.61	CB	40.5
N140	HB2	2.79	CB	40.5
N140	HN	8.00	N	118.4
N140	HD21	6.92	ND2	113.4
N140	HD22	7.46	ND2	113.4
R141	HA	4.43	CA	56.5
R141	HB1	1.74	CB	30.8
R141	HB2	2.10	CB	30.8
R141	HD1	3.11	CD	43.1
R141	HD2	3.19	CD	43.1
R141	HG1	1.60	CG	27.4
R141	HG2	1.71	CG	27.4
R141	HN	8.55	N	119.8
R141	HE	-	NE	-
R141	HH11	-	NH1	-
R141	HH12	-	NH1	-
R141	HH21	-	NH2	-
R141	HH22	-	NH2	-
E142	HA	4.38	CA	56.6
E142	HB1	1.90	CB	31.6
E142	HB2	2.08	CB	31.6
E142	HG1	2.15	CG	36.4
E142	HG2	2.15	CG	36.4
E142	HN	7.92	N	115.9
N143	HA	4.60	CA	52.9
N143	HB1	2.48	CB	40.1
N143	HB2	2.56	CB	40.1
N143	HN	7.56	N	116.8
N143	HD21	6.84	ND2	113.0
N143	HD22	7.42	ND2	113.0
F144	HA	4.44	CA	57.6
F144	HB1	2.90	CB	41.6
F144	HB2	2.90	CB	41.6
F144	HD1	7.22	CD1	131.9
F144	HD2	7.22	CD2	131.9
F144	HE1	7.18	CE1	131.9
F144	HE2	7.18	CE2	131.9
F144	HZ	-	CZ	-
F144	HN	8.41	N	120.4
H145	HA	-	CA	-
H145	HB1	-	CB	-
H145	HB2	-	CB	-
H145	HD2	-	CD2	-
H145	HE1	-	CE1	-
H145	HN	-	N	-
H145	HD1	-	ND1	-
H145	HE2	-	NE2	-
Y146	HA	3.56	CA	60.8
Y146	HB1	2.73	CB	39.6
Y146	HB2	2.87	CB	39.6
Y146	HD1	6.76	CD1	133.2
Y146	HD2	6.76	CD2	133.2
Y146	HE1	6.82	CE1	118.3
Y146	HE2	6.82	CE2	118.3
Y146	HN	8.66	N	121.9
G147	HA1	3.44	CA	44.7
G147	HA2	4.48	CA	44.7
G147	HN	8.89	N	118.4
S148	HA	4.46	CA	61.5
S148	HB1	4.02	CB	64.6
S148	HB2	4.25	CB	64.6
S148	HN	9.04	N	118.8
V149	HA	4.92	CA	61.2
V149	HB	1.96	CB	34.7
V149	HG11	0.94	CG1	20.9
V149	HG12	0.94	CG1	20.9
V149	HG13	0.94	CG1	20.9
V149	HG21	0.94	CG2	21.6
V149	HG22	0.94	CG2	21.6
V149	HG23	0.94	CG2	21.6
V149	HN	8.46	N	122.9

Residue	Atom	Shift	Atom	Shift
V150	HA	4.60	CA	60.9
V150	HB	1.37	CB	34.3
V150	HG11	0.16	CG1	21.5
V150	HG12	0.16	CG1	21.5
V150	HG13	0.16	CG1	21.5
V150	HG21	0.16	CG2	21.6
V150	HG22	0.16	CG2	21.6
V150	HG23	0.16	CG2	21.6
V150	HN	9.02	N	126.7
T151	HA	5.12	CA	61.3
T151	HB	3.74	CB	71.2
T151	HG21	1.19	CG2	21.6
T151	HG22	1.19	CG2	21.6
T151	HG23	1.19	CG2	21.6
T151	HN	8.05	N	120.1
Y152	HA	4.96	CA	58.8
Y152	HB1	2.61	CB	41.7
Y152	HB2	2.71	CB	41.7
Y152	HD1	6.96	CD1	133.3
Y152	HD2	6.96	CD2	133.3
Y152	HE1	6.40	CE1	117.5
Y152	HE2	6.40	CE2	117.5
Y152	HN	9.23	N	126.5
R153	HA	4.57	CA	54.7
R153	HB1	1.77	CB	33.1
R153	HB2	1.77	CB	33.1
R153	HD1	3.06	CD	43.9
R153	HD2	3.06	CD	43.9
R153	HG1	1.53	CG	26.1
R153	HG2	1.53	CG	26.1
R153	HN	8.65	N	115.0
R153	HE	-	NE	-
R153	HH11	-	NH1	-
R153	HH12	-	NH1	-
R153	HH21	-	NH2	-
R153	HH22	-	NH2	-
C154	HA	5.22	CA	53.5
C154	HB1	2.70	CB	39.7
C154	HB2	3.09	CB	39.7
C154	HN	8.85	N	119.5
N155	HA	4.79	CA	51.5
N155	HB1	2.57	CB	37.3
N155	HB2	2.87	CB	37.3
N155	HN	8.92	N	124.1
N155	HD21	7.61	ND2	110.0
N155	HD22	6.90	ND2	110.0
P156	HA	4.65	CA	62.8
P156	HB1	1.97	CB	32.5
P156	HB2	2.22	CB	32.5
P156	HD1	3.65	CD	50.7
P156	HD2	3.92	CD	50.7
P156	HG1	2.01	CG	27.3
P156	HG2	2.01	CG	27.3
G157	HA1	3.55	CA	44.3
G157	HA2	4.22	CA	44.3
G157	HN	8.52	N	107.9
S158	HA	-	CA	-
S158	HB1	-	CB	-
S158	HB2	-	CB	-
S158	HN	-	N	-
G159	HA1	3.85	CA	46.3
G159	HA2	3.97	CA	46.3
G159	HN	-	N	-
G160	HA1	3.67	CA	45.4
G160	HA2	4.12	CA	45.4
G160	HN	8.25	N	107.7
R161	HA	4.16	CA	55.9
R161	HB1	1.68	CB	31.1
R161	HB2	1.68	CB	31.1
R161	HD1	3.12	CD	43.5
R161	HD2	3.12	CD	43.5
R161	HG1	1.53	CG	27.5
R161	HG2	1.53	CG	27.5
R161	HN	7.12	N	120.0
R161	HE	-	NE	-
R161	HH11	-	NH1	-
R161	HH12	-	NH1	-
R161	HH21	-	NH2	-
R161	HH22	-	NH2	-
K162	HA	4.18	CA	57.1
K162	HB1	1.68	CB	32.5
K162	HB2	1.72	CB	32.5
K162	HD1	1.61	CD	29.6
K162	HD2	1.61	CD	29.6
K162	HE1	2.88	CE	42.2
K162	HE2	2.88	CE	42.2
K162	HG1	1.13	CG	24.9
K162	HG2	1.35	CG	24.9
K162	HN	8.26	N	123.3
K162	HZ1	-	NZ	-
K162	HZ2	-	NZ	-
K162	HZ3	-	NZ	-

Residue	Atom	Shift	Atom	Shift
V163	HA	4.07	CA	62.0
V163	HB	1.69	CB	33.8
V163	HG11	0.54	CG1	19.9
V163	HG12	0.54	CG1	19.9
V163	HG13	0.54	CG1	19.9
V163	HG21	0.55	CG2	21.3
V163	HG22	0.55	CG2	21.3
V163	HG23	0.55	CG2	21.3
V163	HN	8.45	N	122.9
F164	HA	4.87	CA	57.1
F164	HB1	2.66	CB	42.3
F164	HB2	3.04	CB	42.3
F164	HD1	7.26	CD1	-
F164	HD2	7.26	CD2	-
F164	HE1	-	CE1	-
F164	HE2	-	CE2	-
F164	HZ	-	CZ	-
F164	HN	8.15	N	121.5
E165	HA	4.42	CA	54.7
E165	HB1	1.82	CB	31.9
E165	HB2	1.94	CB	31.9
E165	HG1	2.17	CG	35.9
E165	HG2	2.17	CG	35.9
E165	HN	9.55	N	122.8
L166	HA	4.87	CA	55.3
L166	HB1	1.66	CB	42.4
L166	HB2	1.80	CB	42.4
L166	HD11	0.72	CD1	26.7
L166	HD12	0.72	CD1	26.7
L166	HD13	0.72	CD1	26.7
L166	HD21	0.72	CD2	26.7
L166	HD22	0.72	CD2	26.7
L166	HD23	0.72	CD2	26.7
L166	HG	1.30	CG	28.1
L166	HN	8.19	N	126.9
V167	HA	4.25	CA	62.5
V167	HB	1.86	CB	33.7
V167	HG11	0.86	CG1	20.9
V167	HG12	0.86	CG1	20.9
V167	HG13	0.86	CG1	20.9
V167	HG21	0.86	CG2	20.9
V167	HG22	0.86	CG2	20.9
V167	HG23	0.86	CG2	20.9
V167	HN	9.55	N	130.6
G168	HA1	3.66	CA	42.6
G168	HA2	4.48	CA	42.6
G168	HN	8.43	N	116.6
E169	HA	4.54	CA	53.7
E169	HB1	1.99	CB	29.6
E169	HB2	2.08	CB	29.6
E169	HG1	2.36	CG	36.4
E169	HG2	2.43	CG	36.4
E169	HN	7.96	N	120.9
P170	HA	4.29	CA	64.8
P170	HB1	2.18	CB	33.1
P170	HB2	2.34	CB	33.1
P170	HD1	3.91	CD	51.1
P170	HD2	3.96	CD	51.1
P170	HG1	2.05	CG	27.4
P170	HG2	2.12	CG	27.4
S171	HA	5.93	CA	57.0
S171	HB1	3.67	CB	65.6
S171	HB2	3.67	CB	65.6
S171	HN	7.78	N	110.8
I172	HA	4.56	CA	60.1
I172	HB	1.98	CB	42.9
I172	HD11	0.46	CD1	13.9
I172	HD12	0.46	CD1	13.9
I172	HD13	0.46	CD1	13.9
I172	HG11	0.91	CG1	25.6
I172	HG12	1.34	CG1	25.6
I172	HG21	0.68	CG2	17.7
I172	HG22	0.68	CG2	17.7
I172	HG23	0.68	CG2	17.7
I172	HN	8.64	N	116.7
Y173	HA	6.00	CA	55.1
Y173	HB1	2.69	CB	42.3
Y173	HB2	2.85	CB	42.3
Y173	HD1	6.83	CD1	134.0
Y173	HD2	6.83	CD2	134.0
Y173	HE1	6.67	CE1	117.8
Y173	HE2	6.67	CE2	117.8
Y173	HN	8.83	N	115.9
C174	HA	3.75	CA	54.6
C174	HB1	1.36	CB	40.0
C174	HB2	2.55	CB	40.0
C174	HN	8.36	N	122.4
T175	HA	4.39	CA	58.1
T175	HB	3.78	CB	70.3
T175	HG21	0.87	CG2	18.5
T175	HG22	0.87	CG2	18.5
T175	HG23	0.87	CG2	18.5
T175	HN	9.09	N	129.0

Residue	Atom	Shift	Atom	Shift
S176	HA	4.92	CA	58.0
S176	HB1	3.58	CB	65.7
S176	HB2	3.92	CB	65.7
S176	HN	8.42	N	112.7
N177	HA	-	CA	-
N177	HB1	-	CB	-
N177	HB2	-	CB	-
N177	HN	-	N	-
N177	HD21	7.50	ND2	111.5
N177	HD22	6.84	ND2	111.5
D178	HA	-	CA	-
D178	HB1	-	CB	-
D178	HB2	-	CB	-
D178	HN	8.31	N	119.0
D179	HA	-	CA	-
D179	HB1	-	CB	-
D179	HB2	-	CB	-
D179	HN	-	N	-
Q180	HA	-	CA	-
Q180	HB1	-	CB	-
Q180	HB2	-	CB	-
Q180	HG1	-	CG	-
Q180	HG2	-	CG	-
Q180	HN	8.42	N	114.4
Q180	HE21	7.34	NE2	111.3
Q180	HE22	6.84	NE2	111.3
V181	HA	4.34	CA	61.0
V181	HB	2.10	CB	34.8
V181	HG11	0.94	CG1	21.0
V181	HG12	0.94	CG1	21.0
V181	HG13	0.94	CG1	21.0
V181	HG21	0.94	CG2	21.4
V181	HG22	0.94	CG2	21.4
V181	HG23	0.94	CG2	21.4
V181	HN	-	N	-
G182	HA1	3.65	CA	45.3
G182	HA2	4.65	CA	45.3
G182	HN	8.54	N	116.4
I183	HA	4.36	CA	58.9035
I183	HB	1.73	CB	42.5
I183	HD11	0.73	CD1	13.9
I183	HD12	0.73	CD1	13.9
I183	HD13	0.73	CD1	13.9
I183	HG11	0.88	CG1	26.2
I183	HG12	1.17	CG1	26.2
I183	HG21	0.78	CG2	18.3
I183	HG22	0.78	CG2	18.3
I183	HG23	0.78	CG2	18.3
I183	HN	8.53	N	122.5
W184	HA	4.70	CA	57.4
W184	HB1	2.95	CB	30.0
W184	HB2	3.18	CB	30.0
W184	HD1	7.07	CD1	126.3
W184	HE3	6.93	CE3	119.8
W184	HH2	6.41	CH2	123.0
W184	HZ2	6.77	CZ2	113.6
W184	HZ3	6.68	CZ3	121.8
W184	HN	8.16	N	122.7
W184	HE1	8.94	NE1	128.1
S185	HA	3.63	CA	61.0
S185	HB1	2.92	CB	62.1
S185	HB2	3.72	CB	62.1
S185	HN	9.04	N	115.8
G186	HA1	4.11	CA	44.8
G186	HA2	4.22	CA	44.8
G186	HN	7.18	N	108.6
P187	HA	4.42	CA	61.6
P187	HB1	2.13	CB	32.4
P187	HB2	1.95	CB	32.4
P187	HD1	3.44	CD	49.1
P187	HD2	3.55	CD	49.1
P187	HG1	1.77	CG	26.3
P187	HG2	1.93	CG	26.3
A188	HA	2.71	CA	50.2
A188	HB1	0.68	CB	16.9
A188	HB2	0.68	CB	16.9
A188	HB3	0.68	CB	16.9
A188	HN	7.85	N	120.9
P189	HA	4.54	CA	62.3
P189	HB1	-	CB	32.1
P189	HB2	-	CB	32.1
P189	HD1	2.13	CD	49.3
P189	HD2	3.03	CD	49.3
P189	HG1	-	CG	-
P189	HG2	-	CG	-
Q190	HA	4.55	CA	54.1
Q190	HB1	1.94	CB	33.3
Q190	HB2	1.68	CB	33.3
Q190	HG1	2.17	CG	34.5
Q190	HG2	2.17	CG	34.5
Q190	HN	8.00	N	112.1
Q190	HE21	7.30	NE2	111.0
Q190	HE22	6.64	NE2	111.0

Residue	Atom	Shift	Atom	Shift
C191	HA	5.31	CA	56.3
C191	HB1	2.22	CB	43.2
C191	HB2	2.71	CB	43.2
C191	HN	8.74	N	121.7
I192	HA	4.41	CA	62.4
I192	HB	1.88	CB	41.1
I192	HD11	0.79	CD1	13.3
I192	HD12	0.79	CD1	13.3
I192	HD13	0.79	CD1	13.3
I192	HG11	1.09	CG1	26.9
I192	HG12	1.41	CG1	26.9
I192	HG21	0.87	CG2	18.2
I192	HG22	0.87	CG2	18.2
I192	HG23	0.87	CG2	18.2
I192	HN	9.04	N	127.1

Table B.1: Chemical shifts for CR1~2-3 spins.

Appendix C

Relaxation data table for CR1~16

The relaxation data for CR1~16 spins are listed in the table below.

Residue	R_1	dR_1	R_2	dR_2	NOE	dNOE
K961						
S962	2.013	0.028	4.334	0.143	0.596	0.052
C963	2.232	0.028	5.556	0.146	0.590	0.051
K964	2.199	0.028	6.134	0.145	0.632	0.055
T965	2.056	0.029	7.155	0.144	0.593	0.052
P966						
P967						
D968	1.996	0.028	4.875	0.144	0.665	0.058
V970	2.136	0.028	5.038	0.140	0.768	0.067
N971	2.292	0.028	5.701	0.145	0.709	0.062
G972	2.277	0.028	5.784	0.147	0.729	0.064
M973						
V974	2.276	0.028	5.217	0.141	0.675	0.059
H975	2.224	0.028	6.188	0.146	0.639	0.056
V976	2.149	0.028	5.377	0.147	0.696	0.061
I977	2.058	0.028	3.381	0.153	0.607	0.053
T978	2.174	0.028	7.127	0.154	0.641	0.056
D979	2.075	0.028	4.625	0.139	0.651	0.057
I980	2.201	0.028	9.790	0.160	0.662	0.058
Q981	2.328	0.028	5.691	0.140	0.670	0.058
V982	2.126	0.028	6.203	0.145	0.762	0.066
G983	2.324	0.028	6.243	0.143	0.733	0.064
S984	2.327	0.028	5.977	0.141	0.733	0.064
R985	2.202	0.028	5.936	0.146	0.625	0.055
I986	2.272	0.028	6.306	0.141	0.695	0.061
T987	2.402	0.028	7.772	0.143	0.660	0.058
Y988	2.220	0.028	5.505	0.140	0.684	0.060
S989	2.303	0.028	4.673	0.150	0.681	0.059
C990						
T991	2.237	0.028	4.420	0.140	0.654	0.057
T992	2.227	0.029	5.585	0.144	0.663	0.058
G993						
H994	2.259	0.028	5.451	0.142	0.699	0.061
R995	2.251	0.028	5.081	0.143	0.696	0.061
L996	2.225	0.028	5.586	0.154	0.687	0.060
I997	2.324	0.028	5.620	0.155	0.693	0.060
G998	2.207	0.028	5.466	0.143	0.766	0.067
H999	2.165	0.028	6.009	0.145	0.703	0.061
S1000	2.229	0.029	5.427	0.145	0.651	0.057
S1001	2.235	0.028	4.895	0.142	0.726	0.063
A1002	2.240	0.028	5.528	0.144	0.708	0.062
E1003	2.282	0.028	5.842	0.144	0.693	0.060
C1004	2.297	0.028	5.944	0.148	0.709	0.062
I1005	2.403	0.028	4.380	0.159	0.682	0.059
L1006	2.148	0.028	5.571	0.140	0.631	0.055
S1007	2.186	0.029	5.895	0.145	0.508	0.044
G1008						
N1009						
T1010	2.059	0.028	5.509	0.150	0.633	0.055
A1011	2.196	0.028	3.152	0.134	0.604	0.053
H1012	2.259	0.028	6.112	0.147	0.662	0.058
W1013	2.249	0.028	5.142	0.140	0.719	0.063
S1014	2.282	0.028	4.812	0.140	0.734	0.064
T1015	2.181	0.028	2.881	0.125	0.637	0.056
K1016	2.191	0.028	4.896	0.140	0.628	0.055
P1017						
P1018						
I1019	2.313	0.028	6.120	0.156	0.623	0.054
C1020	2.188	0.028	4.683	0.138	0.640	0.056
Q1021	2.135	0.028	5.652	0.149	0.637	0.056
R1022	2.356	0.029	6.958	0.157	0.691	0.060
I1023	2.153	0.028	1.863	0.128	0.539	0.047
P1024						

Table C.1: Relaxation data for CR1~16 spins.

Appendix D

Relaxation data table for CR1~2-3

The relaxation data for CR1~2-3 spins are listed in the table below.

Residue	R_1	dR_1	R_2	dR_2	NOE	dNOE
A60	0.677	0.030	4.856	0.148	-0.424	-0.065
K61						
S62	0.984	0.034	9.559	0.218	0.571	0.126
C63						
R64	1.371	0.021	10.656	0.138	0.866	0.039
N65	0.969	0.058	12.237	0.418	0.660	0.072
P66						
P67						
D68						
P69						
V70	1.383	0.017	11.305	0.131	0.876	0.050
N71						
G72	1.284	0.029	11.691	0.216	0.745	0.053
M73	1.477	0.025	10.889	0.158	0.682	0.051
V74						
H75						
V76	0.990	0.059	18.551	0.782	0.881	0.107
I77	1.160	0.026	7.639	0.140	0.532	0.055
K78						
G79						
I80	1.132	0.032	22.483	0.580	0.873	0.074
Q81	1.233	0.031	14.698	0.303	0.806	0.050
F82	0.964	0.021	10.993	0.161	0.622	0.046
G83	1.351	0.030	16.981	0.373	0.752	0.056
S84	1.247	0.017	11.400	0.127	0.811	0.044
Q85	1.318	0.014	8.606	0.092	0.663	0.042
I86						
K87	1.443	0.033	14.970	0.324	0.842	0.058
Y88	1.417	0.019	11.365	0.142	0.876	0.052
S89	1.360	0.036	10.690	0.250	0.616	0.069
C90	1.264	0.030	11.167	0.228	0.653	0.046
T91	1.483	0.021	11.942	0.173	0.712	0.046
K92	1.321	0.022	10.207	0.151	0.734	0.055
G93	1.364	0.042	10.858	0.304	0.859	0.132
Y94	1.415	0.016	10.555	0.107	0.799	0.037
R95	1.496	0.026	10.270	0.173	0.667	0.050
L96	1.369	0.022	9.863	0.151	0.697	0.051
I97						
G98	1.203	0.024	12.516	0.183	0.924	0.059
S99						
S100						
S101	1.363	0.014	10.630	0.090	0.627	0.033
A102						
T103	1.322	0.028	10.801	0.189	0.768	0.057
C104	1.513	0.028	13.055	0.219	0.681	0.060
I105						
I106	1.501	0.025	13.319	0.218	0.622	0.053
S107	1.333	0.015	7.884	0.088	0.502	0.032
G108						
D109	0.826	0.091	12.065	0.745	0.703	0.100
T110	1.171	0.012	8.588	0.074	0.536	0.031
V111	1.326	0.012	9.104	0.078	0.589	0.028
I112						
W113	1.401	0.033	11.429	0.243	0.606	0.065
W113*s*	1.154	0.020	9.266	0.141	0.636	0.045
D114						
T115	1.180	0.014	7.890	0.088	0.442	0.031
E116	1.149	0.013	8.412	0.075	0.507	0.023
T117	1.051	0.014	8.967	0.087	0.661	0.034
P118						
I119	1.484	0.024	11.358	0.168	0.779	0.037
C120	1.458	0.022	9.667	0.150	0.803	0.049
D121	1.424	0.025	9.854	0.172	0.733	0.053
R122	1.384	0.024	10.966	0.170	0.595	0.051

Residue	R_1	dR_1	R_2	dR_2	NOE	dNOE
I123	1.487	0.028	9.516	0.171	0.886	0.070
P124						
C125	1.335	0.022	10.831	0.128	0.707	0.038
G126	1.339	0.020	8.487	0.109	0.751	0.045
L127	1.308	0.017	9.778	0.118	0.902	0.044
P128						
P129						
T130						
I131	1.327	0.018	9.442	0.115	0.705	0.040
T132	1.062	0.019	9.163	0.114	0.741	0.043
N133	0.952	0.076	10.647	0.509	0.885	0.115
G134	1.271	0.016	11.308	0.117	0.901	0.034
D135	1.388	0.016	9.940	0.101	0.805	0.037
F136						
I137						
S138	1.031	0.033	9.874	0.200	0.569	0.054
T139						
N140						
R141						
E142						
N143	0.775	0.063	8.551	0.394	0.424	0.072
F144						
H145						
Y146	1.185	0.020	11.508	0.154	0.733	0.043
G147	1.357	0.024	11.335	0.167	0.711	0.044
S148						
V149						
V150						
T151	1.331	0.015	8.676	0.087	0.683	0.040
Y152						
R153	1.412	0.021	10.489	0.141	0.742	0.046
C154						
N155	1.481	0.022	9.044	0.125	0.638	0.041
P156						
G157	1.063	0.023	8.261	0.129	0.469	0.042
S158						
G159						
G160						
R161	1.140	0.205	10.409	1.594	0.296	0.081
K162	1.078	0.077	11.071	0.575	0.324	0.079
V163						
F164	1.158	0.012	8.101	0.067	0.492	0.031
E165	1.427	0.014	8.620	0.090	0.504	0.033
L166	1.328	0.015	9.638	0.103	0.580	0.045
V167						
G168	1.270	0.018	10.421	0.133	0.768	0.043
E169	1.041	0.011	8.351	0.077	0.541	0.031
P170						
S171	1.394	0.014	9.105	0.092	0.786	0.035
I172	1.426	0.018	10.185	0.134	0.727	0.036
Y173	1.448	0.020	10.801	0.132	0.647	0.036
C174	1.421	0.022	9.878	0.133	0.762	0.044
T175						
S176	1.326	0.018	10.381	0.115	0.869	0.040
N177						
D178	0.878	0.109	10.753	0.800	0.571	0.107
D179						
Q180						
V181						
G182						
I183	1.344	0.026	12.124	0.207	0.641	0.050
W184	1.253	0.015	8.872	0.087	0.455	0.027
W184s	1.167	0.016	9.402	0.109	0.660	0.033
S185						
G186	1.422	0.016	9.750	0.103	0.668	0.031
P187						
A188	1.436	0.017	8.651	0.102	0.574	0.035
P189						
Q190	1.414	0.017	9.066	0.116	0.767	0.041
C191	1.383	0.016	8.900	0.103	0.853	0.041
I192						

Table D.1: Relaxation data for CR1˜2-3 spins.

Appendix E

Modelfree fittings table for CR1~16

The parameters fitted by Modelfree to CR1~16 are listed in the table below.

Residue	Model	S^2	dS^2	t_e	dt_e	R_{ex}	dR_{ex}
K961	111	0.473	0.008	102.487	4.415	0.410	0.158
S962	100	0.724	0.009				
C963	110	0.797	0.013	81.498	25.207		
K964	100	0.829	0.009				
T965	111	0.723	0.014	54.268	16.107	2.312	0.164
P966							
P967							
D968	100	0.733	0.009				
P969							
V970	100	0.780	0.009				
N971	100	0.845	0.009				
G972	100	0.842	0.009				
M973							
V974	100	0.827	0.009				
H975	100	0.837	0.009				
V976	100	0.793	0.009				
I977	Not fitted						
T978	111	0.774	0.015	50.035	22.240	1.970	0.172
D979	100	0.750	0.009				
I980	101	0.815	0.015			4.323	0.175
Q981	100	0.856	0.009				
V982	101	0.780	0.010			1.072	0.160
G983	101	0.861	0.009			0.469	0.159
S984	100	0.863	0.009				
R985	100	0.824	0.009				
I986	101	0.841	0.009			0.662	0.157
T987	101	0.889	0.016				
Y988	100	0.818	0.009				
S989	Not fitted						
C990							
T991	Not fitted						
T992	100	0.823	0.010				
G993							
H994	100	0.828	0.009				
R995	100	0.816	0.009				
L996	100	0.821	0.009				
I997	100	0.853	0.009				
G998	100	0.813	0.009				
H999	101	0.801	0.009			0.632	0.161
S1000	100	0.819	0.010				
S1001	100	0.806	0.009				
A1002	100	0.825	0.009				
E1003	100	0.846	0.009				
C1004	100	0.852	0.009				
I1005	Not fitted						
L1006	100	0.799	0.009				
S1007	111	0.751	0.013	105.451	19.879	0.811	0.163
G1008							
N1009							
T1010	100	0.769	0.009				
A1011	Not fitted						
H1012	100	0.846	0.009				
W1013	100	0.816	0.009				
S1014	100	0.817	0.009				
T1015	Not fitted						
K1016	100	0.792	0.009				
P1017							
P1018							
I1019	100	0.861	0.009				
C1020	100	0.785	0.009				
Q1021	100	0.796	0.009				
R1022	101	0.872	0.016			1.105	0.172
I1023	Not fitted						
P1024							

Table E.1: Modelfree parameter fittings for CR1~16 spins.

Appendix F

Modelfree fittings table for CR1~2-3

The parameters fitted by Modelfree to CR1~2-3 are listed in the table below.

Residue	Model	S^2	dS^2	S_f^2	dS_f^2	S_s^2	dS_s^2	t_e	dt_e	R_{ex}	dR_{ex}
A60	4	0.303	0.015			0.303	0.015	54.287	4.374	0.996	0.234
K61											
S62	1	0.724	0.014			0.724	0.014				
C63											
R64	1	0.877	0.009			0.877	0.009				
N65	4	0.591	0.037			0.592	0.037	17.188	7.486	5.053	0.611
P66											
P67											
D68											
P69											
V70	1	0.907	0.008			0.907	0.008				
N71											
G72	3	0.816	0.019			0.895	0.013			1.830	0.311
M73	2	0.899	0.011			0.916	0.010	112.122	38.770		
V74											
H75											
V76	3	0.629	0.038			0.630	0.038			10.951	0.905
I77	5	0.614	0.021	0.740	0.015	0.831	0.022	641.270	147.58		
K78											
G79											
I80	3	0.720	0.021			0.721	0.021			13.788	0.631
Q81	3	0.783	0.020			0.781	0.022			5.232	0.386
F82	4	0.583	0.014			0.584	0.014	20.662	4.432	3.900	0.235
G83	3	0.858	0.019			0.844	0.022			6.610	0.439
S84	3	0.792	0.011			0.790	0.014			1.826	0.183
Q85	5	0.689	0.015	0.808	0.012	0.853	0.014	1055.457	200.602		
I86											
K87	3	0.917	0.021			0.918	0.021			3.893	0.411
Y88	1	0.921	0.008			0.921	0.008				
S89	2	0.857	0.017			0.876	0.015	73.767	30.191		
C90	1	0.865	0.013			0.865	0.013				
T91	1	0.964	0.010			0.964	0.010				
K92	1	0.842	0.009			0.842	0.009				
G93	1	0.884	0.018			0.884	0.018				
Y94	1	0.885	0.007			0.885	0.007				
R95	5	0.832	0.024	0.939	0.017	0.886	0.020	832.724	239.977		
L96	1	0.840	0.009			0.840	0.009				
I97											
G98	3	0.765	0.015							3.276	0.258
S99											
S100											
S101	2	0.859	0.006			0.874	0.006	63.078	13.184		
A102											
T103	1	0.870	0.012			0.870	0.012				
C104	1		0.013				0.013				
I105											
I106	4	0.903	0.019				0.012	192.600	72.489	2.330	0.308
S107	5	0.618	0.012	0.818	0.010	0.755	0.012	876.367	74.418		
G108											
D109	3	0.525	0.058			0.748	0.042			5.724	1.024
T110	5	0.701	0.005	0.789	0.042	0.895	0.042	48.561	4.956		
V111	5	0.739	0.011	0.849	0.008	0.870	0.010	592.540	89.952		
I112											
W113	1	0.919	0.014			0.919	0.014				
W113s	1	0.752	0.009			0.752	0.009				
D114											
T115	5	0.637	0.012	0.775	0.009	0.823	0.012	437.969	74.464		
E116	5	0.684	0.005	0.789	0.039	0.842	0.035	50.860	3.842		
T117	4	0.642	0.010			0.713	0.006	21.155	4.119	1.172	0.148
P118											
I119	1	0.942	0.010			0.942	0.010				
C120	Not fitted										
D121	5	0.799	0.025	0.880	0.019	0.908	0.021	1164.477	490.445		
R122	1	0.895	0.010			0.895	0.010				

Residue	Model	S^2	dS^2	S_f^2	dS_f^2	S_s^2	dS_s^2	t_e	dt_e	R_{ex}	dR_{ex}
I123	Not fitted										
P124											
C125	1	0.879	0.008			0.878	0.008				
G126	Not fitted										
L127	1	0.819	0.007			0.819	0.007				
P128											
P129											
T130											
I131	5	0.771	0.017	0.838	0.012	0.919	0.015	771.408	249.522		
T132	1	0.727	0.007			0.727	0.007				
N133	3	0.605	0.048							3.333	0.774
G134	Not fitted										
D135	Not fitted										
F136											
I137											
S138	4	0.615	0.021			0.754	0.013	30.189	6.275	2.380	0.323
T139											
N140											
R141											
E142											
N143	4	0.446	0.037			0.622	0.025	24.551	5.475	3.089	0.600
F144											
H145											
Y146	4	0.736	0.014			0.737	0.014	18.432	7.918	2.586	0.228
G147	1	0.904	0.010			0.904	0.010				
S148											
V149											
V150											
T151	5	0.694	0.016	0.810	0.012	0.857	0.014	1165.108	236.680		
Y152											
R153	1	0.881	0.009			0.881	0.009				
C154											
N155	5	0.712	0.019	0.884	0.015	0.805	0.016	1191.698	189.685		
P156											
G157	4	0.618	0.015			0.619	0.015	44.162	5.243	0.697	0.218
S158											
G159											
G160											
R161	2	0.721	0.083			0.792	0.093	123.549	74.526		
K162	4	0.603	0.045			0.803	0.034	61.607	14.570	3.654	0.790
V163											
F164	Not fitted										
E165	5	0.679	0.013	0.883	0.010	0.769	0.011	832.365	74.864		
L166	2	0.792	0.007			0.792	0.007	74.002	11.585		
V167											
G168	1	0.836	0.008			0.836	0.008				
E169	4	0.617	0.008			0.618	0.008	34.098	3.529	0.825	0.124
P170											
S171	Not fitted										
I172	5	0.832	0.017	0.897	0.012	0.927	0.015	863.634	282.965		
Y173	2	0.886	0.009			0.905	0.008	107.862	24.105		
C174	5	0.801	0.022	0.871	0.016	0.919	0.018	1457.027	686.214		
T175											
S176	1	0.853	0.007			0.852	0.007				
N177											
D178	1	0.731	0.048			0.731	0.048				
D179											
Q180											
V181											
G182											
I183	4	0.816	0.019			0.926	0.012	61.792	15.863	2.201	0.301
W184	2	0.725	0.006	0.900	0.204	0.804	0.175	78.411	5.965		
W184s	1	0.762	0.007			0.762	0.007				
S185											
G186	5	0.790	0.014	0.891	0.010	0.886	0.012	840.203	150.735		
P187											
A188	5	0.680	0.015	0.870	0.012	0.782	0.013	1005.440	110.395		
P189											
Q190	Not fitted										
C191	Not fitted										
I192											

Table F.1: Modelfree parameter fittings for CR1~2-3 spins.