Structure Refinement and Dynamics of Proteins Using Residual Dipolar Couplings and NMR Relaxation Data

Graeme Ball

A thesis submitted for the degree of Doctor of Philosophy

The University of Edinburgh
2005

Abstract

NMR spectroscopy can be used to investigate the structure and dynamics of proteins, both of which can provide important insights into their function. The aims of this project were to use residual dipolar coupling (RDC) restraints to refine the existing NMR structures of several proteins and to use relaxation data to investigate their dynamics. Three proteins were studied: two module-pair fragments from human complement regulatory proteins (DAF~2,3 and C4BP~1,2) and a fatty acid binding protein from the nematode parasite *Ascaris suum* (ABA-1A), each of which consists of two structural elements with a poorly defined relative orientation.

Implementing experiments required for the measurement of coupling constants led to the development of new methods for the determination of the one-bond $^1D_{C^\alpha H^\alpha}$ RDC. The two 2D and two 3D experiments that have been developed allow increased $^1D_{C^\alpha H^\alpha}$ sampling times by using C^α -selective pulses to refocus other couplings, thus improving the precision of the measurements.

Extensive sets of RDCs were measured for C4BP~1,2 and ABA-1A using ^{13}C,^{15}N labelled samples whereas in the case of DAF~2,3 only a single set of NH RDCs was measured. The structure refinement protocol incorporating RDCs as restraints was optimized in the process of calculating structures for ABA-1A. In the case of ABA-1A significant improvements in local geometry and positioning of secondary structural elements were obtained by using RDC restraints.

The limited set of RDC restraints measured for DAF~2,3 was found to be insufficient to significantly improve the precision of the structures and incapable of unambiguously defining the orientation of structural elements. In contrast, the much larger set of RDC measurements obtained for C4BP~1,2 defined a unique arrangement of the module-pair and improved overall structural precision.

The backbone dynamics of DAF~2,3 and C4BP~1,2 were probed using relaxation data measured at two fields and the time scale and extent of backbone motions were quantified. Finally, some evidence of small amplitude intermodular motions in DAF~2,3 and C4BP~1,2 was observed. The functional significance of these findings is discussed.

Contents

Declaration i

Abstract ii

List of Figures viii

Symbols and Abbreviations x

1 Introduction 1

2 Background 4

2.1 Basic NMR theory 4

2.1.1 The density matrix and product spin operator formalism 7

2.1.2 Fourier NMR 10

2.2 Pulse sequence building blocks 12

2.2.1 Spin Echoes and INEPT 12

2.2.2 Pulsed field gradients (PFG) 14

2.2.3 BIRD elements 16

2.2.4 Experiments for measuring $^{1}J_{NH}$, $^{1}J_{NC'}$ and $^{1}J_{C\alpha C'}$ 17

2.3 Relaxation and dynamic processes 23

2.3.1 Measurement of ^{15}N relaxation parameters 27

2.3.2 The Model-Free Formalism 29

2.4 Residual dipolar couplings and the molecular alignment tensor 32

2.5 Incorporating RDC restraints into NMR structure calculations 38

2.6 Regulators of Complement Activation 40

3 Measuring $^1D_{C^\alpha H^\alpha}$ **44**

3.1 Review of existing methods 44

3.2 ^{15}N HSQC-like schemes for measuring $^1D_{C^\alpha H^\alpha}$ 46

3.3 Selective inversion of C^α magnetization 50

3.4 Testing DPFGSE and BIRD pulse sequence elements 52

3.5 Sampling the $C^\alpha H^\alpha$ splitting via C^α magnetization 60

3.5.1 2D (HNCO)-(J-CA)NH 61

3.5.2 3D (HN)CO-(J-CA)NH 64

3.6 Sampling the $C^\alpha H^\alpha$ splitting via H^α magnetization 64

3.6.1 2D (J-HACACO)NH 65

3.6.2 3D J-HA(CACO)NH 67

3.7 2D (HNCO)-(J-CA)NH, 2D (J-HACACO)NH and 3D J-HA(CACO)NH spectra of a model peptide 70

3.8 The benefits of an extended $C^\alpha H^\alpha$ sampling period 73

3.8.1 Examples of spectra for an aligned protein: ABA-1A 74

4 ABA-1A **78**

4.1 Background 78

4.2 Experimental 79

4.2.1 Sample preparation 80

4.2.2 Measurement of RDCs 80

4.3 Structure refinement with RDCs 82

4.3.1 Is the use of a single alignment tensor for ABA-1A justified? . . 84

4.4 Effects of RDCs on the structure of ABA-1A 86

4.5 Discussion 89

5 DAF~2,3 **91**

5.1 Background 91

5.2 Experimental 93

5.2.1 Measurement of $^1J_{NH}$ and repurification by Ni-affinity chromatography 93

5.2.2 Measurement of ^{15}N relaxation data 94
5.2.3 Alignment using the CPBr/NaBr/hexanol 96
5.2.4 Measurement of one bond N-H^{N} residual dipolar couplings . . . 97
5.3 RDC refinement of the original NMR structures of DAF~2,3 98
5.3.1 Can a single alignment tensor be used for both modules? 99
5.4 RDC refinement of the current NMR structures of DAF~2,3 101
5.5 Analysis of relaxation data 108
5.6 Discussion 114

6 C4BP~1,2 **117**
6.1 Background 117
6.2 Experimental 118
6.2.1 Measurement of ^{15}N relaxation data 118
6.2.2 C12E5/hexanol as a potential aligning medium 120
6.2.3 Pf1 phage as a potential aligning medium 123
6.2.4 Measurement of RDCs for a Pf1-aligned ^{13}C,^{15} N -labelled C4BP~1,2 sample 124
6.3 RDCs in the structure refinement of C4BP~1,2 126
6.3.1 Can a single alignment tensor be used for both modules? 126
6.3.2 Evaluation of the final ensemble 132
6.4 Analysis of relaxation data 136
6.5 Discussion 141

7 Conclusions **144**

Appendices **148**

A Measured RDC Values **148**

B CNS refinement protocol **153**

Bibliography **160**

List of Figures

2.1 Larmor precession 5

2.2 Offset evolution in a rotating frame 7

2.3 Standard rotations for offset evolution and pulses 9

2.4 Standard rotations for scalar coupling evolution 10

2.5 Heteronuclear spin echoes 13

2.6 Dephasing caused by application of a PFG 15

2.7 WATERGATE 16

2.8 BIRD sandwiches 17

2.9 IPAP editing 18

2.10 ^{15}N -IPAP HSQC pulse sequence 19

2.11 HN(α/β-NC'-J) pulse sequence 20

2.12 HN(α/β-COCA-J) pulse sequence 22

2.13 Transitions in a two spin system 24

2.14 Relaxation delay elements 27

2.15 T_1 decay curve 28

2.16 Diffusion in a cone and two-site jump models. 31

2.17 Histogram determination of the principle order values S_{xx}, S_{yy} and S_{zz} 36

2.18 The Sanson-Flamsteed projection 39

2.19 Complement cascade 41

2.20 CCP domain structure 42

3.1 Basic HN(CO)CA-based scheme for measuring $^1D_{C^\alpha H^\alpha}$ 47

3.2 Basic HA(CACO)NH-based scheme for measuring $^1D_{C^\alpha H^\alpha}$ 49

3.3 Chemical shift distributions for C^{α} and C^{β} 51
3.4 Inversion profile of a 1400 μs Q3 pulse. 51
3.5 Pulse sequence for testing the SPFGSE element 53
3.6 Pulse sequence for testing the DPFGSE element 53
3.7 Properties of SPFGSE and DPFGSE pulse sequence elements 54
3.8 Comparison of SPFGSE and DPFGSE elements with changing ^{13}C 180° Q3 pulse offset. 55
3.9 Pulse sequence for testing the $BIRD_{d,x}$ element 56
3.10 Pulse sequence for testing the $BIRD_{r}$ element 57
3.11 Properties of the $BIRD_{d,x}$ pulse sequence element. 58
3.12 Properties of the $BIRD_{r}$ pulse sequence element. 59
3.13 2D (HNCO)-(J-CA)NH pulse sequence for measuring $^{1}D_{C^{\alpha}H^{\alpha}}$ 62
3.14 Chemical shift distributions for H^{α} and H^{β}. 65
3.15 2D (J-HACACO)NH pulse sequence for measuring $^{1}D_{C^{\alpha}H^{\alpha}}$ 66
3.16 $BIRD_{r}$ element from the final 3D J-HA(CACO)NH pulse sequence. . . . 68
3.17 Test octapeptide molecule structure 70
3.18 Final pulse sequences tested on a labelled peptide 71
3.19 Effects of the $BIRD_{r}$ element 74
3.20 2D and 3D $^{1}J_{C^{\alpha}H^{\alpha}}$ spectra of aligned ABA-1A 76

4.1 ABA-1A structure 79
4.2 ABA-1A structure refinement optimization 84
4.3 Alignment tensor distributions for ABA-1A 86
4.4 Predicted and observed $^{1}D_{C^{\alpha}H^{\alpha}}$ RDCs 88
4.5 ABA-1A structures 90

5.1 Comparison of the original DAF~2,3 NMR structure / x-ray structure . 92
5.2 SDS/PAGE gel of degraded ^{15}N DAF:2,3 sample. 94
5.3 Alignment tensor distributions for DAF23. 101
5.4 Alignment tensor distributions for the latest DAF23 structures. 104

5.5 Observed and back-calculated NH RDCs using the DAF~2,3 x-ray crystal structure . . . 105

5.6 DAF23 intermodular orientation . . . 107

5.7 T_1 and T_2 relaxation times for DAF~2,3 . . . 108

5.8 T_1/T_2 ratio and $\{^1H\}$-^{15}N NOE for DAF~2,3 . . . 109

5.9 Axially symmetric diffusion tensor for DAF~2,3 . . . 110

5.10 DAF~2,3 model free results . . . 112

5.11 DAF~2,3 mobility . . . 113

5.12 DAF~2,3 mutations . . . 116

6.1 ^2H NMR spectra for the PEG/hexanol medium . . . 122

6.2 Alignment tensor distributions for C4BP~1,2. . . . 128

6.3 Relative orientation of the PAS for M1/M2 . . . 130

6.4 T_2 relaxation time comparison, first ^{13}C,^{15}N C4BP~1,2 / Pf1 sample . 131

6.5 C4BP~1,2 structure ensemble overall and NOE energies. . . . 133

6.6 C4BP~1,2 final structures . . . 134

6.7 C4BP~1,2 intermodular orientation . . . 135

6.8 T_1 and T_2 relaxation times for C4BP~1,2 . . . 137

6.9 T_1/T_2 ratio and $\{^1H\}$-^{15}N NOE for C4BP~1,2 . . . 138

6.10 C4BP~1,2 model free results . . . 140

6.11 C4BP~1,2 mobility . . . 141

6.12 C4BP~1,2 putative binding sites . . . 142

Abbreviations

AP	Alternative Pathway
AP	Antiphase (magnetization)
BIRD	Bilinear Rotation Decoupling
C4BP	C4b-Binding Protein
CCP	Complement Control Protein
CP	Classical Pathway
CPBr	Cetylpyridinium bromide
CPMG	Carr-Purcell-Meiboom-Gill
CSA	Chemical Shift Anisotropy
DAF	Decay Accelerating Factor
DPFGSE	Double Pulsed Field Gradient Spin Echo
FID	Free Induction Decay
HSQC	Heteronuclear Single Quantum Coherence
INEPT	Insensitive Nuclei Enhanced by Polarization Transfer
IP	In-Phase (magnetization)
IPAP	In-Phase Antiphase (spectral editing scheme)
ISAC	Implicit Saupe tensor Alignment Constraint
MCP	Membrane Cofactor Protein
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect (enhancement)
NOESY	Nuclear Overhauser Effect spectroscopy
PAS	Principal Axis System
PEG	Polyethylene Glycol
PFG	Pulsed Field Gradient
ppm	parts per milion
RCA	Regulator of Complement Activation
RDC	Residual Dipolar Doupling
RMSD	Root Mean Square Deviation
S/N	Signal-to-Noise ratio
SVD	Singular Value Decomposition
TROSY	Transverse Relaxation Optimised Spectroscopy
VCP	*Vaccinia* complement protein

Names and abbreviations for amino acids

Name	Abbreviation	Single 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

Chapter 1

Introduction

Nuclear magnetic resonance spectroscopy (NMR) is an indirect technique for biomolecular structure determination - it relies largely on the measurement of many short range semiquantitative restraints in the form of Nuclear Overhauser Effect (NOE) enhancements. The process of protein structure determination by NMR [117] involves sequential assignment of specific atoms to the experimentally observed resonance frequencies, followed by the assignment of NOEs to these atoms and finally, restrained molecular dynamics calculations using a hybrid energy function including terms for both; (1) empirically derived parameters describing idealized geometry (the force field) and (2) the difference between experimentally observed and calculated data.

One advantage of NMR structure determination, when compared with a more direct method such as x-ray crystallography, is that structures determined by NMR represent the state of the molecule in solution rather than the ground state in a crystalline solid [9]. Regions of a structure that experience significant internal motions usually exhibit a lack of measurable NOEs, producing less well defined regions in the NMR structure ensemble [1].

The presence of motion in proteins and its importance in understanding function has long been recognized [48] - dynamical processes and conformational changes are ubiq-

[1] The idea of producing an ensemble of structures that represents conformational variability with time is implemented more rigorously in a recently published protocol [60] that uses an experimental energy term based on the ensemble average structure and incorporates the measured degree of structural order as an additional restraint.

uitous and crucial for selective binding or catalytic activity in many cases. Evidence for the presence of motion and information about its time scale and extent can be obtained from NMR: relaxation data have been shown to be sensitive to internal motions over a range of timescales [49, 46]. Residual dipolar couplings (RDC) have also been shown to be sensitive to motions [109], although to a lesser extent than relaxation data [2].

Residual dipolar couplings (RDC) are a relatively recent addition to the collection of experimental restraints used in NMR structure determination [103] and in contrast to NOE restraints, provide information that is of a non-local, i.e. long range nature. The use of RDC restraints has the potential to produce significant improvements in structural precision [106], particularly in the case of larger proteins, and can be used to determine average domain orientations in multidomain proteins [100].

Aims of the project

Using the techniques described above, namely measurement of RDCs and relaxation data, the aims of this project were twofold: (1) the refinement of existing protein structures using RDC restraints to improve structure and the definition of the relative orientation of structural elements and (2) investigation of the dynamical processes within and between these structural elements. The first aim also led to work on the development of methods for measuring $^1D_{C^\alpha H^\alpha}$ RDCs.

Overview

Of the three proteins chosen for investigation, two are module-pair fragments from human RCA proteins (regulators of complement activation): DAF~2,3 and C4BP~1,2 and the other, ABA-1A, is a fatty acid binding protein from the nematode parasite *Ascaris suum*. The chronology of the work that was carried out in this project is as follows:

[2] The dependence of measured RDCs on structural order S scales with S as opposed to S^2 for relaxation data. In addition, a model-free analysis of RDCs requires data recorded in at least 5 different aligning media [84].

- Measurement of $^1D_{NH}$ RDCs and relaxation data for DAF~2,3 in CPBr/hexanol and structure refinement.

- Implementation of NMR methods to measure $^1D_{NC'}$, $^1D_{C^\alpha C'}$ and development of methods to measure $^1D_{C^\alpha H^\alpha}$.

- Measurement of relaxation data for C4BP~1,2 and attempted alignment of C4BP~1,2 using a PEG/hexanol system.

- Alignment of C4BP~1,2 with Pf1 phage and measurement of $^1D_{NH}$, $^1D_{NC'}$, $^1D_{C^\alpha C'}$ and $^1D_{C^\alpha H^\alpha}$ RDCs.

- Alignment of ABA-1A using Pf1 phage, measurement of $^1D_{NH}$, $^1D_{NC'}$, $^1D_{C^\alpha C'}$ and $^1D_{C^\alpha H^\alpha}$ RDCs and evaluation of $^1D_{C^\alpha H^\alpha}$ methods using ABA-1A.

- Optimization of the structure calculation refinement protocol and refinement of ABA-1A structures using RDCs.

- Refinement of C4BP~1,2 structures using RDCs and interpretation of relaxation data.

- Refinement of new DAF~2,3 structures using $^1D_{NH}$ RDCs and interpretation of relaxation data in light of x-ray crystal structures.

Chapter 2

Background

2.1 Basic NMR theory

Nuclei with spin angular momentum $\mathbf{I}$ give rise to a magnetic dipole moment $\boldsymbol{\mu}$ due to the circular current created by a spinning charged particle:

$$\boldsymbol{\mu} = \gamma \mathbf{I} \tag{2.1}$$

where $\mathbf{I}$ and $\boldsymbol{\mu}$ are vector quantities and the proportionality constant γ is the *gyromagnetic ratio* of the nucleus. For a macroscopic sample of many identical nuclei the sum of contributions from individual nuclear magnetic dipoles μ_i can be considered, giving rise to a macroscopic magnetization $\mathbf{M} = \sum_i \boldsymbol{\mu}_i$ which represents an ensemble average.

In the presence of an externally applied magnetic field B_0, conventionally along the z-axis, a small excess of spins are aligned with the field giving a resultant bulk magnetization. The bulk magnetization at equilibrium has no *coherent* non-z component and cannot be observed: magnetization with an x-y component, i.e. non-equilibrium magnetization, is therefore created in an NMR experiment by radiofrequency irradiation at the appropriate frequency. Coherent non-equilibrium bulk magnetization in an applied field mirrors the behaviour of a single nuclear dipole and *precesses* about the field vector (Figure 2.1) with a characteristic angular velocity ω_0: the corresponding frequency $\nu_0 = \omega_0/2\pi$ is referred to as the *Larmor frequency* of the spin:

$$\nu_0 = -\gamma B_0/2\pi \tag{2.2}$$

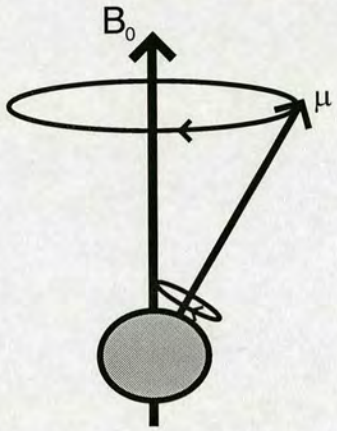

Figure 2.1: **Larmor precession.** Larmor precession of a nuclear magnetic dipole μ about a field B_0 applied along the z-axis. Nuclei with positive γ (e.g. ^{1}H and ^{13}C, but not ^{15}N) have a negative Larmor frequency and precess in a clockwise direction when viewed from a direction that is against the magnetic field.

The magnetic field experienced by each spin in a molecule depends on its local environment as a consequence of *shielding* or *deshielding* arising from nearby electrons circulating in a manner that cancels out or reinforces the applied field. This means that each type of atomic environment in a molecule tends to be associated with a unique Larmor frequency, normally expressed as a *chemical shift* δ defined in equation 2.3 below, where ν_{ref} is the chosen reference frequency:

$$\delta(\text{ppm}) = \frac{(\nu - \nu_{ref}) \times 10^6}{\nu_{ref}} \tag{2.3}$$

Although it is possible to understand many aspects of NMR experiments by treating the bulk magnetization of the sample as a simple vector (the 'vector model'), a quantum-mechanical formulation is needed to fully explain the transition energies observed in an NMR experiment - this is a consequence of the quantisation of spin angular momentum. Of the several basic postulates of quantum mechanics [4], the following are most relevant to NMR:

- The state of a system is fully described by a *wavefunction*, $\psi(r_1, .., r_n, t)$ where r_n are spatial coordinates and t is time.
- Observables are represented by *operators* and the *expectation value* of an operator Ω is denoted $\langle\Omega\rangle$ where $\langle\Omega\rangle = \int \psi^* \Omega \psi \mathrm{d}\tau$.
- If ψ is an eigenfunction of the operator Ω then determination of the property Ω always yields the corresponding eigenvalue ω.
- The wavefunction $\Psi(r_1, .., r_n, t)$ evolves in time according to the *Schrödinger equation* (equation 2.4) below.

$$i\hbar \frac{\partial \Psi}{\partial t} = H\Psi \tag{2.4}$$

where H is the hamiltonian operator; the operator that corresponds to the total energy of the system, $\hbar = h/2\pi$ and h is Planck's constant. In the case of a single spin exposed to an applied field B_0, the free precession or *Zeeman* hamiltonian is:

$$H_{free} = \gamma B_0 \hbar I_z \tag{2.5}$$

where I_z is the operator for the z-component of the spin angular momentum. For spin-$\frac{1}{2}$ nuclei (^{1}H, ^{13}C and ^{15}N), there are $2I + 1 = 2$ eigenvalues with corresponding eigenfunctions usually denoted $|\ \alpha\rangle$ and $|\ \beta\rangle$ for which $I = +\frac{1}{2}$ and $I = -\frac{1}{2}$, respectively: eigenvalues (i.e. energies) are $+\frac{1}{2}\hbar\omega_0$ and $-\frac{1}{2}\hbar\omega_0$, resulting in an energy for transitions between the two levels of $\hbar\omega_0$, which corresponds to the Larmor frequency. The relative populations P of the α and β spin states at equilibrium are given by the Boltzmann distribution, equation 2.6 below:

$$P_\alpha / P_\beta = exp(-\Delta E_{\alpha/\beta}/kT) \tag{2.6}$$

A radiofrequency coil can be used to create a magnetic field oscillating along the x-direction (relative to the static B_0 field) at close to the Larmor frequency of a spin, giving a Hamiltonian in units of angular velocity (radians s^{-1}):

$$H = \omega_0 I_z + 2\omega_1 \cos(\omega_{RF} t) I_x \tag{2.7}$$

ω_{RF} is the carrier frequency of the radiofrequency (RF) pulse and ω_1 is the strength of the RF field $B_1 = \omega_1/\gamma$. This situation can be simplified by defining a quantity called the *offset*, $\Omega = (\omega_0 - \omega_{RF})$. Considering the problem from the perspective of a *rotating frame* at ω_{RF} about the z-axis (Figure 2.2), the Hamiltonian becomes time independent and for pulses satisfying $\omega_1 >> \mid \Omega \mid$ the pulse Hamiltonian is simply $H = \omega_1 I_x$. Short, high power pulses of this kind are referred to as *hard* or *non-selective* because they cover a broad range of frequencies. Calibration of a pulse applied along the x-axis such that $\omega_1 t = \pi/2$, i.e. 90°, gives a pulse that will rotate bulk z magnetization into the x-y plane. Likewise a $\omega_1 t = \pi$ is referred to as a 180° pulse, which causes complete inversion of magnetization.

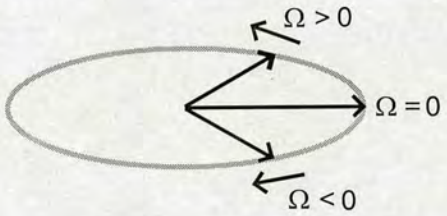

Figure 2.2: **Offset evolution in a rotating frame.** In a rotating frame at the spectrometer frequency, precession occurs at a reduced frequency $\Omega = (\omega_0 - \omega_{RF})$ (where ω_0 is the Larmor frequency of the magnetization): for $\Omega < 0$ and $\Omega > 0$ the sense of the precession is different.

2.1.1 The density matrix and product spin operator formalism

The state of an entire quantum mechanical system can be represented by a density operator, $\sigma(t) = \mid \psi(t)\rangle\langle\psi(t) \mid$, where $\psi(t)$ is a normalized state function for an ensemble of spin systems that are in the same state (this is the case for a homogeneous ensemble at equilibrium at the start of an NMR experiment). Using a density operator to represent the state of the spin system ensemble therefore removes the need for explicit ensemble averaging. Calculations are most conveniently handled in matrix form with a *density matrix*. For example, the equilibrium state of a one spin system with basis

states α and β (setting populations $P_\alpha = P_{av} + \Delta$ and $P_\beta = P_{av} - \Delta$) is represented by equation 2.8:

$$\sigma_{eq} = \begin{pmatrix} P_{av} + \frac{1}{2}\Delta & 0 \\ 0 & P_{av} - \frac{1}{2}\Delta \end{pmatrix} = P_{av} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \Delta \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix} \tag{2.8}$$

This result shows that σ_{eq} can be separated into two elements: only the second can lead to any observables, and represents the excess population of the α spin state and population deficit of the β state that gives the coherent bulk z magnetization at equilibrium: the second term is in fact the matrix representation of the z angular momentum operator I_z (Δ is a scaling factor that can be ignored). The I_x and I_y angular momentum operators likewise have matrix representations, given below in equations 2.9 and 2.10, respectively:

$$I_x = \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix} \tag{2.9}$$

$$I_y = \begin{pmatrix} 0 & -\frac{i}{2} \\ \frac{i}{2} & 0 \end{pmatrix} \tag{2.10}$$

The density matrix for a single spin system after a 90° radiofrequency pulse from the x-direction can be represented by $-Iy$, and similarly a 90° pulse from y gives $\sigma = I_x$. Unlike I_z, these matrix representations have off-diagonal elements that correspond to non-equilibrium magnetization.

The density matrix for a system of several spins can be constructed by combining the basis sets for each spin system in the same density operator matrix: this is achieved by evaluating the *direct product* of the individual basis sets as illustrated in equation 2.11 for a two spin system in which spin 1 has coherent y magnetization (y *coherence*) and spin 2 has no coherent magnetization:

$$I_{1y} \otimes E_2 = \begin{pmatrix} 0 & -\frac{i}{2} \\ \frac{i}{2} & 0 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & -\frac{i}{2} & 0 \\ 0 & 0 & 0 & -\frac{i}{2} \\ \frac{i}{2} & 0 & 0 & 0 \\ 0 & \frac{i}{2} & 0 & 0 \end{pmatrix} \tag{2.11}$$

The **product spin operator formalism** [51] is a simplification of the density operator

treatment - it makes use of the fact that the density operator for a spin can, at any time, be represented as a sum of the angular momentum operators: $\sigma(t) = a(t)I_x + b(t)I_y + c(t)I_z$. Both pulses and offset evolution can be represented as rotations changing the contribution of the three angular momentum operators to the density matrix. The sense of the rotations about z (offset evolution) and about x and y (effects of x and y pulses) are illustrated in Figure 2.3 and the magnitude of the contributions from the initial operator and rotation result can be determined from equation 2.12 below:

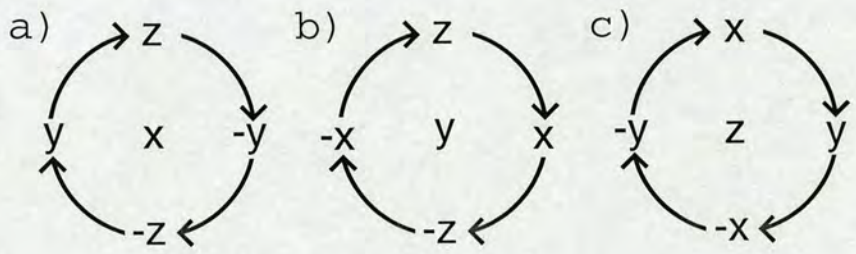

Figure 2.3: **Standard rotations for offset evolution and pulses.** The effects of (a) free precession (operator I_z) and (b,c) pulses (operators I_x and I_y) are summarised above: in each case, rotation by an angle θ by the operator at the centre of the rotation diagram results in $\cos\theta$ times the original operator plus $\sin\theta$ times the new operator. Offset evolution during free precession results in a rotation of Ωt where Ω is the offset frequency of the spin in $rad \cdot s^{-1}$ and t, in s, is the length of time for which free precession occurs. The angle of rotation for a pulse is equal to $\Omega_1 t$ where Ω_1 is the radiofrequency of the pulse in $rad \cdot s^{-1}$ and t is the length of the pulse in s.

$$\sigma = I_{initial} \cdot \cos\theta + I_{rot-result} \cdot \sin\theta \tag{2.12}$$

Systems of two or more spins can also be represented by constructing product operators to represent all of the spins, which greatly simplifies analysis of the effects of couplings between the spins. One type of coupling between spins in liquid NMR is the *scalar* coupling or J coupling, which is mediated by bonding electrons. Direct through-space interaction of the nuclear dipoles leads to dipolar couplings, which are usually averaged out by isotropic molecular tumbling and therefore do not contribute to splitting of the NMR signals [1]. In an important recent development, dipolar coupling constants

[1] In spite of isotropic tumbling, the dipolar coupling does however contribute to relaxation and the Nuclear Overhauser Effect (NOE).

have been reintroduced in high resolution NMR studies of biomolecules [103] - this phenomenon is discussed in detail in section 2.4.

In the case of coherences between two spins, 1 and 2, coherent spin 1 x-magnetization is represented as I_{1x}, which is actually the direct product: $I_{1x} \otimes E_2$ and is referred to as in-phase magnetization for reasons explained below. Antiphase magnetization represented by $I_{1x} \otimes I_{2z} = 2I_{1x}I_{2z}$ is produced by evolution of the coupling, where the '2' before the $I_{1x}I_{2z}$ term is a normalization factor. Application of a pulse from x at the frequency of spin 2 would lead to the creation of a *multiple quantum* term: $-2I_{1x}I_{2y}$ which is not observable. The effects of J-coupling evolution can be calculated using the Hamiltonian $H_J = 2\pi J_{12} I_{1z} I_{2z}$, where the sense of the rotation is given in Figure 2.4.

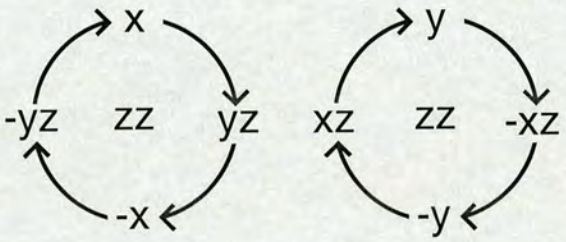

Figure 2.4: **Standard rotations for scalar coupling evolution.** The angle of rotation for coupling evolution (operator I_{zz}, represented by 'zz' in the rotation diagram) with coupling constant J (s^{-1}) for a length of time t results in $\cos(\pi Jt)$ times the original operator plus $\sin(\pi Jt)$ times the new operator.

2.1.2 Fourier NMR

Detection of the transverse magnetization that exists at the end of an NMR experiment relies on measuring the current induced in the receiver coil by the precession of coherent transverse magnetization. Transverse magnetization decays both by relaxation back to z and by the loss of coherence in the x-y plane, resulting in a damped sinusoidal signal as x and y magnetization interconvert according to the offset Ω. This is known as a Free Induction Decay or FID. The signal observed for a coherence I_x in the case of an isolated spin would be:

$$I(t) = k \times [\cos(\Omega t + \phi)I_x + i \cdot \sin(\Omega t + \phi)I_y] \times \exp(-t/\mathrm{T}_2) \tag{2.13}$$

where k is a scaling factor and ϕ is a phase shift introduced by the instrument. The phase shift can be removed later by combining a pair of signals obtained using intermediate frequencies that are 90° phase-shifted with respect to each other - this is known as *quadrature detection.* Fourier transformation of this complex time-domain signal leads to a complex function in frequency ω: $I(\omega) = k \times [A(\omega) + i \cdot D(\omega)]$, where $A(\omega)$ and $D(\omega)$ are Lorentzian absorptive and dispersive lineshape components, respectively.

In the case of a spin that has couplings with other spins, evolution of the transverse magnetization during detection periods is complicated by the evolution of the couplings. The case of a weakly coupled two spin system, as in a heteronuclear pair such as the amide ^{15}N and ^{1}H, is considered in equations 2.14 and 2.15 below for in-phase and antiphase proton magnetization respectively:

$$\mathrm{H}_x \xrightarrow{\pi Jt2\mathrm{H}_z\mathrm{N}_z} \mathrm{H}_x\cos(\pi Jt) + 2\mathrm{H}_y\mathrm{N}_z\sin(\pi Jt) \tag{2.14}$$

$$2\mathrm{H}_x\mathrm{N}_z \xrightarrow{\pi Jt2\mathrm{H}_z\mathrm{N}_z} 2\mathrm{H}_x\mathrm{N}_z\cos(\pi Jt) + \mathrm{H}_y\sin(\pi Jt) \tag{2.15}$$

It can be seen that evolution under the $2\mathrm{H}_z\mathrm{N}_z$ operator during acquisition for a time t results in the interconversion of in-phase and antiphase terms, but that the observable in-phase term is cosine modulated in the case of initial in-phase magnetization (2.14) whereas initial antiphase magnetization results in a sine modulated in-phase term (2.15). Upon Fourier transformation, the $\mathrm{H}_x\cos(\pi Jt)$ term gives a pair of peaks at $\Omega + J/2$ and $\Omega - J/2$ that are both positive, whereas in the case of the $\mathrm{H}_y\sin(\pi Jt)$ term the peaks at $\Omega + J/2$ and $\Omega - J/2$ are of opposite sign. This property will be utilized in the simplification of spectra in Chapter 3.

Multidimensional NMR spectra are extremely useful for a number of reasons: they solve overlap problems and enable the recording of the frequencies of a number of different types of spins in a correlated manner, providing a convenient means of mapping interactions such as NOE enhancements between individual spins. Extra dimensions

are introduced by collecting a series of FIDs in which the offsets of the desired spins are allowed to evolve for an incrementally increasing amount of time: a series of successive Fourier transformations in each of the sampled time domains will then yield the desired spectrum.

2.2 Pulse sequence building blocks

The NMR experiments used to obtain protein spectra are built up from a number of simple elements that allow specific coherences to be created and manipulated. All pulse sequence diagrams presented here use the following symbols: hard 90° pulses are represented by full height, filled rectangles; hard 180° pulses are represented by full height, open rectangles; soft rectangular 90° water-selective pulses are represented by filled half height rectangles; soft 90° Q5 carbon pulses are represented by filled half ellipses; soft 180° Q3 carbon pulses are represented by open half ellipses; decoupling is represented by extended, open rectangles with a label (e.g. DIPSI-2, GARP) and flanking filled rectangles are 90° pulses at the decoupling power level. Each line represents a separate nucleus/atom type, and there is an additional line for the z-gradients G_z. Pulse phases are x unless labelled otherwise, a 'BS' label indicates a Bloch-Siegert [2] compensating pulse and 90° Q5 pulses with a 'tr' label are time-reversed. Gradients labelled 'Gn' are full length (1 ms) while those labelled 'gn' are shorter (0.5 ms), and each is followed by a 0.1 ms recovery delay.

2.2.1 Spin Echoes and INEPT

Spin echo elements are used for a number of purposes in pulse sequences: they can enable evolution of the offset without any coupling evolution, or evolution of a coupling without any offset evolution for the purpose of the interconversion of in-phase and antiphase states in systems of more than one spin. In the case of a heteronuclear two spin system it is possible to invert either of the two spins, or both spins simultaneously, giving three types of spin echo that are illustrated in Figure 2.5.

[2] Bloch-Siegert phase shifts are an off-resonance effect: e.g. $^{13}C^{\alpha}$-selective pulses have an off-resonance component at $^{13}C'$ frequencies (and vice versa) that affect the Larmor frequency unless compensated

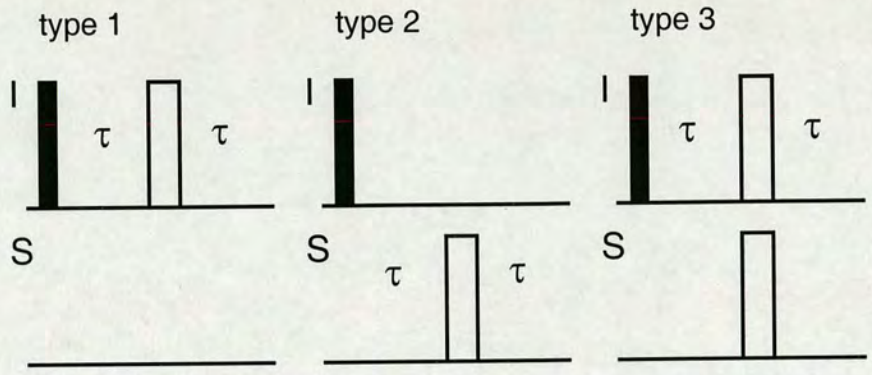

Figure 2.5: **Heteronuclear spin echoes.** The effects of the three types of symmetrical spin echo illustrated here on transverse (e.g. I_y) magnetization can be summarised as follows: the 'type 1' spin echo results in refocussing of both the I offset and IS coupling; the 'type 2' spin echo results in refocussing of the IS coupling but the I offset evolves for the full period; the 'type 3' spin echo allows the IS coupling to evolve for the full period but the I offset is refocussed.

Efficient interconversion of in-phase and antiphase magnetization requires that the overall length of a 'type 3' spin-echo be tuned according to the size of the scalar coupling constant between the two spins: for the amide N-H spin system, conversion of in-phase H_y magnetization to the corresponding antiphase $2\mathrm{H}_x\mathrm{N}_z$ term is outlined below, where the effects of coupling evolution (equation 2.16) and offset evolution (equation 2.17) are considered separately:

$$\mathrm{H}_y \xrightarrow{2\pi J\tau 2\mathrm{H}_z\mathrm{N}_z} \mathrm{H}_y\cos(2\pi J\tau) - 2\mathrm{H}_x\mathrm{N}_z\sin(2\pi J\tau) \tag{2.16}$$

$$\begin{aligned}
\mathrm{H}_y \xrightarrow{\Omega_\mathrm{H}\tau\mathrm{H}_z} & \xrightarrow{\pi\mathrm{H}_x} \xrightarrow{\pi\mathrm{N}_x} \xrightarrow{\Omega_\mathrm{H}\tau\mathrm{H}_z} \\
& -\cos^2(\Omega_\mathrm{H}\tau)\mathrm{H}_y + \cos(\Omega_\mathrm{H}\tau)\mathrm{H}_x\sin(\Omega_\mathrm{H}\tau)\mathrm{H}_x \\
& -\cos(\Omega_\mathrm{H}\tau)\mathrm{H}_x\sin(\Omega_\mathrm{H}\tau)\mathrm{H}_x - \sin^2(\Omega_\mathrm{H}\tau)\mathrm{H}_y \\
& = -\mathrm{H}_y
\end{aligned} \tag{2.17}$$

for.

It can be seen from equation 2.16 that by tuning the length of the delay 2τ to a value of $1/2J_{\mathrm{NH}}$, complete interconversion of the in-phase and antiphase terms is achieved (i.e. the cosine term is zero and the sine term one). The two 180° pulses (H and N) result in an effective 360° pulse experienced by antiphase $2\mathrm{H}_x\mathrm{N}_z$ magnetization, which produces no net effect. Equation 2.17 shows that evolution of the offset for the original in-phase term is refocussed by the 180° proton pulse located between two delays of equal length [3].

Insensitive Nuclei Enhanced by Polarization Transfer (INEPT)

INEPT type elements are commonly used in multidimensional NMR to transfer magnetization between weakly coupled, i.e. heteronuclear, spin pairs: the name has its origin in the fact that INEPT enables magnetization to be transferred from protons to nuclei with a lower gyromagnetic ratio and correspondingly lower sensitivity. An INEPT step consists of a 'type 3' spin echo (see above) of length $1/2J$ which completely interconverts in-phase magnetization to antiphase magnetization, which can then be transferred by means of a pair of 90° pulses from ^{1}H to ^{15}N as shown in equation 2.18 below:

$$2\mathrm{H}_x\mathrm{N}_z \xrightarrow{\frac{\pi}{2}\mathrm{H}_y} \xrightarrow{\frac{\pi}{2}\mathrm{N}_x} 2\mathrm{H}_z\mathrm{N}_y \tag{2.18}$$

This antiphase term can be converted to in-phase ^{15}N magnetization by means of a further spin echo.

2.2.2 Pulsed field gradients (PFG)

Gradient pulses create short-lived magnetic fields that are spatially inhomogeneous - this results in a temporary dephasing of the spins across the sample. For example, applying a z-gradient of strength G for time t results in a variation of the phases of the spins as shown in Figure 2.6. The effects of this gradient can be reversed by means of gradient of equal strength but opposite polarity, or alternatively, a gradient of equal

[3] Delays are of length τ in this case, but the result is general.

strength and the same sign after inversion of the spins by a 180° pulse.

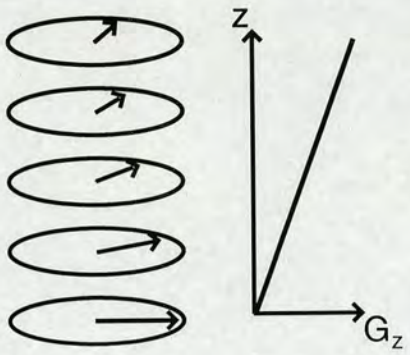

Figure 2.6: **Dephasing caused by application of a PFG.** The effect of a z-gradient on transverse magnetization is shown. Each "slice" of the sample experiences a different effective magnetic field, and the spins within each slice therefore have Larmor frequencies unique to that slice: observable magnetization is destroyed as the vector sum of the magnetization at a given Larmor frequency across the sample volume is zero.

Gradient water signal suppression is used in the majority of the pulse sequences that will be described below: the WATERGATE element (WATER suppression by GrAdient Tailored Excitation - [85]) shown in Figure 2.7 is normally combined with a refocussing of the antiphase magnetization resulting from the last reverse INEPT step in the sequence.

Gradients are also used to dephase unwanted magnetization resulting from pulse imperfections - e.g. when using a 180° pulse for inversion, a pair of identical gradients either side of the pulse can be used to dephase the component that is not inverted by the pulse - a spin echo including such a pair of gradients is termed a Single Pulsed Field Gradient Spin Echo (SPFGSE or PFGSE). It has been found however, that using soft pulses in SPFGSE elements affects the phase of the signal. By using a pair of PFGSE elements (a Double PFGSE - DPFGSE) the phase of the signal is not degraded by any imperfections in the pulses used for inversion, which now create only an attenuation of the detected signal - this is also known as excitation sculpting [45]. Where S represents a unitary transformation caused by any sequence of pulses and delays and G represents a gradient pulse, the sequence $G_1 - S - G_1 - G_2 - S - G_2$ transforms any

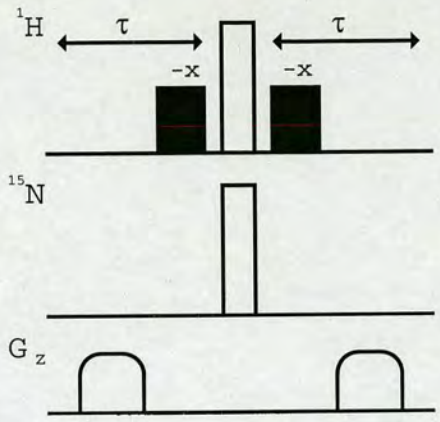

Figure 2.7: **The WATERGATE element.** It can be seen that the central 180° pulse is sandwiched between two water-selective 90° pulses and a pair of identical gradients. This means that water proton spins experience an effective 0° pulse between the two gradients and are dephased, whereas the dephasing caused by the first gradient is refocussed by the second gradient in the case of the non-water proton spins (which are inverted since they are only affected by the central non-selective 180° proton pulse).

initial magnetization $\mathbf{m} = (m_x, m_y, m_z)$ to $\mathbf{M} = (P^2 M_x, P^2 M_y, (1-2P)^2 M_z)$ where P is the probability that a spin is flipped by S - there is no mixing of the transverse and longitudinal components and therefore no possibility of phase distortions. These properties of the DPFGSE element will be utilized in the pulse sequences described in Chapter 3.

2.2.3 BIRD elements

A BIRD sandwich (BIlinear Rotation Decoupling, [36]) is a pulse sequence element used to selectively manipulate protons that are attached to either NMR-active spins (^{13}C,^{15}N) or NMR inactive spins (^{12}C,^{14}N). Several modifications exist however [111], that enable different components of weakly coupled heteronuclear spin systems to be inverted simultaneously. The BIRD sandwich is illustrated in Figure 2.8 for the case of a proton coupled to a heteronucleus X where the phases of the various pulses are denoted ϕ_1 - ϕ_4. The 180° pulse with phase ϕ_4 is not present in the original BIRD sandwich proposed by Garbow *et al.*, but is a necessary part of the BIRD^{r} variant that will be used in a pulse sequence presented later in Chapter 3. A comparison of the

properties of BIRD, BIRD^{r} and BIRD^{d,X} sandwiches is made in Table 2.1.

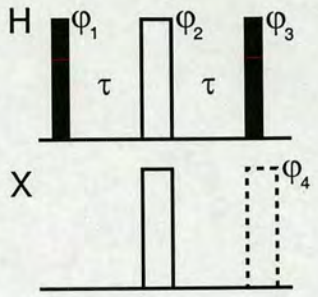

Figure 2.8: **BIRD sandwiches.** The basic layout of the BIRD sandwich: see Table 2.1 for a summary of the properties for selected examples. Note that $\tau = 1/2J_{XH}$ and the pulse with phase ϕ_4 is not present in all variants of the element.

Type	ϕ_1	ϕ_2	ϕ_3	ϕ_4	H_d	H_r	X	ν_d	$J_{d,X}$	$J_{d,r}$
No pulse	-	-	-	-	0	0	0	1	1	1
BIRD^{r}	x	y	x	x	0	π_y	0	1	1	0
BIRD^{d,X}	x	x	x	-	π_y	0	π_y	0	1	0

Table 2.1: **Effects of selected BIRD sandwiches.** Composition and selected properties of the BIRD^{r} and BIRD^{d,X} elements in a weakly coupled heteronuclear system, and the effect of no BIRD element for comparison. The phases of the four pulses ϕ_1-ϕ_4 are shown (a '$-$' for ϕ_4 indicates this pulse is not present), as are the resulting rotations for the directly X-attached proton H_d, remote protons H_r and the X-nucleus. The presence of effective H_d offset evolution, ν_d, $J_{d,X}$ X-H coupling evolution and $J_{d,r}$ H-H coupling evolution are indicated by a '1' and their absence by a '0'. Note that it has been assumed that the X-nucleus is NMR active and has a single attached proton H_d.

2.2.4 Experiments for measuring $^{1}J_{NH}$, $^{1}J_{NC'}$ and $^{1}J_{C^{\alpha}C'}$

Triple resonance experiments on ^{13}C,^{15}N -labelled proteins are used for a variety of purposes including resonance assignment and the measurement of restraints for use in structure calculations. Described below are three experiments used to determine the one-bond $^{1}J_{NH}$, $^{1}J_{NC'}$ and $^{1}J_{C^{\alpha}C'}$ heteronuclear coupling constants: they will be used to determine the residual dipolar couplings (RDC, section 2.4) that are basis of the work in later Chapters (4, 5 and 6).

The ^{15}N -IPAP HSQC experiment

A simple scheme for measuring the scalar coupling constant J in the case of the one bond $^1J_{NH}$ coupling would be to remove the decoupling 180° proton pulse during the t_1 period (nitrogen frequency labelling) of a ^{15}N -IPAP HSQC experiment - however, this would result in a doubling of the number of peaks in the spectrum. The IPAP editing scheme suggested by Ottiger *et al.* [77] prevents an increase in spectral overlap by separating the two components of the $^1J_{NH}$ doublet into two subspectra, exploiting the properties of heteronuclear two spin systems described in section 2.1.2. The name IPAP refers to the fact that spectra in which the doublets are alternately in-phase (IP) and antiphase (AP) are collected - these can either be added together to give one doublet component, or the AP plane subtracted from the IP to give the other component as shown in Figure 2.9. The pulse sequence used to collect these data is shown in Figure 2.10

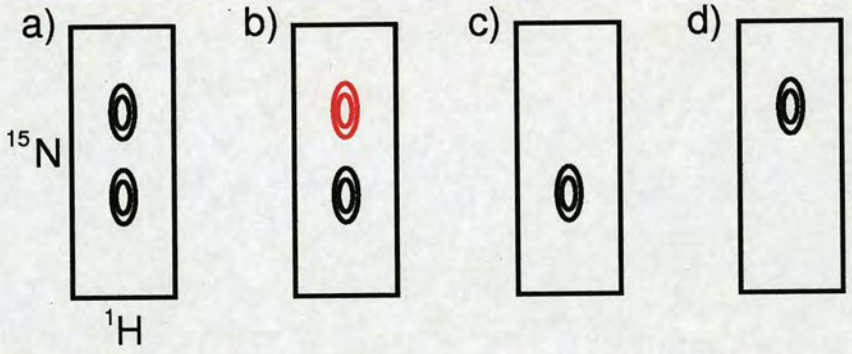

Figure 2.9: **IPAP editing.** Diagram showing the principle of IPAP editing: positive peaks are shown in black and negative peaks in red for the in-phase (a) and antiphase (b) spectra. By adding the two spectra (IP+AP) a single doublet component is obtained (c) and by subtracting the AP spectrum from the IP spectrum (IP-AP) the other doublet component is obtained (d).

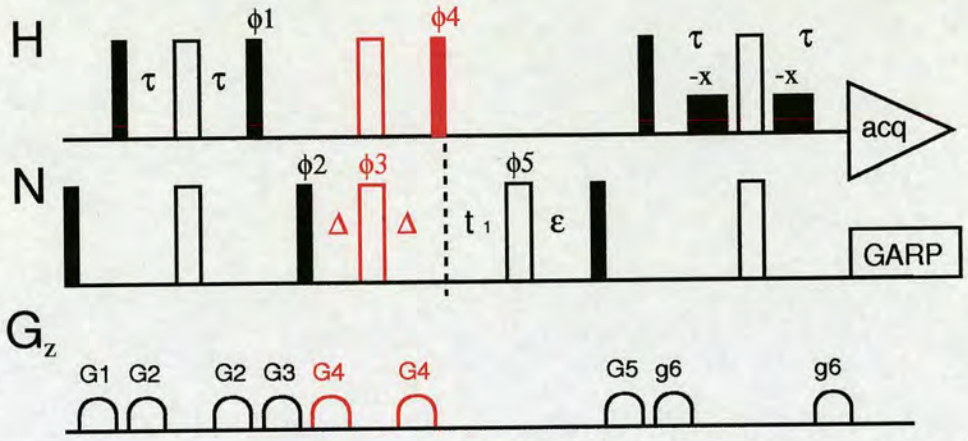

Figure 2.10: ^{15}N **-IPAP HSQC pulse sequence.** See section 2.2 for a description of the symbols and conventions used in the pulse sequence diagram. Delays are as follows: τ=2.67 ms ($1/4J_{NH}$) and Δ=2.67 ms ($1/4J_{NH}$). For double labelled samples there is an additional adiabatic 180° ^{13}C pulse (not shown: 20.1% smoothed 0.5 ms chirp) located centrally within the t_1 period and ϵ=0.5 ms to compensate for evolution during this pulse (for ^{15}N -labelled samples, ϵ is set to the initial t_1 length of 6μs). Gradients are G1=30%, G2=7%, G3=80%, G4=5%, G5=11% and g6=23% and phases are ϕ_1=-y,y and ϕ_5=8(x),8(-x). For the IP spectrum, ϕ_2=2(x),2(-x); the receiver phases are rx=x,-x,-x,x and the spin echo coloured in red is not present; for the AP spectrum ϕ_2=2(-y),2(y); ϕ_3=4(x),4(y),4(-x),4(-y); ϕ_4=8(x),8(-x) and the receiver phases are rx=x,-x,-x,x,-x,x,x,-x. Quadrature detection during t_1 is achieved by incrementing phases ϕ_2 and ϕ_3 according to the States-TPPI protocol.

Experiments for measuring $^1J_{NC'}$ and $^1J_{C\alpha C'}$

Schemes similar to the IPAP HSQC experiment described above can be used to record the $^1J_{NC'}$ and $^1J_{C\alpha C'}$ couplings - these spin-state-selective methods [4] also create pairs of spectra in which the splitting is encoded as a difference between the frequencies of the two doublet components (here corresponding to the $^1J_{NC'}$ or $^1J_{C\alpha C'}$ coupling). The pulse sequence for $^1J_{NC'}$ measurement [81] is shown in Figure 2.11 - it can be seen here that the spin echo prior to ^{15}N frequency labelling creates ^{15}N magnetization that is either in-phase or antiphase with respect to the coupled ^{13}C' spin depending on the positioning of the two C'-selective 180° pulses. Post-acquisitional editing is used to separate the two doublet components into two subspectra in an analogous manner

[4] The spin-states referred to are the α and β spin states of the nucleus to which the observed ^{15}N spin is coupled, where the two spin states correspond to the two doublet components.

to the IPAP HSQC experiment.

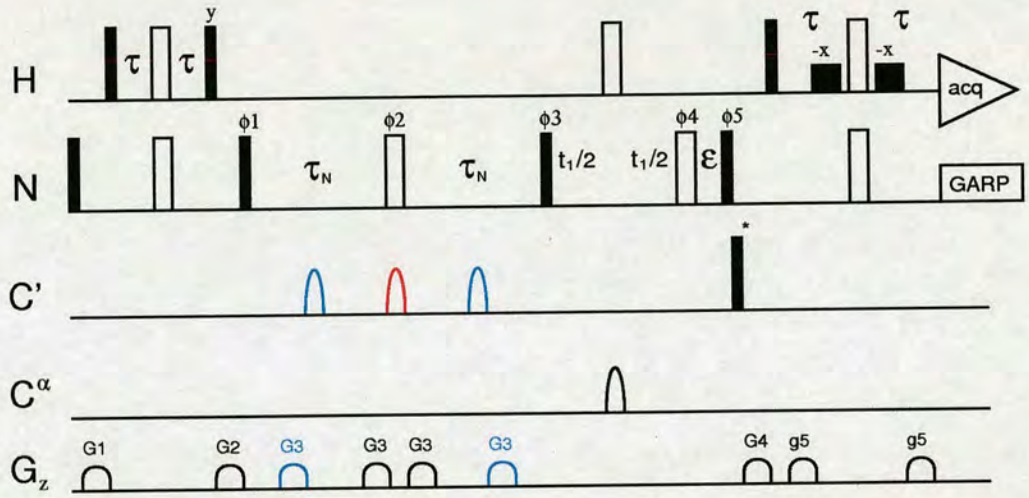

Figure 2.11: **HN(α/β-NC'-J) pulse sequence.** See section 2.2 for a description of the symbols and conventions used in the pulse sequence diagram. Delays are as follows: τ=2.5 ms ($1/4J_{NH}$), τ_N=16.6 ms ($1/4J_{NC'}$) and ϵ=256μs (to compensate for offset evolution during the Cα 180° pulse central to the t_1 period). Gradients are G1=30%, G2=7%, G3=80%, G4=5%, G5=11% and g6=23% and phases are ϕ_1=2(x),2(-x); ϕ_2=x; ϕ_3=2(x),2(-x); ϕ_4=4(x),4(-x) and the receiver phase rx=2(x),2(-x). For the IP spectrum the pulses coloured in blue are present and the pulse phase ϕ_5=x; for the AP spectrum, the pulses coloured in red are present instead of those coloured in blue and the pulse phase ϕ_5=y. The pulse marked with an asterisk (*) is an optional purge pulse for elimination of the unwanted IP/AP component. Quadrature detection during t_1 is achieved by incrementing phases ϕ_1, ϕ_2 and ϕ_3 according to the States-TPPI protocol.

The experiment used to measure the $^1J_{C^{\alpha}C'}$ coupling [83] is a little more complicated than the experiment to measure $^1J_{NC'}$ - the pulse sequence is shown in Figure 2.12. An ^{15}N HSQC-like map correlating the ^{15}N and ^{1}H frequencies but encoding a coupling involving neither of these nuclei can be created using a multiple quantum coherence between ^{15}N and ^{13}C'. After an INEPT step from amide proton to ^{15}N the $2H_zN_y$ magnetization is refocussed to in-phase N_x, antiphase $2N_zC'_y$ is created and the following spin echo or dummy spin echo [5] results in the evolution of C' magnetization that is antiphase ($4N_zC'_yC^{\alpha}_z$) or in-phase ($2N_zC'_y$) with respect to the coupled C$^{\alpha}$ spin. Multiple quantum $2N_yC'_y$ / $4N_yC'_yC^{\alpha}_z$ coherences evolve during the t_1 ^{15}N frequency

[5] The $^1J_{NC'}$ coupling does not evolve during the dummy spin echo since the two 180° C$^{\alpha}$ -selective pulses are located centrally in each of the two τ_C intervals.

labelling period under the effects of both the ^{15}N offset and the $^{1}J_{C^{\alpha}C'}$ coupling and are afterwards converted into observable NH magnetization for detection. Post-acquisitional editing can be used to separate the two doublet components as before.

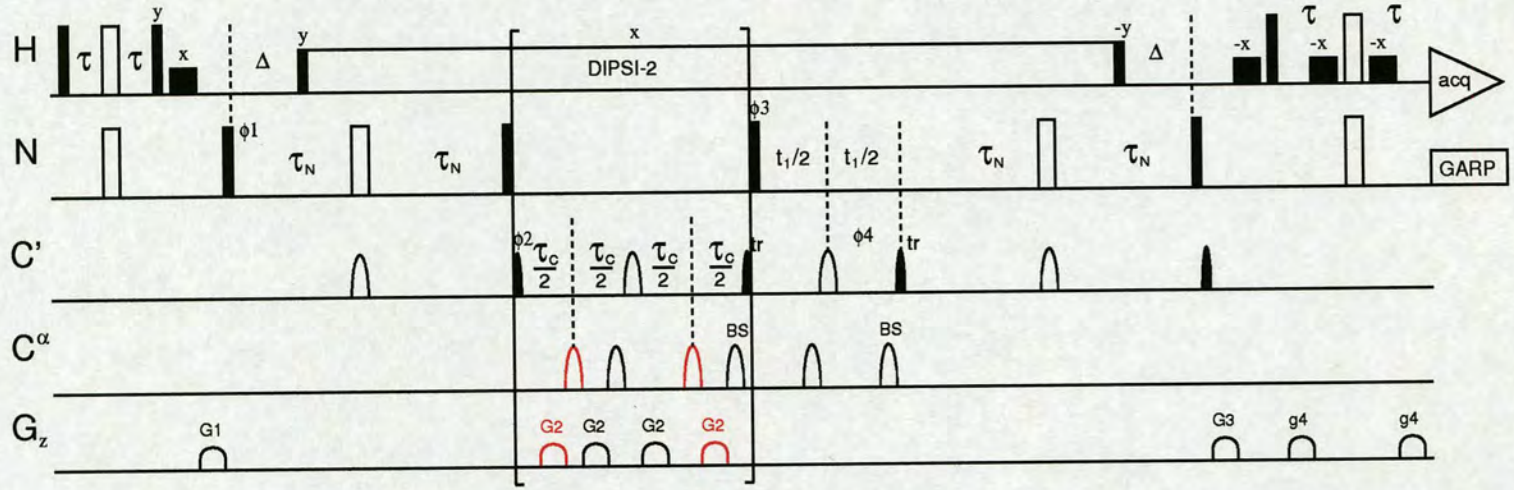

Figure 2.12: **HN(α/β-COCA-J) pulse sequence.** See section 2.2 for a description of the symbols and conventions used in the pulse sequence diagram. Delays are as follows: τ=2.3 ms ($1/4J_{NH}$), τ_N=12 ms, τ_C=4.55 ms and Δ=5.5 ms. Gradients are G1=50%, G2=8%, G3=60%, g4=30% and phases are ϕ_1=x,-x; ϕ_2=2(x),2(-x); ϕ_4=8(x),8(y),8(-x),8(-y) and the receiver phase rx=2(x,-x,-x,x,-x,x,x,-x),2(-x,x,x,-x,x,-x,-x,x). Phase ϕ_3=4(x),4(-x) for the IP spectrum and ϕ_3=4(y),4(-y) for the AP spectrum. The block contained in square brackets prior to t_1 differs for IP and AP spectra: for C^α pulses and gradients, elements coloured black are required for IP spectra and the alternative elements coloured red are required to obtain AP spectra. Quadrature detection during t_1 is achieved by incrementing ϕ_3 according to the States-TPPI protocol.

2.3 Relaxation and dynamic processes

Relaxation processes in an NMR experiment are those processes by which excited spin states return to the equilibrium ground state given by the Boltzmann distribution. Ultimately it is coupling between the spin system and surroundings or *lattice* that is responsible for relaxation. The simplest theoretical treatment of spin relaxation defines two phenomenological relaxation rate constants R_1, which is the *spin-lattice* or *longitudinal* relaxation rate constant, and R_2 which is the *spin-spin* or *transverse* relaxation rate constant: these quantities are defined in the Bloch equations [10] below (equations 2.19 - 2.21):

$$\frac{dM_z(t)}{dt} = \gamma(\mathbf{M}(t) \times \mathbf{B}(t))_z - R_1(M_z(t) - M_0) \tag{2.19}$$

$$\frac{dM_x(t)}{dt} = \gamma(\mathbf{M}(t) \times \mathbf{B}(t))_x - R_2 M_x(t) \tag{2.20}$$

$$\frac{dM_y(t)}{dt} = \gamma(\mathbf{M}(t) \times \mathbf{B}(t))_y - R_2 M_y(t) \tag{2.21}$$

where $\mathbf{M}(t)$ is the nuclear magnetization vector and $\mathbf{B}(t)$ is the applied magnetic field consisting of the static field and radiofrequency field. The longitudinal relaxation rate characterizes processes that lead to energy transfer from the spin system to the surroundings or *lattice*, and is a measure of the rate at which the spin system returns to equilibrium. The transverse relaxation rate constant on the other hand, is a measure of the rate at which transverse magnetization decays to zero. Spin-spin relaxation processes are adiabatic, resulting in dephasing of coherent magnetization without any transfer of energy to the lattice. Both relaxation rates have units of s^{-1} and are the inverse of the corresponding relaxation times: $T_1 = 1/R_1$, $T_2 = 1/R_2$.

The relaxation properties of a two spin system, IS, can be described with reference to the rates of transitions between the energy levels of the system (shown in Figure 2.13). By formulating differential equations to describe the populations of the four spin states $|\alpha\alpha>$, $|\alpha\beta>$, $|\beta\alpha>$ and $|\beta\beta>$ and making the identifications; $\rho_I = W_0 + 2W_I + W_2$, $\rho_S = W_0 + 2W_S + W_2$ (auto-relaxation rate constants) and $\sigma_{IS} = W_2 - W_0$ (cross-relaxation rate constant) one can derive the Solomon equations [102] for a two spin

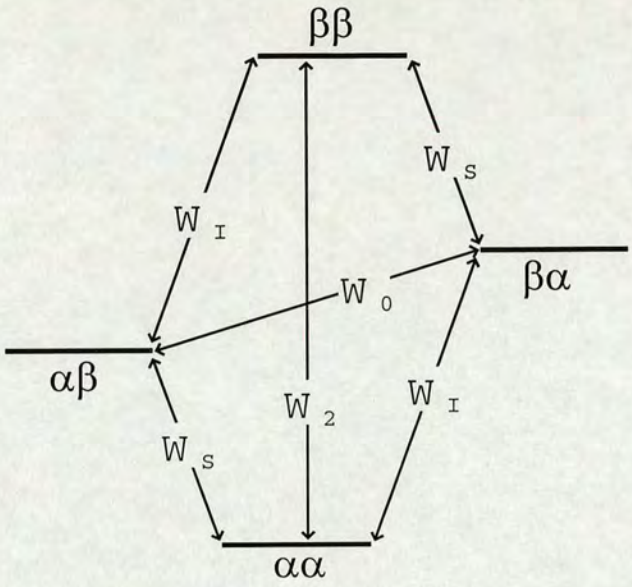

Figure 2.13: **Transitions in a two spin system.** Transition rate constants (W_I, W_S, W_0, W_2) for a two spin system IS: W_0 denotes a transition in which the two spins flip in the opposite sense (i.e. $\alpha \rightarrow \beta$ for one spin, $\beta \rightarrow \alpha$ for the other) whilst W_2 denotes a transition in which both spins flip in the same sense.

system (equations 2.22 - 2.23):

$$\frac{d\Delta I_z(t)}{dt} = -\rho_I \Delta I_z(t) - \sigma_{IS} \Delta S_z(t) \tag{2.22}$$

$$\frac{d\Delta S_z(t)}{dt} = -\rho_S \Delta S_z(t) - \sigma_{IS} \Delta I_z(t) \tag{2.23}$$

where $\Delta I_z(t)$ and $\Delta S_z(t)$ are the non-equilibrium magnitudes of the I_z and S_z operators, respectively. Using terminology introduced above for the Bloch equations, quantities such as R_{1I} and R_{2I} can be defined, which correspond to the observed autorelaxation rates of I_z and $I_{x,y}$ operators, respectively.

The cross-relaxation rate constant, σ_{IS}, relates to the rate of magnetization transfer between spins I and S, which produces an experimentally observable Nuclear Overhauser Effect (NOE) enhancement of the I signal when a non-equilibrium state is created in nucleus S. In a steady state NOE difference experiment, a pair of spectra are recorded: a reference spectrum, and a spectrum in which the S spin is saturated,

producing an enhancement $\eta_{IS} = (I_{sat} - I_{ref})/I_{ref}$ in the I spin signal (equation 2.24).

$$\eta_{IS} = \frac{\sigma_{IS}\gamma_S}{\rho_I\gamma_I} \tag{2.24}$$

where γ_I and γ_S are the gyromagnetic ratios of spins I and S, respectively. In the case of a heteronuclear two spin system, a quantity $NOE = I_{sat}/I_{ref} = \eta_{IS} + 1$, called the heteronuclear NOE, is normally referred to.

Two phenomena are largely responsible for the observed relaxation rates in a two spin system - the dipolar interaction between the spins and the Chemical Shift Anisotropy (CSA), both of which cause fluctuations in local fields. In addition to these two interactions, *chemical exchange* may contribute to the observed transverse relaxation time: $R_2^{obs} = R_2^{DD} + R_2^{CSA} + R_{ex}$. Chemical or conformational exchange processes can lead to severe linebroadening. The effects of chemical exchange depend on the relationship between the exchange rate R_{ex} and the chemical shift difference $\Delta\delta$ between the peaks observed for the two conformers: where $R_{ex} << \Delta\delta$, distinct peaks are observed for the two conformers; where $R_{ex} \sim \Delta\delta$, extensive line-broadening occurs and where $R_{ex} >> \Delta\delta$, a peak is observed at an chemical shift that is the average of the shifts for the two conformers.

Spins relax via fluctuating fields created by molecular tumbling: the component frequencies possessed by fluctuating magnetic fields can be expressed in terms of a *spectral density function* such as that given below in equation 2.25 (which is for a rigid rotor or spherical top).

$$J(\omega) = \frac{2}{5}\frac{\tau_c}{(1 + \omega^2\tau_c^2)} \tag{2.25}$$

where the correlation time τ_c is approximately a measure of the time taken for a molecule undergoing Brownian rotational diffusion to rotate by one radian. The rates of the transitions W_I, W_S, W_0 and W_2 can be expressed in terms of spectral density functions at the frequencies of the two spins: experimentally observable relaxation parameters for a heteronuclear two spin system NH can therefore be expressed in terms of the spectral density functions $J(\omega_H)$ and $J(\omega_N)$ [79]:

$$R_1 = \frac{d^2}{4}\left[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)\right] + c^2 J(\omega_N) \tag{2.26}$$

$$R_2 = \frac{d^2}{8}\left[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + J(\omega_H + \omega_N)\right] + \frac{c^2}{6}\left[4J(0) + 3J(\omega_N)\right] \tag{2.27}$$

$$NOE = \frac{d^2}{4}\left[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)\right]\frac{\gamma_H}{\gamma_N R_1} + 1 \tag{2.28}$$

where $c = \Delta\sigma\omega_N/\sqrt{3}$, $\Delta\sigma$ is the CSA of ^{15}N and $d = \mu_0 h\gamma_N\gamma_H/8\pi^2 r_{NH}^3$ is a constant scaling the strength of the NH dipolar interaction.

In the case of anisotropic rotational diffusion, the relaxation properties of spin pairs also depend on the orientation that the internuclear vector makes with the diffusion axes of the molecule. The spectral density for an ellipsoidal molecule undergoing Brownian rotational motion has been calculated by Woessner [116] and depends on five characteristic correlation times τ_i according to equation 2.29:

$$J(\omega) = \frac{2}{5}\sum_{i=1}^{5}\frac{A_i\tau_i}{1 + (\omega\tau_1)^2} \tag{2.29}$$

where $\tau_{1,2,3} = (4D_{xx} + D_{yy} + D_{zz})^{-1}, (D_{xx} + 4D_{yy} + D_{zz})^{-1}, (D_{xx} + D_{yy} + 4D_{zz})^{-1}$; $\tau_{4,5} = (6D_{iso} \pm 6(D_{iso}^2 - L^2)^{1/2}$, $D_{iso} = (D_{xx} + D_{yy} + D_{zz})/3$; $L^2 = (D_{xx}D_{yy} + D_{xx}D_{zz} + D_{yy}D_{zz})/3$ and A_i are coefficients. For a prolate ellipsoid undergoing axially symmetric rotational diffusion, $D_{xx} = D_{yy}$ and one can define $D_{\parallel} = D_{zz}$ and $D_{\perp} = D_{xx} = D_{yy}$. The coefficients are: $A_{1,2,3} = (3\cos^2\alpha - 1)^2/4,\ 3\sin^2\alpha\cos^2\alpha,\ (3/4)\sin^4\alpha$ where α is the angle between the N-H bond vector and the unique axis of the diffusion tensor (this is $D_{\parallel} = D_{zz}$ in the prolate case [6]). There are dedicated programs available for the fitting of a molecule's diffusion tensor using NMR relaxation data, e.g. *r2r1_diffusion* from the Palmer lab [78]. In this work, the program **Tensor** version 2.0 [31] is used to perform diffusion tensor fitting and for further analysis of the relaxation data.

The diffusional properties of a molecule may also be predicted from the structure without recourse to any experimental data by means of hydrodynamic modelling. In

[6] In the oblate case, since by convention $D_{zz} > D_{yy} > D_{xx}$, then $D_{zz} = D_{yy}$ and $D_{xx} = D_{\parallel}$ is the unique axis.

later Chapters (5 and 6) the program HYDROPRO [27], which employs a shell model to represent the molecule, is used for this purpose.

2.3.1 Measurement of ^{15}N relaxation parameters

Pulse sequences for measuring heteronuclear amide R_1 and R_2 relaxation rates (Figure 2.14, A. and B.) are based on the ^{15}N HSQC experiment: these consist of an initial refocussed INEPT transfer from amide ^{1}H to ^{15}N followed by a relaxation period T, ^{15}N frequency labelling (t_1) and finally, transfer via a reverse INEPT step back to ^{1}H for observation. The period T for the R_1 experiment is an inversion recovery period (initial magnetization -N_z) with ^{1}H decoupling to prevent cross-relaxation and cross-correlation between the dipolar and CSA relaxation mechanisms. In the case of the R_2 experiment T period (initial magnetization N_y), Carr-Purcell-Meiboom-Gill (CPMG) elements [68] are used to remove contributions from magnetic field inhomogeneity and diffusion: the length of the spin-echo delay here is kept intentionally short ($\sim$ 0.5 ms) to prevent evolution of the NH scalar coupling.

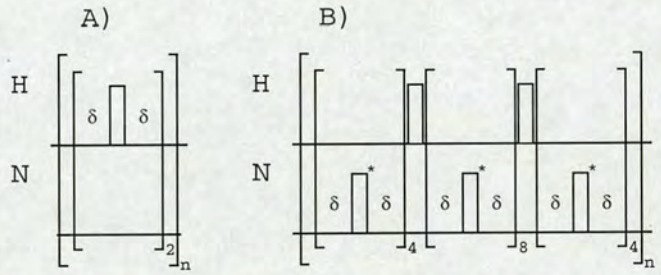

Figure 2.14: **Relaxation delay elements.** A. R_1 relaxation delay: $\delta = 2.5$ ms and 180° proton pulses are amide proton-selective (800 μs Gaussian shaped pulse). B. R_2 relaxation delay: $\delta \sim 0.5$ ms, nitrogen 180° pulses (marked with an asterisk *) are at a reduced power level relative to hard pulses.

By varying the length of the relaxation delay parametrically, the magnetization decay observed for experiments to measure R_1 or R_2 fits an exponential decay function, shown in equation 2.30 below:

$$I(t) = I_0 exp(-t/T_{1/2}) \tag{2.30}$$

where I is the non-zero limiting intensity as $t \rightarrow \infty$ as a consequence of pulse imperfections. Non-linear least squares fitting can be used to obtain the time constant, T_1 or T_2, from the measured intensities at different decay times t (in this work a C program written by Krystyna Bromek, using the *Marquardt method* implementation from Numerical Recipes in C [86], is used for fitting). An example dataset and the resulting fit are shown in Figure 2.15.

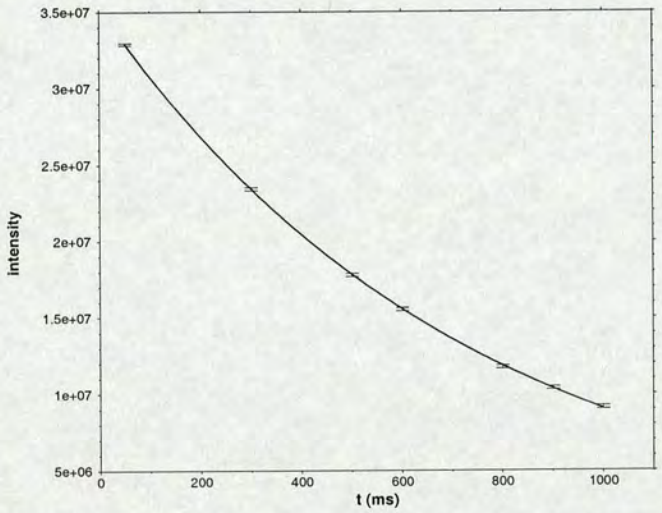

Figure 2.15: T_1 **decay curve.** The measured signal intensities at T_1 decay times of 51.1, 301.1, 501.1, 601.1, 801.1, 901.1 and 1001.1 ms for C122 of C4BP~1,2 (section 6.2), and the resulting fit line are shown - the fit gives a T_1 value of 735 ± 4 ms.

The experiment to measure the $\{^1H\}$-^{15}N NOE is somewhat different from the R_1 and R_2 experiments in that there is no inital INEPT transfer - initial ^{15}N magnetization is directly observed by means of a 90° nitrogen pulse followed by t_1 nitrogen frequency labelling and a reverse INEPT step. The steady state NOE enhancement is measured by recording a reference spectrum and a second spectrum in which protons are saturated during the recovery delay: the saturated spectrum shows an enhancement of the nitrogen amide signal caused by cross-relaxation of the saturated amide protons. The NOE is then calculated according to: $NOE = I_{sat}/I_{ref}$ where I_{sat} and I_{ref} are the NH peak intensities in the saturated and reference spectra, respectively.

2.3.2 The Model-Free Formalism

Nuclear magnetic relaxation experiments can be used to probe the internal motions of proteins in solution, and it is possible to extract the essential information contained is such data without modelling the dynamic processes responsible. In the 'model-free' approach suggested by Lipari and Szabo [61, 62], the essential dynamic information is reduced to two quantities: a generalized order parameter $\mathcal{S}$ which represents the degree to which internal motion is spatially restricted, and an effective correlation time τ_e which quantifies the rate (time scale) of the motion.

If the overall motion of the protein is isotropic and can be represented by a single correlation time τ_M, the following expression is obtained for the spectral density:

$$J(\omega) = \frac{2}{5}\left[\frac{\mathcal{S}^2\tau_M}{1+(\omega\tau_M)^2} + \frac{(1-\mathcal{S}^2)\tau}{1+(\omega\tau)^2}\right] \tag{2.31}$$

where $\tau^{-1} = \tau_M^{-1} + \tau_e^{-1}$. This assumes that internal motions are independent of the overall motion of the molecule, which is true provided the internal motions are so fast as to be close to the extreme narrowing limit ($\omega\tau_e << 1$).

When the tumbling motion of the protein is not isotropic a single overall correlation time τ_M does not exist, although model-free parameters for local motion can still be calculated: it is possible to fit a local rotational correlation time τ_{mi} for each residue independently without making any assumptions about the overall motion of the protein. Alternatively, the rotational diffusion tensor of the molecule can be optimized and related to apparent local rotational correlation times.

A simple physical model which gives a picture of the spatial restriction that $\mathcal{S}$ represents is diffusion within a cone ($\mathcal{S}$ here is equivalent to $\mathcal{S}_f$ in Figure 2.16). If the semiangle of the cone is θ then the order parameter that characterizes this motion, $\mathcal{S}_{cone}$, is given by equation 2.32:

$$\mathcal{S}_{cone} = \frac{1}{2}(cos\theta)(1+cos\theta) \tag{2.32}$$

Values of $\mathcal{S}$ range from zero, which means that internal motions are unrestricted, to unity which corresponds to complete restriction of motion within the molecular reference frame, i.e. a cone angle of zero.

In some cases the internal correlation function of a residue cannot be approximated by a single exponential decay as in the Lipari-Szabo treatment. Some residues are better characterized by two order parameters $\mathcal{S}_f$ and $\mathcal{S}_s$, which correspond to fast (ps-ns) and slow (ns, $< \tau_M$) motions [24]. Assuming the fast and slow motion correlation times are sufficiently different, the total generalized order parameter can be decomposed as follows:

$$\mathcal{S}^2 = \mathcal{S}_f^2 \mathcal{S}_s^2 \tag{2.33}$$

The correlation time for very fast internal motions makes no significant contribution to NMR relaxation parameters but the order parameter $\mathcal{S}_f$ does. The spectral density therefore becomes:

$$J(\omega) = \frac{2}{5}\left[\frac{\mathcal{S}^2\tau_M}{1+(\omega\tau_M)^2} + \frac{(\mathcal{S}_f^2 - \mathcal{S}^2)\tau}{1+(\omega\tau)^2}\right] \tag{2.34}$$

Where motion exists on two time scales, as in the case where the two parameters $\mathcal{S}_f^2$ and $\mathcal{S}_s^2$ are fitted, simple motional models can be used to visualize the meaning of these parameters: e.g. diffusion within a cone for $\mathcal{S}_f^2$ (fast ns motion) and a two-site jump for $\mathcal{S}_s^2$ (slow ns motion) [24] - this is illustrated in Figure 2.16 below.

When fitting R_2, an additional phenomenological exchange term R_{ex} may be required to account for chemical exchange processes (ms-μs) that contribute to the decay of transverse magnetization. An approximate expression for two site chemical exchange in a CPMG experiment [65] can be added for this purpose. A more complicated model including terms for R_{ex} and τ_e can also be considered - the motion on two different time scales can again be visualized by a combination of two different physical models such as diffusion in a cone and two-site jump.

In summary, there are five dynamical models that can be used to fit experimental relaxation data:

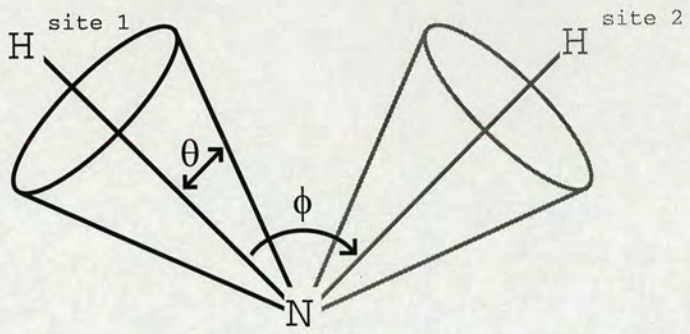

Figure 2.16: **Diffusion in a cone and two-site jump models.** The cone model for $\mathcal{S}_f^2$ gives $\mathcal{S}_f = \frac{1}{2}(cos\theta)(1+cos\theta)$ where θ is the semiangle of the cone about the average NH bond vector. The two-site jump model for $\mathcal{S}_s^2$ gives $\mathcal{S}_s^2 = (1 + 3cos^2\phi)/4$ where ϕ is the angle between NH bond vectors for H at site 1 and site 2.

1) $\mathcal{S}^2$ only : very fast motion, $\tau_f \rightarrow 0$

2) $\mathcal{S}^2$ and τ_e : Lipari-Szabo

3) $\mathcal{S}^2$ and $\mathbf{R_{ex}}$: Model 1 with chemical exchange contribution

4) $\mathcal{S}^2$, τ_e and $\mathbf{R_{ex}}$: Model 2 with chemical exchange contribution

5) $\mathcal{S}^2$, $\mathcal{S}_f^2$ and $\tau_e = \tau_s$: extended Lipari-Szabo

The selection strategy suggested by Mandel et al. [69] can be used to decide which of these models best describe the relaxation data for a particular residue - the quality of the fit between a model and the experimental data can be assessed by using a Monte Carlo procedure to simulate a distribution of sum of squared error (SSE) values that approximate to a χ^2 distribution with $d = n - m$ degrees of freedom (n is the number of relaxation parameters and m is the number of fitted model-free parameters). To assess whether using a more complicated model M2 gives a statistically significant improvement to the fit over less complicated model M1, an F-statistic can be calculated from the simulated SSE values as follows:

$$F = [d_2/(d_1 - d_2)][SSE_1(i) - SSE_2(i)]/SSE_2(i) \tag{2.35}$$

An F-statistic greater than the critical $(1 - \alpha)100\%$ value implies that model M2

describes the data more accurately than M1.

The process of fitting an appropriate diffusional model (isotropic, axially symmetric anisotropic or fully anisotropic) / determining the rotational correlation time and selecting motional models that account for the observed relaxation data has been implemented in a statistically rigorous way in the program **Tensor** version 2.0 [31], which is used in Chapters 5 and 6 to analyze the relaxation data for DAF~2,3 and C4BP~1,2.

2.4 Residual dipolar couplings and the molecular alignment tensor

Dipolar interactions between spins in solution are averaged to zero by isotropic tumbling. Residual dipolar couplings (RDC) [7] can be reintroduced however, by creating a degree of molecular order so that tumbling is no longer entirely isotropic. Although some molecules exhibit an anisotropic magnetic susceptibility and are aligned solely by the magnetic field [108], in most cases the use of an aligning medium is required. Historically, liquid crystalline systems have been used for this purpose: in early work by Saupe and Englert [97] a large degree of order was used to align small molecules. The spectra studied were highly complex due to the large size of the resulting dipolar coupling constants. More recently, Tjandra and Bax [103] used a dilute lipid bicelle system [95] to create a very small degree of order resulting in measurable dipolar couplings in proteins $\sim 10^3 - 10^4$ times weaker than their static values. A variety of aligning systems have since appeared in the literature:

- Filamentous phage [23, 40]
- Polyethylene glycol / alkyl alcohol systems [92]
- Cetylpyridinium chloride (CPCl) / hexanol [90] and CPBr / hexanol [7]
- Strain-induced alignment in strained gels [93, 110]
- Improved ether-linked bicelle systems [76, 20]

[7] A number of detailed reviews covering the theory and practical applications of RDCs can be found in the literature [8, 87, 18].

- Bicelles with paramagnetic ions [89]
- Cellulose crystallites [35]

The alternative media that have been developed overcome some of the limitations of bicelles, which are stable only over narrow ranges of temperature and pH and decompose over time by hydrolysis. Simple chemical surfactants are one such alternative, offering the advantages of greater robustness and inexpensiveness: two such systems have been used in this work - CPBr / hexanol and polyethylene glycol / hexanol.

The use of a quasi-ternary surfactant/salt/alcohol system for alignment was first reported by Prosser *et al.* in 1998 [90]. A mixture of cetylpyridinium chloride (CPCl), NaCl and hexanol forms a lamellar liquid crystalline (LLC) Helfrich phase of bilayers with large interlamellar spacings that are able to accommodate protein molecules of the size range suitable for study by NMR. Typically, high salt concentrations of $\sim$ 200 mM NaCl are required to stabilize the LLC phase for this system, which may be problematic. An alternative system consisting of a mixture of cetylpyridinium bromide (CPBr), NaBr and hexanol has been described by Barrientos *et al.* [7] - in contrast to the CPCl-based system, this system is most stable at low salt concentrations $\sim$ 25 mM NaBr.

Residual dipolar couplings in solution state NMR manifest as a change in the observed coupling constant between a pair of nuclei: the through-space dipolar coupling is additive with the through-bond scalar coupling, i.e. $J_{observed} = J_{scalar} + D_{residual}$. Equation 2.36, taken from Ernst *et al.* [33], gives the Hamiltonian for the dipolar contribution to the splitting observed for a pair of heteronuclear spins I and S:

$$H_D^{IS} = D_{max}^{IS} \langle I_z S_z (3\cos^2\theta - 1) \rangle \tag{2.36}$$

where θ is the angle between the internuclear vector and the magnetic field and D_{max}^{IS} is the maximum value of the static dipolar coupling, the value of which (in units of Hz) is:

$$D_{max}^{IS} = -\mu_0(h/2\pi)\gamma_I\gamma_S/(4\pi^2 r_{IS}^3) \tag{2.37}$$

μ_0 is the magnetic permeability of vacuum, h is Planck's constant, γ_I and γ_S are the gyromagnetic ratios of the two spins and r_{IS} is the effective average internuclear distance. Note that two approximations have been made here: firstly, the high-field approximation means that non-secular terms can be neglected and secondly, since the two spins are heteronuclear, weak coupling means that transverse spin operator terms can be neglected. Effectively, only the z components of the local fields created by the nuclear dipoles are considered. The angular part of the dipolar Hamiltonian can be expressed as a second rank Legendre function $P_2(x) = \frac{1}{2}(3x^2 - 1)$ so that the expression for the dipolar splitting simplifies to $D^{IS} = D_{max}^{IS}\langle P_2(\cos^2\theta)\rangle$.

The angle θ between the internuclear vector and magnetic field B_0 depends on the orientation of the molecular coordinate system with respect to the laboratory frame as well as the orientation of the internuclear vector with respect to the molecular coordinate system. However, if the molecule can be regarded as rigid and its structure is known, then the problem is greatly simplified - in this case the bond orientations are fixed and the average molecular orientation can be described by a 3×3 Saupe order matrix [96] [8]:

$$\langle P_2(\cos^2\theta)\rangle = \sum_{ij=\{x,y,z\}} S_{ij}\cos(\alpha_i)\cos(\alpha_j) \tag{2.38}$$

where the $\cos(\alpha)$ direction cosines define the angles between a given internuclear vector and the molecular coordinate system. A principal axis system $\{x', y', z'\}$ (PAS) in which the order matrix is diagonal can always be found, giving a molecular alignment tensor with three principal values $S_{x'x'}$, $S_{y'y'}$ and $S_{z'z'}$ where $|S_{z'z'}| \geq |S_{y'y'}| \geq |S_{x'x'}|$ by convention. In polar coordinates this can be expressed in the form of equation 2.39 below, where the primes have been dropped:

[8] This is a traceless, symmetric matrix and therefore has only five independent elements.

$$D^{IS} = D^{IS}_{max} \cdot \mathbf{S} \cdot S_{zz} \left[\left(\left(\frac{3}{2}\cos^2\theta - \frac{1}{2} \right) + \frac{\eta}{2}\sin^2\theta\cos 2\phi \right) \right] \tag{2.39}$$

θ and ϕ are the polar angles defining the orientation of the internuclear vector with respect to the PAS and $\eta = (S_{xx} - S_{yy})/S_{zz}$ is the asymmetry parameter [9]. The generalized order parameter $\mathbf{S}$ ($0 \leq \mathbf{S} \leq 1$) has been introduced here to account for internal motions independent of the overall molecular tumbling [10].

The three principal values of the alignment tensor S_{xx}, S_{yy} and S_{zz} can be estimated from a histogram of the measured RDC values [21] as shown in Figure 2.17. In order to create a histogram combining different classes of RDC (e.g. one bond NH, NC', $C^{\alpha}C'$ and $C^{\alpha}H^{\alpha}$) it is necessary to scale each coupling relative to one class; commonly NH, according to the effective internuclear distance r^{eff}_{XY} and the gyromagnetic ratios of the atoms involved (equation 2.40):

$$D^{scaled}_{XY} = D^{unscaled}_{XY} \left(\frac{\gamma_N \gamma_H}{\gamma_X \gamma_Y} \right) \left(\frac{r^{eff}_{XY}}{r^{eff}_{NH}} \right)^3 \tag{2.40}$$

Values of average effective bond lengths, gyromagnetic ratios and the resulting scaling factors for classes of RDC commonly used as restraints in protein structure determination are given in Table 2.2. Empirically determined optimal scaling factors are also shown - these have been determined by Cornilescu and Bax for ubiquitin [25], and include the effects of motions (i.e. $\mathbf{S}$).

Singular value decomposition (SVD)

Five independent RDCs are sufficient to completely define the alignment tensor for a rigid molecular unit of known structure: one numerical method that can be used to determine the alignment order matrix in such a case is singular value decomposition

[9] In diagonalizing the matrix (equivalent to a rotation defined by 3 Euler angles), the alignment description is reduced to two independent quantities: a principal order parameter S_{zz} and an asymmetry parameter η. In other derivations [103] an alignment tensor A is defined such that $A_{ii} = \frac{2}{3}S_{ii}$: the reduced order parameter description in this case consists of an axial component of the tensor $A_a = S_{zz}$ and a rhombicity $R = \frac{2}{3}\eta$.

[10] A crude estimate can be made using the S^2 values obtained from NMR relaxation data (section 2.3.2).

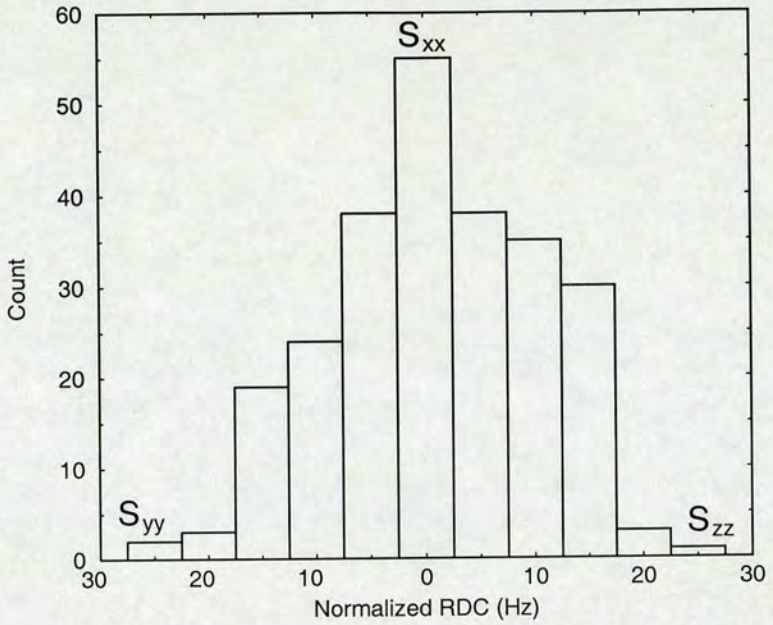

Figure 2.17: **Histogram determination of the principle order values S_{xx}, S_{yy} and S_{zz}.** Histogram of one bond NH, NC', $C^{\alpha}C'$ and $C^{\alpha}H^{\alpha}$ RDCs measured for the protein ABA-1A (Chapter 5). All 249 couplings are scaled relative to the NH RDC and have been grouped into bins of 5 Hz. The RDC distribution resembles a solid state NMR powder pattern and the 3 extrema correspond to $D_{\mathrm{NH}}^{max} S_{yy}$, $D_{\mathrm{NH}}^{max} S_{xx}$ and $D_{\mathrm{NH}}^{max} S_{zz}$ as shown. Hence the value of S_{zz} determined from this plot would be 1.03^{-3} since $D_{\mathrm{NH}}^{max} = 24350$ Hz.

RDC	r_{XY}^{eff} (Å)	$\gamma_X \gamma_Y$ $(10^7 \mathrm{radT^{-1}s^{-1}})^2$	α_{XY}	empirical α_{XY}
NH	1.041	-72.567	1.00	1.00
NC'	1.329	-18.251	8.27	8.33
$C^{\alpha}C'$	1.525	45.270	-5.04	-5.05
$C^{\alpha}H^{\alpha}$	1.117	179.995	-0.498	-0.481

Table 2.2: **Scaling factors for commonly used RDC classes relative to NH.** Effective average bond lengths are taken from Ottiger *et al.* [75] and gyromagnetic ratios from van de Ven [114]. The resulting scaling factor α_{XY} gives: $D_{\mathrm{XY}}^{scaled} = \alpha_{\mathrm{XY}} D_{\mathrm{XY}}^{unscaled}$.

[63]. This requires a set of linear equations of the form $\mathbf{A} \cdot \mathbf{x} = \mathbf{b}$ to be solved. From equations 2.36 and 2.38 above it can be seen that the value of a dipolar coupling is dependent upon the orientation of the internuclear vector with respect to the axis system as well as the order matrix elements representing the alignment tensor: for a column vector of n couplings (equation 2.42b), matrix $\mathbf{A}$ consists of the direction

cosines of the internuclear vectors (equation 2.41) and the column vector $\mathbf{x}$ contains the 5 independent elements of the alignment tensor (equation 2.42a).

$$\mathbf{A} = \begin{pmatrix} \cos^2\phi_y^1 - \cos^2\phi_x^1 & \cos^2\phi_z^1 - \cos^2\phi_x^1 & 2\cos\phi_x^1\cos\phi_y^1 & 2\cos\phi_x^1\cos\phi_z^1 & 2\cos\phi_y^1\cos\phi_z^1 \\ \cos^2\phi_y^2 - \cos^2\phi_x^2 & \cos^2\phi_z^2 - \cos^2\phi_x^2 & 2\cos\phi_x^2\cos\phi_y^2 & 2\cos\phi_x^2\cos\phi_z^2 & 2\cos\phi_y^2\cos\phi_z^2 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cos^2\phi_y^n - \cos^2\phi_x^n & \cos^2\phi_z^n - \cos^2\phi_x^n & 2\cos\phi^n x_{\cos}\phi_y^n & 2\cos\phi^n x_{\cos}\phi_z^n & 2\cos\phi^n y_{\cos}\phi_z^n \end{pmatrix} \tag{2.41}$$

$$a) \quad \mathbf{x} = \begin{pmatrix} S_{yy} \\ S_{zz} \\ S_{xy} \\ S_{xz} \\ S_{yz} \end{pmatrix} \qquad b) \quad \mathbf{b} = \begin{pmatrix} D_{red}^1 \\ D_{red}^2 \\ \cdot \\ \cdot \\ D_{red}^3 \end{pmatrix} \tag{2.42}$$

The inverse of the matrix $\mathbf{A}$ must be found in order to determine the order matrix $\mathbf{x}$ since $\mathbf{x} = \mathbf{A}^{-1} \cdot \mathbf{A} \cdot \mathbf{x} = \mathbf{A}^{-1} \cdot \mathbf{b}$. The singular value decomposition of the $m \times n$ matrix $\mathbf{A}$ is: $\mathbf{A} = \mathbf{U} \cdot \mathbf{W} \cdot \mathbf{V}^T$ where $\mathbf{U}$ is an $m \times n$ column orthogonal matrix, $\mathbf{W}$ is an $n \times n$ diagonal matrix with nonnegative elements ω_i and $\mathbf{V}^T$ is the transpose of an $n \times n$ orthogonal matrix $\mathbf{V}$. The inverse of the matrix $\mathbf{A}$ can therefore be expressed as: $\mathbf{A}^{-1} = \mathbf{V} \cdot [\mathrm{diag}(1/\omega_i)] \cdot \mathbf{U}^T$. Experimental uncertainties are taken into account by solving for a randomly sampled Gaussian distribution of values of $D_{reduced}^n$ about the measured value with a spread depending upon the estimated error. The solutions obtained from the procedure therefore reflect the degree of uncertainty in the order matrix elements as a consequence of errors in the data.

Interpretation of the alignment tensor

The alignment tensor for a rigid molecular fragment contains useful structural information. If the alignment tensors for two or more such fragments can be determined, the relative orientations of the fragments are constrained. There is however, a 4-

fold degeneracy [11] in the information content of RDCs: the values of the couplings are insensitive to inversion of each of the alignment tensor axes (studying equation 2.39 it can be seen that $\cos^2\theta = \cos^2(180 - \theta)$, $\sin^2\theta\cos2\phi = \sin^2(180 - \theta)\cos2\phi$ and $\sin^2\theta\cos2\phi = \sin^2\theta\cos2(180 - \phi)$). In the absence of additional data, such as NOE distance restraints or RDCs measured in a medium with a different alignment tensor, the relative orientation of the fragments will remain ambiguous. In such cases, it is useful to be able to compare the different possibilities and assess how reasonable they are. Two approaches are followed in later Chapters: the first is to use the program **Module** written by Dosset *et al.* [30] to visualize the different relative orientations of molecular fragments and the second is to evaluate numerically the principal values of the alignment order matrix (eigenvalues) and plot the axis directions (eigenvectors) corresponding to each. Figure 2.18 illustrates the *Sanson-Flamsteed* or *sinusoidal* projection used to map the order matrix eigenvectors obtained by the SVD method described above.

2.5 Incorporating RDC restraints into NMR structure calculations

Structure determination by NMR relies on minimizing the energy of a system containing a number of terms representing empirical knowledge of idealized structural geometry in addition to experimental measurements, as in equation 2.43 below:

$$E_{\text{overall}} = E_{\text{FF}} + E_{\text{NMR}} \tag{2.43}$$

where E_{FF} represents the force-field and includes terms for van der Waals forces, ideal bond lengths and the E_{NMR} term ensures that experimentally determined restraints (typically NOE distance restraints and coupling constants) are satisfied. An additional energy term can be added into E_{NMR} to ensure that measured RDC values are satisfied in the final structure - the simplest form of the energy term E_{RDC} is given in equation 2.44 below:

[11] In the case of an axially symmetric alignment tensor, i.e. $\eta = 0$, there are an infinite number of possibilities described by a double cone shape.

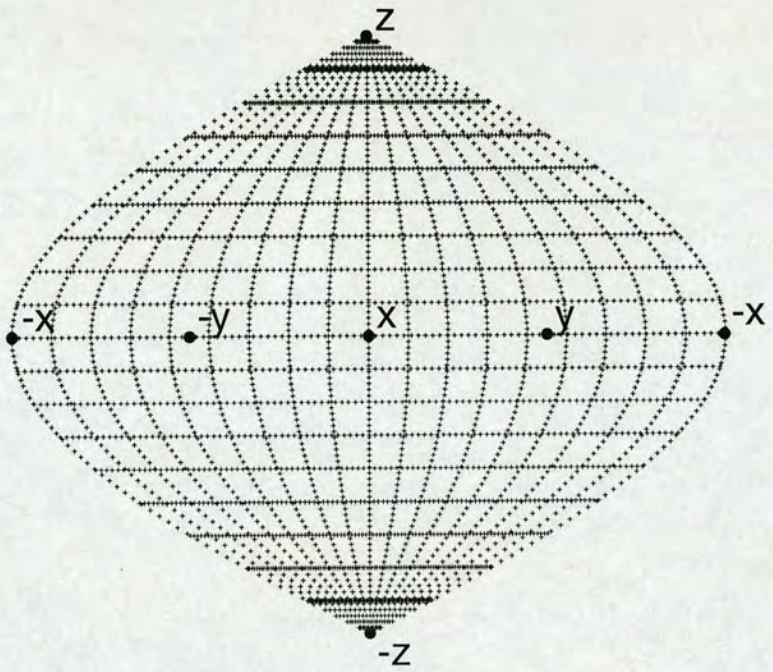

Figure 2.18: **The Sanson-Flamsteed projection.** The directions of the alignment tensor eigenvectors in the arbitrary Cartesian coordinate system x, y, z can be mapped onto a two-dimensional Sanson-Flamsteed projection: the positions of the 3D x, y and z axes on this 2D projection are shown.

$$\mathrm{E_{RDC}} = k_{\mathrm{RDC}} \left(D_{calc} - D_{obs}\right)^2 \tag{2.44}$$

where D_{calc} is the calculated RDC according to the structure, D_{obs} is the measured RDC and k_{rdc} is the force constant for this energy term. Knowledge of the molecular alignment tensor is required in order to be able to calculate D_{calc}. One approach is to determine the axial component of the alignment tensor, $D_a = (D_{max}/2)A_a$ and rhombicity R (see section 2.4) from the distribution of measured RDCs [21] and allow the orientation of the PAS to float: this is achieved in the SANI module [106] of the structure calculation program CNS [17] by creating an artificial tetraatomic molecule with bonds O-X, O-Y and O-Z to represent the PAS. This suffers from the disadvantage that accurate values of D_a and R can be difficult to obtain from a limited RDC distribution, and a search of R values is often performed in repeat sets of calculations [22].

An alternative algorithm allowing RDC measurements to be used as restraints has been implemented in the ISAC module [70, 94] of CNS: in contrast with SANI this does not require prior knowledge of the alignment tensor. The Implicit Saupe tensor Alignment Constraint (ISAC) module makes use of the fact that a RDC can be described by a function that is linear in five parameters representing the alignment tensor and non-linear in parameters describing the internal degrees of freedom of the molecule (torsion angles etc. that define the structure). The alignment tensor can therefore be eliminated from the problem of finding a structure that best fits the measured RDC values: a linear least squares minimization of the alignment tensor is achieved by implicitly adjusting the alignment tensor parameters during a non-linear minimization of the structural parameters representing the internuclear vector for each RDC. The least squares energy penalty for dipolar couplings D_i is given in equation 2.45 where S_m is the alignment order matrix in five-dimensional, irreducible notation and $T_i^{2m}(\alpha)$ is a non-linear function in the internal coordinates α describing the coupling i:

$$\chi^2(\alpha, \mathbf{S}) = \sum_i \left[D_i - \sum_{m=-2}^{2} S_m^* T_i^{2m}(\alpha) \right]^2 \tag{2.45}$$

2.6 Regulators of Complement Activation

The complement system is the principal innate immune response mechanism in humans and is divided into the *classical* and *alternative* pathways, both of which result in the formation of **C3b** from **C3**. The proteins C3 and C4 both belong to the α2-macroglobulin (α2M) family and possess reactive internal thioester bonds that are sequestered and accessible only to small molecules such as water. The thioester bonds are exposed upon cleavage of the C3a and C4a fragments (C3a and C4a are anaphylotoxins promoting inflammation). In the case of C3 a *tickover mechanism* exists in that there is spontaneous, slow hydrolysis of the thioester bond by water. This leads to a conformational change in C3 and allows association of factor B to create an unstable fluid-phase *C3 convertase* which can cleave C3 to C3b + C3a. The lifetime of the unstable C3-H_2O-B convertase is increased by factor D mediated cleavage of factor B

to Bb (releasing Ba) - this is the stable fluid phase C3 convertase C3-H_2O-Bb. The C3b produced by this tickover mechanism is reactive towards nucleophilic attack by hydroxyl and amine groups and is therefore constantly deposited on cell surfaces in a non-specific manner. Binding of factor B to membrane-bound C3b and subsequent cleavage of B to Bb by factor D leads to the formation of the membrane-bound C3 convertase C3bBb, which produces more C3b until it is inactivated. These events, summarized in Figure 2.19, are termed the alternative pathway. In contrast, the classical pathway requires the formation of activated C1 complex (initiated by antigen-bound antibody) or mannose-binding lectin (MBL) [12] in order to cleave C4 and C2 to form the membrane-bound classical pathway C3 convertase C4bC2b.

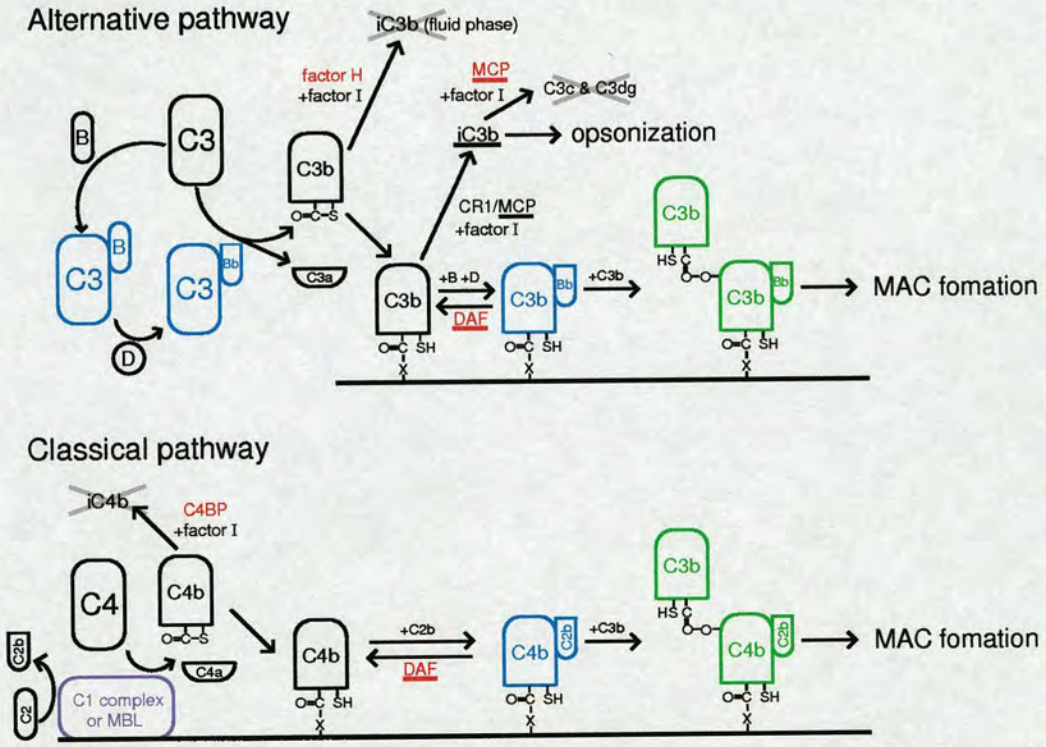

Figure 2.19: **Complement cascade.** The C5 convertase-forming events for the alternative and classical pathways are depicted: C4 convertases are shown in magenta, C3 convertases in blue, C5 convertases in green and complement down-regulators in red; membrane-bound DAF, MCP and iC3b are underlined. Factor H and C4BP are fluid-phase regulators.

Covalent attachment of a C3b molecule via its thioester to an existing membrane-

[12] Sometimes this is referred to as a distinct pathway - the lectin pathway.

bound C3 convertase results in a C5 convertase complex which begins a cascade of catalytic events termed the terminal pathway that amplify the response to non-host cells. The end results of complement activation are assembly of membrane attack complexes (MAC) and opsonization, which marks foreign bodies for phagocytosis. The regulators responsible for preventing assembly of the complement cascade convertases on self cells are known as regulators of complement activation (RCA): they consist mainly or wholly of repeated modular domains called *sushi*, short consensus repeat (SCR) or complement control protein (CCP) domains shown in Figure 2.20 and are encoded by a cluster of genes located on chromosome 1.

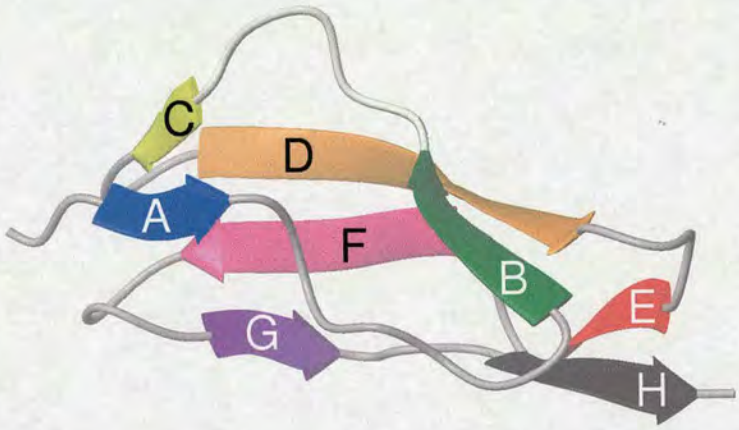

Figure 2.20: **CCP domain structure** β-strands for the x-ray structure of CCP 3 from VCP [73]: the hypervariable loop is located between strands B and C; a β-bulge is located between strands E and F; the N-terminus is to the left of the structure and the C-terminus to the right.

CCP domains consist of 60-70 amino acids arranged in stretches of short β-strands/loops/bulges separated by turns (Figure 2.20), resulting in an ellipsoidal globular/β-barrel structure approximately 30Å in length with N- and C-termini at opposite ends and containing two conserved internal disulfides (in a 1-3, 2-4 pattern) and a buried tryptophan [91, 74]. The β strands of a CCP domain are labelled A-H where strands B and C are separated by a particularly unstructured *hypervariable loop*. Individual domains are joined by 3-8 amino acid linkers in a head to tail arrangement, creating RCA structures that resemble a string of beads. Complete RCA proteins are too large to study

by NMR, but a strategy of dissecting the modular protein into smaller fragments and later combining the results to create a picture of the intact whole [19] must instead be pursued. However, since in many cases more than one CCP module is required for biological activity [52] the study of multiple overlapping constructs, typically of module pairs [101], has proven to be a successful strategy for studying this class of proteins by NMR. Two such constructs are examined in this work (DAF∼2,3 in Chapter 5 and C4BP∼1,2 in Chapter 6).

Chapter 3

Measuring $^1D_{C^\alpha H^\alpha}$

Of the four commonly used one-bond RDCs ($^1D_{NH}$, $^1D_{NC'}$, $^1D_{C^\alpha C'}$ and $^1D_{C^\alpha H^\alpha}$), the $^1D_{C^\alpha H^\alpha}$ constant is the most difficult to measure. The two main problems associated with NMR experiments used to measure $^1D_{C^\alpha H^\alpha}$ are rapid C^α relaxation and overlap of H^α resonances with the residual water signal. The latter hampers acquisition of ^{1}H-^{13}C HSQC spectra, which would otherwise provide a simple means of determining $^1D_{C^\alpha H^\alpha}$ by abolishing decoupling. Overlap with the water signal can be overcome by preparing samples in pure D_2O, although H^α dispersion is still poor and several experiments for the measurement of RDCs utilizing H^N for detection cannot be used: a sample in H_2O would also be required in order to obtain all four couplings. Unfortunately this introduces some uncertainty in that the alignment tensor for the two sets of RDCs could be different. An alternative solution is to use some kind of magnetization transfer scheme and e.g. transfer the $C^\alpha H^\alpha$ splittings to NH resonances.

3.1 Review of existing methods

Several methods for the measurement of the $^1D_{C^\alpha H^\alpha}$ coupling have been published in recent years [88]. These can be broadly grouped into intensity based methods, which involve fitting crosspeak intensities to a function, and frequency based methods in which the couplings are encoded as a splitting of the signals in one dimension.

Intensity based methods can be very precise, providing a precision of $< 1\%$ of the value of the coupling [104, 43], but require long experimental times to collect the necessary

number of points for fitting. This is particularly problematic when measuring the $^1D_{C^\alpha H^\alpha}$ RDC because of the poor sensitivity arising from rapid C^α relaxation.

Frequency based methods are sufficiently precise for the determination of the large $^1D_{C^\alpha H^\alpha}$ coupling and require less experimental time than intensity based methods: this means that more experimental time can be devoted to signal accumulation, which is necessary to compensate for the rapid magnetization decay due to C^α relaxation. Similarly to the NH-IPAP experiment, the increase in spectral overlap caused by a doubling of the number of peaks observed in the spectrum can be overcome by using spin-state selective filters [3] to edit the high-field and low-field components of the split signal into two separate subspectra. Further improvement in the spectral resolution can be obtained by introducing a third dimension, either ^{13}C or ^{15}N , although this significantly increases the experimental time and may compromise the digital resolution in the dimension that is used to measure the coupling constants.

Below a brief review of the experiments available for the measurement of $C^\alpha H^\alpha$ coupling constants is given. A simple modification of the (HA)CA(CO)NH experiment [75], whereby the $C^\alpha H^\alpha$ coupling is allowed to evolve during the constant time C^α frequency labelling period, results in a splitting of the signal in this dimension by $^1J_{C^\alpha H^\alpha}$. Evolution of the $C^\alpha C^\beta$ coupling is suppressed by setting the length of the constant time period to 28 ms ($\sim 1/J_{C^\alpha C^\beta}$). The same idea is used in a modified 3D (HA)CANH experiment [120] in which the $C^\alpha H^\alpha$ coupling is allowed to evolve. A 3D (HACA)CONH experiment [44] samples $^1J_{C^\alpha H^\alpha}$ in a variable-time fashion while magnetization is on H^α and the carbonyl chemical shift is labelled concurrently, resulting in a antiphase doublets in the carbonyl dimension and avoiding the problem of $^1J_{C^\alpha C^\beta}$ coupling evolution. In another experiment [118] based on a 3D HNCO, $C^\alpha H^\alpha$ coupling evolution is combined with carbonyl chemical shift evolution during a semi-constant time period [39, 47] which is limited to 12 ms in order to minimize evolution of C^α-C^β couplings. Subspectra containing the high-field and low-field components of the doublets are produced by IPAP editing [77]. The TROSY principle can also be incorporated into experiments to measure $^1J_{C^\alpha H^\alpha}$: the 3D HN(CO)CA-J-TROSY [82] and 3D HNCA based TROSY experiments [80] show the $C^\alpha H^\alpha$ splitting in the ^{13}C

dimension, although they are primarily designed to measure the $^1J_{NH^N}$ and $^3J_{NH^\alpha}$ couplings. Finally, an experiment published very recently [29] uses the magnetization transfer pathway of the HNCO experiment to produce a 2D ^{1}H-^{15}N correlated spectrum in which evolution of the $C^\alpha H^\alpha$ coupling is superimposed upon ^{15}N chemical shift evolution. The problem of $C^\alpha C^\beta$ coupling evolution is avoided by sampling the $C^\alpha H^\alpha$ coupling using transverse H^α magnetization, and the C'C^α coupling is used to separate the two components of the $C^\alpha H^\alpha$ doublet in an IPAP manner.

3.2 ^{15}N HSQC-like schemes for measuring $^1D_{C^\alpha H^\alpha}$

At the outset of this project, no published experiments existed in which the $C^\alpha H^\alpha$ coupling could be measured from 2D ^{1}H-^{15}N HSQC-like spectra - it was therefore decided to investigate this idea. There are two possible polarization transfer pathways that can be used to correlate H^α /C^α and H^N-N magnetization - an 'out-and-back' pathway starting from and ending on H^N, or alternatively, H^α magnetization can be used as a starting point. Both approaches were investigated, and some of the problems encountered were overcome by means of the Double Pulsed Field Gradient Spin Echo (DPFGSE) and BIlinear Rotation Decoupling (BIRD) elements described in sections 2.2.2 and 2.2.3, respectively, and by use of C^α -selective pulses.

The basic out-and-back scheme (Figure 3.1) is based on a 3D HN(CO)CA [38, 50] but instead of labelling C^α magnetization with the C^α offset at point *a*, the $C^\alpha H^\alpha$ coupling is allowed to evolve incrementally in step with the ^{15}N frequency labelling period at point *c*. This superimposes $C^\alpha H^\alpha$ coupling evolution on evolution of the ^{15}N offset, creating the desired 2D H^N-N correlation spectrum with doublets in the ^{15}N dimension corresponding to the $(i-1)$ $C^\alpha H^\alpha$ coupling constant.

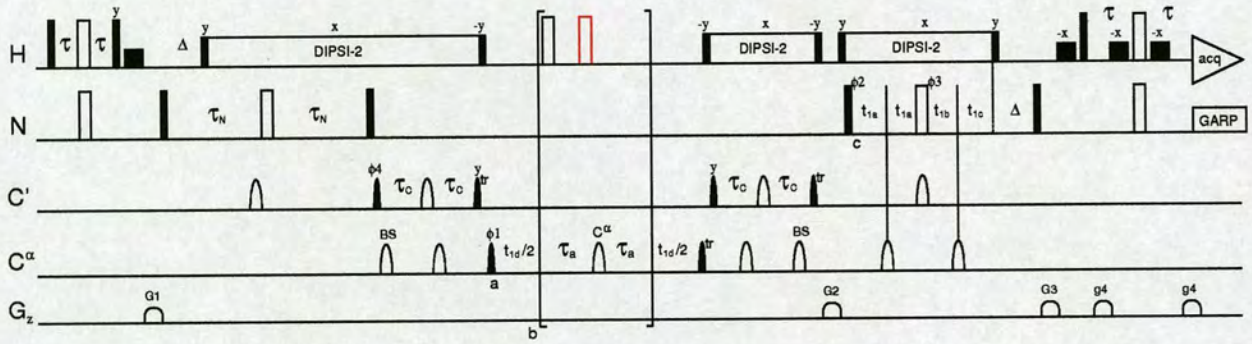

Figure 3.1: **Basic HN(CO)CA-based scheme for measuring $^1D_{C^\alpha H^\alpha}$.** See section 2.2 for a description of the symbols and conventions used in the pulse sequence diagram. Delays are as follows: τ=2.3 ms ($1/4J_{NH}$), τ_a=1.78 ms ($1/4J_{CH}$),τ_C=4 ms ($1/4J_{C^\alpha C'}$), τ_N=12.4 ms ($1/4J_{NC'}$) and Δ=5.3 ms ($1/4J_{NH}$). Gradients are G1=50%, G2=40%, G3=60%, g4=60% and phases are ϕ_1=x,-x; ϕ_3=4(x),4(-x); ϕ_4=8(x),8(-x) rx=2(x,-x,-x,x),2(-x,x,x,-x). For the IP dataset, ϕ_3=2(x),2(-x) and the black 180° proton pulse is applied during the spin echo at point *b*; for the IP dataset, ϕ_3=2(y),2(-y) and the red 180° proton pulse is applied. Quadrature detection is achieved in the ^{15}N dimension by incrementing the phases ϕ_2 and ϕ_3 according to States-TPPI. The constant time ^{15}N frequency labelling periods are of initial length t_{1a}=t_{1b}=6.2 ms ($1/8J_{NC'}$) and t_{1c}=0.7 ms ($1/8J_{NC'}$ - $1/4J_{NH}$), where t_{1b} and t_{1c} are incremented and t_{1a} is decremented.

IPAP editing has been introduced by adding a spin-echo at point *b* in the $C^\alpha H^\alpha$ coupling evolution period. As shown in Figure 3.1, during the spin-echo at point *b* there is evolution of the $C^\alpha H^\alpha$ coupling for a time equal to $1/2J_{C_\alpha H_\alpha}$ in the antiphase experiment - this results in full conversion of the in-phase $C^\alpha H^\alpha$ magnetization to antiphase magnetization provided that the coupling constant matches the value assumed in setting the delay. For the in-phase version of the experiment, a dummy spin-echo [1] is used rather than no spin-echo at all, in an effort to guard against systematic errors.

A major advantage of this scheme over other experiments which utilize C^α magnetization for the sampling of the $^1J_{C^\alpha H^\alpha}$ coupling lies in the suppression of the evolution of the $C^\alpha C^\beta$ couplings during the $^1J_{C^\alpha H^\alpha}$ sampling period, which can be achieved by using a C^α -selective 180° pulse, and removes the limitation on sampling time otherwise imposed by the $C^\alpha C^\beta$ couplings. Suppressing evolution of the $C^\alpha C^\beta$ couplings during the spin-echo also leads to a modest gain in sensitivity over sequences employing a spin-echo with a non-selective ^{13}C pulse (e.g. Permi *et al.* [82]), since the retained component would otherwise be: $\cos(\pi J_{C^\alpha C^\beta}/2J_{C^\alpha H^\alpha}) = 0.92$.

This basic sequence has a number of limitations: (1) since it is impossible to set the length of the spin-echo delay to perfectly match a range of couplings (the range will be large in an aligned sample where $J_{obs} = J + D$), a filter should be added to obtain selection of pure in-phase or antiphase doublets for proper post-acquisitional editing; (2) off resonance effects in the C^α -selective pulse should not manifest as phase errors and (3) the constant time ^{15}N t_1 frequency labelling period should be extended in order to increase digital resolution in the dimension encoding the $^1J_{C^\alpha H^\alpha}$ splitting and to be able to take advantage of the unlimited $^1J_{C^\alpha H^\alpha}$ sampling period allowed by refocussing the $C^\alpha C^\beta$ couplings.

The second scheme (Figure 3.2), in which magnetization is transferred at the start of the sequence from H^α , uses the magnetization transfer pathway of an HA(CACO)NH type experiment. Here the $C^\alpha H^\alpha$ coupling is sampled using H^α magnetization at point *b* and this is superimposed on the ^{15}N offset evolution at point *c* as before, producing a spectrum similar to the out-and-back sequence.

---

[1] A dummy spin-echo element is effectively a 'type 1' spin echo as described in section 2.2.1.

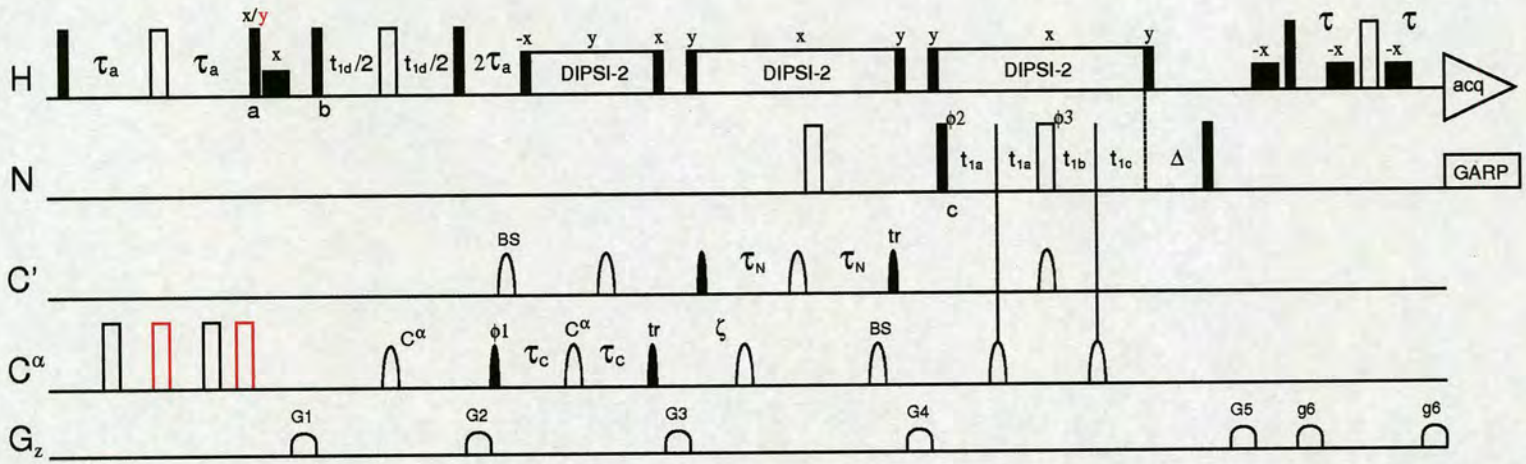

Figure 3.2: **Basic HA(CACO)NH-based scheme for measuring $^1D_{C^\alpha H^\alpha}$** . See section 2.2 for a description of the symbols and conventions used in the pulse sequence diagram. Delays are as follows: τ=2.4 ms ($1/4J_{NH}$), τ_a=1.78 ms ($1/4J_{CH}$),τ_C=4 ms ($1/4J_{C^\alpha C'}$), τ_N=12.4 ms ($1/4J_{NC'}$), δ=2.66 ms ($1/2J_{NH}$), ζ=4.5 ms ($1/4J_{C^\alpha C'}$). Gradients are G1=30%, G2=30%, G3=40%, G4=20%, G5=50%, g6=70%, and phases are ϕ_1=x,-x; ϕ_3=4(x),4(-x); rx=x,-x,-x,x,-x,x,x,-x. For the IP dataset, ϕ_2=2(x),2(-x) and for the AP dataset ϕ_2=2(y),2(-y). Quadrature detection in the ^{15}N dimension is achieved by incrementing the phases ϕ_2 and ϕ_3 according to States-TPPI. IPAP editing is achieved by alternating between the two pairs of ^{1}H 180° pulses during the first spin-echo preceding point *a*: the two black pulses are applied in the IP experiment and the two red pulses are applied in the AP experiment. In addition, the phase of the ^{1}H 90° pulse at *a* is 'x' for the IP experiment and the water flip-back pulse is applied off-resonance; for the AP experiment, the ^{1}H 90° pulse has phase 'y'. The constant time ^{15}N frequency labelling periods are of initial length t_{1a}=t_{1b}=6.2 ms ($1/8J_{NC'}$) and t_{1c}=0.7 ms ($1/8J_{NC'}$ - $1/4J_{NH}$), where t_{1b} and t_{1c} are incremented and t_{1a} is decremented.

IPAP editing has been designed into this sequence by modifying the spin-echo prior to point *a*: antiphase magnetization is created by allowing the $C^\alpha H^\alpha$ coupling to evolve for a length of time equal to $1/2J_{C_\alpha H_\alpha}$ whereas the in-phase experiment uses a dummy spin-echo. At point *a*, the desired in-phase or antiphase component is transferred to H_z / $2H_zC_z$ by a 90° proton pulse and a gradient is used to purge the unwanted term before the magnetization is returned to the transverse plane by a second 90° proton pulse at point *b*. The advantage of this sequence is that the problem of $C^\alpha C^\beta$ coupling evolution is bypassed by sampling the $C^\alpha H^\alpha$ coupling using H^α magnetization. Again, there are a number of limitations that must be overcome: (1) evolution of H^α -H^{other} couplings should be prevented if possible (these couplings become significant when dipolar contributions are present); (2) again, the constant time ^{15}N t_1 frequency labelling period should be extended in order to increase digital resolution in the dimension encoding the $^1J_{C^\alpha H^\alpha}$ splitting and to be able to take advantage of the unlimited $^1J_{C^\alpha H^\alpha}$ sampling period.

A third scheme starting with H^α magnetization was abandoned earlier in development due to poor sensitivity: this scheme involved transfer to C^α , followed by a 28 ms constant time period (to refocus $C^\alpha C^\beta$ couplings) during which the $^1J_{C^\alpha H^\alpha}$ splitting was sampled in step with the ^{15}N t_1 frequency labelling period. C^α relaxation during the 28 ms constant time period was found to result in very poor signal-to-noise ratios.

3.3 Selective inversion of C^α magnetization

In order to implement the ideas outlined above, it is necessary to invert C^α spins selectively without any associated inversion of C^β . The length of the inversion pulse should therefore be calibrated such that, as far as possible, C^α chemical shifts lie within the inverted region while C^β chemical shifts do not. Figure 3.3 shows the distribution of C^α and C^β chemical shifts found in proteins contained within the BMRB database [1], from which it can be seen that inversion of ^{13}C spins in the range 50-68 ppm and excluding < 45 ppm will best satisfy the requirements outlined above.

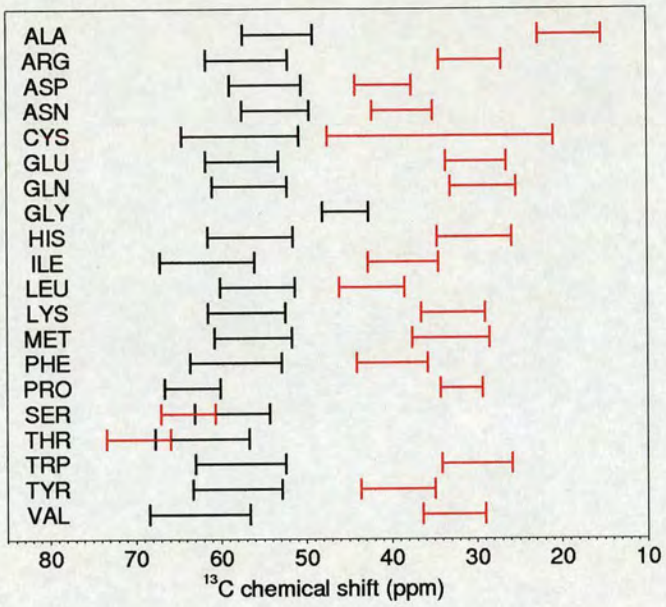

Figure 3.3: **Distribution of chemical shifts for the C^α and C^β resonances of proteins.** Chemical shift ranges for C^α are shown in black, and for C^β in red: the ranges shown correspond to values within two standard deviations of the mean.

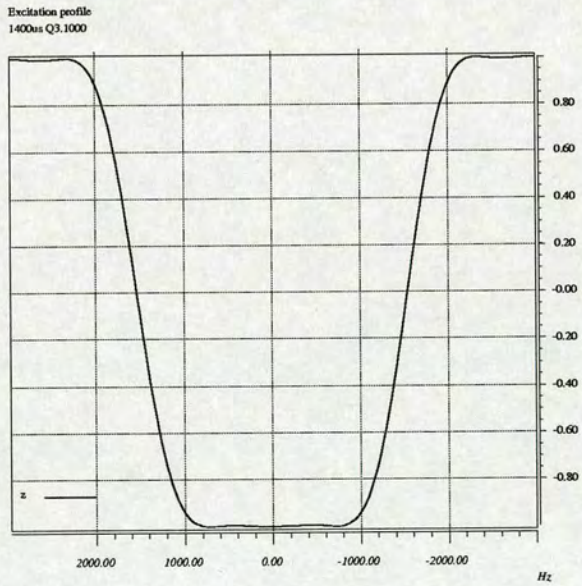

Figure 3.4: **Inversion profile of a 1400 μs Q3 pulse.** > 80% inversion of magnetization within approximately ± 1200 Hz of the offset: at a ^{1}H frequency of 600 MHz and a carbon offset of 58 ppm, this corresponds to the region from 50 ppm to 68 ppm (< 10% inversion outside the region from 45 ppm to 71 ppm).

A Q3 Gaussian cascade [32] was chosen for the purpose of selective inversion of C^α spins, calibrating the length such that inversion is effectively complete within the range 50-68 ppm, and does not occur to any significant extent below 45ppm. The inversion profile for a 1400 μs pulse, which achieves this aim, is shown in Figure 3.4. Unfortunately, serine, and to a lesser extent, threonine C^β spins may fall within the excitation range of this pulse, allowing evolution of the $C^\alpha C^\beta$ coupling for these residues. Incomplete inversion of glycine C^α spins is less of a concern since meaningful couplings cannot be measured for these residues anyway [2].

3.4 Testing DPFGSE and BIRD pulse sequence elements

Preliminary experiments using a simple spin-echo containing a C^α -selective pulse did not always produce pure phase doublets, particularly when the peaks were far from the C^α offset. Pulsed field gradient spin echoes were seen as a way to overcome the problems of incomplete inversion and phase anomalies caused by the 180° C^α pulses: instead of phase distortions, only intensity losses should result from any off-resonance effects. A complete investigation was conducted into the suitability of single and double pulsed field gradient spin-echoes (SPFGSE and DPFGSE, section 2.2.2) for use in ^{13}C -sampled experiments. The BIRD elements (section 2.2.3) used in the H^α -sampled experiments were also investigated. In order to simplify the problem of analyzing the $C^\alpha H^\alpha$ doublets produced by protein samples (peaks are often broad and overlapped due to fast relaxation), C-1 ^{13}C -labelled glucose was chosen as a test molecule and a series of simple pulse sequences were written. The soft 180° Q3 carbon pulses (length 1400 μs) are marked 'C1', signifying the fact that they are centered at the chemical shift of the C-1 carbon. In all cases, the doublet shown is that of the α-anomeric proton at C1

[2] The observed $C^\alpha H^\alpha$ splitting, when measured with experiments that use $^{13}C^\alpha$ magnetization to sample the $^1J_{C^\alpha H^\alpha}$ splitting, is the sum of two coupling constants - one for each H^α proton.

SPFGSE and DPFGSE elements

The test pulse sequences used to invesigate the SPFGSE and DPFGSE elements are shown in Figures 3.5 and 3.6. These experiments start with polarization transfer from ^{13}C and have the J-coupling evolution period separated from the chemical shift evolution period in order to simulate the situation in the equivalent protein experiments. They produce 2D spectra correlating the ^1H shift in the detected dimension with ^{13}C in the indirect dimension. The resonances are split in both dimensions by the $C^\alpha H^\alpha$ coupling - collecting antiphase magnetization removes the need for an extra spin-echo at the end of the sequences, but means that decoupling the detected magnetization is not possible. Long acquisition times appropriate for sampling the signal of small molecules during the directly detected period were used.

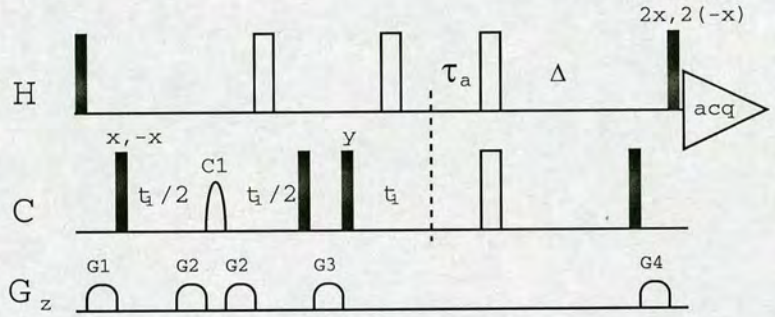

Figure 3.5: **Pulse sequence for testing the SPFGSE element.** $\tau_a = 1/4J_{CH}$ $\Delta = \tau_{180}^{1H} + t_1(0)$ and the receiver phase cycle, rx= $x, -x, -x, x$.

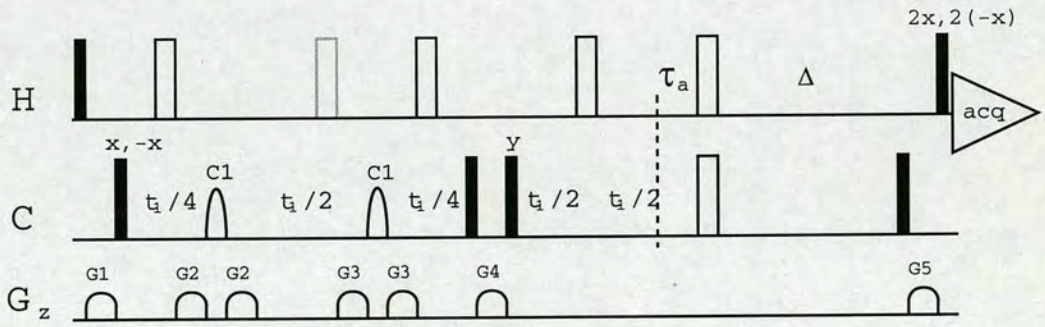

Figure 3.6: **Pulse sequence for testing the DPFGSE element.** See Figure 3.5 - this sequence differs only in that the SPFGSE element has been replaced with a DPFGSE. The position of the second proton pulse in the asymmetric arrangement is shown in grey.

Following an initial proton 90° pulse, unwanted proton magnetization is purged by G1. Transverse ^{13}C magnetization is created by a 90° ^{13}C pulse and the $^1J_{CH}$ coupling is allowed to evolve during the PFGSE element for a period equal to t_1, after which the cosine-modulated in-phase component $C_y\cos(\pi J_{CH} t_1)$ is selected and the antiphase component is purged by G3. ^{13}C frequency labelling proceeds for a period t_1 and the final spin-echo of length $1/2J_{CH}$ creates antiphase magnetization that is transferred to proton for detection. The resulting doublets are in-phase in the ^{13}C dimension, and antiphase in the ^1H dimension.

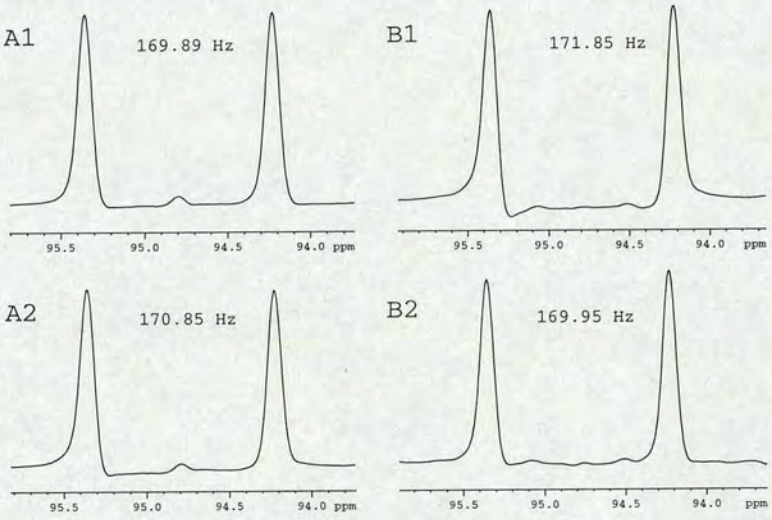

Figure 3.7: **Properties of the SPFGSE (A1-2) and DPFGSE (B1-2) pulse sequence elements.** A1) SPFGSE element with 180° ^{13}C hard pulse, A2) SPFGSE with 180° ^{13}C Q3 soft pulse, B1) unsymmetrical DPFGSE element, B2) symmetrical DPFGSE.

The traces shown in Figures 3.7 and 3.8 compare the doublets produced by these elements. A number of conclusions can be drawn from Figure 3.7: firstly, the SPFGSE element clearly exhibits phase distortions which are particularly pronounced when a soft ^{13}C 180° pulse is used (A2). Secondly, the DPFGSE only produces doublets with phase properties comparable to those of a hard 180° ^{13}C pulse if the 180° proton pulses are symmetrically located about the carbon 180° pulse [3]: indeed, the phase properties are worse for an unsymmetrical DPFGSE than for a SPFGSE (B1).

[3] The position of the second proton pulse in the asymmetric arrangement is shown in Figure 3.6 in grey.

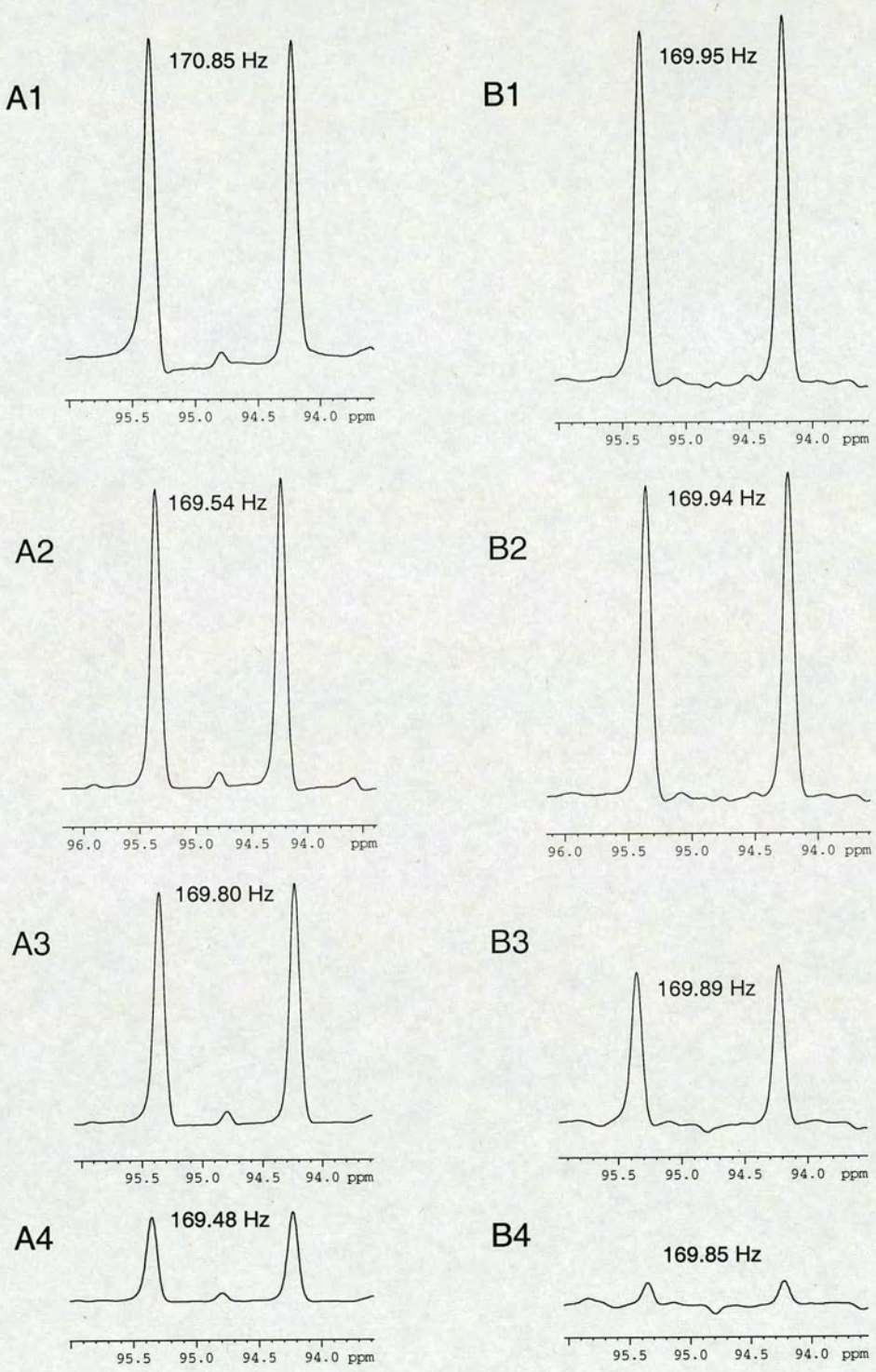

Figure 3.8: **Comparison of SPFGSE (A1-4) and DPFGSE (B1-4) elements with changing ^{13}C 180° Q3 pulse offset.** 1) pulse on-resonance, 2) pulse offset = -2100 Hz, 3) offset = -2550 Hz, 4) offset = -3000 Hz.

Figure 3.8 compares the two elements with varying C^α 180° pulse offset, and again it is apparent that the DPFGSE gives more consistent values for J_{CH}. This also illustrates one disadvantage of the DPFGSE, showing that the intensity loss due to incomplete inversion is greater: it is the square of the intensity loss for the equivalent SPFGSE.

$BIRD_{d,x}$ and $BIRD_r$ elements

Two BIRD elements were also tested - the $BIRD_{d,x}$ element mentioned above, and a $BIRD_r$ element that will be used in a 3D sequence (section 3.6.2). The simplified pulse sequences incorporating $BIRD_{d,x}$ and $BIRD_r$ elements are shown in Figures 3.9 and 3.10, respectively. Note that in the experiment designed for testing the $BIRD_r$ element, both dimensions have proton frequencies. Tests of the $BIRD_{d,x}$ and $BIRD_r$ elements are summarised in Figures 3.11 and 3.12.

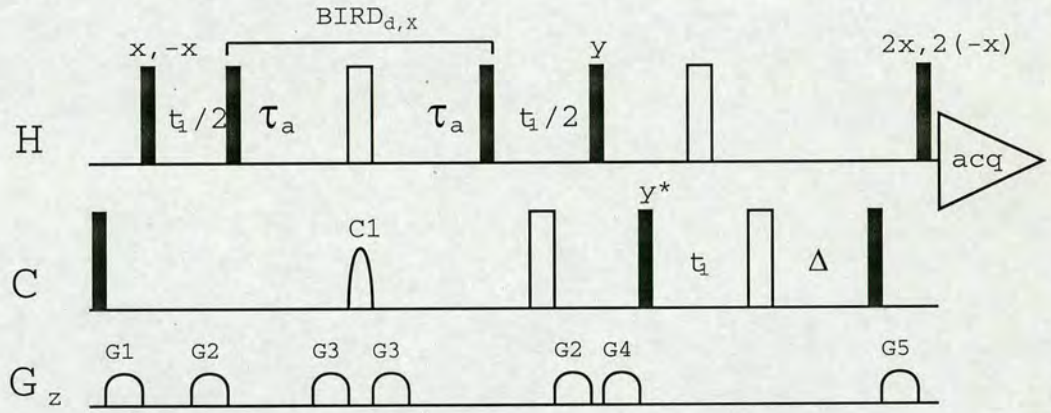

Figure 3.9: **Pulse sequence for testing the $BIRD_{d,x}$ element.** $\tau_a = 1/2J_{CH}$, $\Delta = \tau_{180}^{1H} + t_1(0)$ and the receiver phase cycle, rx= $x, -x, -x, x$.

In the $BIRD_{d,x}$ test sequence, following a purge of ^{13}C magnetization by G1, transverse proton magnetization is created by the first proton 90° pulse. The $BIRD_{d,x}$ element inverts ^{13}C and directly attached protons, resulting in evolution of the $^{13}C^1H$ coupling for a period of t_1 since the BIRD element is surrounded by two periods equal to $t_1/2$. There is no net evolution of the $^{13}C^1H$ coupling during the BIRD element but evolution of the coupling during the surrounding G2 gradients must be refocussed, which is achieved by means of a non-selective 180° ^{13}C pulse immdeiately prior to the second

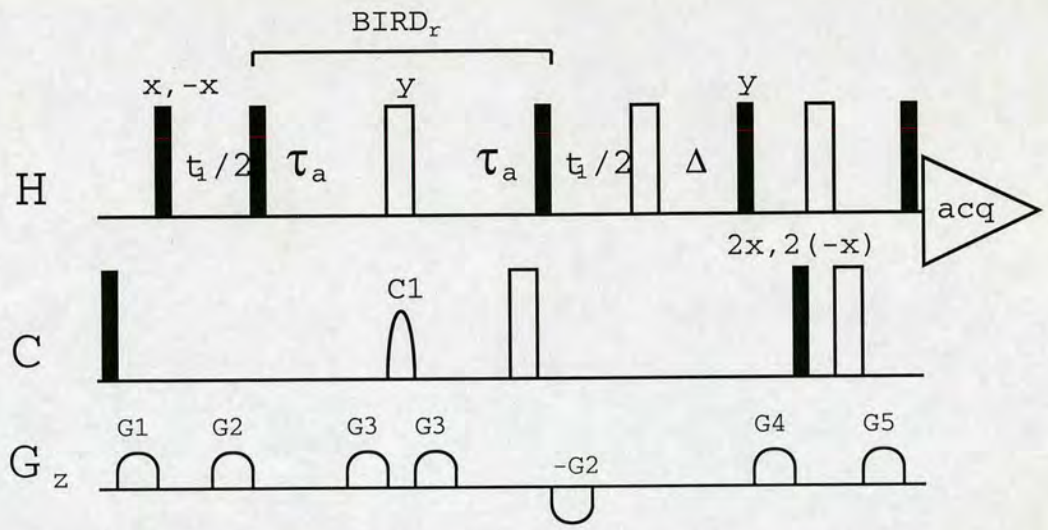

Figure 3.10: **Pulse sequence for testing the BIRD$_r$ element.** $\tau_a = 1/2J_{CH}$, $\Delta = 2\tau_{G2}$ and the receiver phase cycle, rx= $x, -x, -x, x$.

G2. The antiphase $2H_xC_z\sin(\pi J_{CH}t_1)$ term is transferred to carbon $2H_zC_x$ by the 90° carbon pulse marked with phase $(y*)$, and any in-phase magnetization is purged by G4. Carbon frequency labelling then takes place (the proton 180° prevents further $^{13}C^1H$ evolution) and the carbon 180° pulse and period Δ refocus evolution of the carbon offset that would otherwise occur during the proton 180° pulse and initial t_1 delay. Finally, the antiphase magnetization is transferred back to proton for observation. The resulting doublets are antiphase in both the ^{13}C and ^1H dimensions.

In the BIRD$_r$ test sequence, following a purge of ^{13}C magnetization by G1, transverse proton magnetization is created by the first proton 90° pulse. The BIRD$_r$ element situated in the centre of the first t_1 period inverts protons that are not attached to carbon atoms affected by the selective 180° pulse located within the BIRD element (i.e. remote protons): the result is that the $^{13}C^1H$ coupling and offset evolves for C-1. Note that G2 gradients of opposite sign are situated either side of the BIRD element, which ensures that non-inverted antiphase $2H_zC_x$ magnetization is refocussed by the second gradient. The period Δ and the 180° pulse immediately before it prevents evolution of the coupling or offset during the two G2 gradients. Antiphase magnetization is transferred to $2H_zC_z$ by the 90° proton pulse following Δ, and then to $2H_zC_y$ and back (normally further transfers would follow in the case of a protein) before observation of

antiphase $2H_yC_z$ resulting in antiphase doublets in both proton dimensions.

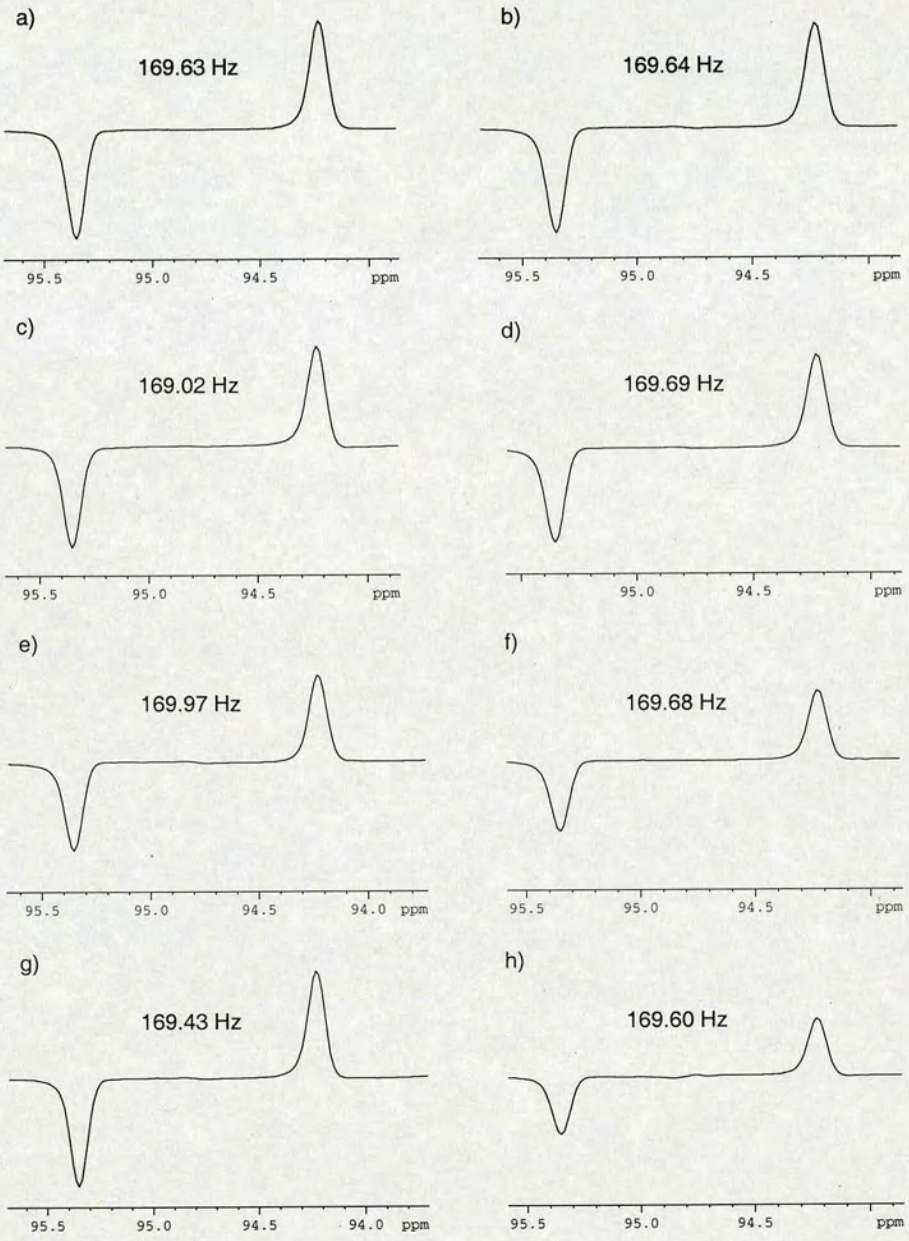

Figure 3.11: **Properties of the BIRD$_{d,x}$ pulse sequence element.** a) BIRD$_{d,x}$ with hard ^{13}C 180° pulse, b) 1.4 ms Q3 ^{13}C 180° pulse, c) no gradients within BIRD$_{d,x}$ element, d), f) and h) Q3 pulse offset of -2100 Hz, -2550 Hz and 3000 Hz, respectively, e) and g) spin-echo delays set assuming J=135 Hz and J=195 Hz, respectively.

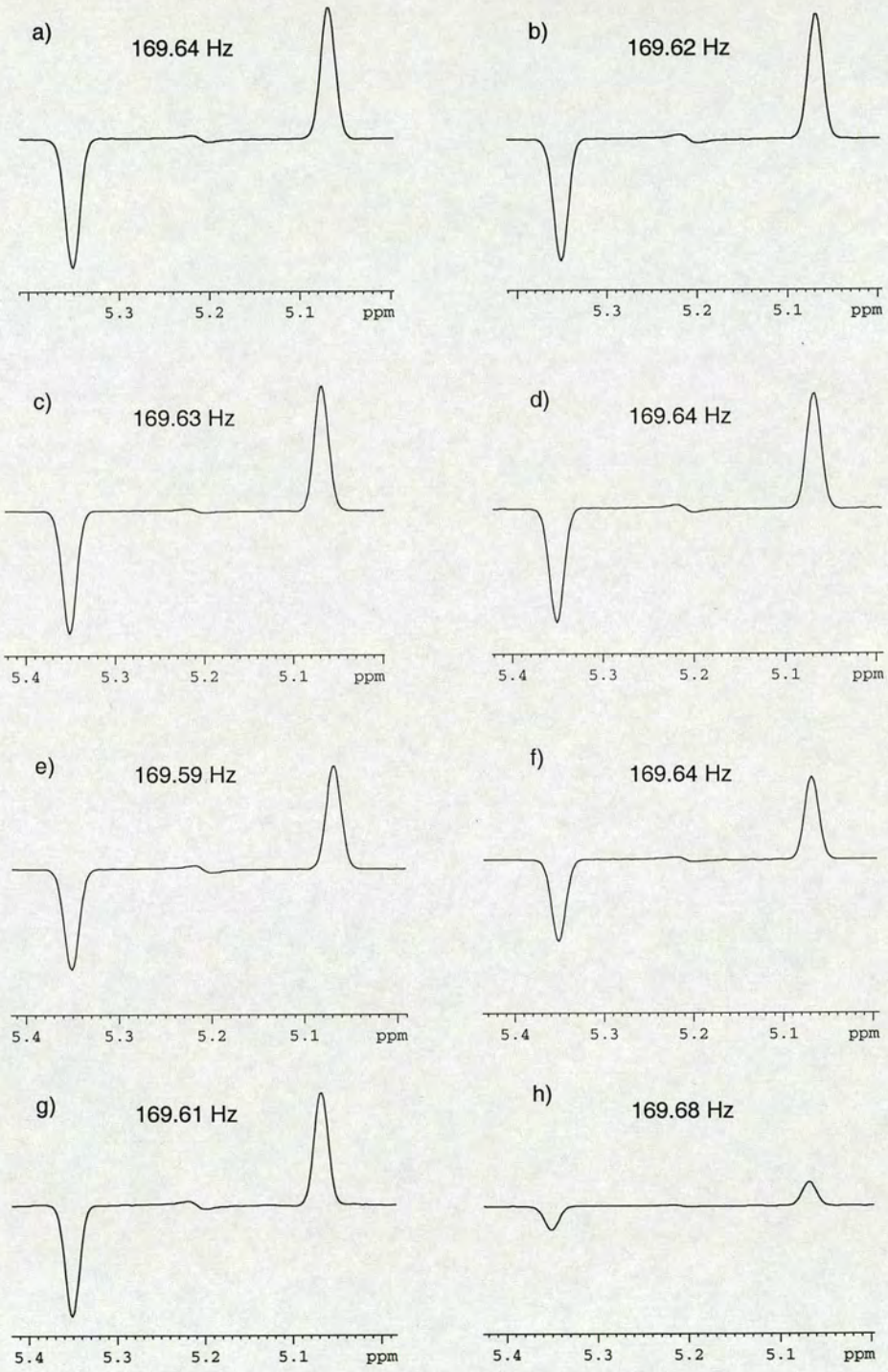

Figure 3.12: **Properties of the BIRD$_r$ pulse sequence element.** a) BIRD$_r$ with hard ^{13}C 180° pulse, b) BIRD$_r$ with 1.4 ms Q3 ^{13}C 180° pulse, c) no gradients within BIRD$_r$ element, d), f) and h) Q3 pulse offset of -2100 Hz, -2550 Hz and 3000 Hz, respectively, e) and g) spin-echo delays set assuming J=135 Hz and J=195 Hz, respectively.

Both BIRD elements give doublets that have good phase properties in spite of any delay/J_{CH} mismatch or off-resonance effects of the C^α selective pulse. It is also apparent that the gradients within the $BIRD_{d,x}$ elements improve the phase properties, since, as seen in part c) of Figure 3.11, setting the strengths of the gradients to zero affects the phases of the signals and therefore the measured splittings. Variation of the measured splitting in the $BIRD_{d,x}$ experiment of < 0.3 Hz was seen when the BIRD period was set according to a $^1J_{CH}$ values ± 30 Hz from the true value for glucose (Figure 3.11 e & g) but no such discrepancy was observed for the $BIRD_r$ test experiments.

In summary, the results obtained using these test sequences for both DPFGSE and BIRD elements with selective ^{13}C pulses show that in principal, they should be sufficiently robust for use in experiments to measure inherently heterogeneous $^1D_{C^\alpha H^\alpha}$ coupling constants in proteins: the values of coupling constants obtained by using the ^1H and ^{13}C -sampled BIRD methods are in excellent agreement (within $\sim 0.01\%$) and are also in accordance with the value obtained using a DPFGSE element (differing by ~ 0.3 Hz or $< 0.2\%$). The differences between the results of the four methods raise the possibility of small systematic errors. However, the dipolar coupling D is determined from the difference $(J + D) - J$ and *if the same method is used to determine* J *and* $(J + D)$, such systematic errors are mitigated to some extent.

3.5 Sampling the $C^\alpha H^\alpha$ splitting via C^α magnetization

The final version of the 2D 'out-and-back' scheme, which has been named 2D (HNCO)-(J-CA)NH, is described here. It incorporates a number of key features including symmetrical DPFGSE elements, filters to purge the unwanted IP or AP component, a modified semi-constant time period in which ^{15}N frequency labelling takes place and lastly, scaling of the $C^\alpha H^\alpha$ coupling constant in order to reduce the time that rapidly-decaying C^α magnetization exists. This 2D experiment can also been modified to incorporate a third dimension, giving the 3D (HN)CO-(J-CA)NH experiment which has the reduced sensitivity associated with adding an extra dimension, but may be of use where there are spectral overlap issues.

3.5.1 2D (HNCO)-(J-CA)NH

The first part of this pulse sequence (Figure 3.13) resembles an HN(CO)CA experiment and transfers the magnetization from H^N protons to C^α carbons via C' carbons. The point at which the (HNCO)-(J-CA)NH sequence diverges from an HN(CO)CA sequence is the point at which a rectangular 90° C^α -selective pulse [4] with phase cycle ϕ_2 transfers the magnetization to $^{13}C^\alpha$. This is the beginning of a spin-echo of duration $1/2J_{C^\alpha H^\alpha}$, during which a pair of 180° proton pulses may be applied either; each in the middle of a τ_a period (shown in black) or one immediately prior to the ^{13}C 180° pulse and the other at the end of the spin-echo (shown in red). The former arrangement results in in-phase $C^\alpha H^\alpha$ magnetization at the end of the spin-echo, while the latter results in antiphase $C^\alpha H^\alpha$ magnetization. The dummy spin-echo which produces in-phase magnetization is not strictly necessary, but has been included in order to eliminate the differences in the intensities of the signals that would otherwise exist due to relaxation.

The relevant spin operators describing the coherences after the 90° C^α pulse with phase cycle ϕ_1 are $4C^\alpha_y C'_z N_z$ for the IP experiment and $8C^\alpha_x H^\alpha_z C'_z N_z$ for the AP experiment. The gradient g3 is used to purge any remaining transverse magnetization, which ensures that the unwanted AP or IP component is destroyed. Pure in-phase or antiphase magnetization is returned to the transverse plane by means of a 90° C^α pulse.

Having created either in-phase or antiphase $C^\alpha H^\alpha$ magnetization, the next block in the pulse sequence (beginning at the C^α pulse with phase cycle ϕ_1) is a DPFGSE (symmetrical) during which the $C^\alpha H^\alpha$ coupling is sampled for a total time of t_{1d}. Short gradients (0.5 ms) are used to further minimize the time that C^α magnetization exists.

---

[4] Pulse length set to $\sqrt{15}/4\Delta\delta$, where $\Delta\delta$ is the difference in chemical shift between C^α and C', giving a null in the excitation profile at the C' frequency.

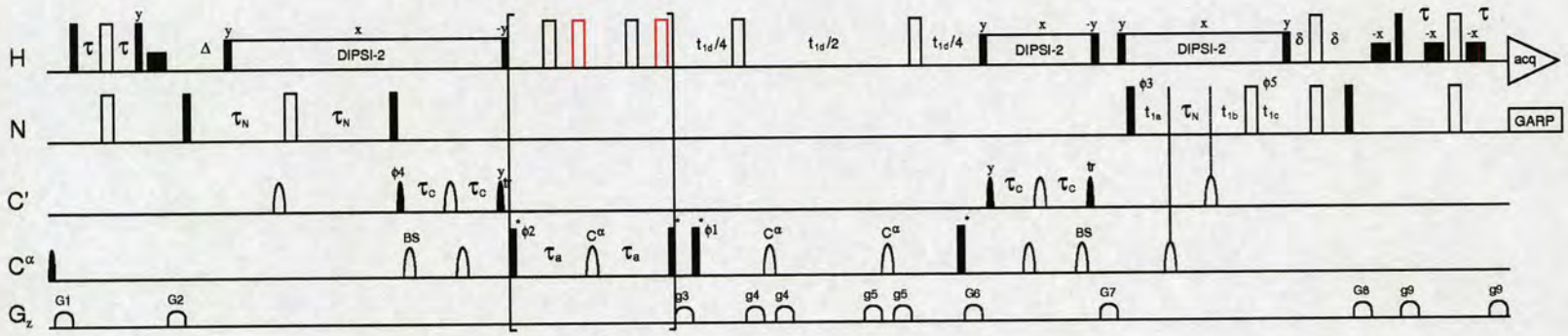

Figure 3.13: **2D (HNCO)-(J-CA)NH pulse sequence for measuring $^1D_{C^\alpha H^\alpha}$** . See section 2.2 for a description of the symbols and conventions used in the pulse sequence diagram. Delays are as follows: τ=2.3 ms ($1/4J_{NH}$), τ_a=1.78 ms ($1/4J_{CH}$),τ_C=4 ms ($1/4J_{C^\alpha C'}$), τ_N=12.4 ms ($1/4J_{NC'}$), Δ=5.3 ms ($1/2J_{NH}$), δ=2.66 ms ($1/4J_{NH}$). Gradients are G1=15%, G2=50%, g3=30%, g4=12%, g5=17%, G6=8%, G7=40%, G8=60%, g9=60% and phases are ϕ_4=8(x),8(-x); ϕ_5=16(x),16(-x); rx=x,-x,-x,x,-x,x,x,-x,-x,x,x,-x,x,-x,-x,x. For the IP dataset, ϕ_1=x,-x; ϕ_2=2(x),2(-x); ϕ_3=4(x),4(-x) and for the AP dataset ϕ_1=y,-y; ϕ_2=2(y),2(-y); ϕ_3=4(y),4(-y). Quadrature detection in the ^{15}N dimension is achieved by incrementing the phases ϕ_3 and ϕ_5 according to States-TPPI. C^α 90° pulses marked with an asterisk are set to a length of $\sqrt{15}/4\Delta\delta$, where $\Delta\delta$ is the chemical shift difference between C^α (58 ppm) and C' (176 ppm), producing a 90° C^α pulse with a null at C'. The IPAP element in brackets contains four ^{1}H 180° pulses, of which the two black pulses are applied in the IP experiment and the two red pulses are applied in the AP experiment. Modifications to the 2D (HNCO)-(J-CA)NH sequence give the 3D (HN)CO-(J-CA)NH alternative - the IPAP element is removed and phase cycles change as follows: ϕ_1=x,-x; ϕ_3=4(x),4(-x); ϕ_4=2(x),2(-x); ϕ_5=8(x),8(-x). In addition, C' constant time chemical shift labelling is implemented using the τ_C delays immediately before the (deleted) IPAP element: the first τ_C is changed to $\tau_C - t_2/2$ and the second becomes $\tau_C + t_2/2$.

Scaling of the $C^\alpha H^\alpha$ coupling constant is achieved by altering the incrementation of t_{1d}, where the coupling is sampled, relative to that of t_{1a-c} which are the semi-constant time delays used to label the ^{15}N frequency. The increment for t_{1d} is reduced by the scaling factor κ relative to that for t_{1a-c}. Setting $\kappa = 1.5$ was found to be a good compromise; scaling down the coupling constant to 0.667× its actual value whilst reducing relaxation losses by decreasing the amount of time that C^α magnetization exists. This magnitude of scaling down of the $^1J_{C^\alpha H^\alpha}$ coupling constants, which are $\sim$ 145 Hz, still gives an acceptably large measured coupling constant of around 95 Hz.

Semi-constant time labelling of the ^{15}N chemical shift [39, 47] is used in this experiment to enable greater digital resolution in the ^{15}N dimension by extending the maximum effective ^{15}N acquisition time beyond the normal limit of $1/2J_{NC'}$ imposed by constant time ^{15}N frequency labelling. It is common practice to use semi-constant time chemical shift labelling for the H^α chemical shift [47], but in applying this method to ^{15}N in a ^{13}C,^{15}N -labelled protein, evolution of both $^1J_{NC'}$ and $^1J_{NC^\alpha}$ couplings needs to be taken into account. Following the 90° ^{15}N pulse with phase ϕ_3, $2N_{x/y}C'$ magnetization exists: in addition to achieving ^{15}N frequency labelling, evolution of the N-C^α coupling must be suppressed but the N-C' coupling must be allowed to evolve for $1/2J_{NC'}$ in order to convert antiphase $2N_{x/y}C'$ to in-phase $N_{x/y}$ magnetization. The scheme shown in Figure 3.13 accomplishes these aims by including the period 'τ_N' (set to a value corresponding to $1/4J_{NC'}$: 12.4 ms is optimal due to ^{15}N relaxation) between t_{1a} and t_{1b}. The lengths of the periods t_{1a}, t_{1b} and t_{1c} for the ith increment are as follows (equations 3.1, 3.2 and 3.3):

$$t_{1a} = \frac{AT}{2} \times \frac{i}{n-1} \tag{3.1}$$

$$t_{1b} = \frac{AT - \delta}{2} \times \frac{i}{n-1} \tag{3.2}$$

$$t_{1c} = \frac{\delta}{2} - \left(\frac{\delta}{2} \times \frac{i}{n-1}\right) \tag{3.3}$$

where $\delta/2$ is set according to $1/4J_{NC'}$, AT is the acquisition time and n the total number of increments. It can be seen from equations 3.1-3.3 that the acquisition time cannot be shorter than δ and the initial length of the delay t_{1c} is $\delta/2$ whereas t_{1a} and t_{1b} are initially 0 ms. For the first t_1 increment therefore, the 180° nitrogen and carbonyl carbon pulses are centrally positioned, allowing the N-C' coupling to evolve for 24 ms but refocussing the ^{15}N offset (i.e. $t_1 = 0$). To achieve t_1 incrementation, t_{1a} and t_{1b} are incremented while t_{1c} is decremented, offsetting the ^{15}N 180° pulse from the centre of the period and refocussing less of the ^{15}N offset. The N-C^α coupling is refocussed during a period of $2t_{1a}$ by the 180° C^α carbon pulse, and for the remainder of the period by the ^{15}N 180° pulse, while the N-C' coupling is free to evolve for a period of $\tau_N + t_{1a} + t_{1c} - t_{1b} = \delta$.

3.5.2 3D (HN)CO-(J-CA)NH

Constant time carbonyl chemical shift labelling can be introduced into the $2\tau_c$ period follwing the carbonyl 90° pulse with phase cycle ϕ_4 in Figure 3.13. It was found that a short carbonyl chemical shift labelling period (t_2) was sufficient to resolve most overlaps observed for the 2D (HNCO)-(J-CA)NH spectra of proteins under investigation, and that IPAP editing was no longer necessary to resolve the majority of peaks: the coupling was able to be measured from the splitting of the antiphase doublets in the ^{15}N dimension. Removing the IPAP element offsets some of the loss of sensitivity consequent upon adding a third dimension.

3.6 Sampling the $C^\alpha H^\alpha$ splitting via H^α magnetization

The alternative scheme presented in 3.2 is expanded upon here and the resulting 2D (J-HACACO)NH and 3D J-HA(CACO)NH experiments employ many of the elements used in the out-and-back experiments: modified semi-constant time ^{15}N frequency labelling, IP / AP selection and scaling of the $C^\alpha H^\alpha$ coupling. The key difference is that these experiments sample the $C^\alpha H^\alpha$ splitting via H^α magnetization.

3.6.1 2D (J-HACACO)NH

The 2D (J-HACACO)NH pulse sequence is shown in Figure 3.15. It is based on the magnetization transfer sheme of an HA(CACO)NH type experiment (a CBCA(CO)NH experiment [39, 72] was in fact used as a starting point) in which the H^α frequency labelling has been replaced with a period during which the $C^\alpha H^\alpha$ coupling is sampled at the same time as the ^{15}N frequency-labelling period is inceremented.

In order to sample the $C^\alpha H^\alpha$ coupling over the t_{1d} period in Figure 3.15, it is necessary to invert H^α magnetization (to refocus the offset) and also to invert the C^α magnetization. Inverting H^α magnetization selectively without inverting H^{other} should ensure that any H^α-H^{other} couplings are refocussed, which will be particularly beneficial in the case of aligned samples where proton-proton dipolar couplings contribute to loss of magnetization. Achieving H^α -selective inversion using H^α -selective pulses is not practical due to the overlap between H^α and H^β chemical shifts, as illustrated in Figure 3.14.

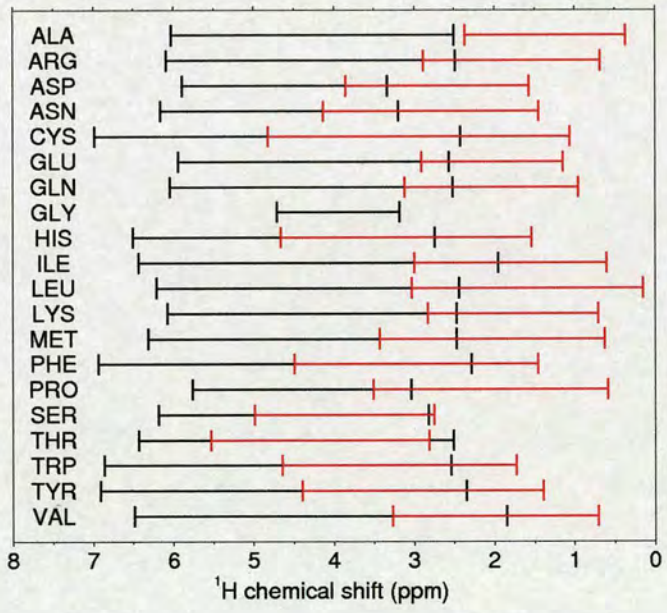

Figure 3.14: **Distribution of chemical shifts for H^α and H^β resonances.** Chemical shift ranges for H^α are shown in black, and for H^β in red: the ranges shown correspond to values within two standard deviations of the mean.

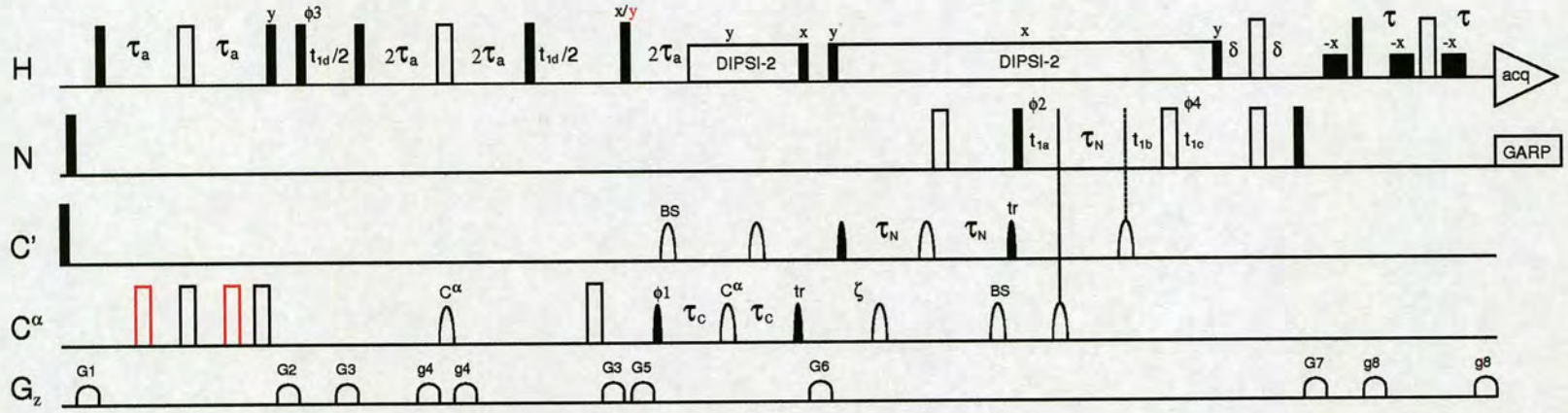

Figure 3.15: **2D (J-HACACO)NH pulse sequence for measuring** $^1D_{C^\alpha H^\alpha}$. See section 2.2 for a description of the symbols and conventions used in the pulse sequence diagram. Delays are as follows: τ=2.4 ms ($1/4J_{NH}$), τ_a=1.78 ms ($1/4J_{CH}$),τ_C=4 ms ($1/4J_{C^\alpha C'}$), τ_N=12.4 ms ($1/4J_{NC'}$), δ=2.66 ms ($1/2J_{NH}$), ζ=4.5 ms ($1/4J_{C^\alpha C'}$). Gradients are G1=8%, G2=30%, G3=12%, g4=17%, G5=32%, G6=40%, G7=50%, g8=70% and phases are ϕ_1=x,-x; ϕ_3=4(x),4(-x); ϕ_4=8(x),8(-x); rx=x,-x,-x,x,-x,x,x,-x. For the IP dataset, ϕ_2=2(x),2(-x) and for the AP dataset ϕ_2=2(y),2(-y). Quadrature detection in the ^{15}N dimension is achieved by incrementing the phases ϕ_2 and ϕ_4 according to States-TPPI. IPAP editing is achieved by alternating between the two pairs of ^{1}H 180° pulses during the first spin-echo of length τ_a-τ_a: the two black pulses are applied in the IP experiment and the two red pulses are applied in the AP experiment. In addition, the phase of the ^{1}H 90° pulse following the second G3 is 'x' for the IP experiment and 'y' for the AP experiment.

An alternative to using H^α -selective pulses, which has been implemented in this experiment, is to use a $BIRD_{d,x}$ element incorporating a C^α -selective pulse: this inverts only C^α -attached protons and C^α [5]. The C^α -selective pulse is surrounded by a pair of short gradients of the same sign (g4) to select only inverted magnetization, which was found to be necessary in the preliminary experiments to maintain the pure phase of the signals. A second pair of same sign gradients (G3) have been placed around the whole $BIRD_{d,x}$ element for the same reason, and evolution of the $C^\alpha H^\alpha$ coupling during the latter gradients is refocussed by means of a non-selective ^{13}C 180° pulse applied at the end of the second $t_{1d}/2$ period. Incrementation of t_{1d} is the same as that for 2D (HNCO)-(J-CA)NH in section 3.5.1: i.e. scaled down by a factor κ relative to the t_{1a-c} period.

To aid in identification of signals in overlapped regions of the ^{15}N HSQC-like spectra which this experiment produces, an IPAP editing scheme has been incorporated for interleaved acquisition of in-phase and antiphase doublets. The IPAP scheme consists of a spin-echo prior to the $BIRD_{d,x}$ element, in which a pair of 180° proton pulses can be applied in either of two configurations: if one of the pulses is applied in a central position at the same time as the 180° ^{13}C pulse and the other at the end of the spin-echo, then antiphase magnetization will be evolved; if, on the other hand, the two 180° proton pulses are applied in the middle of the τ_a periods, then evolution of the coupling over $2\tau_a$ will be refocussed resulting in in-phase magnetization.

3.6.2 3D J-HA(CACO)NH

The 3D J-HACACONH experiment is based on the same CBCA(CO)NH experiment [39, 72] as was used for the 2D (J-HACACO)NH sequence. From the point where magnetization is transferred from C^α to C' onwards the two experiments are basically identical, so only the first part of the 3D sequence (illustrated in Figure 3.16) will be described. Since this is a 3D experiment it is not necessary to create in-phase and antiphase magnetization before sampling the $C^\alpha H^\alpha$ coupling, because the dispersion of the signals into a third dimension removes most of the overlap associated with a

[5] Since no other ^{13}C resonances are inverted this results in non-inversion of their attached protons - i.e. these protons behave as if they were attached to ^{12}C atoms.

doubling of the peaks. The initial IPAP element of the 2D experiment has therefore been removed.

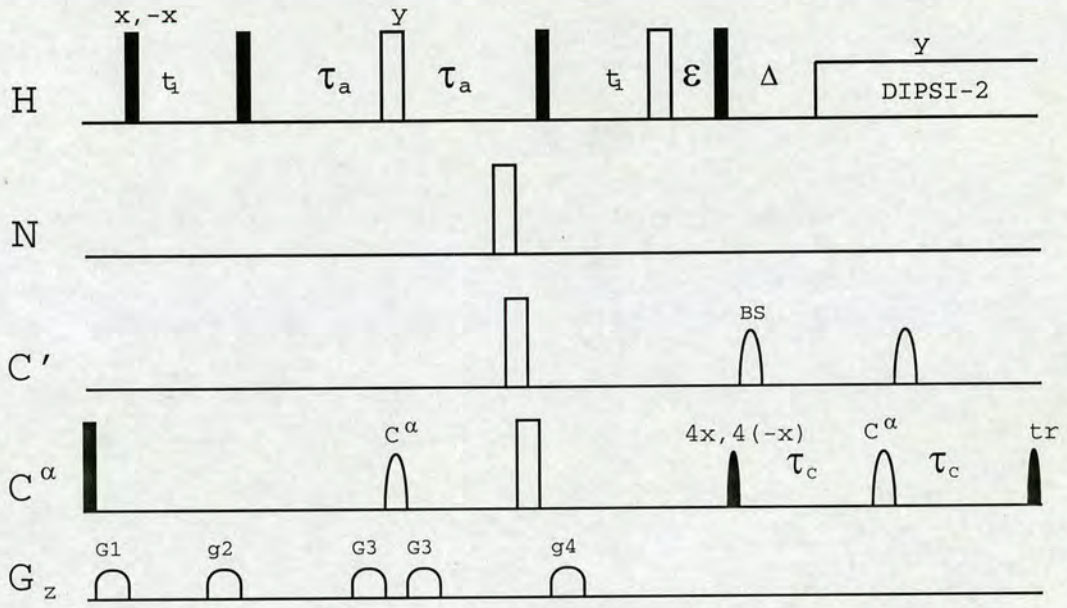

Figure 3.16: $BIRD_r$ **element from the final 3D J-HA(CACO)NH pulse sequence.** See section 2.2 for a description of the symbols and conventions used in the pulse sequence diagram. Delays are as follows: τ_a=1.78 ms ($1/4J_{CH}$), τ_C=4.4 ms ($1/4J_{C^\alpha C'}$), ε=1.2 ms (to compensate for evolution during g2 and g4), Δ=3.4 ms ($1/2J_{CH}$) is optimal for non-Gly residues and Δ=1.7 ms ($1/4J_{CH}$) is optimal for Gly residues. Gradients are G1=12%, g2=9%, G3=14%, g4=-9%. Quadrature detection during t_1 is achieved by incrementing the phase of the first ^{1}H 90° pulse according to States-TPPI.

In the 3D experiment it is still possible to achieve the refocussing of the H^α -H^{other} coupling constants but the problem must be approached differently. In this experiment, the H^α chemical shift is to be recorded simultaneously with $C^\alpha H^\alpha$ coupling sampling: this means that $H^{non-\alpha}$ proton magnetization should be inverted, but not that of H^α , which is achieved by means of a $BIRD_r$ sandwich (see section 2.2.3). $BIRD_r$ inverts protons that are remote to the inverted carbon magnetization, and is normally used to distinguish protons attached to ^{13}C from those attached to ^{12}C. In this case however, by using a C^α -selective 180° pulse only non-H^α protons are inverted as they effectively behave as if they were attached to ^{12}C atoms. In addition, including 180° ^{15}N and C' pulses at the end of the second τ_a period refocusses any $^2J_{H\alpha N}$ or $^2J_{H\alpha C'}$ evolution. Gradients within the $BIRD_r$ sandwich dephase any unwanted magnetization resulting

from imperfections in the C^α -selective pulse, while a pair of gradients of opposite sign outside the sandwich dephase any magnetization inverted by the whole sandwich (i.e. non-H^α proton). Chemical shift and coupling evolution during the 0.5 ms g2 gradients is refocussed by an extra 180° proton pulse and delay ε.

In order to ensure optimal water suppression, the ^1H carrier frequency is set to that of H_2O. This normally results in the sampling of larger ^1H spectral widths than necessary as most of the H^α signals resonate at lower ppm than the H_2O signal. In this pulse sequence, by decreasing the phase of the first ^1H 90° pulse by 90° every t_1 increment (relative to the standard States-TPPI protocol) all H^α resonance frequencies are effectively increased by 0.25× the sweep width (sw). This allows the sweep width in the ^1H dimension to be reduced to two thirds of the range normally required so that the same digital resolution is achieved with fewer increments.

During the period Δ following the $BIRD_r$ element transverse sine-modulated $C^\alpha_y H^\alpha_z$ magnetization is allowed to refocus to in-phase C^α_x. The length of the delay Δ is normally set to 3.4 ms which corresponds approximately to $1/2^1J_{C^\alpha H^\alpha}$, but can also be set to 1.7 ms which is optimal for glycine residues where two $C^\alpha H^\alpha$ couplings evolve simultaneously. Providing the two glycine H^α atoms have sufficiently different chemical shifts this should allow both $C^\alpha H^\alpha$ splittings to be measured. The magnetization transfer scheme of the remainder of the experiment is identical to that of the 3D CBCA(CO)NH sequence on which it is based [39, 72].

3.7 2D (HNCO)-(J-CA)NH, 2D (J-HACACO)NH and 3D J-HA(CACO)NH spectra of a model peptide

For purposes of comparison, the spectra of the labelled octapeptide H-Tyr-Gly-[^{15}N, ^{13}C]Gly-[^{15}N, ^{13}C]Phe-[^{15}N, ^{13}C]Leu-Arg-Arg-Ile-OH (Figure 3.17) obtained using the 2D (HNCO)-(J-CA)NH, 2D (J-HACACO)NH and 3D J-HA(CACO)NH pulse sequences are shown in Figure 3.18 - doublets are for the Leu N-H^N resonance of the peptide (i.e. corresponding to the Phe C$^\alpha$H$^\alpha$ splitting). All spectra were acquired at 298 K in DMSO solvent with a peptide concentration of 1mM.

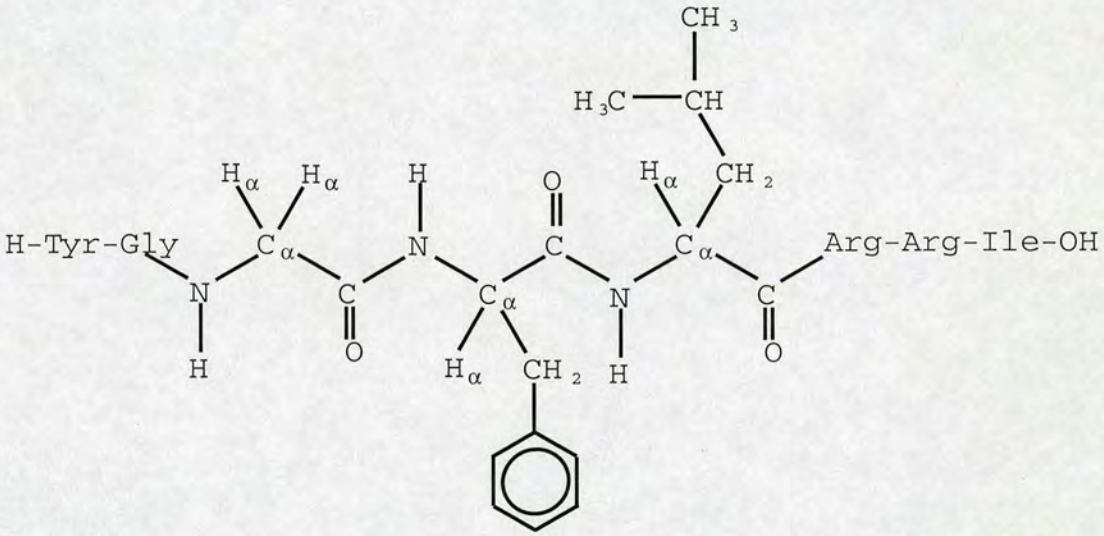

Figure 3.17: **Test octapeptide molecule structure.** Structure of the octapeptide H-Tyr-Gly-[^{15}N, ^{13}C]Gly-[^{15}N, ^{13}C]Phe-[^{15}N, ^{13}C]Leu-Arg-Arg-Ile-OH (Cambridge Isotope Laboratories). The uniformly ^{13}C , n15 enriched portion consists of amino acids Gly, Phe and Leu.

1D traces from the peptide 2D (J-HACACO)NH spectrum are shown in Figure 3.18 A1 (IP) and A2 (AP), for the 2D (J-HACACO)NH spectrum in B1 (IP) and B2 (AP) and for the 3D J-HA(CACO)NH spectrum in C1 and D1 (Δ optimized for glycine and for other amino acids, respectively).

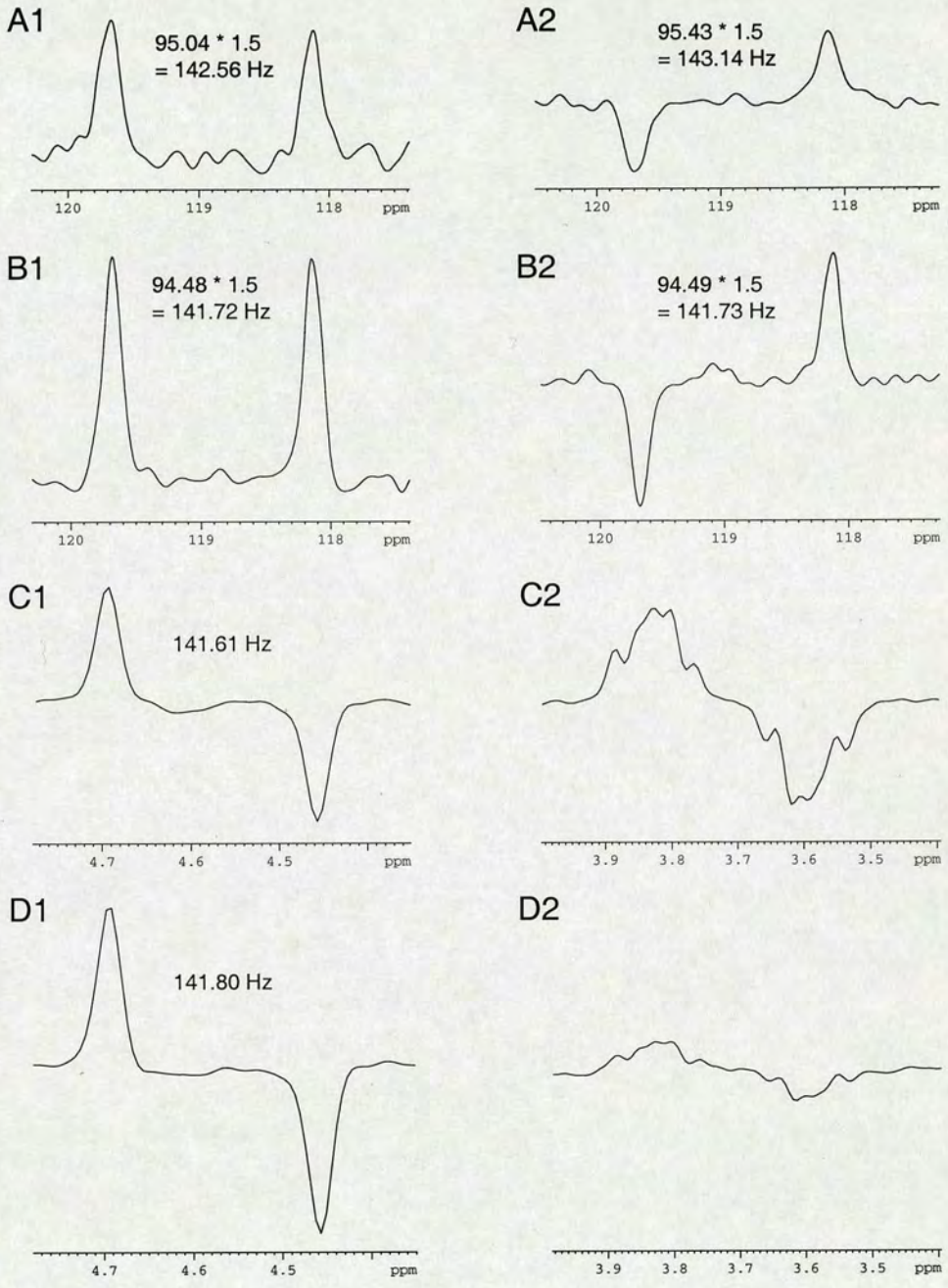

Figure 3.18: **Final pulse sequences tested on a labelled peptide. A1)** in-phase doublet obtained using 2D (HNCO)-(J-CA)NH, **A2)** antiphase doublet obtained using 2D (HNCO)-(J-CA)NH, **B1)** in-phase doublet obtained using 2D (J-HACACO)NH, **B2)** antiphase doublet obtained using 2D (J-HACACO)NH, **C1)** antiphase doublet obtained using 3D J-HA(CACO)NH with $\Delta = 1.7$ ms, **C2)** antiphase multiplet for the two glycine $^1D_{C^\alpha H^\alpha}$ couplings obtained using 3D J-HA(CACO)NH with $\Delta = 1.7$ ms, **D1)** antiphase doublet obtained using 3D J-HA(CACO)NH with $\Delta = 3.4$ ms, **D2)** antiphase multiplet for the two glycine $^1D_{C^\alpha H^\alpha}$ couplings obtained using 3D J-HA(CACO)NH with $\Delta = 3.4$ ms.

It can be seen that there is good agreement between the values obtained from the three experiments; the largest difference is $< 1\%$ of the value of the coupling. It can also be seen that the H^α -sampled spectra show better signal-to-noise (S/N), by a factor of approximately 1.8:1 in fact. This is more than can be accounted for by the extra transfer steps present in the C^α -sampled 2D (HNCO)-(J-CA)NH method using $I_{transferred} = \cos(\pi J \tau)$ for each step at which magnetization is transferred:

2D (HNCO)-(J-CA)NH: $I_{overall} = 0.73$

2D (J-HACACO)NH: $I_{overall} = 0.89$

The observed S/N of the C^α -detected 2D (HNCO)-(J-CA)NH method is due to the relaxation properties of C^α *versus* those of H^α , which the 2D (J-HACACO)NH method uses to sample the $C^\alpha H^\alpha$ splitting. Both spectra were acquired with long t_1 acquisition times (~120 ms) and it was observed that the ^1H magnetization was much longer lived than the ^{13}C magnetization. When the first traces of the 2D experiments were compared however, the theoretical ratio of peak intensities between the two methods was observed.

1D traces from the H^α -H^N plane of the 3D J-HA(CACO)NH experiment recorded for the labelled octapeptide molecule are shown in Figure 3.18 - D1 and D2 show splittings for the Gly and Phe $C^\alpha H^\alpha$ resonances, respectively with $\Delta = 3.4$ ms, i.e. optimized for non-Gly residues. The same slices recorded using $\Delta = 1.7$ ms are shown in C1 and C2 where the reduction in intensity of the doublet for Phe corresponds to $\sin(\pi J \Delta) = 0.70$ where $J = 145$ Hz. Note that in this case, due to the similarity of the two H^α chemical shifts, it is not possible to distinguish the two individual doublets, although these are observed for some glycine residues in the case of the spectrum of the protein ABA-1A discussed later (4.2.2).

The multiplet observed for the glycine splittings (Figure 3.18 C2) indicates that the resonance frequencies for the two H^α atoms are not separated by $> {}^1J_{C^\alpha H^\alpha}$ (higher order effects are present) and the two $^1J_{C^\alpha H^\alpha}$ splittings therefore cannot be determined in this case. It should be possible however, to determine two such couplings for glycine using the 3D J-HA(CACO)NH experiment provided that the resonance frequencies of

the H^α atoms are sufficiently different: in such a case, RDC restraints for both H^α atoms could be used with the necessary stereospecific assignments.

3.8 The benefits of an extended $C^\alpha H^\alpha$ sampling period

The new pulse sequences described here allow the extension of the $^1J_{C^\alpha H^\alpha}$ sampling period beyond the limit of 12 ms normally imposed by the evolution of $C^\alpha C^\beta$ couplings. The improved precision this extension of the $^1J_{C^\alpha H^\alpha}$ sampling period affords is summarized in Table 3.1 - 2D (HNCO)-(J-CA)NH spectra were acquired for the double CCP module construct factor H~2,3 [6] with the following variations: in spectrum A, 192 complex points were acquired in the indirect dimension (64 scans per increment) giving a ^{15}N t_1 acquisition time of 69.9 ms; in spectrum B, 96 complex points were acquired in the indirect dimension (256 scans per increment) giving a ^{15}N t_1 acquisition time of 34.9 ms; in spectrum C, 48 complex points were acquired in the indirect dimension (256 scans per increment) giving a ^{15}N t_1 acquisition time of 17.5 ms. In all cases, 2048 complex points were acquired in the directly detected proton dimension giving t_2 acquisition times of 106.5 ms. Spectum C was limited to a t_1 acquisition time of 17.5 ms (a $^1J_{C^\alpha H^\alpha}$ sampling period of 17.5/1.5 = 11.7 ms) and was acquired using a modified pulse sequence incorporating the nonselective ^{13}C 180° pulse during the $^1J_{C^\alpha H^\alpha}$ sampling period in order to evaluate the effect of neglecting to refocus the $C^\alpha C^\beta$ couplings but using the 2D (HNCO)-(J-CA)NH magnetization transfer scheme. Each spectrum was recorded three times and the average standard deviation observed for the measured $^1J_{C^\alpha H^\alpha}$ values is summarized in the table.

It can be seen that extension of the coupling sampling period from 17.5/1.5 = 11.7 ms (C) to 34.9/1.5 = 23.3 ms (B) is beneficial, lowering the average standard deviation from 1.29 Hz to 0.6 Hz. Sampling the $^1J_{C^\alpha H^\alpha}$ splitting beyond ~20 ms however, is not a worthwhile use of experimental time due to the rapid loss of signal caused by C^α relaxation. Acquiring noisy FIDs for the later t_1 increments degrades the S/N in the 2D spectra and leads to less precise values of the coupling constants.

[6] Prepared by Henry Hocking, University of Edinburgh. Sample conditions were: 10mM potassium phosphate / 50mM Arg / 50mM Glu buffer with a pH of 6.2, 10% D_2O. Protein concentration was 1.5 mM and all spectra were acquired at a temperature of 310 K.

Spectrum	t_1 (ms)	scans/increment	average s.d. (Hz)
A	69.9	64	1.08
B	34.9	256	0.60
C	17.5	256	1.29

Table 3.1: **2D (HNCO)-(J-CA)NH precision** with and without (spectrum C) $C^\alpha C^\beta$ refocussing. Average standard deviations are calculated from 3 repeat measurements of 40 $^1J_{C^\alpha H^\alpha}$ couplings.

3.8.1 Examples of spectra for an aligned protein: ABA-1A

In the case of the H^α -sampled 2D (J-HACACO)NH and 3D J-HA(CACO)NH experiments the issue of $C^\alpha C^\beta$ coupling evolution is bypassed and instead, the evolution of H^α -H^{other} couplings, particularly in aligned samples, is the main problem that must be overcome in order to extend the sampling period. The effects of H^α -H^{other} coupling suppression, or lack thereof, are shown in Figure 3.19 for an aligned sample of the protein ABA-1A (1.5 mM ABA-1A, 4.8 mg/mL Pf1 phage, see section 4.2.1).

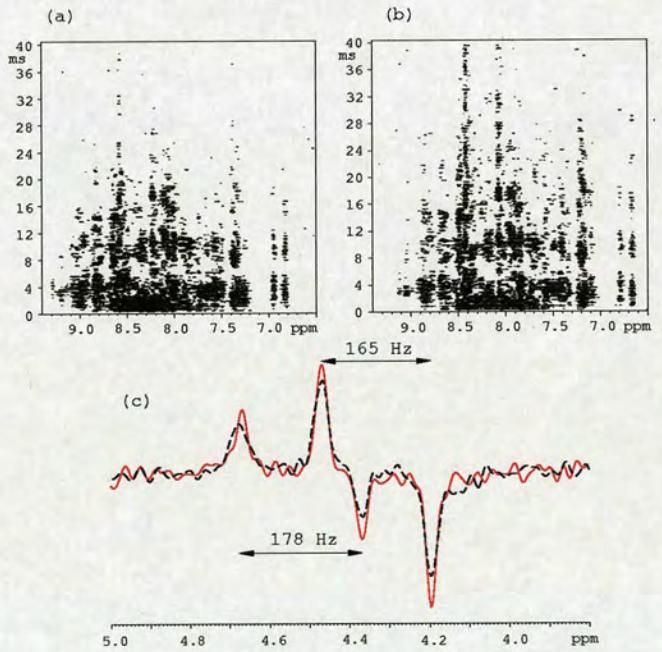

Figure 3.19: **Effects of the BIRD$_r$ element.** a) 3D J-HA(CACO)NH without BIRD$_r$ and b) with BIRD$_r$. c) An overlay of 1D traces from the same 3D J-HA(CACO)NH experiment without (dashed black line) and with (solid red line) BIRD$_r$.

Here, interferograms from the first ^{15}N planes of a 3D J-HA(CACO)NH experiment and 3D J-HA(CACO)NH-like experiment without the $BIRD_r$ element are shown in parts a) and b), respectively. Part c) shows an overlay of two traces extracted from the spectra after the second Fourier transform was performed on the interferograms. Each trace shows two antiphase doublets due to the overlapping NH protons. For both spectra, 88 t_1 increments were acquired with 144 scans resulting in t_1 and t_2 acquisition times of 40.2 and 107.0 ms, respectively.

It can be seen that the $BIRD_r$ element results in slower magnetization decay and therefore larger intensities and narrower lines, providing better definition of the splittings - note that this is in spite of significant deviation of the splittings from the value of $^1J_{C^\alpha H^\alpha}$ used to calculate the BIRD element delay.

Examples of spectra for the 2D and 3D experiments are shown in Figure 3.20. These spectra have been acquired using the Pf1 phage-aligned sample of the protein ABA-1A: for the 2D (HNCO)-(J-CA)NH spectrum, acquisition times of 107 and 84.2 ms in t_2 and t_1 were used accumulating 128 time points in the indirect dimension using 192 scans per increment; for the 2D J-HA(CACO)NH spectrum, acquisition times of 107 and 63.1 ms in t_2 and t_1 were used accumulating 96 time points in the indirect dimension using 288 scans per increment; for the 3D (HN)CO-(J-CA)NH spectrum, acquisition times of 107, 42.1 and 10.6 ms in t_3, t_2 (^{15}N) and t_1 (CO) were used with 16 scans per increment; for the 3D J-HA(CACO)NH spectrum, acquisition times of 107, 17.1 and 42.6 ms in t_3, t_2 (^{15}N) and t_1 (H^α) were used with 16 scans per increment. The window functions in F_1 were chosen such that the later t_1 increments were given very little weight so as not to degrade the S/N.

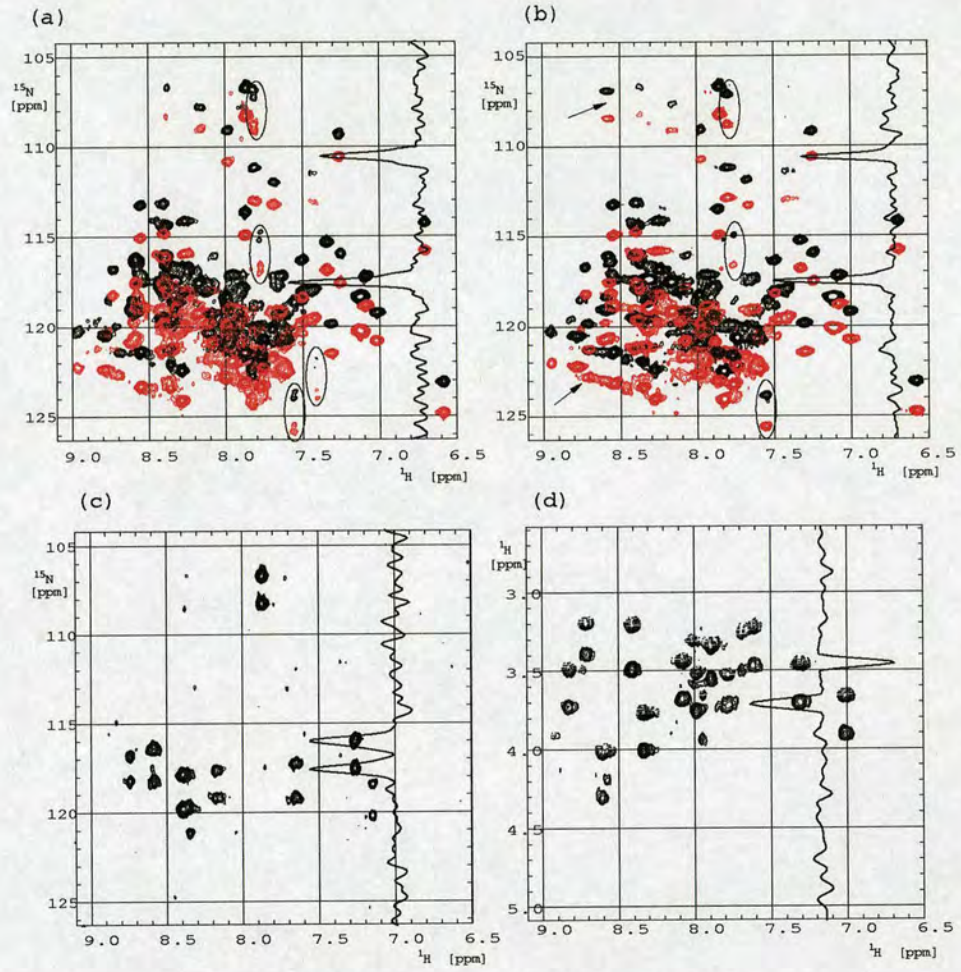

Figure 3.20: **2D and 3D $^1J_{C^\alpha H^\alpha}$ spectra of aligned ABA-1A.** Overlays of IPAP edited spectra acquired using (a) the 2D (HNCO)-(J-CA)NH pulse sequence, and (b) the 2D (J-HACACO)NH pulse sequence. The IP+AP and IP-AP spectra are shown in black and red, respectively, and circled signals correspond to residues preceeded by serines or threonines. Signals that are stronger in spectrum (b) are indicated with arrows. (c) A representative carbonyl carbon plane from the 3D (HN)CO-(J-CA)NH spectrum of aligned ABA-1A: the scaled $C^\alpha H^\alpha$ couplings correspond to in-phase doublets in the ^{15}N dimension. (d) A representative nitrogen plane from the 3D J-HA(CACO)NH spectrum of ABA-1A: the antiphase $C^\alpha H^\alpha$ splittings occur in the H^α dimension (positive intensity solid, negative intensity shaded). An F_1 trace through the NH proton at 7.3 ppm is shown for both (c) and (d).

In all cases, spectra of sufficient quality for the determination of the size of the $C^\alpha H^\alpha$ splittings are obtained and in the case of the 3D experiments, overlap is much reduced. An examination of the 2D spectra also reveals that peaks for residues with a preceeding serine or threonine are of low intensity and show resolved $C^\alpha C^\beta$ doublets - this is to be expected since the C^α -selective pulses on which these schemes rely cannot distinguish between the C^α and C^β resonances for these residues (the shifts overlap - see Figure 3.3) and the $C^\alpha C^\beta$ coupling is therefore not fully refocussed.

Chapter 4

ABA-1A

4.1 Background

ABA-1 is a nematode polyprotein allergen (NPA) with fatty acid binding functionality from the parasitic nematode *Ascaris suum* [71]. The 16 kDa, mostly α-helical ABA-1A construct studied here (a single A-type unit) consists of 140 amino acids numbered from L224 to S357 with the sequence shown below (the seven helical regions, as observed in the final ensemble, are overlined - dots indicate a discontinuity within a helical region):

224 GSPEFH

230 HFTLESSLDT HLKWLSQEQK DELLKMKKDG KAKKELEAKI

270 LHYYDELEGD AKKEATEHLK GGCREILKHV VGEEKAAELK

310 NLKDSGASKE ELKAKVEEAL HAVTDEEKKQ YIADFGPACK

350 KIYGVHTSLC REIL

A number of factors made ABA-1A a good candidate for measurement of RDC restraints with the hope that this would lead to improvements in the structure: firstly, in contrast to CCP module pairs, ABA-1A is a relatively globular, uncharged molecule [1] and would not be expected to bind excessively to, or be too aligned strongly by, charged aligning media such as Pf1 phage. Secondly, ^{15}N T_2 relaxation times were

[1] Predicted pI = 6.09, 27 positively charged amino acids, 29 negatively charged amino acids.

estimated [2] to be > 100 ms, which compares favourably to the CCP module pairs described elsewhere in this work (average values of 87 and 88 ms for DAF~2,3 and C4BP~1,2, respectively). Finally, there was a lack of unambiguous NOEs bridging the interface between the region comprising residues 306 and 328 of helices 5 and 6 and the other helices that make up ABA-1A (helices 1,2,3,4 and 7) as shown in Figure 4.1. The addition of RDC restraints could be expected to improve the definition of the orientation of helices 5 and 6 with respect to the rest of the molecule and also produce a better overall 3D structure.

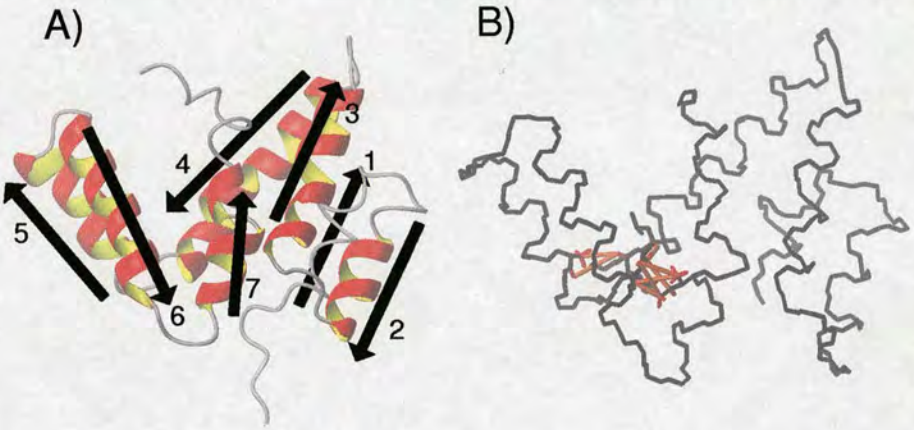

Figure 4.1: **ABA-1A structure** using NOE restraints only (lowest energy structure of the ensemble). **A)** Ribbon diagram showing the location of the 7 helices in the structure of ABA-1A. Note that the region in which the first helix occurs is not well structured here and is therefore not represented by a helix cartoon. **B)** NOEs that can be unambiguously assigned to restraints bridging the interface between helices 5/6 and the rest of the molecule are shown in red/orange. These are: 341HD1#-329HB1, 296HA-329HD2#, 305HB1-3297HA, 329HA-341HD#/HG21, 329HB1-341HD1#/HG2#, 329HB2-341HD1#/HG2#, 341HD1#-329HA, 341HG11-329HD2#, 341HG12-329HD2# and 305NH-297HA.

4.2 Experimental

A double labelled sample of ABA-1A was prepared by Nicola Meenan, who also carried out assignment [67] and initial structure calculations using the program 'Crystallography & NMR System' (CNS) [17].

[2] Based on an analysis of the t_1 decay time in ^{15}N HSQC spectra.

4.2.1 Sample preparation

Unaligned sample conditions were as follows: a ratio of 9:1 H_2O:D_2O , 50 mM phosphate buffer (pH 7) and 50 mM NaCl with a final concentration of ABA-1A of 2.5 mM and a volume of 250 μl (Shigemi tube). In addition, a molar excess of oleic acid was present to stabilize the protein. An aligned sample was prepared using identical conditions but with the addition of 4.8 mg/mL Pf1 phage (Profos AG, Regensburg, Germany) giving a protein concentration of 1.5 mM (550 μl sample volume) and a residual quadrupolar splitting of the D_2O signal of 13.1 Hz.

4.2.2 Measurement of RDCs

Experiments to measure RDC restraints were as follows: N-H^N coupling constants were measured using the IPAP experiment of Ottiger, Delaglio and Bax [77]. A dataset was acquired in the unaligned state consisting of two data matrices (i.e. IP and AP) of 1024 $^1H^N$ times 320 ^{15}N complex points (48 scans per increment) giving t_2 and t_1 acquisition times of 113.9 ms and 105.3 ms, respectively; a dataset in the aligned state consisting of two matrices of 1024 $^1H^N$ times 360 complex points was acquired (32 scans per increment) giving t_2 and t_1 acquisition times of 56.9 ms and 118.4 ms, respectively. Prior to addition / subtraction of the IP and AP planes, a scaling factor of 1.1 was applied to the AP planes to compensate for relaxation losses. Processing was carried out with AZARA [14] using Gaussian window functions and extensive zerofilling to give IP+AP and IP-AP spectra measuring 2048 $^1H^N$ by 4096 ^{15}N real points in the case of the data for the unaligned state and 1024 by 4096 real points in the case of the data for the aligned state.

Three other types of RDCs were measured in a similar manner: $^1D_{NC'}$, $^1D_{C^\alpha C'}$ and $^1D_{C^\alpha H^\alpha}$. In all cases, addition and subtraction of a pair of planes was performed without scaling of one of the subspectra. The $^1D_{NC'}$ and $^1D_{C^\alpha C'}$ constants were measured using the spin-state-selective methods of Permi *et al.* [81, 83]. The N-C' dataset for the unaligned state consisted of two data matrices of 2048 times 384 complex points (64 scans per increment) giving t_2 and t_1 acquisition times of 113.9 ms and 126.3 ms, respectively. These were processed using AZARA with the application of a Gaussian

window function to give a real data matrices of 4096 by 4096 points. The N-C' dataset in the aligned state consisted of two data matrices of 2048 times 320 complex points (72 scans per increment) giving t_2 and t_1 acquisition times of 113.9 ms and 105.3 ms, respectively. These data were processed in the same way to give spectra consisting of 4096 times 4096 real points. The C^{α}-C' dataset in the unaligned state, consisting of data matrices of 2048 times 288 complex points (160 scans per increment, t_2 and t_1 acquisition times of 122.1 ms and 94.7 ms, respectively) was processed as before to give spectra of 2048 times 4096 real points; the corresponding dataset in the aligned state consisting of data matrices of 2048 times 256 complex points (192 scans per increment, t_2 and t_1 acquisition times of 122.1 ms and 84.2 ms, respectively) was processed in an identical manner to give spectra measuring 2048 times 2048 real points. Finally, the 2D (HNCO)-(J-CA)NH method described in 3.5.1 was used to acquire spectra for the determination of $^1D_{C^{\alpha}H^{\alpha}}$. Data matrices consisting of 2048 times 256 complex points in the case of the unaligned sample (192 scans per increment, t_2 and t_1 acquisition times of 106.5 ms and 84.2 ms, respectively) and 2048 times 220 complex points in the case of the aligned sample (288 scans per increment, t_2 and t_1 acquisition times of 106.5 ms and 65.8 ms, respectively) were processed with AZARA as before, using a Gaussian window function, to give spectra of 2048 times 2048 complex points in both cases. The length of the time period during which the $C^{\alpha}H^{\alpha}$ coupling constant was sampled was scaled down by a factor of 1.5 times, i.e. to $t_{1max}/1.5$, as described in section 3.5.1.

All spectra were loaded into the program CCPN Analysis [15] and after peak picking (using the parabolic fit of Analysis to determine peak centres), the coupling constants were determined as the difference between the ^{15}N frequency of the high-field and low-field peaks of each peak pair in the ^{15}N dimension (in Hz). The RDCs were then determined by subtraction: $^1D = {}^1(J + D)$ - ^1J, and following determination of $^1D_{C^{\alpha}H^{\alpha}}$, this coupling constant was scaled up by a factor of 1.5. The approach of Kontaxis *et al.* [53] (detailed more fully later in Chapter 5) was used to estimate the errors in the couplings, giving:

$^1D_{NH}$: 0.50 Hz

$^1D_{NC'}$: 0.55 Hz, scaled error = 4.6 Hz

$^1D_{C^\alpha C'}$: 0.71 Hz, scaled error = 3.6 Hz

$^1D_{C^\alpha H^\alpha}$: 2.80 Hz, scaled error = 1.3 Hz

where the scaled error refers to the error after scaling each coupling relative to $^1D_{NH}$ according to gyromagnetic ratio and bond length (scaling factor values are summarized in Table 2.2) - this is necessary in order to use the different types of measured RDC as restraints with the same alignment tensor.

4.3 Structure refinement with RDCs

Relaxation data (not shown) were used to identify flexible regions of the protein using the $\{^1H\}$-^{15}N NOE values measured at a ^1H frequency of 600 MHz: data for residues preceding T232 and following G353, as well as G278 and K309 were therefore not used as restraints. RDC restraints were added to the structure calculation protocol using the TENSO module of the program Crystallography & NMR System (CNS, [17]) which contains an implementation of the Implicit Saupe tensor Alignment Constraint (ISAC) method [94, 70] [3] - this is described in section 2.5. In all sets of structure calculations, the tenso force constant was increased geometrically from 0.001 kcal·(mol·Hz^2)$^{-1}$ to 1 kcal·(mol·Hz^2)$^{-1}$ during the first cooling stage of refinement and fixed at the final value throughout the second cooling stage.

In an effort to improve the convergence of the structure calculations, a number of alterations to the refinement script were made and tested (see Appendix B for the structure calculation protocol used). By increasing the time available to search the conformational space corresponding to different internuclear vector orientations, it was hoped that more refined structures would emerge that were able to satisfy both the NOE distance restraints and the RDC restraints. To this end, the following changes

[3] The advantage of using the ISAC algorithm rather than SANI for example, is that no prior estimates of the alignment tensor's magnitude and rhombicity are required (see section 2.5).

were tested independently starting from the same set of 30 'rrsa' structures (using all RDC restraints and assuming a single alignment tensor for the molecule):

- First cooling stage 'tempstep' reduced from 50 K to 10 K
- First cooling stage 'tempstep' reduced from 50 K to 25 K
- First cooling stage 'timestep' reduced from 0.003 ps to 0.001 ps
- Second cooling stage 'tempstep' reduced from 50 K to 10 K
- Second cooling stage 'timestep' reduced from 0.003 ps to 0.001 ps
- Length of first cooling stage dynamics increased by five times

The 'tempstep' variable is the size of the step by which the temperature is reduced as the molecule is cooled down during the simulated annealing process and the 'timestep' is the length of time after which calculations are performed during the Cartesian dynamics run at each temperature. Note that reducing the size of these steps leads to an increase in the overall calculation time if the total length of the Cartesian dynamics run at each temperature is to be kept the same. In order to distinguish the effects of a reduced temperature step from the effects of the consequent increase in overall calculation time, the last change tested is an increase in the length of each Cartesian dynamics run in the first cooling stage. The results are illustrated in Figure 4.2.

It can be seen that largest improvement in convergence is obtained by changing the 'tempstep' in first cooling stage from 50 K to 10 K - this means that the structures cool more slowly after the high temperature step of the refinement protocol and are therefore less likely to become trapped in local minima: simply increasing the length of the Cartesian dynamics during the first cooling stage does not lead to such an improvement. A 10 K 'tempstep' was therefore used in all structural refinements for which RDC restraints were available. In addition, the error in $^1D_{\mathrm{NH}}$ was increased from 0.5 Hz to 1.0 Hz: this was in an effort to increase the contribution of the other RDC classes to the RDC energy term, since the energy penalty for each restraint is divided by the square of the error estimate when using the harmonic ISAC potential.

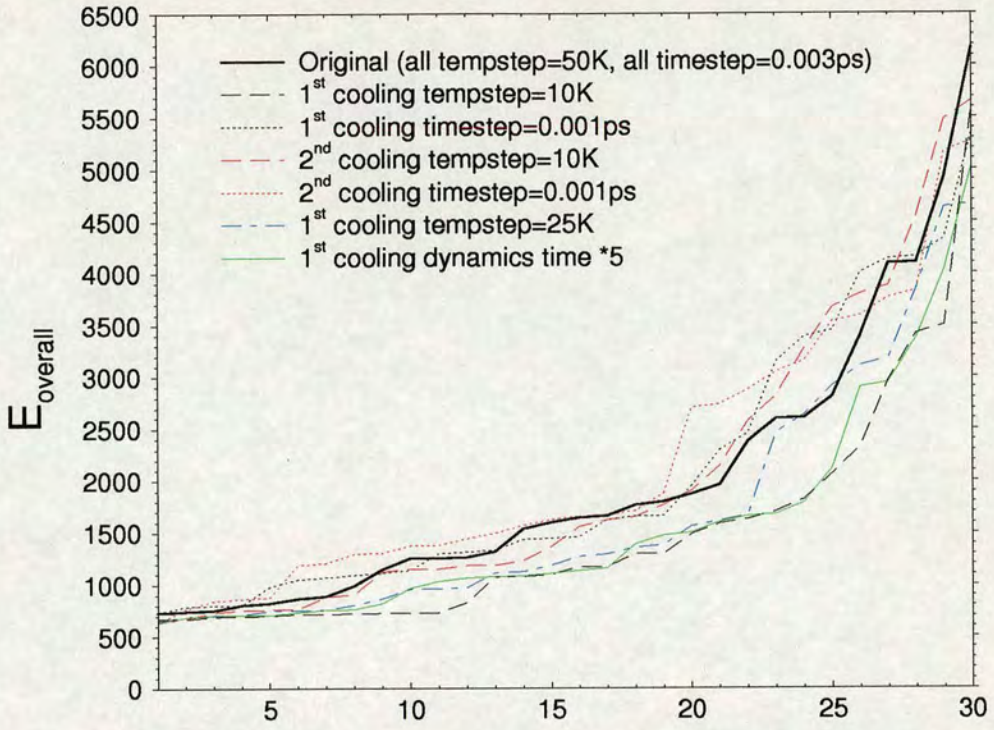

Figure 4.2: **ABA-1A structure refinement optimization.** Graph of $E_{overall}$ vs. energy-ranked structure number for the original and altered refinement protocols.

4.3.1 Is the use of a single alignment tensor for ABA-1A justified?

Four sets of structures were calculated: one set containing all RDC restraints (set 'ALL'), one set excluding RDC restraints for helices 5-6 (set 'MAIN') one set set using RDC restraints for helices 5-6 only (set 'HEL56') and finally, a set including RDC restraints for the whole molecule but excluding the $^1D_{C^\alpha H^\alpha}$ RDCs (set 'noCH'). Total numbers of RDC restraints used are:

Set 'ALL' - $^1D_{NH}$: 73, $^1D_{NC'}$: 70, $^1D_{C^\alpha C'}$: 66, $^1D_{C^\alpha H^\alpha}$: 40

Set 'MAIN' - $^1D_{NH}$: 53, $^1D_{NC'}$: 53, $^1D_{C^\alpha C'}$: 52, $^1D_{C^\alpha H^\alpha}$: 32

Set 'HEL56' - $^1D_{NH}$: 16, $^1D_{NC'}$: 15, $^1D_{C^\alpha C'}$: 13, $^1D_{C^\alpha H^\alpha}$: 8

Set 'noCH' - $^1D_{NH}$: 73, $^1D_{NC'}$: 70, $^1D_{C^\alpha C'}$: 66, $^1D_{C^\alpha H^\alpha}$: 0

The spread of alignment tensor values can be determined for each ensemble by Singular Value Decomposition (SVD - see section 2.4) using the calculated structures and experimental dipolar coupling values. The program Orderten-SVD [63] was used for this purpose. All structures were overlaid with reference to the lowest energy structure of the set 'ALL' ensemble so that the alignment tensor eigenvectors shown in Figure 5.3 are in the same (arbitrary) coordinate system. The spread of alignment tensors within the 'ALL' (overlaid on residues 232-353), 'MAIN' (overlaid on residues 232-299) and 'HEL56' (overlaid on residues 302-332) ensembles was sampled by performing 100 SVD computations for each converged structure of the ensemble. The resulting eigenvalues are summarized in Table 4.1 and the corresponding eigenvector orientations are mapped onto Sanson-Flamsteed plots (described in section 2.4) in Figure 4.3.

Set	Szz	Syy	Sxx	η
'ALL'	8.52 (0.27)	-8.31 (0.25)	-0.21 (0.27)	0.93 (0.05)
'MAIN'	8.42 (0.29)	-8.45 (0.28)	0.03 (0.29)	0.94 (0.05)
'HEL56'	8.76 (0.61)	-8.10 (0.69)	-0.66 (0.46)	0.84 (0.08)

Table 4.1: **Alignment tensor eigenvalues for ABA-1A.** for 'ALL', 'MAIN' and 'HEL56' structure ensembles. Sii values shown $\times 10^{-4}$ are average values in the PAS followed by the standard deviation in parentheses. The majority of solutions for the 'ALL' ensemble give positive Szz, negative Syy and negative Sxx; but Szz and Syy are interchanged in some solutions because $\eta \sim 1$ (the 'MAIN' Szz and Syy entries should be swapped to agree with convention).

These results show that the alignment tensor eigenvalues for helices 5 and 6 are the same as those for the rest of the molecule or 'MAIN' part (helices 1,2,3,4 and 7) within the spread of calculated solutions. The eigenvector orientations for the 'MAIN' and 'HEL56' RDC subsets shown in Figure 4.3 are also in good agreement: therefore the molecule can be treated as a rigid entity with a single alignment tensor. Although there are other relative orientations of the 'MAIN' and 'HEL56' alignment tensors that are equally valid (see section 2.4), the covalent bonds and NOE restraints must preclude these other possibilities since only one orientation is seen in the set 'ALL' ensemble.

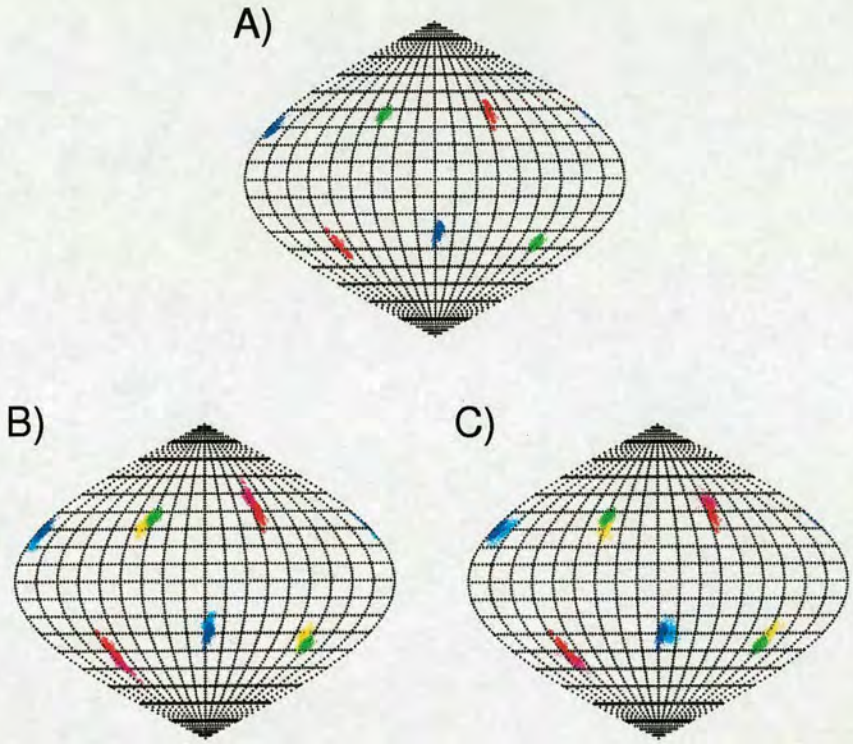

Figure 4.3: **Alignment tensor distributions for ABA-1A.** Sanson-Flamsteed plots showing the orientation of the alignment tensor eigenvectors in coordinate systems coincident with that of the lowest energy set 'ALL' structure. **A)** The eigenvectors for the 'ALL' ensemble using the set 'ALL' RDCs: colours for the eigenvectors corresponding to the largest positive, largest negative and smallest eigenvalues are shown in red, blue and green, respectively. **B)** Eigenvectors for the 'MAIN' ensemble calculated using set 'MAIN' RDCs: the eigenvectors corresponding to the largest positive, largest negative and smallest eigenvalues are shown in magenta, cyan and yellow, respectively. For comparison, the set 'ALL' eigenvectors are overlaid in red, blue and green. **C)** Eigenvectors for the 'HEL56' ensemble calculated using set 'HEL56' RDCs: the eigenvectors corresponding to the largest positive, largest negative and smallest eigenvalues are shown in magenta, cyan and yellow, respectively. For comparison, the set 'ALL' eigenvectors are overlaid in red, blue and green.

4.4 Effects of RDCs on the structure of ABA-1A

Tables 4.2 and 4.3 summarize the CNS energy terms and backbone RMSD values for the converged structures - for comparison, the statistics for a set of structures without RDC restraints (set 'NOE') are also shown.

ensemble	Etot	Enoe	Etenso
'ALL'	781.6	143.4	144.9
'MAIN'	675.9	113.2	117.1
'HEL56'	350.4	70.0	24.7
'noCH'	647.4	131.9	79.3
'NOE'	237.2	43.4	-

Table 4.2: **Average CNS energy terms for ABA-1A structures.**

Ensemble/No.	$\text{RMSD}_{overall}$	RMSD_{main}	RMSD_{hel56}
'ALL'/28	0.90	0.64	0.58
'MAIN'/32	1.80	0.74	0.84
'HEL56'/25	1.42	1.01	0.52
'noCH'/32	1.04	0.74	0.70
'NOE'/37	1.30	0.90	0.80

Table 4.3: **RMSD values for ABA-1A structures.** The structures were overlaid on the backbone N, C and CA atoms of the following residues (denoted overall, main and hel56): 232-299 and 335-353 for the main body of the molecule (i.e. excluding flexible tails and helices 5 and 6), 302-332 for the helix 5-6 region and 232-353 for the whole molecule. All RMSD values are given in angstroms (Å), and the number of structures in each ensemble is shown in the first column.

Inclusion of the RDC restraints in addition to the NOE data gives rise to two noticeable improvements: firstly, the orientation of helices 5 and 6 relative to the rest of the molecule is better defined and the accuracy of the structure as a whole is improved: the backbone pairwise RMSD over residues 232-353 in the NOE only ensemble is 1.30Å (over helices 5 and 6, 0.80Å) and for ensemble with set 'ALL' RDCs it is 0.90Å (over helices 5 and 6, 0.58Å). Secondly, the local geometry of the structures improves as judged by the positions of the residues on a Ramachandran plot (see Table 4.4).

According to both the Ramachandran statistics and RMSD values, the set 'noCH' structures are an improvement on the NOE only structures, but are less accurate than the structures including $^1D_{C^\alpha H^\alpha}$ RDCs. An independent means of checking the improved accuracy of the structures obtained by refining with RDC restraints is to check whether an improved fit with data not used in the refinement is obtained: to this end, the $^1D_{C^\alpha H^\alpha}$ RDCs have been calculated for the closest-to-mean structures

Region	set 'ALL'	set 'noCH'	'NOE'
Most favoured	79.7	77.3	75.5
Additionally allowed	18.7	20.1	21.6
Generously allowed	1.1	2.1	2.2
Disallowed	0.5	0.5	0.8

Table 4.4: **ABA-1A Ramachandran plot statistics.** Percentages of residues with ϕ-ψ angle pairs in the most favoured, additionally allowed, generously allowed and disallowed regions of a Ramachandran plot calculated without (NOEs only) and with RDC restraints.

of the set 'NOE', 'noCH' and 'ALL' ensembles using $^1D_{\mathrm{NH}}$ RDCs to determine the alignment tensor (a program provided by Tran Pham was used to carry out these predictions). The results are summarized in Figure 4.4.

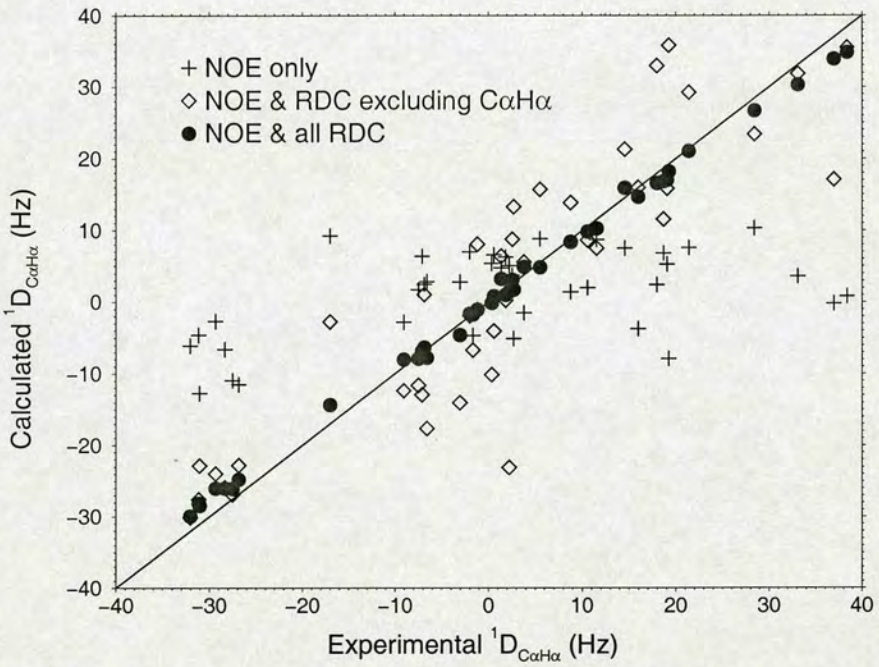

Figure 4.4: **Predicted and observed $^1D_{C^\alpha H^\alpha}$ RDCs.**

While there is little or no observable correlation between the predicted and observed $^1D_{C^\alpha H^\alpha}$ RDCs in the case of the 'NOE' structures (correlation coefficient $r = 0.51$), the incorporation of three other types of RDCs ($^1D_{\mathrm{NH}}$, $^1D_{\mathrm{NC'}}$ and $^1D_{C^\alpha C'}$) clearly leads to some agreement ($r = 0.90$). When the $^1D_{C^\alpha H^\alpha}$ couplings are also used as

restraints very good agreement is achieved ($r = 1.00$).

A comparison between the structure ensembles with and without RDC restraints is shown in Figure 4.5.

4.5 Discussion

The refinement protocol including the 'tenso' energy term for RDC restraints can be optimized to improve the convergence of the structure calculations by reducing the temperature step on cooling from 2000 K to 1000 K. By including RDC restraints, not only has the orientation of the two distinct structural elements (helices 5-6 and the main body of the protein) become better defined, but the secondary structure has also improved slightly. The close agreement of the alignment tensors obtained for the helix 5-6, main body and whole protein, and the improvement of the secondary structure upon inclusion of RDC restraints are evidence that the whole protein experiences the same alignment tensor and possesses a meaningful average structure. The small discrepancies between the alignment tensors for the two structural elements and the protein as a whole could be due to restricted motion between the elements and/or due to structural 'noise' in the structure determined using NOE data. The $^1D_{C^\alpha H^\alpha}$ RDCs measured using the 2D (HNCO)-(J-CA)NH method described in section 3.5.1 have been shown to produce an improvement in RMSD and secondary structure over and above that obtained with $^1D_{NH}$, $^1D_{NC'}$ and $^1D_{C^\alpha C'}$ RDCs only, providing evidence for the validity of the method.

Figure 4.5: **ABA-1A structures. A.** The converged structures (best 31) from the ensemble calculated without RDC restraints are shown in black, with helices 5 and 6 coloured blue. The best 28 from the ensemble calculated with set 'ALL' RDC restraints are coloured red for the main part of the molecule and orange for helices 5 and 6. The backbone pairwise RMSD over residues 232-353 in the NOE only ensemble is 1.31Å (over helices 5 and 6, 0.74Å) and for ensemble with set 'ALL' RDCs it is 0.90Å (over helices 5 and 6, 0.58Å). **B.** As for A. but viewed from the top.

Chapter 5

DAF~2,3

5.1 Background

Decay accelerating factor (DAF) or CD55 is an ~ 80 kDa RCA protein comprising 4 CCP modules inserted via a glycophosphatidyl inositol (GPI) anchor into host tissue membranes and protects self cells from autologous complement activation by dissociating C3 convertases assembled on the membrane [66] (see section 2.6). The complement regulatory activity of DAF has been localised to CCPs 2-4 [16]: CCPs 2-3 have been found to be necessary and sufficient for acceleration of the decay of the classical pathway C3 convertase [16] and CCP 4 is additionally required in the case of the alternative pathway C3 convertase [16]. DAF is of therapeutic interest in suppressing xenograft hyperacute rejection [5] and is targeted by a number of pathogenic organisms including Dr^+ *E. coli* and echovirus 7 [59].

The construct studied here consists of CCPs 2-3 and comprises the following amino acid sequence (native sequence from F61-E188):

60 EFRSC

65 EVPTRLNSAS LKQPYITQNY FPVGTVVEYE CRPGYRREPS

105 LSPKLTCLQN LKWSTAVEFC KKKSCPNPGE IRNGQIDVPG

145 GILFGATISF SCNTGYKLFG STSSFCLISG SSVQWSDPLP

185 ECREHHHHHH

A site-directed mutagenesis study probing the functional role of 24 substituted residues in CCPs 2-3 has been published [56], raising the possibility of obtaining a better understanding of the complement regulatory activity of DAF by mapping these mutations onto an experimentally determined structure.

The publication of the first DAF~2,3 NMR structures [112] was followed by the publication of crystal structures of a complete 4 module DAF~1-4 construct [64]. Comparing modules 2 and 3 in the NMR and x-ray structures (Figure 5.1, parts a) and b), respectively), it became clear that there were differences in the structure of the intermodular junction and in particular the F-G loop of module 3.

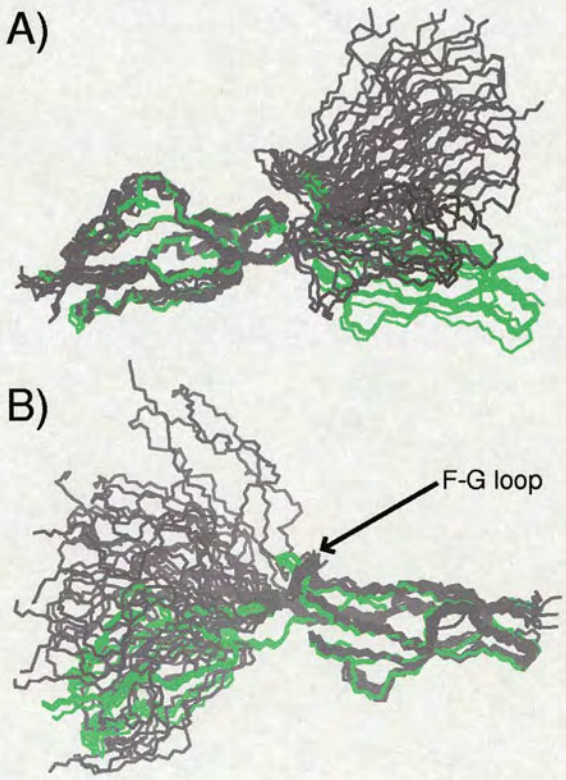

Figure 5.1: **Comparison of the original DAF~2,3 NMR structure / x-ray structure.** a) module 2 overlay of the original NMR (black) and x-ray (green) structures of DAF~2,3 and b) module 3 overlay of the NMR and x-ray structures.

The DAF~2,3 NMR data are therefore currently being re-examined. Although the second determination of the DAF~2,3 NMR structure is a work in progress, the $^1D_{NH}$ RDC data are analyzed here using the new structures in their current state in addition to presenting the results obtained from the first structure determination.

5.2 Experimental

An ^{15}N -labelled sample of DAF~2,3 was obtained from Feng Lin, Cleveland, Ohio, prepared as described by Lin *et al.* [58]. Initial assignment [113] and structural work were carried out by Stanislava Uhrinova and Eve Brook: sample conditions for these experiments were 0.9 mM DAF~2,3, 10% D_2O , 50 mM NaOAc(d3) buffer pH5, 0.03% NaN_3 and 5 mM EDTA(d12).

5.2.1 Measurement of $^1J_{NH}$ and repurification by Ni-affinity chromatography

The IPAP-HSQC method of Ottiger, Delaglio and Bax [77] was used to measure $^1J_{NH}$ couplings for later determination of the corresponding RDCs. A single unaligned dataset consisting of data matrices of 2048 by 224 complex points was acquired with a 600 MHz Varian Inova spectrometer with 48 transients per point, giving an acquisition time in the indirect dimension of 50.9 ms.

It became obvious monitoring the state of the ^{15}N sample originally used for the structural work that degradation was occurring (Figure 5.2). In order to restore the original quality of the spectra, some repurification was necessary. Since the protein was originally purified by means of Ni-affinity chromatography, and since the His-tag still remained, this was the method chosen to recover the sample.

The first step was exchange into phosphate buffered silane (PBS): 0.02 M sodium phosphate buffer with 0.15 M sodium chloride, pH adjusted to 7.4, which was accomplished by four rounds of dilution to 5 mL and concentration to 500 μL using a centrifuge concentrator with 5kDa cutoff membrane. This was followed by loading onto a Ni-affinity column (Quiagen), washing with 3×2 mL PBS and elution with 2×2 mL PBS contain-

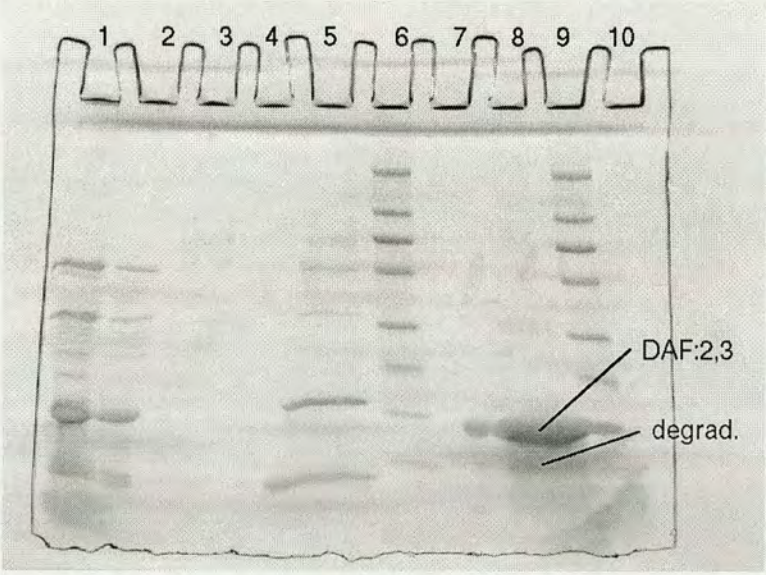

Figure 5.2: **SDS/PAGE gel of degraded DAF:2,3 sample.** Lane 8 shows the condition of the sample before repurification - note the degradation product band at ~6.5 kDa.

ing 0.3 M imidazole. EDTA and PMSF in EtOH were added in order to remove Ca^{2+} and render any serine proteases inactive: 20 μL 0.5 M EDTA and 25 μL PMSF/EtOH (7 mg/mL and 0.04 M, respectively) were added to each of the 2 mL eluted fractions. Finally, the fractions were combined and exchanged into a 50 mM NaOAc(d3) buffer, pH5.5, by the process of dilution and concentration described above. Additionally, deuterated EDTA(d12) and sodium azide were added to give final concentrations of 5 mM and 0.03% w/v, respectively.

5.2.2 Measurement of ^{15}N relaxation data

For the ^{15}N T_1 data measured at a ^1H frequency of 600 MHz, 2048 complex points in the directly detected proton dimension and 226 complex points in the indirect ^{15}N dimension were collected (number of transients per time point shown below in brackets). Processing was carried out using AZARA with sinbell squared window functions shifted by 80° in the direct dimension and 90° in the indirect dimension. After zero-filling and Fourier transformation the resulting 2048 by 256 data matrices were loaded into ANSIG for analysis. The T_2 raw data consisting of 2048 by 256 points were

processed in the same way (time points marked * were repeated for purposes of error estimation):

T_1 **ms/(nt):** 6.75*(32), 337.5(32), 472.5(32), 675(64), 945(64), 1080(64), 1147.5(64)

T_2 **ms/(nt):** 16.03*(40), 32.06(40), 48.10(40), 80.16(80), 96.19(80), 112.22(80), 128.26(80)

$\{^1H\}$-^{15}N NOE data consisting of unsaturated and saturated spectra measuring 2048 by 256 points (64 transients) were also collected and processed using AZARA to with sinbell squared window functions shifted by 80° in the direct dimension and 90° in the indirect dimension. After zerofilling and Fourier transformation the resulting 2048 by 256 data matrices were loaded into ANSIG for analysis.

For the T_1 relaxation data measured at a ^1H frequency of 800 MHz, a data matrix of 1024 points by 192 points was collected with 56 transients per point. Processing was carried out using AZARA with sinbell squared window functions shifted by 80° in the direct dimension and 90° in the indirect dimension. After zerofilling and Fourier transformation the resulting 1024 by 256 data matrices were loaded into ANSIG for analysis. The T_2 raw data consisting of 1024 by 192 points (56 transients) were processed in the same way (time points marked * were repeated for purposes of error estimation):

T_1 **ms:** 8.625, 193.190*, 393.805, 795.035, 1196.265

T_2 **ms:** 8.528, 17.056*, 42.640, 76.752, 119.392

$\{^1H\}$-^{15}N NOE data consisting of unsaturated and saturated spectra measuring 3600 by 192 points (48 transients) were also collected and processed using AZARA to with sinbell squared window functions shifted by 80° in the direct dimension and 90° in the indirect dimension. After zerofilling and Fourier transformation the resulting 4096 by 256 data matrices were loaded into ANSIG for analysis.

In all cases, assignments were copied from an assigned ^{15}N -HSQC spectrum in ANSIG and integration was performed. Repeated time points in the T_1 and T_2 spectra allowed

estimation of the error in measured intensity. In the case of the $\{^1H\}$-^{15}N NOE spectra, the root mean square of intensity of 20 peaks at random positions in the noise region of the spectrum was instead used as a measure of the error. For a description of the fitting procedure used see section 2.3.1. The resulting values of T_1, T_2 and $\{^1H\}$-^{15}N NOE are shown in Figures 5.7 and 5.8 (section 5.5).

5.2.3 Alignment using the CPBr/NaBr/hexanol

A simple chemical system consisting of a mixture of cetylpyridinium bromide (CPBr), NaBr and hexanol [7] was decided upon for this work because of the advantages of inexpensiveness and robustness with respect to pH outlined in section 2.4.

Before attempting to prepare an aligned sample using the CPBr/NaBr/hexanol system, the compatibility of the system with the conditions used for previous NMR work was tested: 840 μL of 6.5% w/v CPBr/hexanol (w/w ratio 1/1.33) was prepared ([NaBr] = 25 mM) by mixing the components, vortexing, heating to 70°C and centrifugation for 2 minutes at 4000 g resulting in a single, homogeneous LLC phase. 260 μL of 100 mM pH 5 NaOAc buffer ([NaBr] = 25 mM) was then added in five steps of approximately 50 μL interspersed with vortexing in order to prevent disruption of the LLC phase. The resulting liquid crystalline phase was $\sim$ 5% w/v in CPBr/hexanol, 25 mM NaBr and $\sim$ 25 mM NaOAc buffer. A broad proton spectrum and residual quadrupolar splitting of the D_2O signal of 12.5 Hz (symmetrical doublet, linewidth at half height of 2.7 Hz) verified that the CPBr/hexanol system was compatible with the original DAF~2,3 sample conditions.

An aligned sample of DAF~2,3 was prepared using ^{15}N -labelled sample with a protein concentration of $\sim$ 0.9 mM: as a first step, 140 μL ^{15}N -labelled DAF~2,3 was added to 420 μL of 6.5% w/v CPBr/hexanol to give a CPBr/hexanol concentration of 5% and a protein concentration of 0.15 mM. The residual quadrupolar splitting of the water signal observed for this sample was $\sim$ 4 Hz but the broad proton spectrum and correspondingly poor ^{15}N HSQC indicated that alignment was too strong to be able to measure $^1D_{NH}$ for individual residues. At this stage, an additional 60 μL of DAF~2,3 solution was added. After vortexing and centrifugation, the resulting LLC

phase with a protein concentration of 0.2 mM and 4% w/v CPBr/hexanol gave a residual quadrupolar splitting of 2.5 Hz and ^{15}N HSQC spectra appeared to be of sufficient quality to enable $^1D_{\mathrm{NH}}$ determination.

5.2.4 Measurement of one bond N-H$^{\mathrm{N}}$ residual dipolar couplings

The IPAP-HSQC method of Ottiger, Delaglio and Bax [77] was used to measure $^1D_{\mathrm{NH}}$ RDCs. Two aligned datasets were acquired for the purposes of error evaluation: one consisting of 1280 by 200 complex points (^{15}N acquisition time 45.5 ms) with 352 transients per point and the other of 1280 by 256 points (^{15}N acquisition time 58.2 ms) with 320 transients per point. The IP and AP planes were processed with AZARA (Wayne Boucher, University of Cambridge): after zerofilling and application of a sinbell squared window function in the directly detected dimension, processing in the indirect dimension was performed with the maximum entropy routine. In each case the resulting spectra measured 2048 by 4096 real points. Prior to addition / subtraction of the IP and AP spectra, the AP spectrum was scaled up by a factor of 1.1 to compensate for the signal lost due to relaxation.

$^1D_{\mathrm{NH}}$ RDCs were first determined using peak frequencies obtained by manually picking the peak centres in ANSIG [55] and copying assignments from an assigned ^{15}N -HSQC spectrum. The largest contribution to the error was assumed to be from the poorer quality (in terms of signal-to-noise) aligned dataset: the pairwise RMSD between values of $^1J_{\mathrm{NH}} + ^1D_{\mathrm{NH}}$ measured from the two aligned datasets, 2.3 Hz, was therefore used as an estimate for the error in the $^1D_{\mathrm{NH}}$ RDCs.

For the second attempted refinement of DAF~2,3 structures, the $^1D_{\mathrm{NH}}$ RDCs were re-measured using the parabolic fit routine of CCPN Analysis to determine the peak centres. Estimates for the errors in the splittings were obtained using observed S/N ratios, acquisition times and the average value of R_2 according to the work of Kontaxis *et al.* [53]. Signal-to-noise ratios for the unaligned and aligned spectra were approximately 20:1 and 10:1, respectively, and the average transverse relaxation rate $R_2 = 11.7s^{-1}$ for unaligned DAF~2,3. For the aligned sample, the R_2 relaxation rates were faster: $\sim 20s^{-1}$. Using the above parameters, the uncertainty in the position

of the downfield peak in these aligned spectra would be expected to be 0.6 Hz and 0.7 Hz. Examining the doublets prior to addition/subtraction it is apparent that the upfield and downfield ^{15}N-{^{1}H^{N}} doublet components differ in intensity and linewidth by a factor of approximately 1.5:1 - i.e. the upfield component is 1.5 times weaker and broader than the downfield component. If the uncertainty in the position of the downfield component is U then the resulting uncertainty in the splitting between the two components is $\sqrt{1+(1.5^2)^2}U$ since the uncertainty in the upfield component is $(1.5^2)U$. The uncertainty in measured $^{1}J_{\mathrm{NH}}+^{1}D_{\mathrm{NH}}$ splitting values would therefore be expected to be approximately 1.5 Hz for the aligned dataset with the longer t_1 acquisition time and 1.7 Hz in the other case: combining the two measurements yields an average value of $^{1}J_{\mathrm{NH}}+^{1}D_{\mathrm{NH}}$ with an error of 1.1 Hz. The pairwise RMSD between $^{1}J_{\mathrm{NH}}+^{1}D_{\mathrm{NH}}$ splitting values measured for the two aligned datasets should be 2.3 Hz, which was in fact observed. In the same way the uncertainty in the position of the downfield peaks in the unaligned IPAP spectra is 0.3 Hz giving an estimated error in $^{1}J_{\mathrm{NH}}$ of 0.8 Hz. The final error in the value of $^{1}D_{\mathrm{NH}}$ obtained using the two aligned datasets and the single unaligned dataset is therefore 1.3 Hz.

5.3 RDC refinement of the original NMR structures of DAF~2,3

Structures generated using $\sim$ 2400 NOE distance restraints ($\sim$ 1200 per module) were provided by Stanislava Uhrinova and refinement with RDC restraints carried out with the assistance of Krystyna Bromek [112]: $^{1}D_{\mathrm{NH}}$ RDCs measured as described above were added to the refinement step of the structure calculation protocol using the TENSO module of the program Crystallography & NMR System (CNS, [17]) which contains an implementation of the Implicit Saupe tensor Alignment Constraint (ISAC) method [94, 70]. The tenso force constant was increased geometrically from 0.001 kcal·(mol·Hz^2)$^{-1}$ to 0.6 kcal·(mol·Hz^2)$^{-1}$ during the first cooling stage of refinement, and fixed at the final value throughout the second cooling stage.

5.3.1 Can a single alignment tensor be used for both modules?

The structures generated using NOE restraints alone showed considerable variation in intermodular orientation, which could be indicative of intermodular motions in solution or simply reflect the lack of NOE data defining the intermodular orientation. While it has been shown that significant flexibility in the linker between a module pair can produce different diffusion characteristics and alignment tensors [34], others have noted that RDCs measured in multidomain proteins are relatively insensitive to small amplitude interdomain motions [100, 107]: therefore, a set of structures was calculated in which the whole molecule was assumed to have a single alignment tensor (set '1AT', 59 $^1D_{\mathrm{NH}}$ RDCs) in addition to a set of structures in which each module was assigned a unique alignment tensor (set '2AT' - 28 $^1D_{\mathrm{NH}}$ RDCs for module 2, 31 $^1D_{\mathrm{NH}}$ RDCs for module 3). See Tables 5.1 and 5.2 for a summary of the CNS energy terms and backbone RMSD values for the converged structures - for comparison, the statistics for a set of structures without RDC restraints (set 'NOE') are also shown.

ensemble	Etot	Enoe	Etenso
'1AT'	1261.0	161.8	24.7
'2AT'	1213.0	154.3	11.8
'NOE'	1136.5	128.4	-

Table 5.1: **Average CNS energy terms for early DAF23 structures.**

Ensemble/No.	$\mathrm{RMSD}_{overall}$	RMSD_{m2}	RMSD_{m3}
'1AT'	1.57	0.63	0.61
'2AT'	2.44	0.67	0.57
'NOE'	2.37	0.66	0.58

Table 5.2: **RMSD values for early DAF23 structures.** The structures were overlaid on the backbone N, C and CA atoms of the following residues: 64-75 and 86-124 for module 1 (i.e. excluding the hypervariable loop), and 129-140, 147-156, 160-170 and 178-186 for module 2 (excluding the hypervariable loop, the β-bulge between strands E and F and the loop between strands F and G). All RMSD values are given in angstroms (Å).

There is a significant increase in energy associated with the use of a single alignment tensor for the whole molecule, from 11.8 to 24.7 kcal·mol^{-1}. However, the '1AT' struc-

ture set shows a marked improvement in the definition of the intermodular orientation (backbone RMSD over both modules is 1.57Å for the '1AT' structure set compared to 2.44Å for the '2AT' structure set). A slight improvement in the convergence of the module 2 structure was observed although the convergence of module 3 is somewhat worse (see Table 5.2). Compared with the NOE only structures, the set '2AT' structures showed little improvement in convergence and significant increases in NOE and overall energy.

The spread of alignment tensors in each ensemble was calculated by Singular Value Decomposition (SVD - see section 2.4) using the calculated structures and experimental dipolar coupling values. The program Orderten-SVD [63] was used for this purpose. The following ensemble alignment tensor distributions were calculated, using the coordinate system of the lowest energy structure from the 'ALL' set as a frame of reference:

- Ensemble '1AT' overlaid on both modules: A-tensor calculated using all RDCs.
- Ensemble '2AT' overlaid on module 2: A-tensor calculated using module 2 RDCs.
- Ensemble '2AT' overlaid on module 3: A-tensor calculated using module 3 RDCs.

The resulting ensemble alignment tensor eigenvector distributions are shown in Figure 5.3. The eigenvalues of the solutions are summarised in Table 5.3.

Set	Szz	Syy	Sxx	η
'1AT'	7.3 (0.6)	-6.7 (0.6)	-0.5 (0.2)	0.85 (0.06)
'2AT(m2)'	9.2 (1.7)	-8.5 (1.6)	-0.7 (0.4)	0.84 (0.08)
'2AT(m3)'	6.8 (1.3)	-6.0 (1.0)	-0.8 (1.1)	0.78 (0.21)

Table 5.3: **Alignment tensor eigenvalues for the first DAF~2,3 structures.** Values shown $\times 10^{-4}$ are average values followed by the standard deviation in parentheses, where m2 denotes module 2 and m3 denotes module 3.

It can be seen that the three alignment tensors are not well defined by the 59 NH RDCs [1]: the eigenvectors of the alignment tensors for the individual modules are not exactly

[1] The orientational space sampled by the couplings is smaller than would be sampled by an isotropic

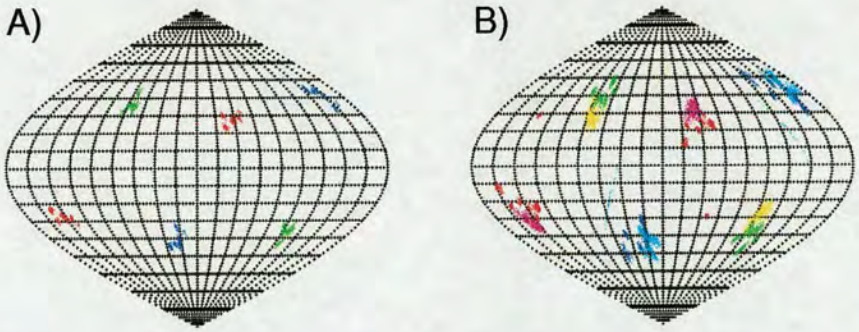

Figure 5.3: **Alignment tensor distributions for DAF23.** Sanson-Flamsteed plots showing the orientation of the alignment tensor eigenvectors in coordinate systems coincident with that of the lowest energy set '1AT' structure. **A)** The eigenvectors for the '1AT' ensemble using all RDCs. Colours for the eigenvectors corresponding to the largest positive, largest negative and smallest eigenvalues are red, blue and green, respectively. **B)** Eigenvectors for the module 2 and 3 alignment tensors of the '2AT' ensemble: colours for the eigenvectors corresponding to the largest positive, largest negative and smallest eigenvalues are red, blue and green, respectively for module 2; magenta, cyan and yellow, respectively for module 3.

coincident with those for the '1AT' structures, but it is not possible to tell whether this is due to the small $^1D_{\mathrm{NH}}$ RDC dataset or a genuine difference in the alignment of the two modules. Likewise, the eigenvalues for modules 2 and 3 are substantially different but still agree within one standard deviation due to the large spread of values calculated (up to 20% of the average in some cases). The increase in the 'tenso' energy term associated with the use of a single alignment tensor for the whole molecule does however appear to be significant (~50%), suggesting that the two modules are aligned independently and the use of a single alignment tensor is not justified in the case of DAF~2,3.

5.4 RDC refinement of the current NMR structures of DAF~2,3

The current NMR structures using NOE restraints show RMSD values for individual modules of 0.81Å and 0.75Å, respectively and are therefore of sufficient quality for

distribution of vectors since few NH bonds are close to colinear with the long axis of the molecule due to the molecule's β-strand composition.

refinement using RDCs. These new structures include several intermodular NOEs that were not assigned in the original data analysis, resulting in better definition of the intermodular orientation and therefore considerably lower RMSD values when both modules are overlaid.

RDCs measured for residues undergoing significant ps-ns time scale motion were excluded on the basis of $\{^1H\}$-^{15}N NOE values below 0.6 (measured at a ^1H frequency of 600 MHz):

ps-ns motion: 61, 63, 77, 90, 93, 104, 120, 121, 134, 173

In addition, the criterion of Tjandra *et al.* [105] was used to exclude residues indicating exchange on a ms-μs time scale:

$$\frac{(\langle T_2 \rangle - T_2)}{T_2} - \frac{(\langle T_1 \rangle - T_1)}{T_1} > 1.5 \times \text{S.D.} \tag{5.1}$$

where S.D. is the standard deviation of the left-hand side of equation 5.1. This gives the following additional excluded residues:

R_{ex} **term (Tjandra** *et al.*)**:** 61, 85, 93, 173

As before, RDC restraints were added to the structure calculation protocol using the TENSO module of the program Crystallography & NMR System (CNS, [17]). The tenso force constant was increased geometrically from 0.001 kcal·(mol·Hz^2)$^{-1}$ to 1 kcal·(mol·Hz^2)$^{-1}$ during the first cooling stage of refinement, and fixed at the final value throughout the second cooling stage. The modification to the refinement protocol described in section 4.3 (a reduced temperature step in the first cooling stage) was used here.

The validity of using a single alignment tensor for the whole construct was checked; this time calculating the following structure sets: a set including all RDCs (set 'ALL'), RDCs for module 2 only (set 'M2'), or for module 3 only (set 'M3'). Statistics for the best structures from each set are shown in Tables 5.4 and 5.5 below:

ensemble	RDCs	Etot	Enoe	Etenso
'ALL'	62	891.7	75.7	2.3
'M2'	29	852.9	67.0	1.2
'M3'	30	851.3	66.8	0.9
'NOE'	-	850.2	64.0	-

Table 5.4: **Average CNS energy terms** for the latest DAF23 structures with and without RDCs.

Ensemble/No.	$RMSD_{overall}$	$RMSD_{m2}$	$RMSD_{m3}$
'ALL'/13	1.27	0.72	0.71
'M2'/17	1.46	0.77	0.70
'M3'/21	1.45	0.76	0.73
'NOE'/13	1.45	0.81	0.75

Table 5.5: **RMSD values for the latest DAF23 structures.** The structures were overlaid on the backbone N, C and CA atoms of the following residues: 64-75 and 86-124 for module 1, and 129-140, 147-156, 160-170 and 178-186 for module 2. All RMSD values are given in angstroms (Å), and the number of structures in each ensemble is shown in the first column.

The use of a single alignment tensor for both modules is accompanied by a small increase in the energy terms (Etenso, Enoe and Etot) versus the energy terms for set 'M2', 'M3' and 'NOE' structures: the average 'tenso' energy term for the set 'ALL' structures is ~10% greater than the sum of the terms for the set 'M2' and 'M3' sets. In contrast, the set 'M2' and 'M3' structures show little or no increase in Enoe and Etot energies when compared with the set 'NOE' structures. Inclusion of the RDC restraints leads in all cases to improvement in the pairwise RMSD for the area of structure for which RDCs have been used, with the largest improvement in precision seen in the set 'ALL' structures. It should also be noted that the NOE only structures show a much lower RMSD when overlaid on both modules (1.45Å) than was seen in the original NMR structures (2.37Å), which is entirely due to the extra intermodular NOEs that have now been assigned.

The spread of alignment tensors in each ensemble was calculated by Singular Value Decomposition (SVD - see section 2.4) using the calculated structures and experimental dipolar coupling values with the program Orderten-SVD [63]. Using the coordinate

system of the lowest energy structure from the set 'ALL' as a reference, 100 SVD computations were carried out for each converged structure of the set 'ALL', 'M2' and 'M3' ensembles, producing the alignment tensor distributions shown in Figure 5.4 with eigenvalues as summarized in Table 5.6.

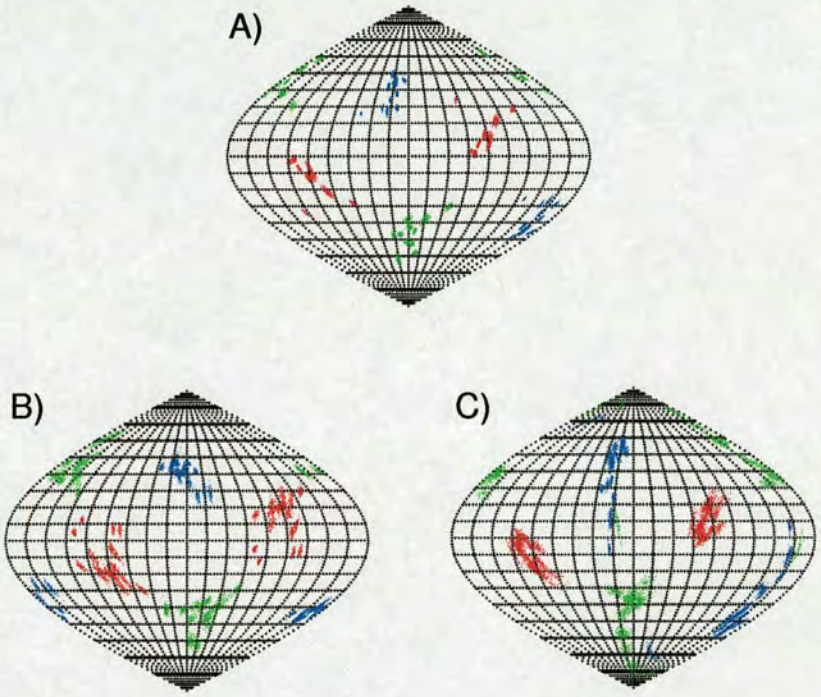

Figure 5.4: **Alignment tensor distributions for the latest DAF23 structures.** Sanson-Flamsteed plots showing the orientation of the alignment tensor eigenvectors in coordinate systems coincident with that of the lowest energy set 'ALL' structure. **A)** The eigenvectors for the 'ALL' ensemble using RDCs for both modules. Colours for the eigenvectors corresponding to the largest positive, largest negative and smallest eigenvalues are red, blue and green, respectively. **B)** Eigenvectors for the 'M2' ensemble using module 2 RDCs: colours for the eigenvectors corresponding to the largest positive, largest negative and smallest eigenvalues are red, blue and green, respectively. **C)** Eigenvectors for the 'M3' ensemble using module 3 RDCs: colours for the eigenvectors corresponding to the largest positive, largest negative and smallest eigenvalues are red, blue and green, respectively.

Again it would appear that the three alignment tensors are not well defined by the data and the spread of eigenvalues is slightly larger than was observed in the original structures (Table 5.3). The eigenvalues for the set 'ALL', 'M2' and 'M3' structures are distributed about different averages, but are within a single standard deviation due to

Set	Szz	Syy	Sxx	η
'ALL'	-6.9 (1.4)	6.7 (1.7)	0.3 (0.8)	0.79 (0.10)
'M2'	-7.3 (1.3)	7.1 (1.5)	0.2 (1.2)	0.75 (0.14)
'M3'	-5.6 (1.0)	5.4 (1.7)	0.2 (1.1)	0.70 (0.17)

Table 5.6: **Alignment tensor eigenvalue magnitudes** for 'ALL', 'M2' and 'M3' structure ensembles. Values shown $\times 10^{-4}$ are average values followed by the standard deviation in parentheses.

the spread of solutions. The eigenvectors plotted in Figure 5.4 are also scattered over a range of orientations. It is uncertain from these data whether both modules experience the same alignment and one can therefore conclude that a single set of NH RDCs is insufficient for the analysis of the intermodular orientation in CCP module pairs.

The program MODULE [30] has been used to back-calculate the NH RDCs using the x-ray crystal structure; both assuming the whole molecule has a single alignment tensor and assuming separate alignment tensors for each module - this is shown in Figure 5.5.

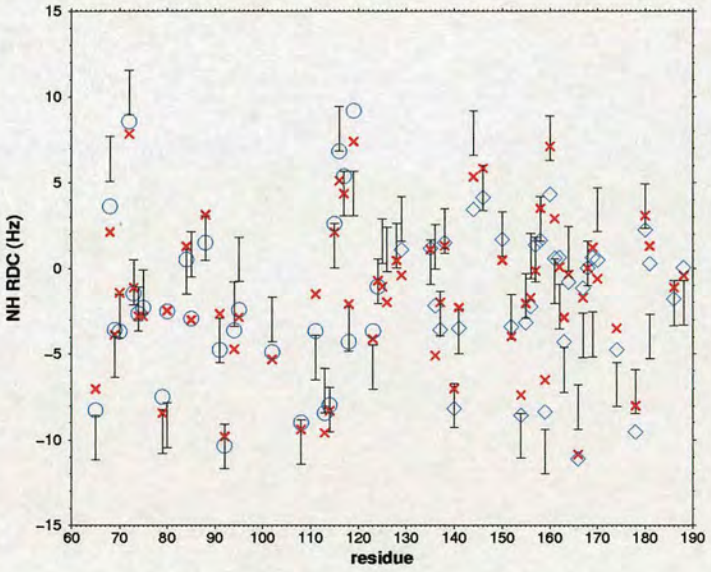

Figure 5.5: **Observed and back-calculated NH RDCs using the DAF~2,3 x-ray crystal structure.** Measured $^1D_{\mathrm{NH}}$ values and the estimated experimental error are represented by black error bars. The back-calculated values obtained assuming a single alignment tensor for the whole molecule are shown as red crosses and the back-calculated values obtained considering the two modules independently are shown in blue (circles for module 2, diamonds for module 3).

It can be seen that there is reasonable agreement between the measured and back-calculated values: the pairwise RMSD between the measured and back-calculated values in the case of a single alignment tensor for both modules is 2.5 Hz [2]. The agreement is only slightly better if the the back-calculation is performed for the two modules independently (2.2 Hz). This suggests that to a good approximation the two modules can be considered to have the same alignment tensor.

The intermodular orientations for the NOE only, NOE+RDC (set 'ALL') and x-ray structures have been summarized for comparison in a plot of the tilt, twist and skew angles made between the two modules (see Figure 5.6). The average orientation produced by using RDC restraints agrees less well with the orientation found in the crystal structure than does the average orientation obtained using NOEs only - in particular, a larger average tilt angle is observed in the structures refined with RDCs. This difference in the intermodular orientation could be due to a genuine difference in the average solution conformation (more bent than the x-ray crystal structure), but could also be due to 'structural noise' / poor definition of the alignment tensor.

[2] The pairwise RMSD between the measured and back-calculated values in the case of the lowest energy set 'ALL' structure calculated using RDCs is 0.5 Hz.

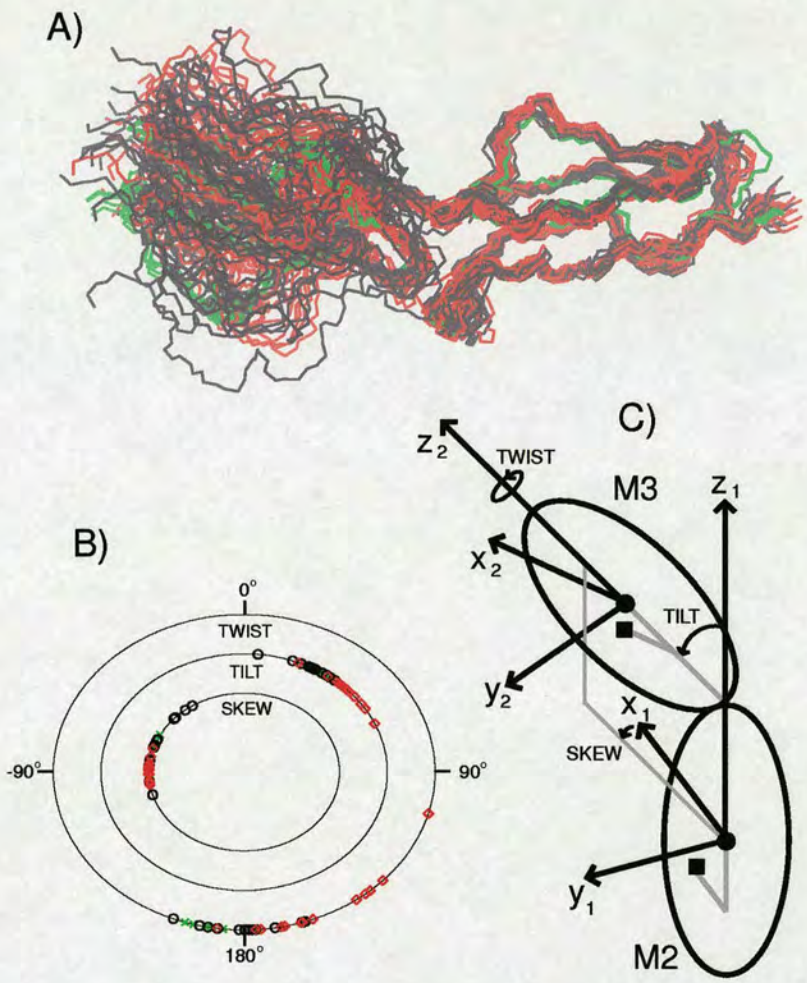

Figure 5.6: **DAF23 intermodular orientation. A)** Comparison of structures using set 'ALL' RDCs (best 13, in red) with NOE only structures (best 13, in black) and x-ray structures (in green). All structures have been overlaid on module 3. **B)** Tilt, twist and skew angles for the converged structures of the NOE only (black), set 'ALL' RDC (red) and x-ray (green) structure ensembles. The average angles for the ensembles are for the NOE set: skew=-70°(25), tilt=28°(9) and twist=179°(13), for the set 'ALL' RDC set: skew=-88°(7), tilt=46°(11) and twist=156°(22) and for the x-ray set: skew=-69°(4), tilt=33°(2) and twist=-168°(4) where the values in parentheses are standard deviations. **C)** Diagram of tilt, twist and skew angle definitions [6]: the centre of mass of each module (represented by a filled circle) defines the origin of the frame for the module and the z-axis points in the direction of the C-terminus aligned with the principal component of the inertia tensor. The direction of the x-axis is defined as being perpendicular to the z-axis and in the direction of the CA atom of the conserved tryptophan residue (W117 for module 1 and W170 for module 2) and is shown as a filled square. The limits of modules 1 and 2 are defined as being C64-C124 and C129-C186, respectively.

5.5 Analysis of relaxation data

Plots of T_1 and T_2, the T_1/T_2 ratio and $\{^1H\}$-^{15}N NOE are shown in Figures 5.7 and 5.8, respectively.

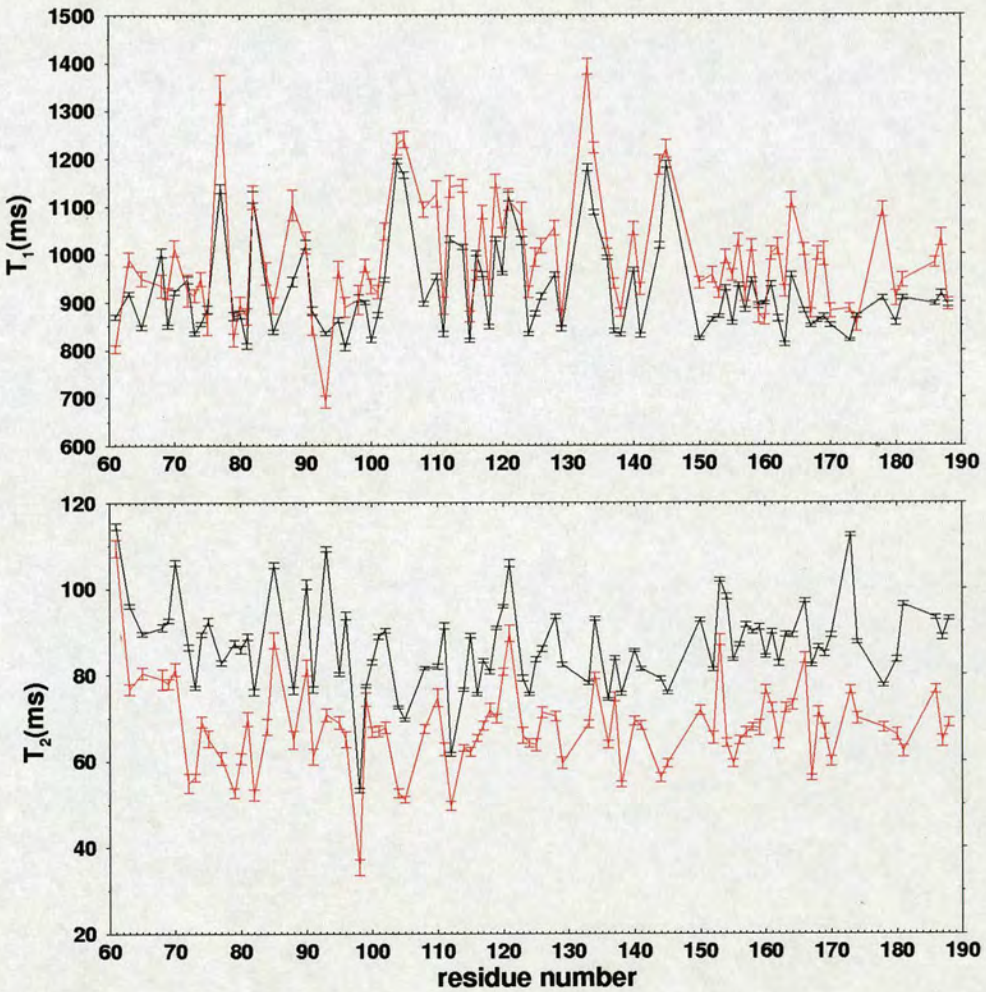

Figure 5.7: **T_1 and T_2 relaxation times for DAF~2,3.** Relaxation data at 600 MHz are shown in black, data at 800 MHz in red.

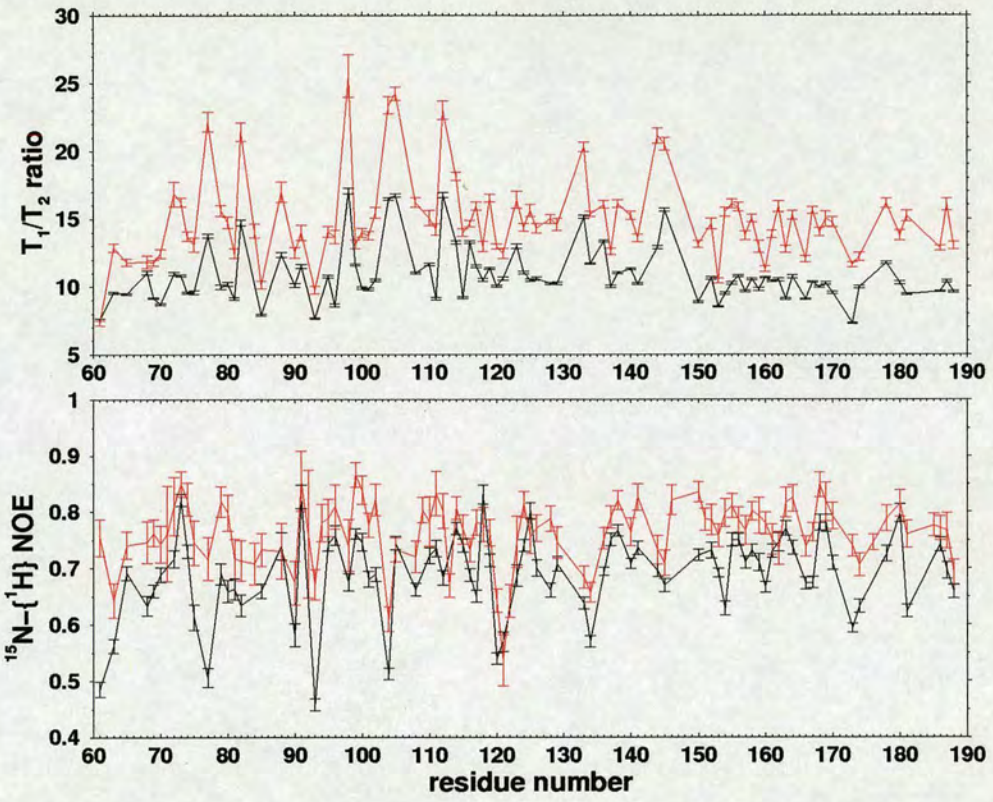

Figure 5.8: T_1/T_2 **ratio and** $\{^1H\}$-^{15}N NOE **for DAF~2,3.** Relaxation data at 600 MHz are shown in black, data at 800 MHz in red.

Estimates for the correlation time of the molecule and of its two modules using the R_2/R_1 ratios at 600 MHz and 800 MHz, obtained as described in section 2.3, are:

τ_c, **600 MHz :** 9.8 ± 0.4 ns

τ_c, **800 MHz :** 8.8 ± 0.5 ns

A similar analysis for the two modules separately gives:

τ_c, **M2, 600 MHz :** 9.9 ± 0.5 ns

τ_c, **M2, 800 MHz :** 8.8 ± 0.6 ns

τ_c, **M3, 600 MHz :** 9.7 ± 0.4 ns

τ_c, **M3, 800 MHz :** 8.8 ± 0.4 ns

Whilst there is a clear qualitative agreement between the trends in the 600 MHz and 800 MHz relaxation data (Figures 5.7 and 5.8), the values of the correlation time obtained from the 600 MHz and 800 MHz data do not quite agree within one standard deviation, suggesting the possibility of systematic errors in one or both sets. Unfortunately, it was not possible to repeat the relaxation experiments because the sample was used to produce an aligned sample for measurement of $^1D_{\mathrm{NH}}$ RDCs.

The program Tensor [31] was used to analyze the relaxation data and due to the systematic difference in implied correlation time, the 600 MHz and 800 MHz data were considered separately (with both the x-ray and new NMR solution structures) to check the agreement between the two datasets. Only in the case of the 600 MHz data set and using the x-ray crystal structure was an intuitively reasonable axially symmetric model for diffusion obtained: 1st minimum $D_{||}/D_{\perp} = 1.91$, D_z close to coincident with I_z of the inertia tensor, as shown in Figure 5.9 (a correlation plot between the observed and predicted R_2/R_1 values according to the angle between the NH bond vector and principle axis of the diffusion tensor, α, is also shown). However, using this anisotropic model there was no statistically significant improvement to the fit of the relaxation data over the fit obtained using the isotropic model for diffusion.

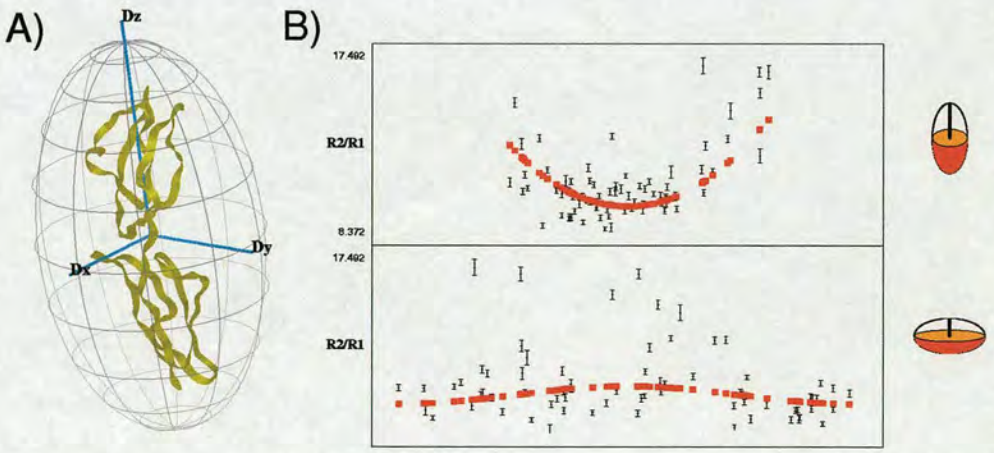

Figure 5.9: **Axially symmetric diffusion tensor for DAF~2,3. A)** Grid and axes illustrating the fitted diffusion tensor using the DAF~2,3 x-ray crystal structure and 600 MHz relaxation data. **B)** Plot of R_2/R_1 versus the angle α made between $D_{||}$ and the NH bond vectors.

The diffusion tensors for three structures were calculated using a shell model with the program HYDROPRO [27]: the results for the x-ray crystal structure and NMR structures with the smallest and largest tilt angles are summarized in Table 5.7.

Structure	$D_{\|\|}$	$D_{\perp}$	$D_{\|\|}/D_{\perp}$	τ_3 (ns)	tilt (°)
x-ray	2.605×10^7	8.086×10^6	3.22	8.3	35.5
NMR	2.345×10^7	7.300×10^6	3.21	9.2	22.4
NMR	2.142×10^7	8.810×10^6	2.43	9.7	65.9

Table 5.7: **Predicted diffusional properties of DAF~2,3.**

In all cases, D_{xx} and D_{yy} differ by $< 5\%$ and axially symmetric anisotropic diffusion is therefore assumed ($D_{||} = D_{zz}$ and $D_{\perp} = D_{xx} = D_{yy}$). There are a number of characteristic correlation times in the case of axially symmetric anisotropic diffusion, but in the case of NH bond vectors in CCP modules, the angle α is distributed around $\alpha = \pi$, i.e. perpendicular to $D_{||}$, where $\tau_3 = (2D_{\perp} + 4D_{||})$ is the dominant term contributing to the spectral density function. Even the largest tilt angle observed in the NMR structure ensemble does not explain the low $D_{||}/D_{\perp}$ value of 1.91 indicated by the relaxation data, which would imply a structure with a tilt angle of $> 65.9°$ - other possible explanations are the presence of intermodular motions, errors in the relaxation data or poor sampling of the diffusion tensor by the NH bond vectors.

Model-free analyses of the 600 MHz and 800 MHz relaxation data were conducted separately [3]. Since the correlation times for the two modules appear to be the same (there is no evidence of a difference in the average values) and the attempt to fit a diffusion tensor using the relaxation data yielded better results in the case of the crystal structure, the crystal structure of the whole two-module construct was used for this analysis assuming an isotropic model for diffusion - the results are shown in Figures 5.10 and 5.11

[3] Normally it would be advantageous to combine data at all fields for a given residue since models 4 and 5 require the fitting of 3 parameters.

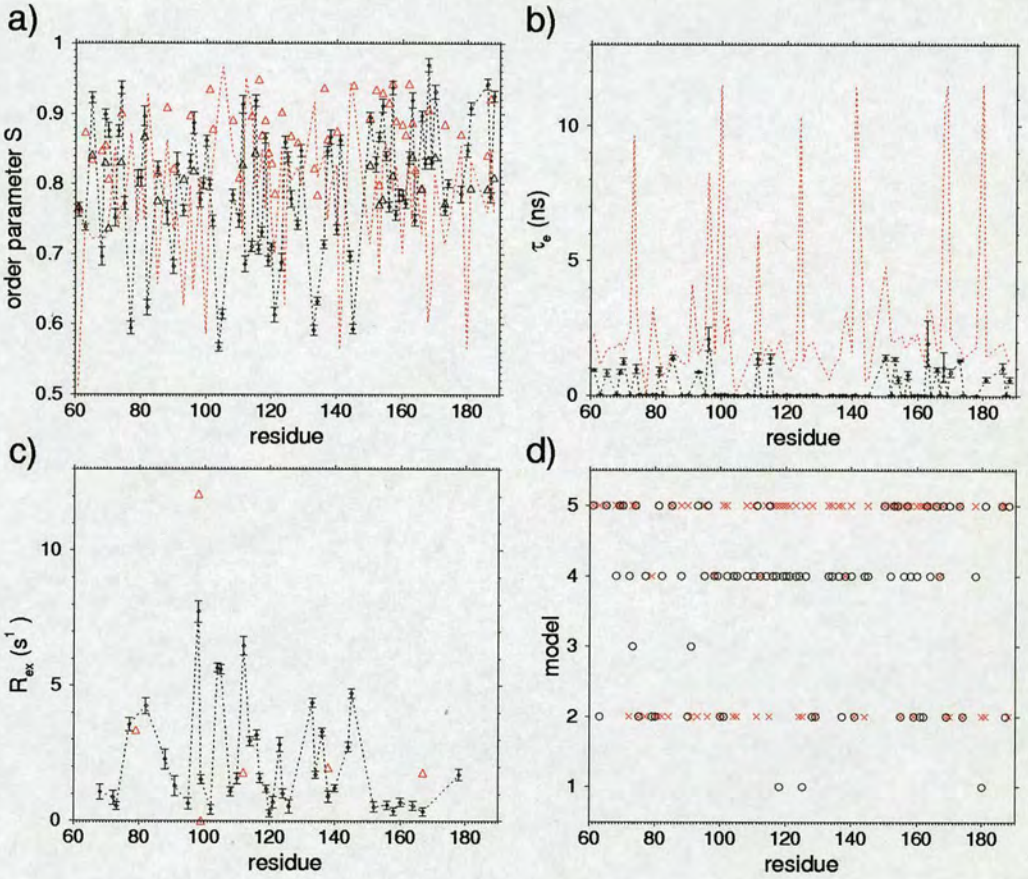

Figure 5.10: **DAF~2,3 model free results. A)** Order parameters: 600 black, 800 red, dotted lines are S or Ss and triangles are Sf where present (model 5). **B)** Effective correlation time on ns time scale (black 600, red 800). **C)** Exchange rates, Rex. **D)** Model number fitted to each residue.

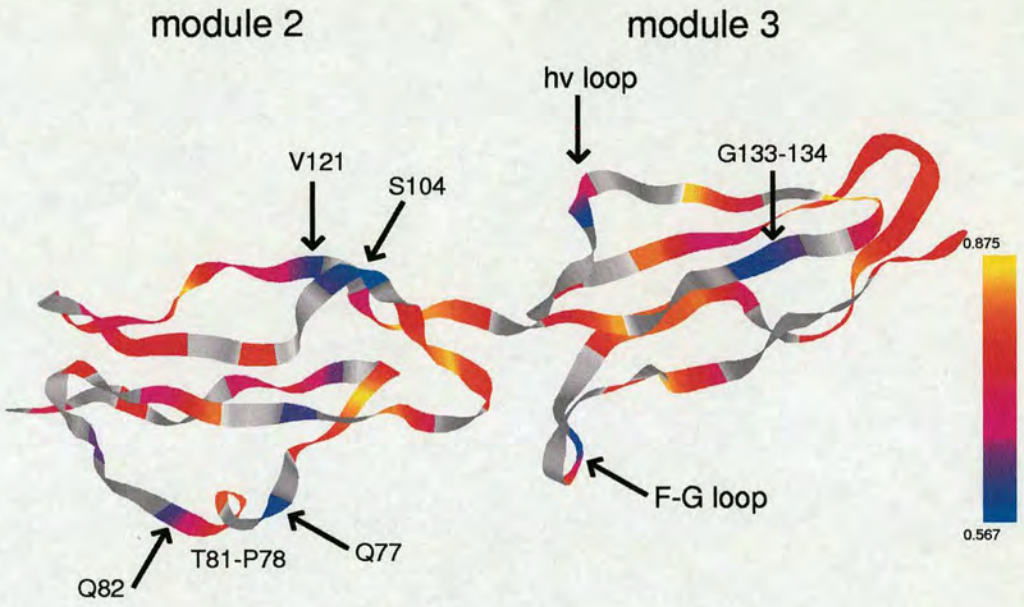

Figure 5.11: **DAF~2,3 mobility.** Plot of backbone S values for DAF~2,3 using 600 MHz data: module 2 is to the left, module 3 to the right.

The agreement between the model-free analyses using 600 MHz and 800 MHz data is rather poor - the models selected are different in most cases as are the parameters fitted, particularly in terms of time scale of the motions. However, it is still possible to interpret some of the results by checking for agreement between the two datasets and checking for supporting evidence in the $\{^1H\}$-^{15}N NOE and R_2/R_1 plots (i.e. reduced $\{^1H\}$-^{15}N NOE indicating ns time scale motion, anomalous R_2/R_1 indicating exchange). In module 2, low order parameters indicate ns time scale motion in the region of residues Q77 and Q82, which is part of the hypervariable loop for this module - it is interesting to note that the single turn of α-helix present in the crystal structure (P78-T81) exhibits higher order parameters, implying that it is indeed structured, but that the loop as a whole is flexible with respect to the rest of the module. There is evidence of ms time scale motion at residues Q77 and Q82, which may or may not be related to the fact that major (*trans* - in calculations this is the fixed isomer) and minor (*cis*) conformer populations of P78 are present. S104, located at the β-bulge between strands E and F, also shows evidence of ns and ms time scale motions, and the reduced $\{^1H\}$-^{15}N NOE and low order parameter at V121 indicate ns time scale

motion. There is no strong evidence of motion in the linker region (K125-S128). In module 3 regions G133-E134 (between strands A and B), G144-G145 (hypervariable loop) and S173-G174 (loop between strands F and G extending towards module 2) show evidence of motion on the ns time scale and G133 and G145 show evidence of exchange.

5.6 Discussion

Ill-defined intermodular orientation in the original solution structures did not lead to an increased number of violations upon inclusion of RDC restraints, suggesting that there is a large amount of flexibility in the structures calculated using NOE data to accommodate the small changes induced by including RDC restraints. A second set of solution structures has also been investigated in which the intermodular orientation is changed and again the single set of $^1D_{\mathrm{NH}}$ RDC restraints are accommodated. The orientation of the F-G loop in the old and new NMR structures is also different in spite of the RDC data: in the old NMR structures, the orientation of the F-G loop is markedly different from that found in the x-ray structures (Figure 5.1) whereas the orientation of the F-G loop in the new NMR structures and x-ray structures is similar (Figure 5.6). This is most likely due to the fact that only a single NH RDC was obtained for the F-G loop (G174). The single set of $^1D_{\mathrm{NH}}$ RDCs measured for DAF~2,3 appear to be insufficient to adequately specify alignment tensors for the individual modules or for the whole two-module construct. The fact that the intermodular orientation produced by the use of a single alignment tensor for both modules does not lead to an average orientation that is coincident with the orientation observed in the crystal structure could be due to one of three possibilities [100, 28]: (1) the data are not adequate to define a unique alignment tensor, or (2) the two modules have separate alignment tensors and the orientation obtained by assuming they are the same is artefactual [28], or (3) the average intermodular orientation in solution does not correspond to the orientation in the crystal [4]. Certainly, the alignment tensors *are* poorly defined by the small number of couplings measured, and the average intermodular orientation obtained using RDCs

[4] Crystallization conditions were 0.2 M ammonium sulfate, 20% w/v monomethoxypolyethylene glycol 5000, 0.1 M sodium acetate (pH 4.6) and 10% w/v glycerol.

is different from that found in the crystal structure.

The poor fit to an anisotropic diffusion tensor obtained using the relaxation data imply that either there is intermodular motion (although there is no direct evidence for this on the ps-ns or ms-μs time scales) or that there are significant errors in the relaxation data (poor agreement between the data acquired at 600 MHz and 800 MHz strongly suggest this as a possibility). In another two CCP module construct, VCP∼2,3 [41], the correlation times for the two modules are appreciably different, and the fact that this is not the case for DAF∼2,3 is an argument against a very flexible linker. Furthermore, reasonable agreement between the calculated and observed $^1D_{\mathrm{NH}}$ RDCs using the x-ray crystal structure for both modules implies that any intermodular motions are likely to be limited in extent, although smaller amplitude motions are still a possibility since RDCs relatively insensitive to such motions [100] (less so than relaxation data which scales according to S^2 rather than S).

Mutagenesis data for DAF∼2,3 [56, 115] identify R69, R96, R100, K127, F148 and L171 as important for both classical *and* alternative pathway regulatory activity, K126 and F169 for alternative pathway regulatory activity and E134 for classical pathway regulatory activity. It has additionally been noted that the mutations S165A and T166A *increase* classical pathway regulatory activity. These mutations have been mapped onto the crystal structure and NMR structure from the set '1AT' set with the closest-to-average intermodular orientation (Figure 5.12), and it can be seen that excepting the exact orientation of the sidechains, the residues involved occupy similar positions in both structures: most of the important residues appear to form either a positive band of charge lying at the C-terminal end of module 2 or a hydrophobic patch at the N-terminal end of module 3. Some of these residues are also be involved in stabilizing the intermodular junction (e.g. K127) as indicated by their involvement in intermodular NOEs - the loss-of-function mutations associated with these residues are further evidence for a reasonably rigid linker: the positioning of residues over a number of modules may be important for binding. Most of the residues known from mutagenesis studies to be involved in complement regulation appear to be located in regions of rigid structure, with the exception of E134.

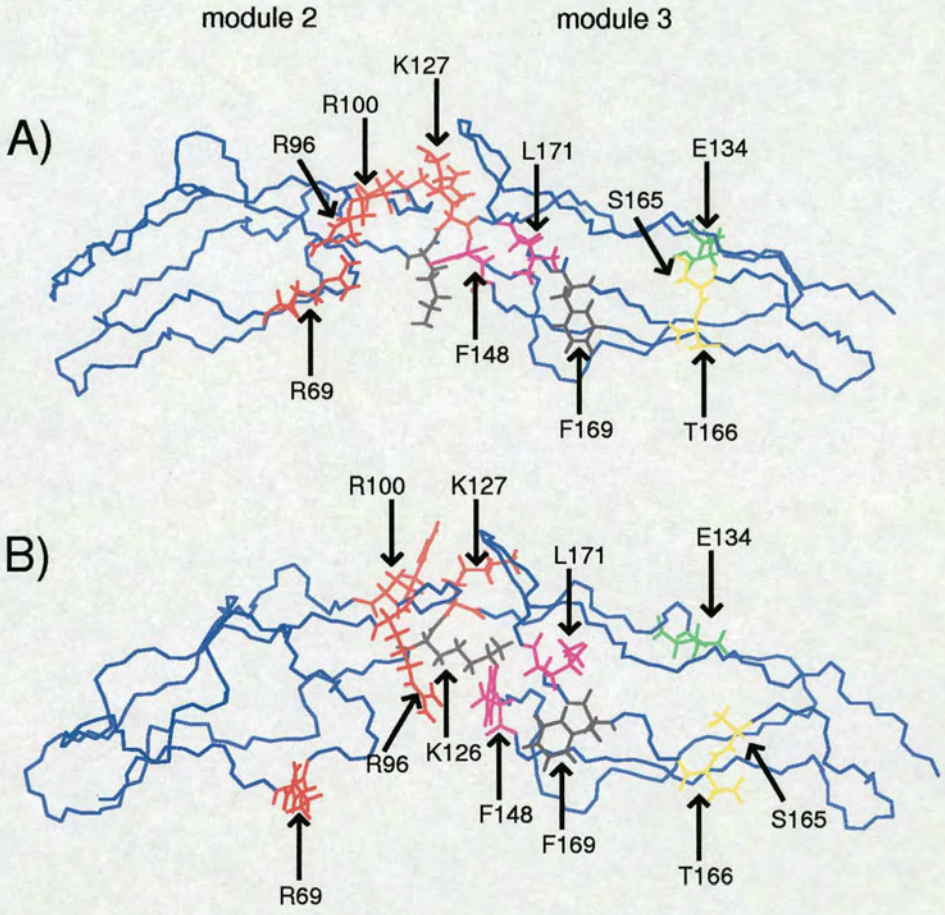

Figure 5.12: **DAF~2,3 mutations.** DAF~2,3 backbone for A) the x-ray crystal structure and B) the NMR structure with closest-to-average tilt angle - sidechains shown for residues of functional significance (mutagenesis studies): red = charged residues important for CP and AP regulatory activity, magenta = hydrophobic residues important for CP and AP regulatory activity, black = residues important for AP regulatory activity, green = residues important for CP regulatory activity and yellow = residues associated with an *increase* in CP regulatory activity upon mutation (CP = classical pathway, AP = alternative pathway).

Chapter 6

C4BP~1,2

6.1 Background

C4b-binding protein (C4BP) is a fluid phase regulator of complement activation (RCA - see section 2.6) and consists, in its predominant 570 kDa isoform, of seven identical α-chains and a single β-chain linked at a central core region [26]: the α-chains are composed of eight complement control protein (CCP) modules and the β chain is composed of three CCP modules. The 15 kDa C4BP~1,2 construct examined here consists of the N-terminal first and second module pair from the α-chain, which in addition to CCP 3, is the region thought to effect binding to C4b [12]. The amino acid sequence of the C4BP~1,2 construct is given below:

0 M

1 NCGPPPTLSF AAPMDITLTE TRFKTGTTLK YTCLPGYVRS

41 HSTQTLTCNS DGEWVYNTFC IYKRCRHPGE LRNGQVEIKT

81 DLSFGSQIEF SCSEGFFLIG STTSRCEVQD RGVGWSHPLP

121 QCEILEHHHH HH

The native sequence from N1 to I124 contains 11 positively charged and 13 negatively charged residues and the molecule has a predicted pI of 6.48 [37]. The positively charged cluster of residues in the linker region of the molecule is particularly interesting

since there is evidence of its importance to C4b binding [13]. There is evidence that a C4BP heparin binding site exists [98] and that there is overlap with the C4b binding site [42] since heparin binding abolishes C4b binding and vice versa.

In addition to its complement regulatory function which protects the host, C4BP has been shown to be recruited by a number of pathogens for the purpose of evading attack by the complement system: an example of a pathogen protein able to bind C4BP for this purpose in the CCP1-2 region is the M protein M4 (Arp4) of *S. Pyogenes* [2, 11].

6.2 Experimental

Labelled samples of C4BP~1,2 (both ^{15}N and ^{13}C,^{15}N labelled) were obtained from Linda Mark, Lund University, Sweden and assignment and initial structural studies were carried out by Huw Jenkins, University of Edinburgh. Buffers of varying pH and ionic strength were tested in order to determine the best conditions for NMR studies, and for the structural work the following conditions were chosen: 20 mM sodium acetate buffer (pH 4.5), 0.03% sodium azide. Note that the concentration of NaCl was 0 mM. It was also observed that good spectra, as judged by the appearance of the ^{15}N -HSQC, could be obtained over a range of pH values from 4 to 6.

6.2.1 Measurement of ^{15}N relaxation data

^{15}N relaxation data were recorded for a ^{15}N labelled sample of approximately 1.3 mM strength as judged by the A280 value recorded for the solution. Heteronuclear ^{15}N relaxation parameters were measured at ^{1}H frequencies of 600 MHz and 800 MHz using the 2D experiments described in section 2.3.1.

For the 600 MHz T_1 data, 2048 complex points in the directly detected proton dimension and 128 complex points in the indirect ^{15}N dimension were collected (16 scans per increment) giving t_2 and t_1 acquisition times of 122.1 ms and 34.2 ms, respectively. Processing was carried out using XWIN-NMR to zero-fill and apply a Gaussian window function to the data prior to Fourier transformation. After Fourier transformation, the real 2048 times 2048 spectrum was blocked using AZARA [14] in order to render it

suitable for analysis with ANSIG. It was found that better filtering of the water signal could be achieved by this method than by processing with AZARA alone. The 600 MHz T_2 raw data consisted of matrices of 2048 times 256 complex points acquired using 32 scans per increment, giving t_2 and t_1 acquisition times of 122.1 ms and 68.3 ms, respectively. This was processed with AZARA using sine bell squared window functions, shifted by 70° and 80° for the directly and indirectly detected dimensions, respectively, to obtain a real spectrum of 2048 times 512 points. The 600 MHz $\{^1H\}$-^{15}N NOE raw data consisted of data matrices of 2048 times 196 complex points (t_2 and t_1 acquisition times of 122.1 ms and 52.3 ms, respectively) acquired using 64 scans per increment. This was also processed with AZARA using sine bell squared window functions shifted by 70° and 90° for the directly and indirectly detected dimensions, respectively, to obtain a spectrum consisting of 2048 times 256 real points. The time points sampled in the T_1 and T_2 experiments at 600 MHz were as follows (time points marked * were repeated for purposes of error estimation):

T_1 **(ms):** 51.1*, 301.1, 501.1, 601.1, 801.1, 901.1, 1001.1

T_2 **(ms):** 16.96*, 50.88, 84.8, 118.72, 135.68

For the 800 MHz T_1 data, 2048 complex points in the directly detected proton dimension and 192 complex points in the indirect ^{15}N dimension were acquired (16 scans per increment) giving t_2 and t_1 acquisition times of 91.8 ms and 38.4 ms, respectively. Processing was carried out using XWIN-NMR to zero-fill and apply a Gaussian window function to the data prior to Fourier transformation. After Fourier transformation, the real 2048 times 1024 spectrum was blocked using AZARA. Similarly, 800 MHz T_2 data matrices consisting of 2048 times 192 complex points were collected (32 scans per increment) giving t_2 and t_1 acquisition times of 91.8 ms and 38.4 ms, respectively. This was processed using XWIN-NMR and then blocked with AZARA giving real spectra of 2048 times 1024 points. Finally, 800 MHz $\{^1H\}$-^{15}N NOE raw data consisting of data matrices of 2048 times 256 complex points was collected (64 scans per increment) giving t_2 and t_1 acquisition times of 91.8 ms and 51.2 ms. This was processed using AZARA with sine bell squared window functions, shifted by 80° and 90° for the di-

rectly and indirectly detected dimensions, respectively, to obtain spectra consisting of 2048 times 256 real points. The time points sampled in the T_1 and T_2 experiments at 800 MHz were as follows (time points marked * were repeated for purposes of error estimation):

T_1 (ms): 51.1*, 401.1, 701.1, 851.1, 1101.1, 12501.1, 1401.1

T_2 (ms): 16*, 48, 64, 80, 96, 112, 128

In all cases, assignments were copied from an assigned ^{15}N -HSQC spectrum in ANSIG and integration was performed. Repeated time points in the T_1 and T_2 spectra allowed estimation of the error in measured intensity. In the case of the $\{^1H\}$-^{15}N NOE spectra, the root mean square of intensity of 20 peaks at random positions in the noise region of the spectrum was instead used as a measure of the error. The procedure outlined in section 2.3.1 was followed for the calculation of relaxation parameters. The resulting values of T_1, T_2 and $\{^1H\}$-^{15}N NOE are shown in Figures 6.8 and 6.9 (section 6.4).

6.2.2 C12E5/hexanol as a potential aligning medium

Alignment of the DAF~2,3 module pair using the CPBr/hexanol/NaBr system, as described in section 5.2.3, did not result in an ideal sample. It was apparent from the asymmetry and linewidth of the residual quadrupolar splitting of the D_2O signal ($\sim$ 2 Hz) that the aligned phase was not very homogeneous. In addition, the splitting of the D_2O signal was only 2.5 Hz, which was obtained by using a concentration of aligning medium so low that the aligned phase was on the verge of breaking up. The stability of the phase was poor and crystallization of the CPBr component occurred after a period of some months. For these reasons, other alignment systems were investigated for use with the C4BP~1,2 construct.

The main advantage of simple chemical systems such as the CPBr/hexanol/NaBr system described in section 5.2.3 is that they are inexpensive and commercially available. Also, they tend to be relatively insensitive to pH whereas lipid bicelles undergo acid or base catalyzed hydrolysis and bacteriophage precipitate outside narrow pH limits

(e.g. Profos guarantee stability of their Pf1 phage from pH 6 to 8). An alternative to the CPBr/hexanol/NaBr system are mixtures of *n*-Alkyl-poly(ethylene glycol)/*n*-alkyl alcohol [92] which form an ordered, lamellar liquid crystalline phase. An additional advantage of these systems over the CPBr/hexanol/NaBr system is that they have been found to be much less sensitive to salt concentration.

The stability of the aligned phase depends on the ratio of glycol:alcohol components. The initial structural work on C4BP~1,2 was carried out at 37°C and the system that was reported to be stable at this temperature was a mixture of C12E5/hexanol with a molar ratio of C12E5 to hexanol, $r = 0.96$ [92]. It was reported that surfactant/water ratio of 5 wt % resulted in a stable aligned phase over a temperature range from 22-39°C giving a residual quadrupolar splitting of the D_2O signal of 25-30 Hz. The C12E5/hexanol system was therefore tested with the buffer previously used for structural work on C4BP~1,2 with the intention of verifying stability of the liquid crystalline phase and testing alignment of a ^{15}N -labelled sample.

Initial testing of the PEG/hexanol medium was performed at a concentration of 5.2 wt % with a D_2O concentration of 10% and PEG/hexanol molar ratios of $r = 0.7$ to $r = 1.5$. In recording the deuterium spectra of the samples over a range of temperatures, the following trends were observed: for $r = 0.7$, alignment was not achieved; for $r = 0.9$, a stable aligned phase existed between 22-30°C and showed a residual quadrupolar splitting of $\sim$ 15-20 Hz; for $r = 1.1$, a stable aligned phase existed between 24-32°C and showed a residual quadrupolar splitting of $\sim$ 12 Hz; for $r = 1.3$ a stable aligned phase existed between 20-36°C and showed a residual quadrupolar splitting of $\sim$ 6 Hz; for $r = 1.5$ a stable aligned phase existed between 24-38°C and showed a residual quadrupolar splitting of $\sim$ 10 Hz. Some of the deuterium spectra illustrating the trends are shown in Figure 6.1 (A, B and C), where it can also be seen that: (1) at the high-temperature limit of the region over which the medium is stable, greater homogeneity (i.e. narrower signals) are obtained and (2) the break-down of the aligned phase is accompanied by the appearance of a single peak for the unaligned phase located centrally within the doublet observed for the aligned phase.

A sample of 7 wt % PEG/hexanol medium was prepared for the purpose of aligning

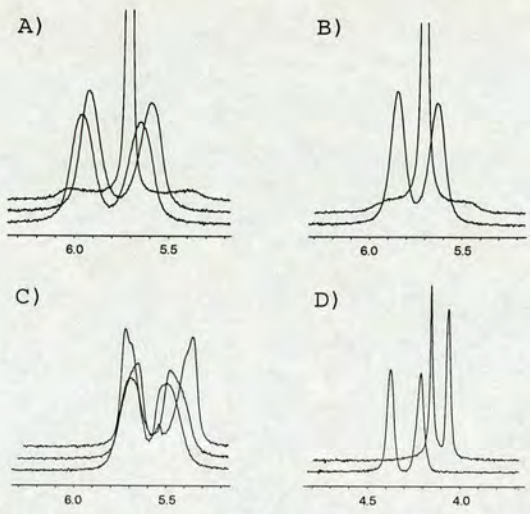

Figure 6.1: **^{2}H NMR spectra for the PEG/hexanol medium.** In spectra A-C, samples containing 20 mM sodium acetate buffer (pH 4.5) but no protein. All spectra were acquired at a field strength corresponding to a ^{1}H frequency of 600 MHz. **A)** Molar ratio r=0.9, 1D deuterium spectra are from bottom to top: at 28°C, at 30°C and at 32°C. Splittings at 28°C and 30°C were 17.0 Hz, 18.1 Hz, respectively. **B)** Molar ratio r=1.1, 1D deuterium spectra are from bottom to top: at 32°C and at 34°C. The splitting at 32°C was 12.1 Hz. **C)** Molar ratio r=1.5, 1D deuterium spectra are from bottom to top: at 34°C, at 36°C and at 38°C. Splittings at the three temperatures were 11.3 Hz, 9.7 Hz and 20.2 Hz, respectively. **D)** Deuterium spectra for ^{15}N -labelled C4BP~1,2 aligned sample using 4.5 wt % PEG/hexanol (r=1.5). The top spectrum (splitting 11.4 Hz) was acquired a week after the bottom spectrum (splitting 20.4 Hz), illustrating the instability of the aligned phase.

a ^{15}N -labelled sample of C4BP~1,2 - the molar PEG/hexanol ratio of $r = 1.5$ was used since this had been observed to have greater temperature stability. Alignment of the medium was verified by the presence of a 67 Hz residual quadrupolar splitting in the deuterium spectrum. Addition of the protein solution was subsequently carried out in 50 μL steps and the deuterium and ^{15}N HSQC spectra were monitored: adding 250 μL (resulting PEG/hexanol concentration 4.5 wt %) gave a D_2O splitting of 20.4 Hz but ^{15}N -HSQC spectra of a quality inadequate for NH RDC determination: alignment was too strong for this particular protein. The aligned phase was found to be unstable over the course of a week at 37°C (Figure 6.1 D) which may be due to hexanol separation and evaporation, as noted by others [57]. Further reduction of the PEG/hexanol concentration in an effort to reduce alignment to the point that RDCs could be measured led to increased temperature instability and break-up of the aligned

medium, ruling out its use for this protein.

6.2.3 Pf1 phage as a potential aligning medium

Pf1 bacteriophage can be used as an aligning medium [40] and is commercially available from a number of suppliers - for the work described below, 50 mg of protease-free Pf1 was purchased from Profos AG, Regensburg, Germany. The virus particles are filamentous (approximately 2000 by 6 nm) and negatively charged consisting of circular DNA surrounded by coat protein molecules. The partial alignment of protein molecules in solution by Pf1 phage is a result of both steric and electrostatic interactions [119]. The electrostatic component can cause problems if the protein being studied has significant positively charged patches on its surface since binding of the phage not only results in excessive alignment, but is accompanied by severe line broadening and in extreme cases causes aggregation. In addition, Pf1 virus particles may not be stable outside a pH range of 6 to 8.

In an effort to find suitable conditions for alignment of C4BP~1,2 using phage, a sample of 16.7 mg/mL Pf1 was prepared at pH 5.5 using a 20 mM acetate buffer. The residual quadrupolar splitting of the D_2O signal was 14.5 Hz with a linewidth at half height of 3.2 Hz and doublet components of equal height indicating that the phage forms an aligned liquid crystalline phase under these conditions. However, upon adding 100 μL of ^{15}N labelled C4BP~1,2 to the 300 μL of phage solution, precipitation was observed.

Aggregation most likely occurred due to the presence of exposed, positively charged residues on the surface of C4BP~1,2. To test this hypothesis, and in an effort to recover the protein, NaCl was added to the precipitate with the aim of increasing the ionic strength in order to break up charge-charge interactions and release the protein and phage back into solution. It was found that a NaCl concentration of 90 mM was sufficient to redissolve the precipitate but the resulting solution lacked homogeneity. A new 250 μL sample of Pf1 phage (16 mg/mL) in pH 5.5 acetate buffer was prepared, with the addition this time of 90 mM NaCl. Adding 50 μL of 1.5 mM ^{15}N C4BP~1,2 resulted in the formation of a short-lived precipitate that redissolved with mixing, giving a homogeneous sample with a residual D_2O splitting

of 12 Hz. Unfortunately the sample yielded an extremely poor ^{1}H spectrum due to the broadening caused by large proton-proton dipolar couplings. Further additions of 1.5 mM ^{15}N C4BP~1,2 and buffer were required to reduce the alignment to the degree that good quality ^{15}N -HSQC spectra were recovered (~ 7 mg/mL Pf1 phage). Having thus found suitable conditions for alignment, a double-labelled ^{13}C,^{15}N C4BP~1,2 sample (0.7 mM) was prepared with a final Pf1 phage concentration of 7.3 mg/mL, giving a residual quadrupolar splitting of the D_2O signal of 0.9 Hz.

6.2.4 Measurement of RDCs for a Pf1-aligned ^{13}C,^{15}N -labelled C4BP~1,2 sample

The following experiments were used to measure RDC restraints (all spectra processed with AZARA): N-H^N coupling constants were measured using the IPAP experiment of Ottiger, Delaglio and Bax [77]. A dataset was acquired for the unaligned state consisting of data matrices of 1024 $^{1}H^N$ times 256 ^{15}N complex points (16 scans per increment) giving t_2 and t_1 acquisition times of 56.9 ms and 50.1 ms, respectively. A dataset for the aligned state consisting of data matrices of 1024 times 512 complex points was also collected using 32 scans per increment, giving t_2 and t_1 acquisition times of 56.9 ms and 58.5 ms, respectively. Both the unaligned and aligned datasets were processed using the maximum entropy algorithm of AZARA to yield processed spectra of 1024 times 4096 real points. For the NH-IPAP data the antiphase spectrum was scaled by a factor of 1.1 prior to addition or subtraction to obtain the upfield and downfield doublet components - no such scaling was applied to the datasets for other coupling constants. N-C' coupling constants were measured using the spin-state selective method of Permi *et al.* [81]: a dataset for the unaligned state consisting of data matrices of 2048 times 480 complex points was collected (48 scans per increment) giving t_2 and t_1 acquisition times of 113.9 ms and 109.7 ms, respectively. A 70° shifted sine bell squared window function was applied to both dimensions prior to Fourier transformation with AZARA, giving real spectra of 2048 times 4096 points. A dataset for the the aligned state consisting of data matrices of 2048 times 384 complex points was acquired (80 scans per increment) giving t_2 and t_1 acquisition times of 113.9 ms and 102.4 ms, respectively. Fourier transformation was carried out using AZARA following

application of 70° shifted sine bell squared window functions in both dimensions, giving real spectra of 2048 times 4096 points. C^{α}-C' coupling constants were measured using the spin-state selective method of Permi *et al.* [83]: for the unaligned state dataset, data matrices of 2048 times 384 complex points were collected using 32 scans per increment (t_2 and t_1 acquisition times of 122.1 ms and 87.7 ms, respectively). 70° shifted sine bell squared and Gaussian window functions were applied to the direct and indirect dimensions, respectively and after processing real data matrices of 2048 times 4096 points were obtained; for the aligned state dataset, data matrices of 2048 times 338 complex points were acquired using 96 scans per increment (t_2 and t_1 acquisition times of 122.1 ms and 77.2 ms, respectively) and 80° shifted sine bell squared and Gaussian window functions were applied to the direct and indirect dimensions prior to processing with AZARA to obtain real data matrices of 2048 times 4096 points. Finally, C^{α}-H^{α} coupling constants were measured using the 2D (HNCO)-(J-CA)NH experiment described in section 3.5.1: data matrices of 1024 times 256 complex points were collected using 96 scans per increment for the unaligned state dataset, giving t_2 and t_1 acquisition times of 53.2 ms and 58.5 ms, respectively. These data were processed using the maximum entropy algorithm of AZARA to obtain real data matrices of 1024 times 4096 points; similarly the aligned state dataset consisting of data matrices of 2048 times 64 complex points was acquired using 1024 scans per increment (t_2 and t_1 acquisition times of 106.5 ms and 14.6 ms, respectively) and processed using the maximum entropy method to give real data matrices of 2048 times 4096 points.

Processed spectra were loaded into the program CCPN Analysis. After peak picking (using parabolic interpolation) the $^1J_{\mathrm{NH}}$ coupling constants were determined as the difference between the ^{15}N frequency of the high-field and low-field peaks of each peak pair in the ^{15}N dimension. The $^1D_{\mathrm{NH}}$ residual dipolar coupling constants (RDC) were then determined by subtraction: $^1D = {}^1(J + D)$ - ^1J. The other coupling constants were determined in the same way, and in the case of the $^1J_{\mathrm{C^{\alpha}H^{\alpha}}}$ coupling, scaling by a factor of 1.5 was applied to obtain the true value of the coupling (see section 3.5.1). Errors were estimated using the approach of Kontaxis *et al.* [53], explained previously in section 5.2.4 where the same approach has been used to estimate the errors in $^1D_{\mathrm{NH}}$ couplings for DAF~2,3. Using acquisition times and signal/noise ratios for the three

sets of couplings considered here, the following errors are estimated: $^1D_{NH}$ ±0.68 Hz, $^1D_{NC'}$ ±0.77 Hz, $^1D_{C^\alpha C'}$ ±1.38 Hz and $^1D_{C^\alpha H^\alpha}$ (unscaled) ±4.15 Hz. Before using the measured couplings as restraints, scaling relative to the $^1D_{NH}$ coupling by gyromagnetic ratio and bond length was performed using the values in Table 2.2.

6.3 RDCs in the structure refinement of C4BP~1,2

Before proceeding with a refinement step including RDC restraints, it is necessary to exclude RDCs measured for residues undergoing significant motion: measured RDCs for these residues are scaled by an unknown order parameter and refinement with the measured RDC value could result in incorrect geometry. Motion on two separate timescales can be detected. Motion on a ps-ns timescale can be inferred when the measured $\{^1H\}$-^{15}N NOE at 600 MHz is below a value of 0.6. In addition, the criterion suggested by Tjandra *et al.* [105] was used in order to exclude residues on the basis of probable ms-μs timescale motion as before (see equation 5.1). This means that RDCs measured for the following residues are not used as restraints:

excluded RDCs: 1, 10, 15, 17, 18, 19, 20, 21, 25, 43, 45, 79, 110, 111, 112, 125, 126

RDC restraints were added to the structure calculation protocol using the TENSO module of CNS, which is contains an implementation of the Implicit Saupe tensor Alignment Constraint (ISAC) method [94, 70]. The tenso force constant was increased geometrically from 0.001 kcal·(mol·Hz^2)$^{-1}$ to 1 kcal·(mol·Hz^2)$^{-1}$ during the first cooling stage of refinement, and fixed at the final value throughout the second cooling stage. The modification to the refinement protocol described in section 4.3 (a reduced temperature step in the first cooling stage) was used here.

6.3.1 Can a single alignment tensor be used for both modules?

In an effort to improve the definition of the intermodular orientation and the convergence of the structure calculations, the possibility of using a single alignment tensor for the whole construct was explored. To check the validity of the assumption that

both modules experience the same alignment, several sets of structures were calculated including all RDCs (set 'ALL'), RDCs for module 1 only (set 'M1'), or for module 2 only (set 'M2'). Statistics for the best structures from each set are shown in Tables 6.1 and 6.2 below:

Structure ensemble	Number of RDC restraints					CNS energy terms		
	NH	NC'	$C^{\alpha}C'$	$C^{\alpha}H^{\alpha}$	Total	Etot	Enoe	Etenso
'ALL'	62	65	41	24	192	1256.3	133.8	67.6
'M1'	24	27	19	11	81	976.5	87.1	28.1
'M2'	38	38	22	13	111	1081.7	112.7	29.4
'NOE'	-	-	-	-	-	913.5	89.9	-

Table 6.1: **Numbers of RDC restraints and average CNS energy terms for 4 sets of C4BP~1,2 structures.**

Ensemble/No.	$RMSD_{overall}$	$RMSD_{m1}$	$RMSD_{m2}$
'ALL'/40	0.96	0.85	0.49
'M1'/39	1.93	0.75	0.57
'M2'/40	2.49	0.93	0.50
'NOE'/31	2.82	0.96	0.62

Table 6.2: **RMSD values for 4 sets of C4BP~1,2 structures.** The structures were overlaid on the backbone N, C and CA atoms of the following residues: 2-14, 23-42 and 46-60 for module 1, and 64-78, 81-109 and 113-122 for module 2. All RMSD values are given in angstroms (Å), and the number of structures in each ensemble is shown in the first column.

It can be seen that the use of a single alignment tensor for both modules is accompanied by a slight increase in energy: Enoe for the set 'ALL' structure set is higher than for the 'M1' or 'M2' set, and Etenso('ALL') > [Etenso('M1') + Etenso('M2')]. Convergence of the structure calculation is also improved by using the RDC restraints, giving ~ 40 converged structures with RDCs as opposed to ~ 30 with NOEs only. Backbone RMSDs obtained by overlaying on module 1, on module 2, and on both modules simultaneously are improved upon inclusion of RDC restraints: set 'M1' RDCs give the best module 1 overlay, set 'M2' RDCs give the best module 2 overlay and set 'ALL' RDCs give the best overlay for the whole protein (due to a greatly improved definition of the intermodular orientation).

The spread of alignment tensors in each ensemble was calculated by Singular Value Decomposition (SVD - see section 2.4) using the calculated structures and experimental dipolar coupling values. For each ensemble, the program Orderten-SVD [63] was used to perform 100 SVD computations for each converged structure in a coordinate system coincident with that of the lowest energy set 'ALL' structure: the orientation of the resulting solution eigenvectors are mapped onto Sanson-Flamsteed plots (described in section 2.4) in Figure 6.2 and the eigenvalues are summarized in Table 6.3.

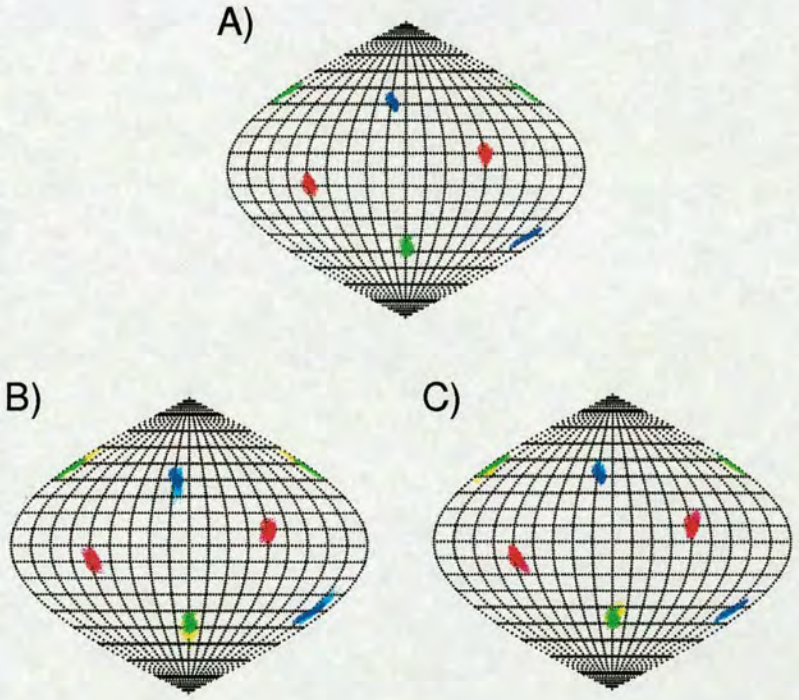

Figure 6.2: **Alignment tensor distributions for C4BP~1,2.** Sanson-Flamsteed plots showing the orientation of the alignment tensor eigenvectors. **A)** The eigenvectors for the 'ALL' ensemble using RDCs from both modules. Colours for the eigenvectors corresponding to the largest positive, largest negative and smallest eigenvalues are red, blue and green, respectively. **B)** Eigenvectors for the 'M1' ensemble using RDCs from module 1: colours for the eigenvectors corresponding to the largest positive, largest negative and smallest eigenvalues are magenta, cyan and yellow, respectively. For comparison, the set 'ALL' eigenvectors are overlaid in red, blue and green. **C)** Eigenvectors for the 'M2' ensemble using RDCs from module 2: colours for the eigenvectors corresponding to the largest positive, largest negative and smallest eigenvalues are magenta, cyan and yellow, respectively. For comparison, the set 'ALL' eigenvectors are overlaid in red, blue and green.

Set	Szz	Syy	Sxx	η
'ALL'	-11.18 (0.34)	9.91 (0.42)	1.26 (0.3)	0.77 (0.06)
'M1'	-8.74 (0.52)	9.08 (0.66)	-0.34 (0.62)	0.88 (0.08)
'M2'	-11.86 (0.50)	9.31 (0.64)	2.55 (0.50)	0.57 (0.08)

Table 6.3: **Alignment tensor eigenvalue magnitudes** for 'ALL', 'M1' and 'M2' structure ensembles. Values shown $\times 10^{-4}$ are average values followed by the standard deviation in parentheses. All solutions for the 'ALL' ensemble give negative Szz, positive Syy and positive Sxx. For the 'M1' structure set Szz and Syy are interchanged in some solutions because $\eta \sim 1$: the entries for 'M1' Szz and Syy given here should be swapped to agree with convention.

The directions of the alignment tensor eigenvectors shown in Figure 6.2 are in excellent agreement: i.e. the alignment tensor for module 1 is the same as that for the whole construct, which is the same as that for module 2. There is however a difference in the magnitude of the eigenvalues obtained for the two modules (Table 6.3): in the case of the largest negative eigenvalue, which is Szz in set 'ALL', the corresponding 'M1' structure set eigenvalue is significantly different. This lack of agreement could be due to a limited sampling of three-dimensional space by the bond vectors for which RDC data were obtained (note also that the number of RDCs in the 'M1' set is $\sim$ 20% smaller than in the 'M2' set). Small amplitude intermodular motion may also be present, meaning that the alignment tensors for the two modules are similar but subtly different: there is some evidence for this in that the structure of module 1 is better defined in structures calculated using only module 1 RDCs.

The set 'ALL' structures display a very limited distribution of intermodular orientations which corresponds to a single relative orientation of the alignment tensor PAS for modules 1 and 2: alternative relative orientations are theoretically possible but the other possible solution, shown in Figure 6.3, is ruled out by the NOE data.

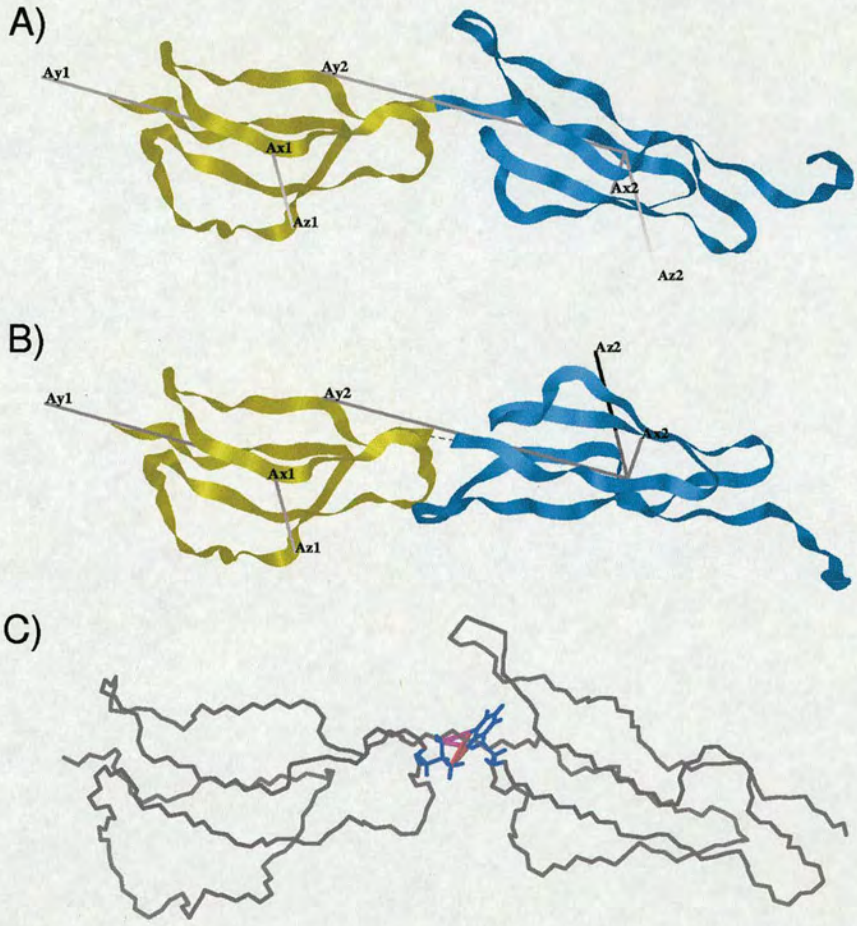

Figure 6.3: **Relative orientation of the PAS for M1/M2. A)** The relative orientation found in the converged set of structures obtained using set 'ALL' RDCs. **B)** Alternative relative orientation of M1 and M2 PAS obtained by 180° rotation about Syy axis; equivalent to inversion of Szz and Sxx (obtained using the program MODULE - see text). **C)** The intermodular NOEs between the HG1# or HG2# of valine 38 and the HD1/2 or HE1/2 of phenylalanine 84, in combination with the module/linker NOEs, are sufficient to discriminate between the two relative orientations of the PAS for modules 1 and 2 shown in A) and B).

It can also be seen that the alignment tensor eigenvector with the largest eigenvalue is perpendicular to the long axis of the molecule, suggesting that the negatively charged phage are involved in an electrostatic aligning interaction with C4BP~1,2 (note also the positive patch at the interface between modules 1 and 2 in Figure 6.12). Transient binding interactions between the phage and protein would also explain the reduced T_2 values observed for the aligned sample (Figure 6.4).

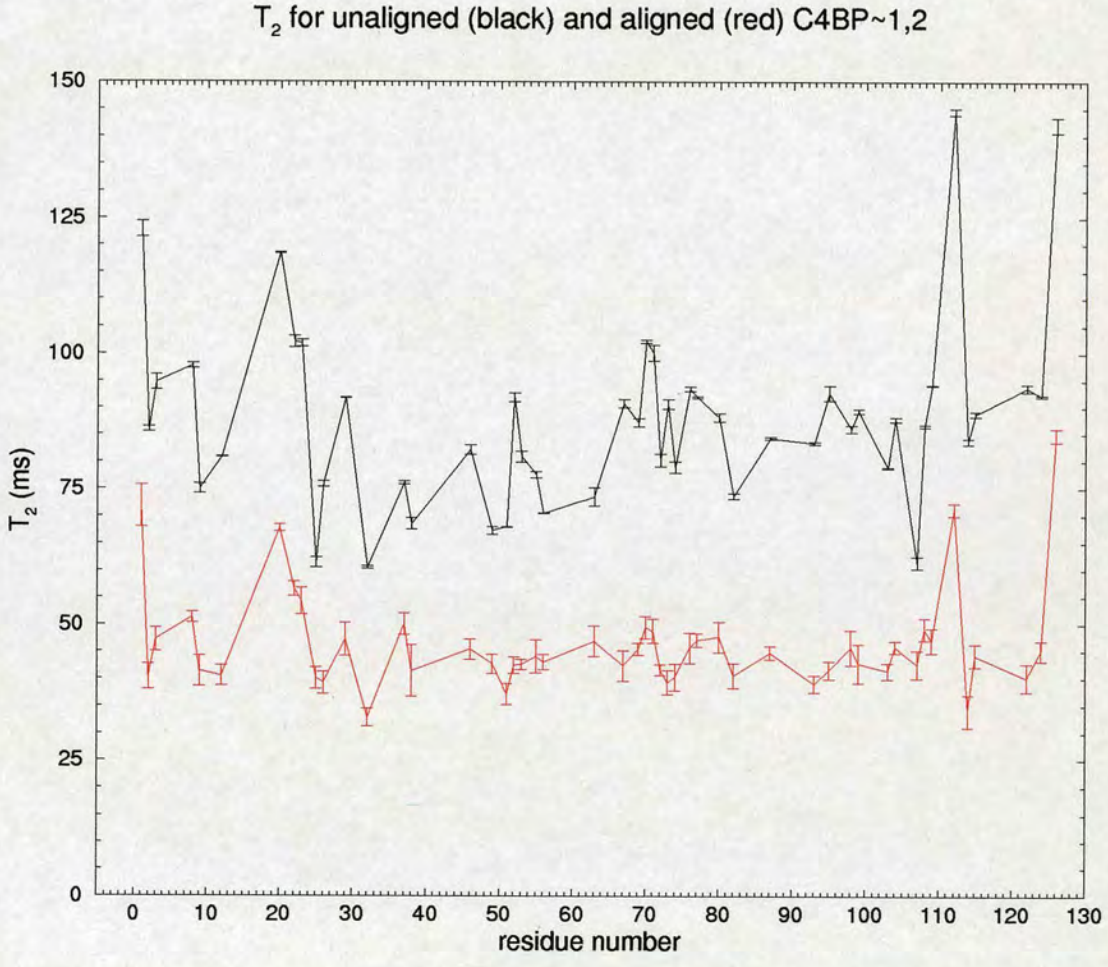

Figure 6.4: **T_2 relaxation time comparison.** Graph illustrating the reduction in T_2 relaxation time in a sample aligned with 7.3 mg/mL Pf1 phage where [NaCl] = 90 mM. Data measured at 600 MHz for the unaligned sample in black and for the aligned sample in red: data for the aligned sample follow the same trend as the for unaligned sample, but T_2 values for all residues are reduced by ~25-50 ms.

For the final ensemble, the set 'ALL' RDCs were used with single alignment tensor in order to produce a structure with well-defined intermodular orientation. Although slightly better structures can be obtained for the individual modules by specifying separate alignment tensors for modules 1 and 2, the good agreement between the set 'M1', 'M2' and 'ALL' alignment tensors supports the validity of making the approximation that the whole molecule experiences the same alignment tensor. It is uncertain whether the intermodular orientation obtained using these RDC data reflect the average orientation in solution or whether the interactions with the aligning medium perturb the structure. RDC data acquired with a different aligning system would need to be used as an independent check on the intermodular orientation obtained here.

6.3.2 Evaluation of the final ensemble

After the first round of calculations including RDC restraints for both modules (set 'ALL'), a number of new NOE violations were observed, and those that were found to be mis-assigned were corrected. An ensemble of 100 structures was calculated and the overall and NOE energies sorted according to total energy are shown in Figure 6.5.

The best 40 structures have been selected according to overall energy and overlaid on the N, C^{α}, and C' atoms of the following residues using MOLMOL [54]:

module 1: 2-14,22-24,26-42,46-60

module 2: 65-78,80-109,113-122

Overlays on module 1 and on module 2, as well as on both modules together are shown in Figure 6.6 below.

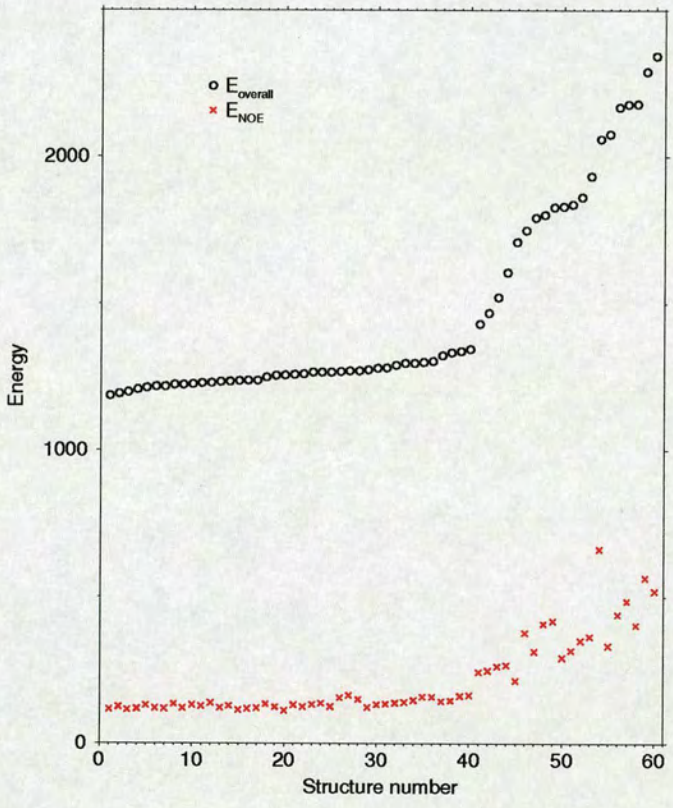

Figure 6.5: **C4BP~1,2 structure ensemble overall and NOE energies.** Plot of energy vs. structure number sorted according to overall energy for 60/100 C4BP~1,2 structures. The first 40 were judged to have converged.

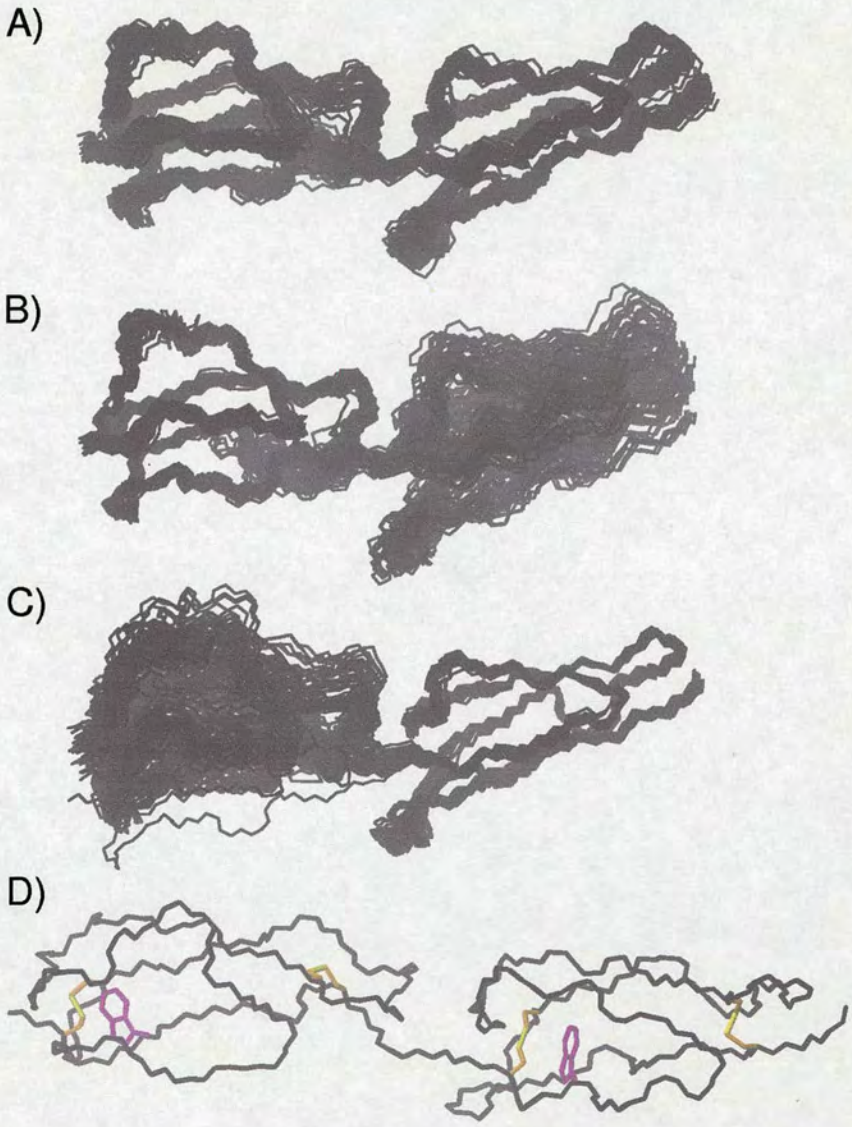

Figure 6.6: **C4BP~1,2 final structures** obtained using set 'ALL' RDCs. **A)** Best 40 final structures with RDCs overlaid on modules 1 and 2 (backbone only). **B)** Best 40 final structures with RDCs overlaid on module 1 only. **C)** Best 40 final structures with RDCs overlaid on module 2 only. **D)** Lowest energy structure with disulfide bridges shown in yellow and conserved tryptophan residues in purple.

Comparing the structures calculated using NOE restraints only versus those including RDC restraints, it can be seen that in addition to improved local structure RDC restraints lead to much better definition of the intermodular orientation, which is very poorly defined by the NOE data (overlay on both modules for the NOE only structure set is 2.82Å compared with 0.96Å for the structures using set 'ALL' RDCs). This is clearly seen in Figure 6.7 where the intermodular orientations of the NOE only and RDC structure ensembles are expressed in terms of tilt, twist and skew angles - a much narrower distribution of these angles is seen in the RDC structure ensemble (particularly tilt and twist).

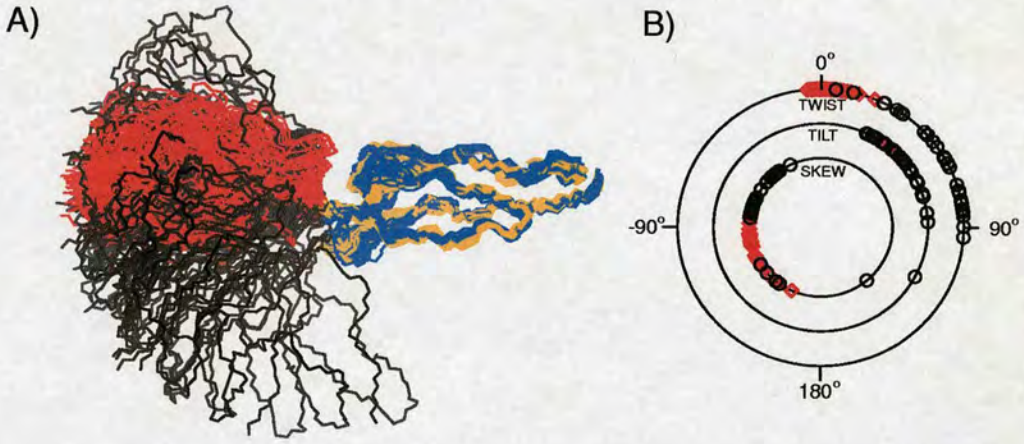

Figure 6.7: **C4BP~1,2 intermodular orientation. A)** Comparison of final structures using RDCs (best 40) with NOE only structures (best 31): for the structures with RDCs, module 1 is coloured red and module 2 orange; for the NOE only structures module 1 is coloured black and module 2 blue. All structures have been overlaid on module 2. **B)** Tilt, twist and skew angle distributions for the converged structures of the NOE only (black) and set 'ALL' RDC (red) structure ensembles. The average values for the ensemble including RDCs are: skew=-110°(18), tilt=38.4°(4.4) and twist=3.9°(5.7) where the values in parentheses are standard deviations. For the definition of tilt, twist and skew angles [6] see Figure 5.6. The conserved tryptophan residues for C4BP~1,2 are W54 for module 1 and W115 for module 2; the limits of modules 1 and 2 are defined as being C2-C60 and C65-C122, respectively.

6.4 Analysis of relaxation data

Plots of T_1 and T_2, the T_1/T_2 ratio and $\{^1H\}$-^{15}N NOE are shown in Figures 6.8 and 6.9, respectively.

Analysis of the 600 MHz and 800 MHz ^{15}N relaxation data according to the extended model-free formalism described in section 2.3.2 was performed using the program Tensor [31]. It is apparent in Figure 6.8 that the estimated errors for the 800 MHz T_1 data are substantially larger relative to the errors in T_2 than those estimated for the 600 MHz data, which is a consequence of a systematic difference between the intensities measured for the repeated time point. Estimates for the correlation time of the whole molecule, and of its two modules, can be obtained using the R_2/R_1 ratio at 600 MHz and 800 MHz in the manner described in section 2.3:

τ_c, **600 MHz** : 8.7 ± 0.5 ns

τ_c, **800 MHz** : 7.3 ± 0.5 ns

τ_c, **M1, 600 MHz** : 8.9 ± 0.6 ns

τ_c, **M1, 800 MHz** : 7.4 ± 0.5 ns

τ_c, **M2, 600 MHz** : 8.6 ± 0.4 ns

τ_c, **M2, 800 MHz** : 7.3 ± 0.4 ns

There appears to be a systematic difference between the 600 MHz and 800 MHz data (the values clearly do not agree within one standard deviation), but there is also a noticeable trend in that the correlation time for module 1 appears to be higher than the correlation time for module 2 (although the differences are well within the limit of variations seen in a single module). This would imply a degree, at least, of independent tumbling - i.e. some flexibility in the linker. The 600 MHz and 800 MHz relaxation data (Figures 6.8 and 6.9 below) clearly follow the same trends but there is a possible systematic error in the 800 MHz T_1 data as indicated by the larger error bars: the 600 MHz and 800 MHz relaxation data were therefore analysed independently.

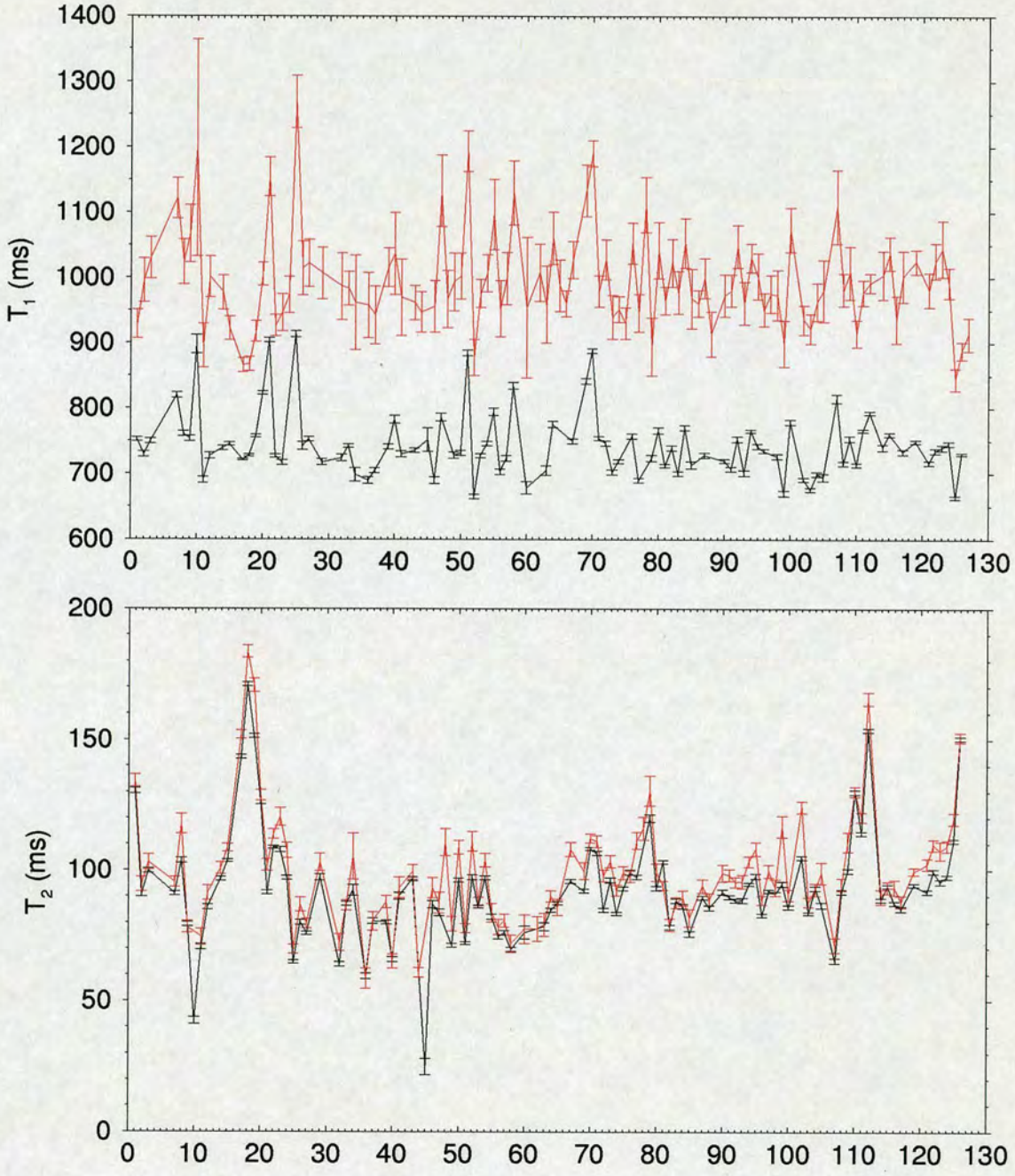

Figure 6.8: **T_1 and T_2 relaxation times for C4BP~1,2.** Relaxation data at 600 MHz are shown in black, data at 800 MHz in red.

Diffusion tensors were predicted using HYDROPRO [27] for the lowest energy structure calculated with RDCs (elongated: tilt angle=38.3°) and the most bent structure calculated using NOEs alone (tilt angle=118.3°). The results are shown in Table 6.4.

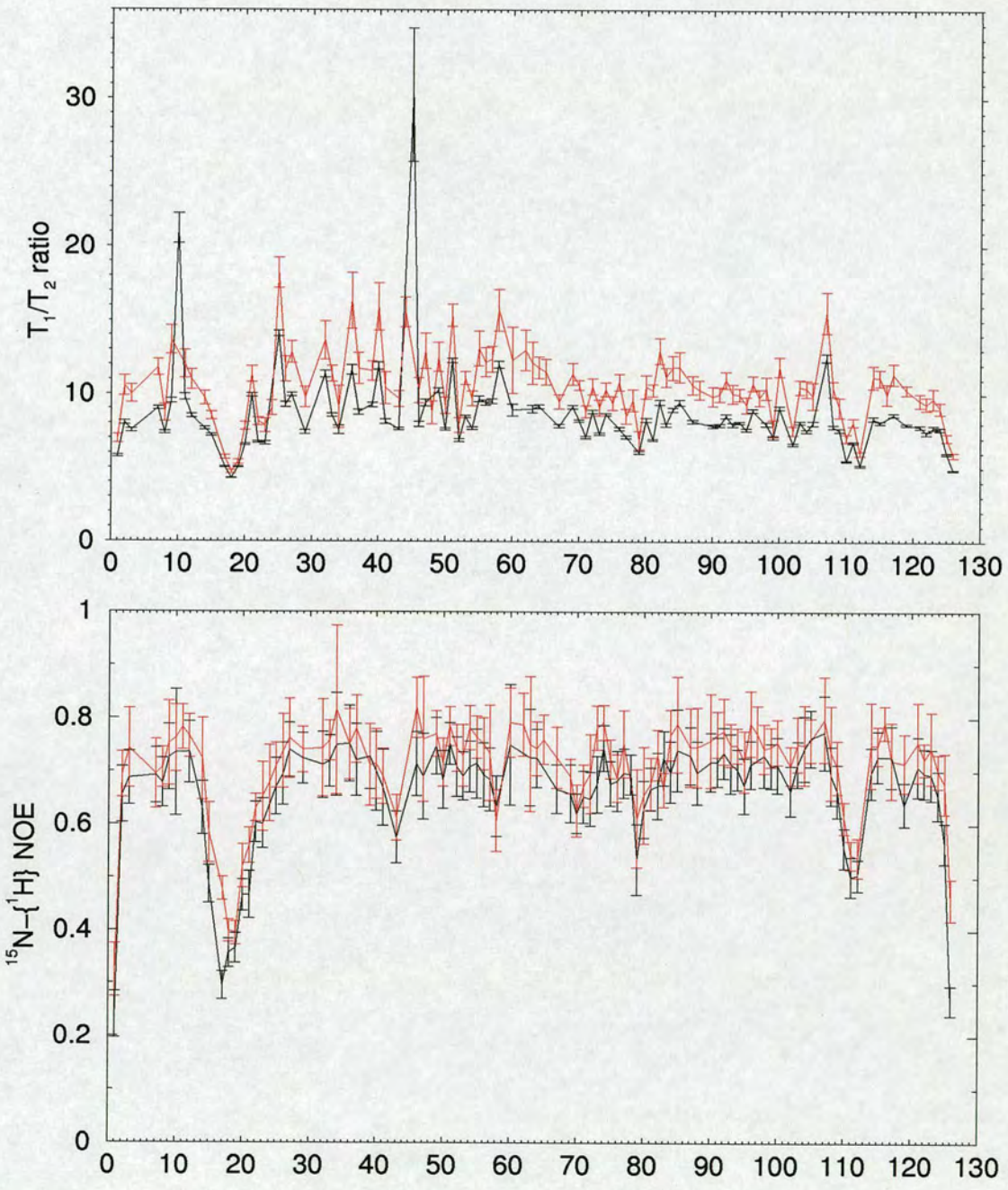

Figure 6.9: T_1/T_2 **ratio and** $\{^1H\}$-^{15}N NOE **for C4BP~1,2.** Relaxation data at 600 MHz are shown in black, data at 800 MHz in red.

Structure	$D_{\parallel}$	$D_{\perp}$	$D_{\parallel}/D_{\perp}$	τ_3 (ns)	tilt (°)
RDC(lowest energy)	2.277×10^7	7.177×10^6	3.17	9.5	38.3
NOE(most bent)	1.773×10^7	1.090×10^7	1.63	10.8	118.3

Table 6.4: **Predicted diffusional properties of C4BP~1,2.** Calculations carried out using the program HYDROPRO.

In spite of the elongated shape of the C4BP~1,2 structure there was no significant improvement in the fit of the relaxation data (Tensor) using an anisotropic model, either axially symmetric or fully anisotropic, over the isotropic model of diffusion. Furthermore, the results of fitting an anisotropic model did not appear to be meaningful (i.e. D_{zz} did not coincide with the long axes of the modules). Given the nature of the predicted diffusion tensor, the lack of improvement in fit of the relaxation data using an anisotropic diffusional model is surprising, but could be explained by 'structural noise' (i.e. poor N-H bond vector definition), poor sampling of the diffusion tensor by the NH bond vectors or motion throughout/between the module pair.

Assuming an isotropic model of diffusion and using the lowest energy structure calculated with RDCs a model-free analysis of the relaxation data was carried out using the program Tensor [31]. The majority of residues are best described by the more complicated models (3, 4 and 5) where motion on the ns and ms-μs time scales is present (Figure 6.10).

The results obtained using the 600 MHz and 800 MHz data follow broadly the same trends and there is a significant amount of agreement in the model number selected for each residue and the time scale of the motions. Although some of the motions may be spurious, reflecting the fact that the effects of anisotropic tumbling of the molecule could not be taken into account [99], it would appear that there are significant motional processes occurring in this molecule on a range of timescales. It can be seen in a colour plot of the principal order parameter values, S, obtained using the 600 MHz relaxation data (Figure 6.11), that the flexible loops of C4BP~1,2 as judged by poor overlay of the structures and low $\{^1H\}$-^{15}N NOE show low order parameter values (< 0.7); particularly the hypervariable loops for modules 1 and 2, and the loop in the region of

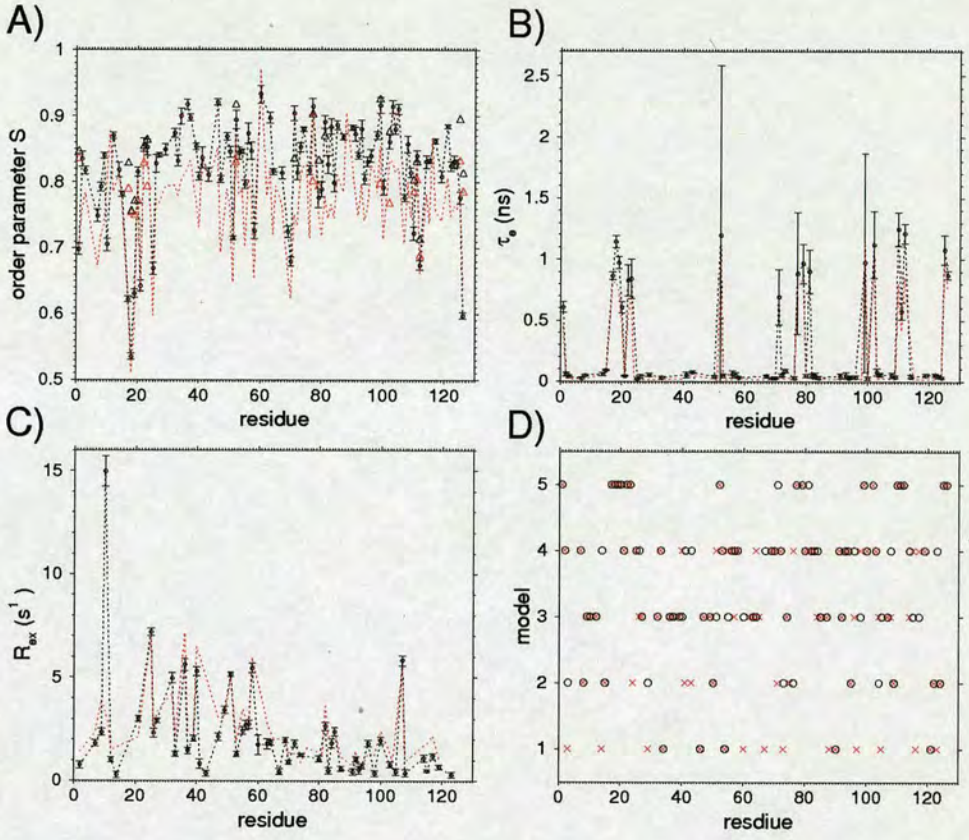

Figure 6.10: **C4BP~1,2 model free results. A)** Order parameters: 600 black, 800 red, dotted lines are S or Ss and triangles are Sf where present (model 5). **B)** Effective correlation time on ns time scale (black 600, red 800). **C)** Exchange rates, Rex. **D)** Model number fitted to each residue.

residues 110-112 in module 2.

The model-free analysis implies motion throughout the protein on 1-2 s^{-1} and 0.5 ns time scales. In module 1, the hypervariable loop, and in particular residues 17-21 exhibit low order parameters ~0.6 and effective correlation times of 0.5-1 ns, while N1 has an effective correlation time of 0.6 ns and G52 has an effective correlation time of ~1-1.5 ns. Residue F10 shows evidence of exchange on a time scale of 15 s^{-1} (600 MHz data only), T25 shows evidence of exchange on a 7 s^{-1} time scale and several other residues; T32, G36, S40, D51 and T58, show evidence of motion on a time scale of ~5 s^{-1}. In module 2, residue L71 has an effective correlation time of ~0.5-1 ns and E70 has a low order parameter; E77, K79 and D81 (module 2 hypervariable loop) have effective

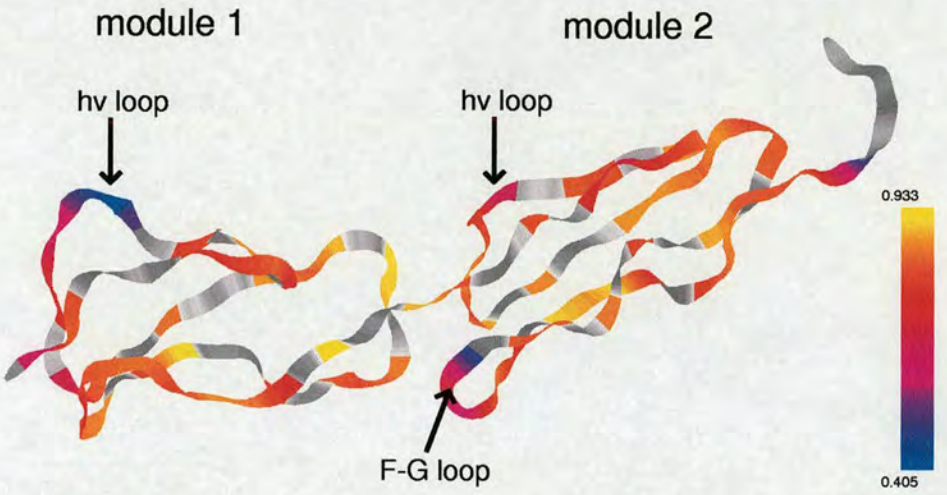

Figure 6.11: **C4BP~1,2 mobility.** Plot of backbone S values for C4BP~1,2 using 600 MHz data: module 1 is to the left, module 2 to the right.

correlation times of ~0.5-1 ns; residues I99 and T102 have effective correlation times of ~1 ns; residue E107 shows evidence of exchange on a time scale of ~6 s^{-1}; in the region of D110-G112 which forms part of a loop between strands F and G extending towards module 1, residues D110 and G112 have effective correlation times of ~1 ns, residue R111 has an effective correlation time of ~0.5 ns and G112 has a low order parameter; residues L125 and E126 at the C-terminal end of the protein have low order parameters and effective correlation times of ~1 ns. The presence of the motions outlined above is for the most part supported by a simple analysis of the relaxation data in which a dip in the $\{^1H\}$-^{15}N NOE is taken to indicate ns time scale motion and anomalous R_2/R_1 values indicate ms-μs time scale motion.

6.5 Discussion

The RDC and heteronuclear relaxation data presented here can be used to draw some useful conclusions regarding the structure and dynamics of C4BP~1,2. Overall, the NMR data point to the fact that module 1 is less well structured than module 2, since the backbone RMSD is higher and the $\{^1H\}$-^{15}N NOE values are lower for module 1. A number of mobile regions (on the ns time scale) can be identified: the module 1

hypervariable loop T17-T21, the module 2 hypervariable loop E77-D81, the module 2 β-bulge I99-T102 and the module 2 loop D110-G112 joining strands F and G. There are several residues that show evidence of exchange: F10, T25, T32, G36, S40, D51, T58 and E107, most of which belong to module 1. None of these flexible regions of the protein correspond to residues known to be involved in interactions with C4b (R39, R64 and especially R66 [13]) / heparin [42] or the M protein M4 (Arp4) of *S. pyogenes* (R66/H67 [2]).

The residues that have been identified as important in binding interactions with C4b, heparin and the M protein M4 (Arp4) of *S. Pyogenes* have been mapped onto a surface plot of the lowest energy structure from the set 'ALL' ensemble calculated using RDCs - this is shown in Figure 6.12, with a surface potential plot for comparison. The interaction between C4BP and C4b is known to have an electrostatic component since it is disrupted by high salt concentrations [13], whereas the interaction with the *S. Pyogenes* M protein is not in spite of some overlap of the two binding sites [11].

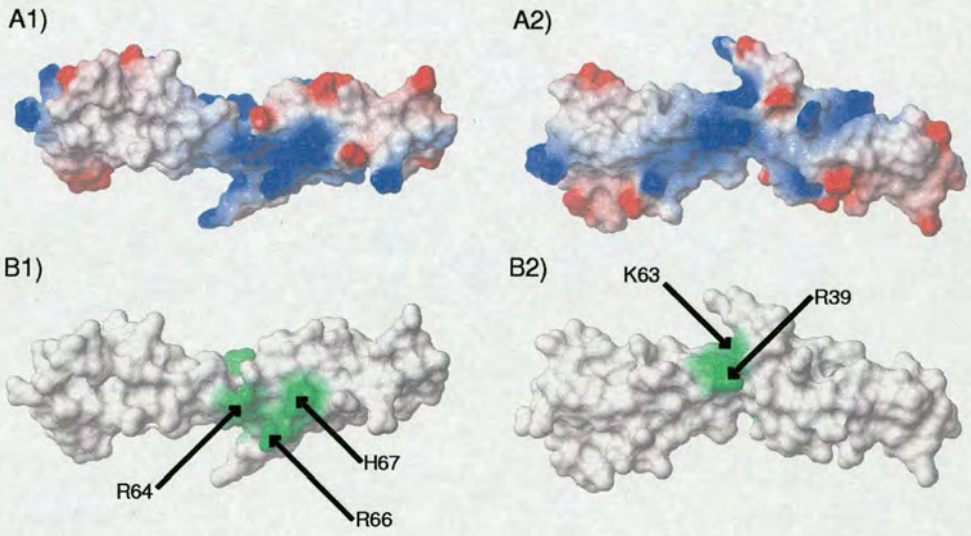

Figure 6.12: **C4BP~1,2 putative binding sites.** A1) Surface potential plot for the lowest energy final structure: positively charged regions are coloured blue and negatively charged regions red. A2) As for A1 but rotated 180° about a horizontal axis. B1) Surface plot with labelled residues of interest coloured green (see discussion). B2) As for B1 but rotated 180° about a horizontal axis.

As in the case of DAF~2,3, intermodular motional processes of limited extent can

account for the fact that it is not possible to fit an anisotropic diffusion tensor using the relaxation data, since the molecule is normally assumed to be rigid in such models - i.e. intramolecular motions must not be coupled to the overall molecular tumbling as is probably the case here. The apparent ~1 ns time scale dynamics observed here may be a manifestation of such intermodular motions. Slightly different estimated τ_c values for the two modules may also be a consequence of some degree of independent tumbling (although the errors in estimated τ_c are too large to draw any definite conclusions).

The use of RDC restraints improves the accuracy of the structure: such improvement is seen on the level of individual modules and especially on the level of the whole protein, leading to a reasonably well-defined module pair orientation (RMSD over the module pair of < 1Å and a tighter distribution of tilt, twist and skew angles than obtained using NOEs alone). The structure obtained using RDCs and the assumption that the whole molecule experiences the same alignment tensor is almost certainly an average structure, about which significant motions can still occur in solution. However, a significant degree of rigidity is implied (but not proven) by the fact that the alignment tensors of the individual modules are unchanged by making the assumption that the whole molecule experiences the same alignment tensor. Mutagenesis data also support the existence of a somewhat rigid linker since insertion of two extra alanine residues into the linker between modules 1 and 2 abolishes the C3-convertase decay acceleration activity of C4BP [12]. Whether or not the average intermodular orientation observed here reflects the true average conformation in solution is unclear. Motional processes and interactions with the aligning medium could both potentially lead to an artefactual average orientation and additional data, e.g. RDCs measured in a different aligning medium, would be required to rule out this possibility.

Chapter 7

Conclusions

Residual dipolar coupling constants make a valuable addition to the NOE restraints that are commonly used for biomolecular structure determination by NMR. In contrast with NOEs, RDCs contain information that is of a non-local nature, allowing accurate orientation of bonds with reference to the structure as a whole. Given a double-labelled protein sample, it is a relatively straightforward procedure to create an aligned sample and measure several sets of one-bond RDCs (typically $^1D_{\mathrm{NH}}$, $^1D_{\mathrm{NC'}}$, $^1D_{\mathrm{C^\alpha C'}}$ and $^1D_{\mathrm{C^\alpha H^\alpha}}$).

In this work, four new pulse sequences were developed for the measurement of the one-bond $^1D_{\mathrm{C^\alpha H^\alpha}}$ RDCs: the 2D (HNCO)-(J-CA)NH and (J-HACACO)NH methods and the 3D (HN)CO-(J-CA)NH and J-HA(CACO)NH methods. The 2D (HNCO)-(J-CA)NH and 3D (HN)CO-(J-CA)NH methods use $^{13}C^\alpha$ magnetization to sample the $C^\alpha H^\alpha$ splitting and feature DPFGSE elements incorporating C^α -selective pulses for the refocussing of $C^\alpha C^\beta$ coupling evolution together with the use of novel semi-constant time ^{15}N frequency labelling periods, thus extending the length of time for which the $C^\alpha H^\alpha$ splitting can be sampled. In the 2D (J-HACACO)NH and 3D J-HA(CACO)NH methods, the $C^\alpha H^\alpha$ splitting is sampled using H^α magnetization: these methods circumvent the problem of $C^\alpha C^\beta$ coupling evolution and solve the problem of H^α -H^{other} coupling evolution using novel BIRD elements, again incorporating C^α -selective pulses. The 2D (J-HACACO)NH also uses semi-constant time ^{15}N frequency labelling. All of the new methods allow extended sampling of $C^\alpha H^\alpha$ splittings, providing improved precision within the limits imposed by C^α relaxation (up to $\sim$ 20 ms $C^\alpha H^\alpha$ sampling

was found to be beneficial) and H^{α} relaxation. Sensitivity was increased by scaling down the $^{1}J_{C^{\alpha}H^{\alpha}}$ sampling period relative to the ^{15}N frequency labelling period by a factor of 1.5 times.

Using the protein ABA-1A, for which $^{1}D_{NH}$, $^{1}D_{NC'}$, $^{1}D_{C^{\alpha}C'}$ and $^{1}D_{C^{\alpha}H^{\alpha}}$ RDCs were measured, the structure refinement protocol was optimized - it was found that better convergence of the structures could be obtained by reducing the temperature step of the first cooling stage from 50 K to 10 K. The quality of the ABA-1A structures was improved by the addition of RDC restraints as judged by the presence of more residues in favourable regions of a Ramachandran plot and by reduced RMSD values, particularly due to the improved structural definition of a pair of helices for which only a few NOE restraints were available. In addition, it was found that structures refined using $^{1}D_{NH}$, $^{1}D_{NC'}$ and $^{1}D_{C^{\alpha}C'}$ restraints gave predicted $^{1}D_{C^{\alpha}H^{\alpha}}$ values much closer to the observed values than was observed for structures determined using NOE data only. Thus, both the local geometry and positioning of secondary structural elements were improved upon incorporation of RDC restraints into the structure calculations.

The CCP module pairs DAF~2,3 and C4BP~1,2 were investigated using a combination of RDC and relaxation data in an effort to improve the quality of the structures and to shed light on the nature of the intermodular interfaces, either to establish relative orientations that were poorly defined by the NOE data alone, or to investigate any dynamical processes responsible for this poor definition. It was hoped that this would lead to a better understanding of the relevance of the intermodular orientation and / or flexibility to the binding events known to occur in the regulation of complement activation.

In the case of DAF~2,3, the limited set of RDCs measured ($^{1}D_{NH}$ only), in combination with the NOE data, were found to be inadequate to significantly improve the definition of the intermodular interface, or to establish whether the module-pair behave as a single rigid entity. Furthermore, the original NMR structures refined using the RDC data were able to accommodate the $^{1}D_{NH}$ restraints without producing violations, in spite of a probably erroneous arrangement of the individual modules. The RDC restraints are in reasonable agreement with a published x-ray structure of DAF~2,3

but the relaxation data fit rather poorly to the expected axially symmetric anisotropic diffusional model, perhaps implying a structure with a degree of rigidity but still subject to intermodular motions. Current NMR structures of DAF~2,3 resulting from an ongoing re-investigation, do not as yet reproduce the intermodular arrangment seen in the crystal structure, which could reflect a different average conformation in solution or structural noise in the current set of NMR structures. Consideration of the published mutagenesis data for DAF~2,3 in light of the measured relaxation data suggest that the functionally important interactions occur almost entirely in non-flexible regions of the structure.

The extensive set of ${}^1D_{\mathrm{NH}}$, ${}^1D_{\mathrm{NC'}}$, ${}^1D_{\mathrm{C^\alpha C'}}$ and ${}^1D_{\mathrm{C^\alpha H^\alpha}}$ RDCs measured for C4BP~1,2 are sufficient to specify a unique intermodular orientation assuming that the alignment tensors for the two modules are identical (the evidence is that they are very similar, if not identical). Using this larger set of RDCs, in which a correspondingly better sample of internuclear vectors is present, errors in the NOE data were highlighted as violations and the structures obtained by refinement with the RDCs showed better convergence. Again, the relaxation data did not fit an axially symmetric anisotropic diffusion model in spite of the large number of RDCs used as restraints: this is likely a consequence of intermodular motions (the correlation times for the two modules provide some tentative evidence for the presence of such motions). As in the case of DAF~2,3, the residues known to be important in binding interactions appear to be in rigidly structured regions of the protein.

The homology of DAF~2,3 and C4BP~1,2, both in terms of structure (DAF~2 / C4BP~1, and particularly DAF~3 / C4BP~2 which are both 'C' cluster CCP modules with a pronounced F-G loop) and function (both bind and accelerate the decay of C3 convertases) is intriguing considering the difference between the structures: in the case of DAF~2,3, the two modules are oriented such that conserved features (e.g. hypervariable loop) are on opposite faces viewing the module pair from the side, while in the case of C4BP~1,2 they are on the same face.

This work has shown that RDCs can be of great value for the refinement of the NMR structures of proteins provided that a sufficiently large set of couplings is measured. It

has also highlighted the problems arising in the interpretation of RDC and relaxation data for multimodular proteins that are potentially flexible on timescales close to the NMR timescale. To address the questions that remain regarding the structures of the multimodular proteins investigated here, possible approaches include the measurement of RDC restraints in multiple aligned media or the use of a paramagnetic spin label to obtain additional long-range distance restraints.

Appendix A

Measured RDC Values

The RDCs below are unscaled and in the case of $^1D_{NC'}$, the quoted residue number is that of the amide nitrogen.

ABA-1A

residue	$^1D_{NH}$ (Hz)	$^1D_{NC'}$ (Hz)	$^1D_{C^\alpha C'}$ (Hz)	$^1D_{C^\alpha H^\alpha}$ (Hz)
232	-14.9	0.8	-	-
233	-9.9	-	1.2	-
234	-4.4	-0.3	-0.7	-29.3
235	-6.0	2.8	-0.3	-
236	-9.1	-0.4	-2.1	-
237	-2.0	1.3	0.8	21.4
238	6.4	-0.7	-2.1	-31.0
239	-12.4	-0.1	1.9	-
240	6.1	0.1	-1.8	3.8
241	-6.6	0.6	-2.5	14.5
242	14.7	-2.0	-1.5	-7.5
243	14.0	-0.6	0.7	10.6
244	8.6	-0.0	-1.3	-
245	-16.4	0.6	-	-
246	-	-	3.5	-26.8
247	2.2	0.5	-1.2	-2.0
248	-7.9	0.9	-1.0	33.0
249	-2.2	-2.1	0.1	-
250	3.1	2.2	7.5	-1.6
251	-1.3	-0.7	-3.3	-3.0
252	-7.8	0.7	-	-
253	-	-	1.4	-
254	3.2	1.1	1.2	1.3
255	2.2	0.4	-2.6	15.9
256	-6.2	-1.1	-0.1	0.6
257	2.3	1.5	-	-
258	-	-	-0.3	0.4
259	-16.6	1.3	1.2	-
260	-3.8	0.4	-3.4	36.9
261	-9.6	-0.5	-	-

262	-	-	0.9	5.5
263	0.8	-1.8	-3.4	-
264	-1.0	1.1	1.7	-
265	8.2	-0.4	-0.6	-
266	2.7	1.4	-	-
267	16.5	0.6	-3.0	-7.1
268	2.5	-0.2	-	-
270	-	-	-1.3	-
274	-	-	-3.0	-6.5
275	-1.1	1.1	3.2	19.0
277	0.2	-	-	-
279	-	-	0.8	-32.0
280	7.8	-	-1.7	28.4
282	-	-0.5	1.2	-9.0
283	16.3	0.9	0.9	-6.8
284	9.6	-1.0	-3.0	18.7
285	8.3	-0.6	-0.4	-
286	13.8	-1.7	-0.1	-31.1
287	-	1.5	-	-
288	-	-	-3.3	-
289	9.8	-0.1	-	-
290	-	-1.5	-	-
291	11.6	-	-	-
297	16.2	-0.8	-	-
298	-3.6	-1.5	-	-
300	12.1	-	2.5	-
301	19.6	-	-	-
303	-	-	0.4	17.9
304	-0.1	-1.6	-	-
305	-	1.1	-	-
307	-	-	-0.4	-
308	-0.9	1.2	-	-
310	-	-	2.3	8.7
311	-0.5	-1.8	-	-
312	-2.9	-	-	-
313	0.4	-	-2.3	-28.3
314	-2.1	0.9	2.0	-
315	4.0	0.7	2.9	-
316	-11.1	0.3	-2.9	-17.0
317	14.7	-	-	-
319	-	-	-1.8	2.7
320	0.5	-1.0	-	-
321	4.3	2.1	5.0	-
322	1.2	-1.5	-	-
323	-	-	-0.4	19.2
324	2.1	-1.3	0.5	-
325	2.4	2.5	-	-
327	-	0.5	-	-
330	-	-	-0.3	-1.2
331	-0.9	-0.8	0.2	2.2
332	3.0	1.9	-	-
333	-	-0.8	0.6	-
334	-6.5	-1.1	-2.8	-

335	7.7	-	-2.8	-
336	5.0	-1.6	-	-
337	-	1.1	-	-
338	-	-	-1.7	1.9
339	-1.0	-0.2	-1.1	38.4
340	8.5	1.7	-	-
342	-	-	-2.7	11.5
343	4.4	-1.4	-	-
344	15.6	-0.9	-1.2	-27.5
345	0.8	1.5	-	-
347	-	-	-	2.6
348	2.0	0.9	-	-
349	-	-	-2.1	-
350	-	1.4	4.7	-
351	-1.2	-	1.0	-
353	5.4	-	-	-

DAF~2,3

residue	$^1D_{NH}$ (Hz)	residue	$^1D_{NH}$ (Hz)
65	-9.9	128	1.3
68	6.4	137	-2.6
69	-5.1	138	2.2
70	-2.8	140	-8.0
72	10.3	141	-3.7
73	-0.8	144	7.9
74	-2.4	146	4.7
75	-1.4	150	2.0
79	-9.5	152	-2.8
80	-9.1	154	-9.8
84	-0.2	155	-1.6
85	0.8	156	0.7
88	1.8	157	0.5
91	-4.2	158	2.9
92	-10.4	159	-10.7
94	-2.1	160	7.6
95	0.5	161	-0.8
102	-3.0	162	-2.2
108	-10.1	163	-5.9
111	-5.2	164	1.1
113	-7.1	166	-8.1
114	-8.2	167	-3.9
115	1.3	168	0.3
116	8.2	169	-3.9
117	4.4	170	3.4
118	-3.5	174	-6.8
119	4.4	178	-7.2
123	-5.8	180	3.6
124	-0.8	181	-4.0
125	1.6	186	-2.0
126	1.1	188	-2.0

C4BP~1,2

residue	$^1D_{NH}$ (Hz)	$^1D_{NC'}$ (Hz)	$^1D_{C^\alpha C'}$ (Hz)	$^1D_{C^\alpha H^\alpha}$ (Hz)
2	-10.3	0.1	2.5	-
3	1.2	-1.4	-	-
6	-	-	-1.7	-
7	-15.2	-1.2	0.3	20.2
8	-2.6	-2.2	-	-
9	3.5	-0.7	-	-
11	-16.3	1.4	2.8	-10.1
12	-15.8	-0.5	-	-
13	-	-	1.1	-
14	-	-2.4	-0.8	-
22	-2.2	-1.0	4.8	1.8
23	-5.2	-0.8	-	-
24	-	-	-0.4	14.1
26	-0.3	-1.2	-	-
27	-7.8	-	-	-
28	-	-	-3.6	21.3
29	-2.1	-0.8	-	-
32	-19.5	0.4	-	-
33	-	-	-6.8	-
34	2.9	-4.0	-	-
36	-14.5	-0.9	-	-
37	2.5	-	-	-
38	-2.4	-4.2	-3.0	-
39	-	-1.9	-	-
40	-	0.5	-	-
46	-0.5	-1.5	-3.2	-
47	-	-0.6	-	-
48	-	-	0.6	18.4
49	-5.2	-1.5	-	-
50	-	-	2.2	-
51	5.8	-0.2	2.4	-15.4
52	-16.0	-0.9	-2.4	-
53	-6.0	-1.2	-	-
54	-	-	-1.7	6.6
55	0.2	-2.7	-0.7	11.4
56	-6.9	-0.0	-4.5	25.1
57	-	-1.5	-	-
58	-	1.5	-	-
62	-	-	-	0.7
63	-11.6	0.6	-	-
64	-21.3	-	-	-
66	-	-	-2.2	0.8
67	-15.2	-1.1	-	-
68	-	-	0.2	-
69	8.8	-0.9	-1.3	-
70	-18.6	-0.4	1.0	13.1
71	4.3	-2.3	-2.6	-
72	4.9	-0.8	2.3	-5.7
73	-15.9	1.2	0.4	-10.6
74	-17.5	1.8	-	-

75	-18.6	-	-	-
76	-15.5	-0.2	-1.1	25.5
77	-6.5	-3.4	-	-
80	3.0	-0.3	-	-
81	-	-	-1.7	-0.3
82	-8.0	-2.5	-7.5	-
83	-2.0	-1.8	-	-
84	-	0.2	-	-
85	6.8	-1.8	-	-
87	2.4	-0.3	-	-
88	2.2	-	-	-
91	-22.5	-	-	-
92	-24.6	-0.8	-2.3	48.1
93	-17.6	-3.6	-	-
94	-	-	-0.9	46.7
95	-21.1	-2.1	-2.3	-14.0
96	7.2	1.2	-	-
97	-	-	-1.7	-
98	-17.0	0.4	-	-
99	-22.4	-0.9	0.7	-
100	-	-3.2	-	-
102	-20.0	-	-4.7	-
103	1.4	-	3.2	-
104	5.3	-0.0	-	-
105	6.5	1.2	-	-
106	-	-	-3.2	6.8
107	0.3	-0.1	-	-
108	-24.1	0.9	-1.8	-
109	-9.1	-3.0	-	-
113	-	-	-	32.9
114	-19.8	-3.2	-1.8	-
115	-16.6	1.4	-	-
121	-23.6	1.3	-	-
122	-25.6	-2.3	-4.0	-
123	-20.3	-1.7	-0.0	40.3
124	-16.4	0.2	-	-

Appendix B

CNS refinement protocol

```
REMARK refine_partial.inp
REMARK refinement protocol with ambiguous restraints
REMARK essentially as published in
REMARK Nilges, M. (1993), A calculation strategy for the structure
REMARK   determination of symmetric dimers by 1H NMR.
REMARK   Proteins 17, 297-309.
REMARK Nilges, M. (1994)
REMARK   Calculation of three-dimensional structures of proteins from
REMARK   ambiguous distance restraints. JMB in press

if ($exist_start_count = FALSE) then
    {====>} evaluate ($start_count=0)
end if
if ($exist_end_count = FALSE) then
    {====>} evaluate ($end_count=5)          {* number of structures *}
end if

if ($exist_seed= FALSE) then
    evaluate ($seed = 8764)
end if

evaluate ($seed  = $seed * ($start_count + 1))

{* Doubled high_steps and cool_steps 21/8/95 *}
evaluate ($high_steps   = 10000)
evaluate ($cool_steps_1 = 10000)
evaluate ($cool_steps_2 = 4000)
evaluate ($init_t  = 2000.01)

!--------------------------------------------------------------------------
! set the weights for the experimental energy terms

evaluate ($knoe  = 25.0)   ! noes
evaluate ($asym  = 0.1)   ! slope of NOE potential
evaluate ($kcdi  = 0.0)   ! torsion angles

noe
   averaging  * sum
   potential  * soft
   scale      * 25
   sqconstant * 1.0
   sqexponent * 2
   soexponent * 1
   rswitch    * 0.5
   sqoffset   * 0.0
```

```
   asymptote  * 0.1

   ?
end

restraints dihedral
    ?
end

evaluate ($ktenso = 0.001) ! at high temp. almost no weight
tenso
    class dummy
       force 0.001
    class pf1
       force $ktenso
end

evaluate ($rcon  = 0.003)

flags
  exclude *
  include angl bond impr vdw noe cdih dihe tenso
end

evaluate ($count = $start_count)

while ($count <= $end_count) loop stru
   evaluate ($filename="../round5/rrsa_" + $fileroot + "_" + encode($count)+".pdb")

   coor init end
   coor @@$filename

evaluate ($bath = $init_t)
do (mass = 100.0) (all)
do (fbeta = 10.0) (all)

{ define prochiral centres to be swapped & floated }
{ store2 is set AND ids are stored in bcomp & qcomp}
@PROTOCOLS:setup_swap.cns
  (
    swap_sel=&swap_sel;
  )

@PROTOCOLS:float_prochirals.cns

   {* ========================== high temperature stage (reduced non-bonded) *}

   evaluate ($knoe  = 25.0)   ! noes
   evaluate ($asym  = 0.1)   ! slope of NOE potential
   noe
      averaging  * sum
      potential  * soft
      scale      * $knoe
      sqconstant * 1.0
      sqexponent * 2
      soexponent * 1
      rswitch    * 0.5
      sqoffset   * 0.0
      asymptote  * $asym
   end

   parameters
      nbonds
         nbxmod 3
         wmin   =  0.01  ! warning off
         cutnb  =  100   ! nonbonded cutoff
         tolerance 45
```

```
        repel  =  1.54  ! scale factor for vdW radii = 1 ( L-J radii)
        rexp   =  4     ! exponents in (r^irex - R0^irex)^rexp
        irex   =  1
        rcon   =  4.0   ! actually set the vdW weight
      end
    end

inline @PROTOCOLS:setup_reduced.cns

    igroup
       interaction (not store1) (all)
       weights * 1 angl 0.4 impr 0.1 dihe 0.1 vdw 0 elec 0 end
       interaction (store1) (store1)
       weights * 1 angl 0.4 impr 0.1 dihe 0.1 vdw 0.1 end
    end

    dynamics  cartesian
       nstep=$high_steps
       timestep=0.005

       tcoupling = true
       temperature = $bath
       nprint=200

       ntrfr = 99999999
    end

    parameters { starting conditions for cooling }
       nbonds
          atom
          nbxmod 5
          wmin   =  0.01  ! warning off
          cutnb  =  7.0   ! nonbonded cutoff
          tolerance 0.5
          repel  =  1.28  ! scale factor for vdW radii = 1 ( L-J radii)
          rexp   =  4     ! exponents in (r^irex - R0^irex)^rexp
          irex   =  1
          rcon   =  4.0   ! actually set the vdW weight
       end
    end

    {* ================================================== first cooling stage *}

    evaluate($red_f = 5)        {*** reduction factor for temp step ***}

    evaluate ($final_t = 1000)                  { K }
    evaluate ($tempstep =  int(50/$red_f))      { K }

    evaluate ($ncycle = ($init_t-$final_t)/$tempstep)
!   evaluate ($nstep = int($cool_steps_1/$ncycle))
    evaluate ($nstep = int(($cool_steps_1*$red_f)/$ncycle))
    evaluate ($pstep = int($nstep/20))

    evaluate ($ini_rad = 1.28)         evaluate ($fin_rad = 1.0)
    evaluate ($ini_con = 0.012)        evaluate ($fin_con = 16.0)
    evaluate ($ini_ang = 0.25)         evaluate ($fin_ang = 1.0)
    evaluate ($ini_imp = 0.1)          evaluate ($fin_imp = 1.0)
    evaluate ($ini_dihe = 0.1)         evaluate ($fin_dihe = 1.0)
    evaluate ($ini_asy = 0.1)          evaluate ($fin_asy = 1.0)
    evaluate ($ini_flt = 25.0)         evaluate ($fin_flt = 500.0)
    evaluate ($ini_tenso = 0.001)     evaluate ($fin_tenso = 1.5 )

    evaluate ($k_vdw = $ini_con)
    evaluate ($k_vdwfact = ($fin_con/$ini_con)^(1/$ncycle))
    evaluate ($radius=    $ini_rad)
    evaluate ($radfact = ($fin_rad/$ini_rad)^(1/$ncycle))
    evaluate ($k_ang = $ini_ang)
```

```
evaluate ($ang_fac = ($fin_ang/$ini_ang)^(1/$ncycle))
evaluate ($k_imp = $ini_imp)
evaluate ($imp_fac = ($fin_imp/$ini_imp)^(1/$ncycle))
evaluate ($k_dihe = $ini_dihe)
evaluate ($dihe_fac = ($fin_dihe/$ini_dihe)^(1/$ncycle))
evaluate ($asym = $ini_asy)
evaluate ($asy_fac = ($fin_asy/$ini_asy)^(1/$ncycle))
evaluate ($k_flt = $ini_flt)
evaluate ($flt_fac = ($fin_flt/$ini_flt)^(1/$ncycle))
evaluate ($ktenso = $ini_tenso)
evaluate ($tenso_fac = ($fin_tenso/$ini_tenso)^(1/$ncycle))

do (vx = maxwell($bath)) (all)
do (vy = maxwell($bath)) (all)
do (vz = maxwell($bath)) (all)

evaluate ($i_cool = 1)
while ($i_cool < $ncycle) loop cool1
   evaluate ($i_cool=$i_cool+1)

   evaluate ($bath  = $bath  - $tempstep)
   evaluate ($k_vdw=min($fin_con,$k_vdw*$k_vdwfact))
   evaluate ($radius=max($fin_rad,$radius*$radfact))
   evaluate ($k_ang = $k_ang*$ang_fac)
   evaluate ($k_imp = $k_imp*$imp_fac)
   evaluate ($k_dihe = $k_dihe*$dihe_fac)
   evaluate ($asym  = $asym*$asy_fac)
   evaluate ($k_flt = $k_flt*$flt_fac)
   evaluate ($ktenso = $ktenso*$tenso_fac)

   tenso       ! assign $ktenso to actual force constant
      class pf1
         force $ktenso
   end

   igroup interaction (all) (all) weights
      * 1 angles $k_ang improper $k_imp dihedral $k_dihe
   end end
   parameter nbonds
      cutnb=7.0 rcon=$k_vdw nbxmod=5 repel=$radius
   end       end
   parameter
     angle (store3) (all) (store3) $k_flt TOKEN
     angle (all)    (all) (store3) $k_flt TOKEN
   end
   noe
       asymptote * $asym
   end

   { Do swaps }
   @PROTOCOLS:swap_metropolis.cns
     (
       toswap = methyl amido methylene;
       swap_sel=&swap_sel;
       bath=$bath;          {needs $bath}
       swap_stats=$swap_stats;
     )
   flags
     exclude *
     include angl bond impr vdw noe cdih dihe tenso
   end

   dynamics  cartesian
      nstep=$nstep  timestep=0.003
      tcoupling = true  temperature = $bath   nprint=$pstep
      ntrfr = 99999999
   end
```

```
    end loop cool1

    {* ================================================== second cooling stage *}

    evaluate ($final_t = 100)     { K }
    evaluate ($tempstep = 50)     { K }

    evaluate ($ncycle = ($bath-$final_t)/$tempstep)
    evaluate ($nstep = int($cool_steps_2/$ncycle))
    evaluate ($pstep = int($nstep/20))

    evaluate ($i_cool = 1)
    while ($i_cool < $ncycle) loop cool2
       evaluate ($i_cool=$i_cool+1)

       evaluate ($bath  = $bath  - $tempstep)

       { Do swaps }
       @PROTOCOLS:swap_metropolis.cns
         (
           toswap = methyl amido methylene;
           swap_sel=&swap_sel;
           bath=$bath;          {needs $bath}
           swap_stats=$swap_stats;
         )
      flags
         exclude *
         include angl bond impr vdw noe cdih dihe tenso
       end
       dynamics  cartesian
          nstep=$nstep  timestep=0.003
          tcoupling = true  temperature = $bath   nprint=$pstep
          ntrfr = 99999999
       end
    end loop cool2

    evaluate ($stats = 2 * $nstep)
    evaluate ($pstep = 2 * $pstep)

    dynamics  cartesian
          nstep=$stats  timestep=0.003
          tcoupling = true  temperature = $bath   nprint=$pstep
          ntrfr = 99999999
    end

{* ================ Powell energy minimization ======================== *}
{ Do swaps }
@PROTOCOLS:swap_downhill.cns
  (
    toswap = methyl amido methylene;
    swap_sel=&swap_sel;
    bath=$bath;          {needs $bath}
    swap_stats=$swap_stats;
  )
flags
  exclude *
  include angl bond impr vdw noe cdih dihe tenso
end

   mini powell nstep= 500 nprint= 50 end

{ Do swaps }
@PROTOCOLS:swap_downhill.cns
  (
    toswap = methyl amido methylene;
    swap_sel=&swap_sel;
    bath=$bath;          {needs $bath}
```

```
    swap_stats=$swap_stats;
  )
flags
  exclude *
  include angl bond impr vdw noe cdih dihe tenso
end

   mini powell nstep= 500 nprint= 50 end

{ Do swaps }
@PROTOCOLS:swap_downhill.cns
  (
    toswap = methyl amido methylene;
    swap_sel=&swap_sel;
    bath=$bath;          {needs $bath}
    swap_stats=$swap_stats;
  )
flags
  exclude *
  include angl bond impr vdw noe cdih dihe tenso
end

   mini powell nstep= 500 nprint= 50 end

{ lash up to avoid ids overflowing b & q fields of pdb and producing havoc in PROCHECK with q=0.0}
    do (b = 0.0) (all)
    do (q = 1.0) (all)

   evaluate ($violfile = $fileroot + encode($count) + ".viol_rdcs")
   set print $violfile end
     print threshold=0.3 noe
     evaluate ($rms_noe=$result)
     evaluate ($violations_noe=$violations)
     print threshold=5. cdih
     evaluate ($rms_cdih=$result)
     evaluate ($violations_cdih=$violations)
     print thres=0.018 bonds         {6*sigma Engh & Huber}
     evaluate ($rms_bond=$result)
     print thres=3.04 angles         {6*sigma Engh & Huber}
     evaluate ($rms_angl=$result)
     print thres=5. impropers
     evaluate ($rms_impr=$result)
     print thres=5. dihedrals
     evaluate ($rms_dihe=$result)
     tenso print threshold=1.00 class pf1 end
       evaluate ($rms_tenso=$result)
       evaluate ($viol_tenso=$violations)
  close $violfile end

  evaluate ($dipfile = $fileroot + encode($count) + ".dips_rdcs")
  set print $dipfile end
    tenso print threshold=0.0001 class pf1 end
  close $dipfile end

  set print_file=OUTPUT end

  remarks ============================================================
  remarks            overall,bonds,angles,improper,vdw,noe,cdih,dihe,tenso
  remarks energies: $ener, $bond, $angl, $impr, $vdw, $noe, $cdih, $dihe, $tens
  remarks ============================================================
  remarks            bonds,angles,impropers,noe,cdih,dihe,tenso
  remarks rms-d: $rms_bond,$rms_angl,$rms_impr,$rms_noe,$rms_cdih,$rms_dihe,$rms_tenso
  remarks ============================================================
  remarks               noe,  cdih, tenso
  remarks violations.: $violations_noe, $violations_cdih, $viol_tenso
  remarks ============================================================
  remarks seed = $seed
```

```
   evaluate ($file = "refine_rdcs" + $fileroot + "_" + encode($count) + ".pdb")
   write coor output= $file end

   evaluate ($count = $count + 1)
end loop stru
```

Bibliography

[1] BioMagResBank (BMRB) web site: http://www.bmrb.wisc.edu.

[2] P. Accardo, P. Sanchez-Corral, O. Criado, E. Garcia, and S. Rodriguez de Cordoba. Binding of human complement component C4b-binding protein (C4BP) to Streptococcus pyogenes involves the C4b-binding site. *J. Immunol.*, 157:4935–4939, 1996.

[3] P. Andersson, J. Weigelt, and G. Otting. Spin-state selection filters for the measurement of heteronuclear one-bond coupling constants. *J. Biomol. NMR*, 12:435–441, 1998.

[4] P. W. Atkins and R. S. Friedman. *Molecular Quantum Mechanics.* Oxford University Press, 1997.

[5] H. Jr. Auchincloss and D. H. Sachs. Xenotransplantation. *Annu. Rev. Immunol.*, 16:433–470, 1998.

[6] P. N. Barlow, A. Steinkasserer, D. G. Norman, B. Kieffer, A. P. Wiles, 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.

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

[8] A. Bax, G. Kontaxis, and N. Tjandra. Dipolar Couplings in Macromolecular Structure Determination. *Meth. Enzym.*, 339:127–174, 2001.

[9] M. Billeter, J. Vendrell, G. Wider, F. X. Aviles, M. Coll, A. Guasch, R. Huber, and K. Wüthrich. Comparison of the NMR solution structure with the X-ray crystal structure of the activation domain from procarboxypeptidase B. *J. Biomol. NMR*, 2:1–10, 1992.

[10] F. Bloch. Nuclear Induction. *Phys. Rev.*, 70:460–474, 1946.

[11] A. M. Blom, K. Berggard, J. H. Webb, G. Lindahl, B. O. Villoutreix, and B. Dahlbäck. Human C4b-Binding Protein Has Overlapping, But Not Identical, Binding Sites for C4b and Streptococcal M Proteins. *J. Immunol.*, 164:5328–5336, 2000.

[12] A. M. Blom, L. Kask, and B. Dahlbäck. Structural requirements for the complement regulatory activities of C4BP. *J. Biol. Chem.*, 276:27136–27144, 2001.

[13] A. M. Blom, J. Webb, B. O. Villoutreix, and B. Dahlbäck. 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.

[14] W. Boucher and Department of Biochemistry. Azara, v2.7, 1993. Department of Biochemistry at the University of Cambridge.

[15] W. Boucher and T. Stevens. CcpNmr Analysis v1.0.x, 2003. Department of Biochemistry at the University of Cambridge.

[16] W. G. Brodbeck, D. Liu, J. Sperry, C. Mold, and M. E. Medof. Localization of classical and alternative pathway regulatory activity within the decay-accelerating factor. *J. Immunol*, 156:2528–2533, 1996.

[17] A. T. Brünger, P. D. Adams, G. M. Clore, W. L. Delano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, and R. J. Read et al. Crystallography & NMR System. *Acta. Crystallogr. D Biol. Crystallogr.*, pages 905–921, 1998.

[18] E. Brunner. Residual Dipolar Couplings in Protein NMR. *Concepts in Magnetic Resonance*, 13:238–259, 2001.

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

[20] S. Cavagnero, H. J. Dyson, and P. E. Wright. Improved low pH bicelle system for orienting macromolecules over a wide temperature range. *J. Biomol. NMR*, 13:387–391, 1999.

[21] G. M. Clore, A. M. Gronenborn, and A. Bax. A Robust Method for Determining the Magnitude of the Fully Asymmetric Tensor of Oriented Macromolecules in the Absence of Structural Information. *J. Magn. Reson.*, 133:216–221, 1998.

[22] G. M. Clore, A. M. Gronenborn, and N. Tjandra. Direct Structure Refinement against Residual Dipoalr Couplings in the Presence of Rhombicity of Unknown Magnitude. *J. Magn. Reson.*, 131:159–162, 1998.

[23] G. M. Clore, M. R. Starich, and A. M. Gronenborn. Measurement of Residual Dipolar Couplings of Macromolecules Aligned in the Nematic Phase of a Colloidal Suspension of Rod-Shaped Viruses. *J. Am. Chem. Soc.*, 120:10571–10572, 1998.

[24] G. M. 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.

[25] G. Cornilescu and A. Bax. Determination of Relative N-H^N, N-C', C^α-C' and C^α-H^α Effective Bond Lengths in a Protein by NMR in a Dilute Liquid Crystalline Phase. *J. Am. Chem. Soc.*, 122:10143–10154, 2000.

[26] B. Dahlbäck, C. A. Smith, and H. J. Müller-Eberhard. Visualization of human C4b-binding protein and its complexes with vitamin K-dependent protein S and complement protein C4b. *Proc. Natl. Acad. Sci. U.S.A.*, 80:3461–3465, 1983.

[27] J. Garcia de la Torre, M. L. Huertas, and B. Carrasco. Calculation of Hydrodynamic Properties of Globular Proteins from Their Atomic-Level Structure. *Biophys. J.*, 78:719–730, 2000.

[28] H. Demene, T. Ducat, P. Barthe, M. Delsuc, and C. Roumestand. Structure refinement of flexible proteins using dipolar couplings: Application to the protein $p8^{MTCP1}$. *J. Biomol. NMR*, 22:47–56, 2003.

[29] K. Ding and A. M. Gronenborn. Sensitivity-enhanced IPAP experiments for measuring one-bond ^{13}C-$^{13}C^\alpha$ and $^{13}C^{\alpha 1}H^\alpha$ residual dipolar couplings in proteins. *J. Magn. Reson.*, 167:253–258, 2004.

[30] P. Dosset, J-C. Hus, and M. Blackledge. A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings. *J. Biomol. NMR*, 20:223–231, 2001.

[31] P. Dosset, J-C. Hus, M. Blackledge, and D. Marion. Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J. Biomol. NMR*, 16:23–28, 2000.

[32] L. Emsley and G. Bodenhausen. Gaussian pulse cascades: New analytical functions for rectangular selective inversion and in-phase excitation in NMR. *Chem. Phys. Lett.*, 165:469–476, 1990.

[33] R. R. Ernst, G. Bodenhausen, and A. Wokaun. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions.* Oxford University Press, 1997.

[34] M. W. F. Fischer, J. A. Losonczi, J. L. Weaver, and J. H. Prestegard. Domain Orientation and Dynamics in Multidomain Proteins from Residual Dipolar Couplings. *Biochemistry*, 38:9013–9022, 1999.

[35] K. Fleming, D. Gray, S. Prasannan, and S. Matthews. Cellulose Crystallites: A New and Robust Liquid Crstalline Medium for the Measurement of Residual Dipolar Couplings. *J. Am. Chem. Soc.*, 122:5224–5225, 2000.

[36] J. R. Garbow, D. P. Weitekamp, and A. Pines. Bilinear rotation decoupling of homonuclear scalar interactions. *Chem. Phys. Lett.*, 93:504–508, 1982.

[37] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, and A. Bairoch. *Protein Identification and Analysis Tools on the ExPASy Server.* Humana Press, 2005.

[38] S. Grzesiek and A. Bax. Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *J. Magn. Reson.*, 96:432–440, 1992.

[39] S. Grzesiek and A. Bax. Amino acid type determination in the sequential assignment procedure of uniformly ^{13}C,^{15}N -enriched proteins. *J. Biomol. NMR*, 3:185–204, 1993.

[40] M. R. Hansen, L. Mueller, and A. Pardi. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat. Struct. Biol.*, 5:1065–1074, 1998.

[41] C. E. Henderson, K. Bromek, N. P. Mullin, B. O. Smith., D. Uhrin, and P. N. Barlow. Solution Structure and Dynamics of the Central CCP Module Pair of a Poxvirus Complement Control Protein. *J. Mol. Biol.*, 307:323–339, 2001.

[42] M. Hessing, R. A. A. Vlooswijk, T. M. Hackeng, D. Kanters, and B. N. Bouma. The Localization of Heparin-binding Fragments on Human C4b-binding Protein. *J. Immunol.*, 144:204–208, 1990.

[43] T. K. Hitchens, S. A. McCallum, and G. S. Rule. A $^{1}J_{C^{\alpha}H^{\alpha}}$ -Modulated 2D (HACACO)NH Pulse Scheme for Quantitative Measurement of $C^{\alpha}H^{\alpha}$ Couplngs in ^{15}N , ^{13}C -Labeled Proteins. *J. Magn. Reson.*, 140:281–284, 1999.

[44] W. Hu, Z. Zhang, and Y. Chen. A high sensitivity 3D experiment for measuring $C^{\alpha}H^{\alpha}$ residual dipolar coupling constants. *J. Magn. Reson.*, 165:248–252, 2003.

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

[46] R. Ishima and D. Torchia. Protein dynamics from NMR. *Nat. Struct. Biol.*, 7:740–743, 2000.

[47] R. Jerala and G. S. Rule. A Sensitive Triple-Resonance Pulse Sequence to Elucidate Correlations between H^{i}_{α}, N^{i+1} and H^{i+1}_{N}. *J. Magn. Reson.*, 1993.

[48] M. Karplus and J. A. McCammon. Dynamics of Proteins: Elements and Function. *Ann. Rev. Biochem.*, 53:263–300, 1983.

[49] L. E. Kay. Protein dynamics from NMR. *Nat. Struct. Biol.*, 5:513–517, 1998.

[50] L. E. Kay, G. Y. Yu, and T. Yamazaki. Enhanced-sensitivity triple-resonance spectroscopy with minimal H_2O saturation. *J. Magn. Reson.*, pages 129–133, 1994.

[51] J. Keeler. *Understanding NMR Spectroscopy.* Wiley, 2005.

[52] M. D. Kirkitadze and P. N. Barlow. Structure and flexibility of the multidomain proteins that serve as regulators of complement activation. *Immunol. Rev.*, 180:146–161, 2001.

[53] G. Kontaxis, G. M. Clore, and A. Bax. Evaluation of Cross-Correlation Effects and Measurement of One-Bond Couplings in Proteins with Short Transverse Relaxation Times. *J. Magn. Reson.*, 143:184–196, 2000.

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

[55] P. J. Kraulis. ANSIG: A Program for the Assignment of Protein 1H 2D NMR spectra by Interactive Graphics. *J. Magn. Reson.*, 24:627–633, 1989.

[56] L. A. Kuttner-Kondo, L. Mitchell, D. E. Hourcade, and M. E. Medof. Characterization of the Active Sites in Decay-Accelerating Factor. *J. Immunol.*, 167:2164–2171, 2001.

[57] T. C. Leeper, M. B. Martin, H. Kim, S. Cox, V. Semenchenko, F. J. Schmidt, and S. R. Van Doren. Structure of the UGAGAU hexaloop that braces Bacillus RNase P for action. *Nat. Struct. Biol.*, 9:397–403, 2002.

[58] F. Lin, R. M. Immormino, M. Shoham, and M. E. Medof. Bulk production and functional analyses of mouse CD55's native and deglycosylated active domains. *Arch. Biochem. Biophys.*, 393:67–72, 2001.

[59] G. Lindahl, U. Sjobring, and E. Johnsson. Human complement regulators: a major target for pathogenic microorganisms. *Curr. Opin. Immunol.*, 12:44–51, 2000.

[60] K. Lindorff-Larsen, R. B. Best, M. A. DePristo, C. M. Dobson, and M. Vendruscolo. Simultaneous determination of protein structure and dynamics. *Nature*, 433:128–132, 2005.

[61] G. Lipari and A. Szabo. Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1. Theory and Range of Validity. *J. Am. Chem. Soc.*, 104:4546–4559, 1982.

[62] 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.

[63] J. A. Losonczi, M. A. Andrec, M. W. F. Fischer, and J. H. Prestegard. Order Matrix Analysis of Residual Dipolar Couplings Using Singular Value Decomposition. *J. Magn. Reson.*, 138:334–342, 1999.

[64] P. Lukacik, 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, R. A. G. Smith, and S. M. Lea. Complement regulation at the molecular level: The structure of decay-accelerating factor. *Proc. Natl. Acad. Sci. U.S.A.*, 101:1279–1284, 2004.

[65] Z. Luz and S. Meiboom. Nuclear Magnetic Resonance Study of the Protolysis of Trimethylammonium Ion in Aqueous SolutionOrder of the Reaction with Respect to Solvent. *J. Chem. Phys.*, 39:366–370, 1963.

[66] M. E. Medof, T. Kinoshita, and V. Nussenzweig. Inhibition of Complement Activation on the Surface of Cells After Incorporation of Decay-accelerating factor (DAF) Into Their Membranes. *J. Exp. Med.*, 160:1558–1578, 1984.

[67] N. A. Meenan, A. Cooper, M. W. Kennedy, and B. O. Smith. Resonance assignment of ABA-1A, from Ascaris suum nematode polyprotein allergen. *J. Biomol. NMR*, 32:176, 2005.

[68] S. Meiboom and D. Gill. Modified spin-echo method for measuring nuclear relaxation times. *Rev. Sci. Instr.*, 29:688–691, 1958.

[69] A. M. Mendel, M. Akke, and A. G. Palmer III. Backbone Dynamics of Escherichia coli Ribonuclease HI: Correlations with Structure and Function in an Active Enzyme. *J. Mol. Biol.*, 246:144–163, 1995.

[70] S. Moltke and S. Grzesiek. Structural constraints from residual tensorial couplings in high resolution NMR without an explicit term for the alignment tensor. *J. Biomol. NMR*, 246:77–82, 1999.

[71] J. Moore, L. McDermott, N. C. Price, S. M. Kelly, A. Cooper, and M. W. Kennedy. Sequence-divergent units of the ABA-1 polyprotein array of the nematode *Ascaris suum* have similar fatty-acid- and retionl-binding properties but different binding site environments. *Biochem. J.*, 340:337–343, 1999.

[72] D. R. Muhandiram and L. E. Kay. Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. *J. Magn. Reson.*, 103, 1994.

[73] K. H. Murthy, S. A. Smith, V. K. Ganesh, K. W. Judge, N. Mullin, P. N. Barlow, C. M. Ogata, and G. J. Kotwal. Crystal structure of a complement control protein that regulates both pathways of complement activation and binds heparan sulfate proteoglycans. *Cell*, 104:301–311, 2001.

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

[75] M. Ottiger and A. Bax. Determination of Relative N-H^N, N-C', C^α-C' and C^α-H^α Effective Bond Lengths in a Protein by NMR in a Dilute Liquid Crystalline Phase. *J. Am. Chem. Soc.*, 120:12334–12341, 1998.

[76] M. Ottiger and A. Bax. Bicelle-based liquid crystals for NMR-measurement of dipolar couplings at acidic and basic pH values. *J. Biomol. NMR*, 13:187–191, 1999.

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

[78] A. G. Palmer III. r2r1_diffusion, 1996. Department of Biochemistry and Molecular Biophysics at Columbia University.

[79] A. G. Palmer III. NMR Probes of Molecular Dynamics: Overview and Comparison with Other Techniques. *Annu. Rev. Biophys. Biomol. Struct.*, 30:129–155, 2001.

[80] P. Permi. Measurement of residual dipolar couplings from $^1H^\alpha$ to $^{13}C^\alpha$ and ^{15}N using a simple HNCA-based experiment. *J. Biomol. NMR*, 27:341–349, 2003.

[81] P. Permi, S. Heikkinen, I. Kilpeläinen, and A. Annila. Measurement of $^1J_{NC'}$ and $^2J_{NC'}$ Couplings from Spin-State-Selective Two-Dimensional Correlation Spectrum. *J. Magn. Reson.*, 140:32–40, 1999.

[82] P. Permi, I. Kilpeläinen, and A. Annila. Determination of Backbone Angle ψ in Proteins Using a TROSY-Based α/β-HN(CO)CA-J Experiment. *J. Magn. Reson.*, 146:255–259, 2000.

[83] P. Permi, T. Sorsa, I. Kilpeläinen, and A. Annila. HN(α/β-COCA-J) Experiment for Measurement of $^1J_{C'\alpha}$ Couplings from Two-Dimensional ^{15}N,^1H) Correlation Spectrum. *J. Magn. Reson.*, 141:44–51, 1999.

[84] W. Peti, J. Meiler, R. Bruschweiler, and C. Griesinger. Model-Free Analysis of Protein Backbone Motion from Residual Dipolar Couplings. *J. Am. Chem. Soc.*, 124:5822–5833, 2002.

[85] M. Piotto, V. Saudek, and V. Sklenar. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR*, 2:661–665, 1992.

[86] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C. 2^{nd} ed.* Cambridge Univaersity Press, 1997.

[87] J. H. Prestegard, H. M. Al-Hashimi, and J. R. Tolman. NMR structures of biomolecules using field oriented media and residual dipolar couplings. *Q. Rev. Biophys.*, 33:371–424, 2000.

[88] J. H. Prestegard, C. M. Bougalt, and A. I. Kishore. Residual Dipolar Couplings in Structure Determination of Biomolecules. *Chem. Rev.*, 104:3519–3540, 2004.

[89] R. S. Prosser, S. A. Hunt, J. A. DiNatale, and R. R. Vold. Magnetically Aligned Membrane Model Systems with Positive Order Parameter: Switching the Sign of S_{zz} with Paramagnetic Ions. *J. Am. Chem. Soc.*, 118:269–270, 1996.

[90] R. S. Prosser, J. A. Losonczi, and I. V. Shiyanovskaya. Use of a Novel Aqueous Liquid Crystalline Medium for High-Resolution NMR of Macromolecules in Solution. *J. Am. Chem. Soc.*, 120:11010–11011, 1998.

[91] K. B. M. Reid and A. J. Day. Structure-function relationships of the complement components. *Immunol. Today*, 10:177–180, 1988.

[92] M. Rückert and G. Otting. Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments. *J. Am. Chem. Soc.*, 122:7793–7797, 2000.

[93] H. Saas, G. Musco, S. J. Stahl, P. T. Wingfield, and S. Grzesiek. Solution NMR of proteins within polyacrylamide gels: Diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes. *J. Biomol. NMR*, 18:303–309, 2000.

[94] H. Saas, G. Musco, S. J. Stahl, P. T. Wingfield, and S. Grzesiek. An easy way to include weak alignment constraints into NMR structure calculations. *J. Biomol. NMR*, 21:275–280, 2001.

[95] C. R. Sanders and J. P. Schwonek. Characterization of magnetically orientable bilayers in mixtures of dihexanoylphosphatidylcholine and dimyristoylphosphatidylcholine by solid-state NMR. *Biochemistry*, 31:8898–8905, 1992.

[96] A. Saupe. Recent results in the field of liquid crystals. *Angew. Chem. Intern. Engl.*, 7:97–112, 1968.

[97] A. Saupe and G. Englert. High-resolution Nuclear Magnetic Resonance Spectra of Orientated Molecules. *Phys. Rev. Lett.*, 11:462–464, 1963.

[98] J. Scharfstein, A. Ferreira, I. Gigli, and V. Nussenzweig. Human C4-binding protein. 1. Isolation and characterization. *J. Exp. Med.*, 148:207–222, 1978.

[99] J. M. Schurr, H. P. Babcock, and B. S. Fujimoto. A Test of the Model-Free Formulas. Effects of Anisotropic Rotational Diffusion and Dimerization. *J. Magn. Reson.*, 105, 1994.

[100] N. R. Skrynnikov, N. K. Goto, D. Yang, W. Y. Choy, J. R. Tolman, G. A. Müller, and L. E. Kay. Orienting Domains in Proteins Using Dipolar Couplings Measured by Liquid-state NMR: Differences in Solution and Crystal Forms of Maltodextrin Binding Protein Loaded with β-Cyclodextrin. *J. Mol. Biol.*, 295:1265–1273, 2000.

[101] B. O. Smith, R. L. Mallin, M. Krych-Goldberg, X. Wang, R. E. Hauhart, K. Bromek, D. Uhrin, J. P. Atkinson, and P. N. Barlow. Structure of the C3b binding site of CR1 (CD35), the immune adherence receptor. *Cell*, 108:769–780, 2002.

[102] I. Solomon. Relaxation Processes in a System of Two Spins. *Phys. Rev.*, 99:559–565, 1955.

[103] N. Tjandra and A. Bax. Direct Measurement of Distances and Angles in Biomolecules by NMR in a Dilute Liquid Crystalline Medium. *Science*, 278:1111–1114, 1997.

[104] N. Tjandra and A. Bax. Measurement of Dipolar Contributions to $^1J_{C^\alpha H^\alpha}$ Splittings from Magnetic-Field Dependence of J Modulation in Two-Dimensional NMR Spectra. *J. Magn. Reson.*, 124:512–515, 1997.

[105] N. Tjandra, S. E. Feller, R. W. Pastor, and A. Bax. Rotational diffusion anisotropy of human ubiquitin from ^{15}N NMR relaxation. *J. Am. Chem. Soc.*, 117:12562–12566, 1995.

[106] N. Tjandra, J. G. Omichinski, A. M. Gronenborn, G. M. Clore, and A. Bax. Use of dipolar ^{1}H-^{15}N and ^{1}H^{13}C couplings in the structure determination of magnetically oriented macromolecules in solution. *Nat. Struct. Biol.*, 4:732–728, 1997.

[107] J. R. Tolman, H. Al-Hashimi, L. E. Kay, and J. H. Prestegard. Structural and Dynamic Analysis of Residual Dipolar Coupling Data for Proteins. *J. Am. Chem. Soc.*, 123:1416–1424, 2001.

[108] J. R. Tolman, J. M. Flanagan, M. A. Kennedy, and J. H. Prestegard. Nuclear magnetic dipolar ineractions in field-oriented proteins: information for structure determination in solution. *Proc. Natl. Acad. Sci. U.S.A.*, 92:9279–9283, 1995.

[109] J. R. Tolman, J. M. Flanagan, M. A. Kennedy, and J. H. Prestegard. NMR evidnece for slow collective motions in cyanometmyoglobin. *Nat. Struct. Biol.*, 4:292–297, 1997.

[110] R. Tycko, F. J. Blanco, and Y. Ishii. Alignment of Biopolymers in Strained Gels: A New Way To Create Detectable Dipole-Dipole Couplings in High-Resolution Biomolecular NMR. *J. Am. Chem. Soc.*, 122:9340–9341, 2000.

[111] D. Uhrin, T. Liptaj, and K. E. Köver. Modified BIRD Pulses and Design of Heteronuclear Pulse Sequences. *J. Magn. Reson. Ser. A*, 122:41–46, 1993.

[112] S. Uhrinova, F. Lin, G. Ball, K. Bromek, D. Uhrin, M. E. Medof, and P. N. Barlow. Solution structure of a functionally active component of decay accelerating factor. *Proc. Natl. Acad. Sci. U.S.A.*, 100:4718–4723, 2003.

[113] S. Uhrinova, F. Lin, D. Uhrin, M. E. Medof, and P. N. Barlow. Resonance assignments of the central complement control protein module pair of human decay accelerating factor. *J. Biomol. NMR*, 23:167–168, 2003.

[114] F. J. M. van de Ven. *Multidimensional NMR in Liquids.* Wiley-VCH, 1995.

[115] P. Williams, Y. Chaudhry, I. G. Goodfellow, J. Billington, R. Powell, O. B. Spiller, D. J. Evans, and S. Lea. Mapping CD55 Function: The Structure of Two Pathogen-Binding Domans at 1.7Å. *J. Biol. Chem.*, 278:10691–10696, 2003.

[116] D. E. Woessner. Nuclear Spin Relaxation in Ellipsoids Undergoing Rotational Brownian Motion. *J. Chem. Phys*, 37:647–654, 1962.

[117] K. Wüthrich. Protein structure determination in solution by nuclear magnetic resonance spectroscopy. *Science*, 243:45–50, 1989.

[118] D. Yang, J. R. Tolman, N. K. Goto, and L. E. Kay. An HNCO-based Pulse Scheme for the Measurement of 13Cα-1Hα One-bond Dipolar couplings in 15N, 13C Labeled Proteins. *J. Biomol. NMR*, 12:325–332, 1998.

[119] M. Zweckstetter and A. Bax. Characterization of molecular alignment in aqueous suspensions of Pf1 bacteriophage. *J. Biomol. NMR*, 20:365–377, 2001.

[120] M. Zweckstetter and A. Bax. Single-Step Determination of Protein Substructures Using Dipolar Couplings: Aid To Structural Genomics. *J. Am. Chem. Soc.*, 123:9490–9491, 2001.