Real-time NMR of the Transient States of Proteins

A thesis submitted for the degree of

Doctor of Philosophy

by

Iain J. Day

Hertford College

Michaelmas Term 2004

Real-time NMR of the Transient States of Proteins

A thesis submitted for the degree of Doctor of Philosophy

Iain J. Day
Hertford College
Michaelmas Term 2004

Abstract

The work described in this thesis is concerned with the development and application of real-time photo-CIDNP (Chemically Induced Dynamic Nuclear Polarisation) to the study of protein structure and folding.

Chapters 1 and 2 introduce the protein folding problem, and its study by NMR, then go on to elucidate the mechanisms behind the photo-CIDNP phenomenon.

Chapter 3 applies photo-CIDNP spectroscopy to the study of a small cytochrome protein. The difficulties of performing these experiments on chromophore-containing proteins are discussed.

Chapter 4 begins with the development of a rapid mixing device for use in real-time NMR and CIDNP studies. Experiments used to characterise the device are presented. This chapter then goes on to describe CIDNP pulse labelling experiments, used to investigate the surface structure of some molten globule states of two α-lactalbumins. This chapter concludes with an application of the rapid mixing device to the real-time refolding of hen egg white lysozyme.

Chapter 5 extends the work of the previous chapter, studying the real-time refolding of bovine pancreatic ribonuclease A. Refolding studies are performed from different denaturing conditions, and the effects of sample heating during the real-time CIDNP experiment are discussed.

Chapter 6 describes the use of illumination during an NMR experiment to study the conformational changes in a plant blue light receptor protein, phototropin. The structural changes are characterised with 2-dimensional NMR spectroscopy and photo-CIDNP. The kinetics of the ground state recovery are also investigated by real-time NMR spectroscopy.

Chapter 7 uses calculated hyperfine coupling constants and a radical pair diffusion model from the literature to simulate the nuclear polarisation obtained for the amino acid tryptophan. Comparisons are made between theory and experiment.

Chapter 8 describes the structural characterisation of a homologous series of *de novo* peptides, designed for subsequent use in EPR experiments when derivatised with a suitable spin label.

Acknowledgements

First and foremost, it is an immense pleasure to thank Prof. Peter Hore for providing interesting problems and a great group to work with. I am grateful for the freedom to pursue such a wide variety of projects and for his trusting me with the running of the new spectrometer.

There are a number of people who have helped me immensely over the past three years, without whom I wouldn't be writing this! I am extremely grateful to Dr. Ken Hun Mok for being a constant source of advice, inspiration and always being able to answer my questions with good humour. Dr. Toshio Nagashima was a great help in the early days of my work, teaching me about NMR spectrometers, lasers and injectors. Dr. Jonathan Jones has provided numerous answers to even the most naïve questions about gradients, product operators, pulse sequences and spectrometers. I'm still not sure I understand all the answers!

I am extremely grateful for the expert technical knowledge and experience of Nick Soffe, Dr. Johnathan Boyd and Dr. Christina Redfield for assistance in using the Omega spectrometers in the Rex Richards Building. More recently, I would like to thank Andrew Martin, Chris Kember and Dr. Paul Bowyer of Varian Inc. for their expert installation of our new spectrometer, patiently answering my questions and teaching me about the inner workings of an NMR console.

In the PTCL there are several people who deserve a mention, especially Charlie Jones of the mechanical workshops, who constructed the injection device used in Chapter 4, and was always good humoured when we had numerous leaks and problems. I thank Paul Mitchell and the lab services workshop for tirelessly working to help us build the new spectrometer room. In the stores, John Crews and Mervyn Rees run an emporium of "useful things", and always had something useful, even if I couldn't explain what I needed. Also I would like to thank Dr. Pete Biggs for keeping all the various computers running smoothly.

On the work front, it has been a great pleasure to collaborate with several people, both in Oxford and further afield. The cytochrome work in Chapter 3 was performed with the kind assistance of Rachel Wain and Dr. Lorna Smith. Chapter 6 describes the phototropin work, which would not have been possible without Dr. Kevin Gardner of UT Southwestern Medical Center in Dallas. I thank him for allowing me to work on an extremely interesting project and for a great two weeks spent in Dallas. For that I must also thank Shannon Harper and the rest of the Gardner group for looking after me in Dallas, and Shanon and Lori Neil for fun times at the ICMRBS in Toronto. I'd like to thank Janet Banham for asking me to run "a quick NMR spectrum" on her peptides, which turned into a challenging project. I am also grateful to Dr. Jakob Lopez for inviting me to Frankfurt, and to Melanie Hertel and Prof. Dr. Clemens Glaubitz for their hospitality.

I would particularly like to thank the members of the Hore and Timmel groups over the past three years, for cakes, pub challenges, Christmas dinners, conferences, and much more. In alphabetical order: Janet Banham, Maša Čemažar, Filippo Cintolesi, Ailsa Curtis, Rob Evans, Kate Hall, Kevin Henbest, Ilya Kuprov, Paul McKeating, Ken Hun Mok, Toshio Nagashima, Stuart Norman, Anthony O'Dea, Howard Paisley, Chris Rodgers, Liz Sutton, Phil Thomas, Chris Timmel and Nicola Wagner. It was great working with you. I am especially grateful to Nic, Chris and Ken for proof-reading this thesis and spotting countless errors.

I must also mention my friends in and around Oxford, and the residents of 189 Banbury road for an interesting and lively three years. My family: Mum, Dad and David have provided a huge amount of support, not only over the last three years, but over the past 25! Finally, I must thank Lucy for being an amazing and wonderful person, thank you!

Contents

Chapter 1

Introduction

Protein folding has received a large amount of attention recently [1–5], following the discovery of a link between protein misfolding and disease [4]. There are currently more than 15 human diseases associated with the misfolding and aggregation of proteins *in vivo* [4]. The causes of protein misfolding are still largely not understood. This thesis is concerned with the development of real-time NMR methodology and its application to the study of protein structure, folding and conformational changes.

This chapter outlines the background of the protein folding problem, then goes on to introduce the basic theory of nuclear magnetic resonance and the CIDNP phenomenon. An outline of the remainder of the thesis is then presented.

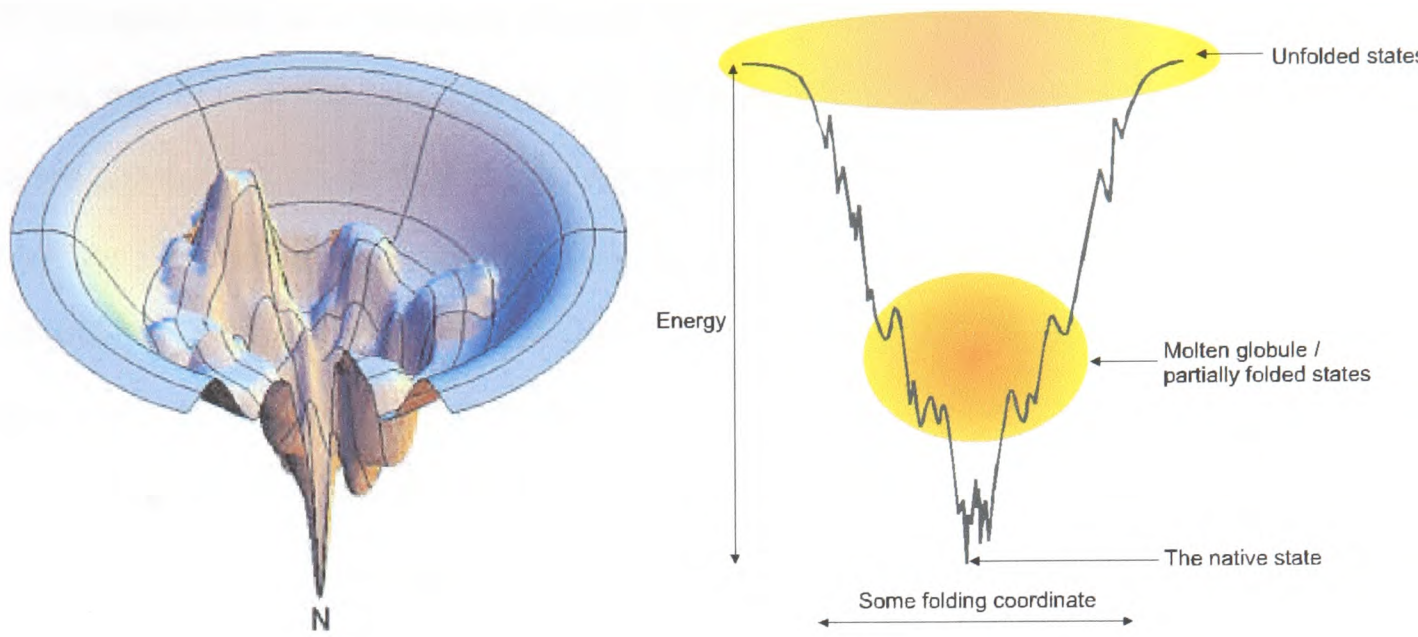
1.1 The Protein Folding Problem

How does a linear chain of amino acids fold to form a well defined, three dimensional structure following synthesis on the ribosome? This is one of the major puzzles of modern biochemistry and molecular biology. If a protein were to attempt an exhaustive search of all possible conformations in order to find the correct native fold, it would take a geological timescale! This is the basis of Levinthal's paradox [6, 7]. In reality, proteins fold on a timescale ranging from microseconds

for the Trp-cage peptide [8] and the B-domain of protein A [9, 10] up to tens to hundreds of seconds for larger, more complex systems [11]. There are three major mechanisms proposed for protein folding. The *framework model* states that local elements of secondary structure are formed independently of the tertiary structure. These secondary structure elements then coalesce to form the native state [12–14]. The *nucleation model* argues that neighbouring residues form native contacts, and hence secondary structure, and act as nucleation sites from which folding occurs in a stepwise manner [15, 16]. The third proposed mechanism is the *hydrophobic collapse*. In this postulate, the hydrophobic residues are quickly sequestered from the solvent, followed by a reorganisation of the protein to adopt the native fold [17, 18]. In reality protein folding *in vitro* is likely to be a subtle combination of these three models. *In vivo*, the presence of large molecular structures, called chaperones, aid the protein in its attainment of the correct functional fold and act as a cellular quality control mechanism [2].

1.1.1 The Free Energy Landscape

The concept of a free energy landscape has long been used in the explanation and interpretation of chemical reaction dynamics. Recently, the idea has been applied to the protein folding problem [19]. Figure 1.1 shows a schematic free energy funnel associated with protein folding, and a slice through that funnel. The native state of the protein is found at the global minimum of this free energy landscape, with partially folded species such, as molten globules often occupying local minima *en route* to this native state. Clearly any refolding experiment which starts from the denatured state will initially span an ensemble of conformational states which must all eventually end up at the native state.



(a) A schematic free energy funnel, taken from [19]

(b) A slice through the funnel

Figure 1.1: A representation of the free energy landscape associated with protein folding.

1.1.2 Following Protein Folding

The range of experimental techniques used to study proteins and follow their folding is vast. The biophysical techniques applied include UV/visible absorption spectroscopy, stopped-flow fluorescence, Fluorescence Resonance Energy Transfer (FRET), Circular Dichroism (CD), Infrared Spectroscopy (IR), Differential Scanning Calorimetry (DSC), Nuclear Magnetic Resonance (NMR) and Electron Paramagnetic Resonance (EPR) to name only the major methods.

A variety of real-time NMR methods have been used in the study of protein folding [20–23]. The principal approach is the use of one-dimensional NMR spectroscopy, due to the time constraints required to acquire a two-dimensional spectrum. Examples include the slow unfolding of various mutants of barnase and chymotryspin inhibitor 2 [24] in which unfolding is initiated by injection of native protein into a high concentration of chemical denaturant. The unfolding was followed over a period of hundreds of seconds by observing changes in the methyl and amide regions

of the spectrum as a function of time following the injection. A second example shows the use of ^{19}F labelling in the refolding of dihydrofolate reductase [25] followed again by stopped-flow NMR over hundreds of seconds. The advantage of using fluorine-19 labelled tryptophan residues is the large spectral simplification achieved, allowing truly site specific data to be obtained, along with the greater chemical shift dispersion offered by the ^{19}F nuclei.

The sensitivity of these one-dimensional experiments can be improved by application of CIDNP (Chemically Induced Dynamic Nuclear Polarisation). Since only histidine, tyrosine and tryptophan residues are observed in the CIDNP experiment, this also allows site specific information to be obtained without the need for potentially complex isotopic labelling strategies. Several proteins have been studied by this technique, including hen egg white lysozyme, on which real-time stopped-flow folding studies have been performed [26, 27]. Also, by production of a nitrogen-15 labelled variant, a two-dimensional CIDNP experiment allowed more detailed information about the number of solvent exposed tryptophan residues in the denatured state to be obtained [28]. A classic system for the study of protein folding is bovine α-lactalbumin. This has recently been studied with respect to the folding of the *apo* protein following the photochemical release of calcium ions into solution [29, 30]. Photo-CIDNP has also been implemented as a technique to probe the structure of a partially folded state of α-lactalbumin, by the transfer of nuclear polarisation from the partially folded state to the native state for detection [31, 32]. Site directed mutagenesis allows the replacement of specific amino acids in a protein sequence. In the case of the histidine-containing phosphocarrier protein from *Escherichia coli*, phenylalanine residues were selectively replaced with tryptophan to allow a stopped-flow CIDNP investigation [33].

More recently, two-dimensional techniques have been applied to slower folding proteins such as ribonuclease T1 [34] which was followed using a fast ^{15}N-HMQC

method allowing changes at the backbone amides to be followed. Over much longer time periods it is possible to record a series of conventional ^{15}N-HSQC spectra to follow folding. An example of this is the folding of the β-sheet protein apoplastocyanin [35]. The LOV2 domain from the signalling protein phototropin undergoes a marked conformational change upon exposure to blue light [36, 37]. The ability to trigger a reversible cycle between conformational states allows a modified time-resolved ^{15}N-HSQC experiment to be performed [38].

The remainder of this chapter outlines the basis of the NMR experiment, and the origins of the photo-CIDNP effect, in the context of proteins and biological systems.

1.2 Nuclear Magnetic Resonance

Atomic nuclei possess nuclear spin angular momentum which can be characterised by a nuclear spin quantum number, I, which can take integer values, including zero, or half-odd integer values. Common nuclear spin quantum numbers include $I = 0$ for ^{12}C and ^{16}O, $I = \frac{1}{2}$ for ^{1}H, ^{13}C, ^{15}N and ^{19}F, and $I = 1$ for ^{2}H, ^{14}N. There are $2I + 1$ projections (m_I) of this spin angular momentum onto the z axis, defined by the following series:

$$m_I = I, I-1, I-2, \cdots, -I \tag{1.1}$$

In the absence of a magnetic field, these projections are degenerate, however, the application of a magnetic field lifts this degeneracy, with the energy of each state being given by:

$$E(m_I) = -\hbar\gamma B_0 m_I \tag{1.2}$$

where γ is the magnetogyric ratio, defined as $\vec{\mu} = \gamma \cdot \vec{I}$, and $\vec{\mu}$ is the nuclear magnetic moment vector and $\vec{I}$ is the nuclear spin vector. B_0 is the applied static magnetic field. The separation of the nuclear spin states by the magnetic field will also result in a population difference between the levels at thermal equilibrium. Consider the

case when $I = \frac{1}{2}$: this implies that there are two spin states, commonly referred to as the α state ($m_I = +\frac{1}{2}$) and the β state ($m_I = -\frac{1}{2}$). In the magnetic field there will be a population difference between these two states given by the Boltzmann distribution:

$$\frac{n_\beta}{n_\alpha} = \exp\left(-\frac{E_\beta - E_\alpha}{k_B T}\right) \tag{1.3}$$

where n_β/n_α is the ratio of the populations of the α and β spin states, with energies E_α and E_β respectively.

The selection rule in magnetic resonance is that $\Delta m_I = \pm 1$, therefore resonant transitions will occur when the frequency of the applied radiation matches the energy gap between spin states separated by $\Delta m_I = \pm 1$. Therefore:

$$\omega = -\gamma B_0 \tag{1.4}$$

where ω is known as the nuclear Larmor frequency. The frequency of electromagnetic radiation required to induce these transitions is usually in the radio frequency range. Nuclear magnetic resonance was first demonstrated experimentally in 1946 by the groups of Bloch [39, 40] and Purcell [41]. The experiments described in this thesis were performed using modern Fourier Transform methods, the basics of which are outlined below.

1.2.1 Fourier Transform NMR

Since there is a population difference between the nuclear spin states in the magnetic field, this will result in a small bulk magnetisation in the sample, proportional to the expectation value $\langle I_z \rangle$. This is so-called longitudinal magnetisation. It can be "tipped" into the transverse (xy) plane by the application of a short, strong radio frequency field (B_1), perpendicular to the main static field. The flip angle (in

degrees) of the pulse is defined as:

$$\theta = -\frac{\gamma B_1}{2\pi}\tau_p \times 360^\circ \tag{1.5}$$

where τ_p is the duration of the pulse. For example, a 90° pulse applied along the x-axis, in a reference frame rotating at the spectrometer transmitter frequency, will result in the magnetisation vector being tipped into the transverse plane, and produce $-I_y$ magnetisation. Observed from this rotating frame the transverse magnetisation will then precess about the static field B_0, at a frequency corresponding to the difference between the Larmor frequency of the nucleus of interest and the spectrometer transmitter frequency. This frequency difference is known as the offset. As this magnetisation precesses it is detected in quadrature, resulting in a complex signal, or Free Induction Decay (FID), of the following form:

$$s(t) = \exp(\mathrm{i}\Omega t)\exp(-R_2 t) \tag{1.6}$$

with Ω being the offset, and R_2 is the transverse relaxation rate ($R_2 = \frac{1}{T_2}$), which governs the decay of the FID envelope.

The free induction decay is related to the NMR spectrum by a Fourier transform [42]. The spectrum is obtained by computation of the following integral:

$$s(\omega) = \int_0^\infty s(t)\exp(\mathrm{i}\omega t)\mathrm{d}t \tag{1.7}$$

$$= \frac{R_2}{R_2^2 + (\omega - \Omega)^2} - \frac{\mathrm{i}(\omega - \Omega)}{R_2^2 + (\omega - \Omega)^2} \tag{1.8}$$

The real part of this expression corresponds to an in-phase absorptive Lorentzian line shape, whilst the imaginary part is an in-phase dispersive Lorentzian.

The principal advantage of Fourier transform NMR over continuous wave (CW) methods is that several FT experiments can be performed in the same time required

to acquire a single CW spectrum, thus allowing extensive signal averaging, significantly improving the signal to noise ratio. By the choice and application of suitable radio frequency pulses and delays, Fourier transform NMR allows the separation and analysis of different interactions in the nuclear spin system. This allows powerful experiments to be designed which can be applied to the determination of molecular structures and dynamics.

1.3 Chemically Induced Dynamic Nuclear Polarisation

Chemically Induced Dynamic Nuclear Polarisation (CIDNP) was first observed experimentally in 1967 by the groups of Bargon and Fischer [43] and Ward and Lawler [44]. They observed anomalous intensities in the NMR spectra of radicals produced during chemical reactions within the NMR magnet. The initial explanations for these observations invoked Overhauser-type arguments of relaxation interactions between the nuclei and the unpaired electron in the radicals, hence the name "chemically induced dynamic nuclear polarisation" [45–47]. These explanations were incomplete, and in 1969 the CIDNP phenomenon was explained, independently by Closs [48] and by Kaptein and Oosterhoff [49], in terms of the Radical Pair Mechanism[1].

Immediately following its discovery, the major application of CIDNP was in mechanistic organic chemistry. The NMR enhancements observed could be used to elucidate the pathways of reactions which proceeded via a radical pair intermediate [52]. In 1978 photo-CIDNP was applied to the study of biopolymers by Kaptein [53]. This section outlines the theoretical background of the radical pair

[1]This mechanism is also responsible for the electron equivalent of CIDNP, Chemically Induced Dynamic Electron Polarisation (CIDEP) [50], and for Magnetic Field Effects (MFEs) in radical pair recombination reactions [51].

mechanism and its application to the CIDNP study of proteins.

1.3.1 Radical Pairs

Radicals are chemical species which possess an unpaired electron. When two radicals are closely associated, they can form a spin-correlated radical pair, that is the overall spin state of the pair is well defined. The total electron spin angular momentum of this radical pair is given by a Clebsch-Gordon series:

$$S = s_1 + s_2, s_1 + s_2 - 1, \cdots, |s_1 - s_2| \tag{1.9}$$

where s_j is the electron spin angular momentum of radical j in the pair. Therefore S can take the values 0 or 1. As is usual for angular momenta, there are $2S + 1$ projections of this angular momentum onto the z axis. Therefore, the case $S = 1$ has $M_S = 0, \pm 1$ and is known as the triplet state. The singlet state has $S = 0$ and correspondingly $M_S = 0$.

The spin parts of the singlet and triplet wavefunctions can be defined in terms of products of the individual electron spin wavefunctions, $|\alpha\rangle$ and $|\beta\rangle$. These wavefunctions are given below:

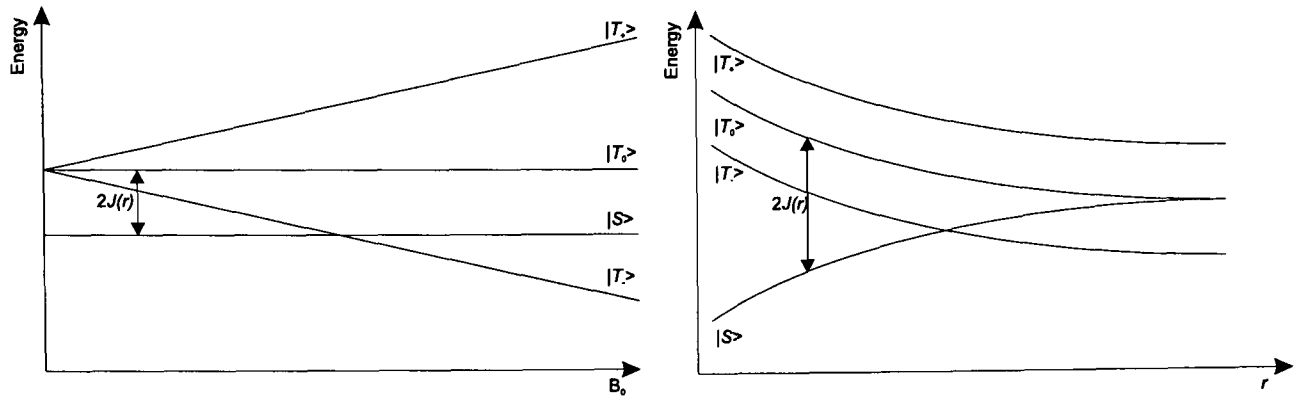
$$|S\rangle = \frac{1}{\sqrt{2}}(|\alpha_1\beta_2\rangle - |\beta_1\alpha_2\rangle) \tag{1.10}$$

$$|T_+\rangle = |\alpha_1\alpha_2\rangle \tag{1.11}$$

$$|T_0\rangle = \frac{1}{\sqrt{2}}(|\alpha_1\beta_2\rangle + |\beta_1\alpha_2\rangle) \tag{1.12}$$

$$|T_-\rangle = |\beta_1\beta_2\rangle \tag{1.13}$$

Figure 1.2 shows how the relative energies of these four spin states depend on the strength of the applied magnetic field (Figure 1.2(a)) and the inter-radical separation (Figure 1.2(b)). The $|T_+\rangle$ and $|T_-\rangle$ states are strongly dependent on the applied field



(a) As a function of applied field strength, B_0

(b) As a function of inter-radical separation, r for $B_0 \neq 0$

Figure 1.2: The relative energies of the singlet and triplet states. $J(r)$ is the electron-electron exchange interaction, described in more detail below.

strength since both these states possess a resultant spin angular momentum either opposed to, or parallel with, the field direction. The $|S\rangle$ and $|T_0\rangle$ states have no resultant spin angular momentum along the field direction, hence their energies are independent of the magnitude of the applied field. The separation of the $|S\rangle$ and $|T_0\rangle$ states, even in zero field, arises due to the electron-electron exchange interaction, which is described in more detail below.

1.3.1.1 The Spin Hamiltonian

The spin dynamics of the radical pair in a magnetic field can be described in terms of a spin Hamiltonian:

$$\hat{H} = \hat{H}_Z + \hat{H}_{hf} + \hat{H}_J \tag{1.14}$$

where $\hat{H}_Z$ describes the electron Zeeman interaction, $\hat{H}_{hf}$ describes the electron-nuclear hyperfine interaction and $\hat{H}_J$ the inter-electron exchange interaction. As the radical pair is free to undergo tumbling and diffusion in solution, all of the spin interactions in the radical pair are assumed to be fully isotropic, that is, there is no orientation dependence. Each of the terms in the spin Hamiltonian is discussed in

more detail below.

The electronic Zeeman interaction is the interaction of electron spin angular momentum with the applied magnetic field, and is given by the following Hamiltonian:

$$\hat{H}_Z = \sum_{j=1}^{2} g_j \hat{S}_{jz} \frac{\mu_B B_0}{\hbar} \tag{1.15}$$

where $\hat{S}_{jz}$ is the z component of the electron spin angular momentum on radical j, which has an isotropic g-value given by g_j. The nuclear Zeeman interaction is assumed to be negligible since, for protons, it is approximately 660 times smaller than the electronic equivalent.

The interaction of electronic and nuclear spin angular momenta is described by the hyperfine Hamiltonian:

$$\hat{H}_{hf} = \sum_{j=1}^{2} \sum_{i} a_{ij} \hat{I}_i \cdot \hat{S}_j \tag{1.16}$$

where a_{ij} is the isotropic hyperfine coupling between nuclear spin i and electron spin j. The isotropic hyperfine coupling arises from the Fermi Contact interaction. Hyperfine couplings between the unpaired electron on one radical and the magnetic nuclei on the other radical are usually negligible.

The two unpaired electrons can interact with each other via the quantum mechanical exchange interaction which arises from overlap of the electronic orbital wavefunctions and is described by the following Hamiltonian:

$$\hat{H}_J = -J(r) \left(\frac{\hat{1}}{2} + 2\hat{S}_1 \cdot \hat{S}_2 \right) \tag{1.17}$$

where $\hat{S}_j$ is the electron spin angular momentum of radical j and $J(r)$ is a distance-dependent function describing the strength of the exchange interaction. The exact form of this function is not known, but it is assumed to decay approximately ex-

ponentially with increasing radical separation. The manifestation of the exchange interaction is to remove the degeneracy between the $|S\rangle$ and $|T_{0,\pm}\rangle$ states, even in zero field, as shown in Figure 1.2.

Expressing Equation 1.14, constructed in the singlet-triplet basis, results in the following matrix form of the spin Hamiltonian:

$$\hat{H} = \begin{array}{c} \\ \langle T_+| \\ \langle S| \\ \langle T_0| \\ \langle T_-| \end{array} \begin{array}{c} \begin{array}{cccc} |T_+\rangle & |S\rangle & |T_0\rangle & |T_-\rangle \end{array} \\ \begin{pmatrix} \omega - J(r) & 0 & 0 & 0 \\ 0 & J(r) & Q & 0 \\ 0 & Q & -J(r) & 0 \\ 0 & 0 & 0 & -\omega - J(r) \end{pmatrix} \end{array} \tag{1.18}$$

where:

$$\omega = \frac{1}{2}(\omega_1 + \omega_2), \qquad Q = \frac{1}{2}(\omega_1 - \omega_2) \tag{1.19}$$

with

$$\omega_j = g_j \frac{\mu_B B_0}{\hbar} + \sum_i m_{Ii} a_{ij} \tag{1.20}$$

This spin Hamiltonian can now be used to explain the salient features of the radical pair mechanism responsible for the observation of chemically induced dynamic nuclear polarisation.

1.3.1.2 Singlet-Triplet Mixing and Spin Sorting

The wavefunctions describing the $|S\rangle$ and $|T_0\rangle$ states are similar, differing only by a phase factor of π radians. They are also connected by off-diagonal elements in the spin Hamiltonian, given in Equation 1.18. This leads to mixing of the two spin states, and a mechanism by which interconversion of the $|S\rangle$ state and the $|T_0\rangle$ state can occur. This is shown in terms of the classical vector model in Figure 1.3. The frequency of this interconversion is given by the difference in the Larmor precession

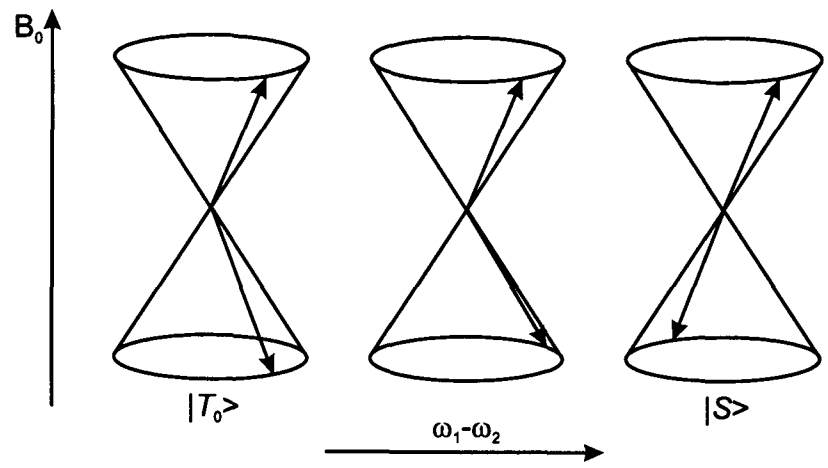


Figure 1.3: The vector model of ST_0 mixing. The cones represent the individual electron spin states and arrows the electron spins.

frequencies of the two electron spins:

$$\begin{aligned} \omega_1 - \omega_2 &= (g_1 - g_2)\frac{\mu_B B_0}{\hbar} \\ \omega_{ST_0} &= \Delta g \frac{\mu_B B_0}{\hbar} \end{aligned} \tag{1.21}$$

This is the so-called Δg mechanism of ST_0 mixing and is generally of greatest effect at high magnetic field strengths, such as those found in an NMR spectrometer, or if the difference in electronic g-values is large, for example if one of the radicals possesses a metal centre.

Nuclear spins coupled to the electron spins via hyperfine interactions will influence the rate of interconversion between the $|S\rangle$ and $|T_0\rangle$ states. For simplicity, consider the effects of a single hyperfine coupling, from a spin-$\frac{1}{2}$ nucleus. The rate of singlet-triplet mixing then becomes:

$$\omega^{\pm}_{ST_0} = \Delta g \frac{\mu_B B_0}{\hbar} \pm \frac{1}{2}a \tag{1.22}$$

It is clear from this that the rate of interconversion between the $|S\rangle$ and $|T_0\rangle$ states is dependent on the relative signs of the difference in radical g-values and the hyperfine

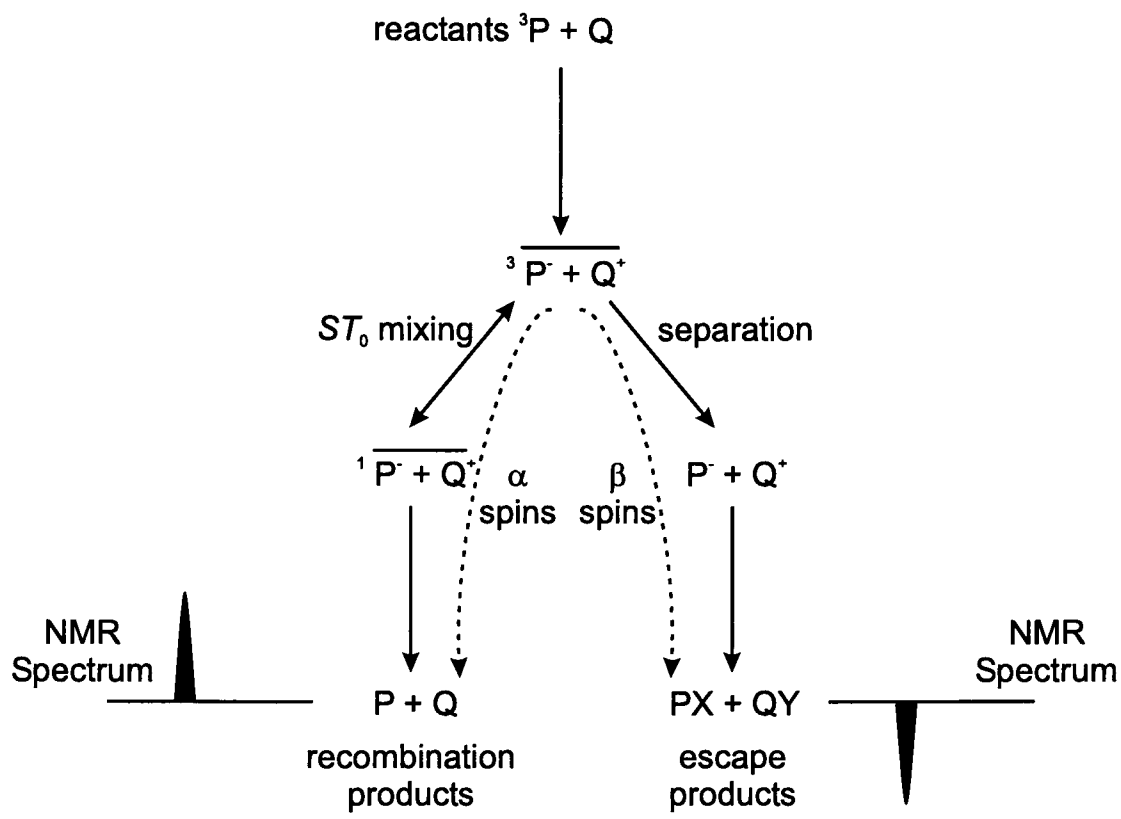


Figure 1.4: The radical pair mechanism and spin sorting process. Δg and a have the same sign as drawn.

coupling. In general, the α nuclear spin state will produce the more rapid ST_0 mixing if Δg and a have the same sign, whereas the β nuclear spin state will produce greater mixing if they have opposite sign.

The principles of the Δg mechanism as outlined above lead into the idea that the radical pair mechanism acts as a *spin sorting* process which ultimately results in nuclear spin polarisation. Figure 1.4 shows this process in more detail, starting from an initial triplet state precursor molecule. This results in the formation of an initially triplet-state spin correlated radical pair, via for example electron transfer. This radical pair then has two fates. It can either separate and diffuse apart, leading to the escape products, or undergo ST_0 mixing and convert into a singlet state radical pair. This then has two fates, it can undergo further spin mixing and reform the triplet correlated pair, or can form recombination products. Recombination can usually only occur via the singlet channel due to the Pauli principle. As drawn, Figure 1.4 shows that the recombination products will contain an excess of α nuclear

spins, while the escape products will contain an excess of β spins. Thus the spin sorting process has resulted in nuclear spin polarisation, which will be detected as enhancements to the NMR spectrum. The radical pair mechanism, recast in terms of the NMR product operator formalism [54], is shown in Appendix A [55].

The exchange interaction, which is responsible for lifting the degeneracy of the $|S\rangle$ and $|T_0\rangle$ states in zero field, also has a pronounced effect on the spin evolution of the radical pair. If the magnitude of this interaction is larger than the average hyperfine coupling in the radical pair then it will tend to inhibit ST_0 mixing driven by the hyperfine coupling. If it is larger than $\Delta g \frac{\mu_B B_0}{\hbar}$ it will inhibit the spin mixing via the Δg mechanism. Figure 1.2(b) shows how the energies of the singlet and triplet states vary as a function of radical pair separation. Therefore, if a radical pair is formed in an initial triplet state, then it must separate to a distance at which the exchange interaction has decreased to less than $\Delta g \frac{\mu_B B_0}{\hbar}$, and to less than the average hyperfine coupling, which is typically of the order of one nanometre, in order for spin mixing to occur. The diffusion of the primary radical pair and its subsequent re-encounter is termed *secondary recombination.* The timescales for these two events, spin mixing on the order of 10^{-9} to 10^{-8} s, and re-encounter on a timescale of 10^{-10} to 10^{-7} s, are such that spin mixing can occur while the radicals are separated to a distance at which the exchange interaction is sufficiently decreased.

The loss of electron spin correlation in the radical pair through electronic relaxation is usually negligible since the electron spin relaxation times are $T_1^{\mathrm{e}} \approx 10^{-6}$ s. This is a slower process than the spin evolution and diffusion required to generate the nuclear spin polarisation.

1.3.1.3 Kaptein's Sign Rules

The phase of the nuclear polarisation, Γ, produced for a particular atom, i, in a molecule can be predicted by a set of empirical rules devised by Kaptein [56], following analysis of the spin sorting mechanism:

$$\Gamma(i) = \mu \cdot \epsilon \cdot \text{sign}\,(\Delta g) \cdot \text{sign}\,(a_i) \begin{cases} +A, \text{ absorptive enhancement} \\ -E, \text{ emissive enhancement} \end{cases} \tag{1.23}$$

where:

$$\mu = \begin{cases} + \text{ for a triplet precursor} \\ - \text{ for a singlet precursor} \end{cases}$$

$$\epsilon = \begin{cases} + \text{ for a recombination product} \\ - \text{ for an escape product} \end{cases}$$

sign (Δg) is the sign of the difference in g-value between the radical on which nucleus i resides and the partner in the radical pair, and sign (a_i) is the sign of the hyperfine coupling at nucleus i. It is worth noting that this sign rule is only valid at high field [57], and for the so-called *net effect*, when all components of a multiplet show the same phase of polarisation.

There is a similar rule for the CIDNP multiplet effect, in which different components of a J-coupled multiplet show differing enhancements. This effect is not considered in this thesis, since it is principally operative at low field.

1.3.2 Biological Photo-CIDNP

The application of photo-CIDNP to the study of biological systems involves the generation of the radical pair consisting of certain aromatic amino acid side chains and a suitable photosensitiser, usually the tricyclic flavin mononucleotide (FMN). The

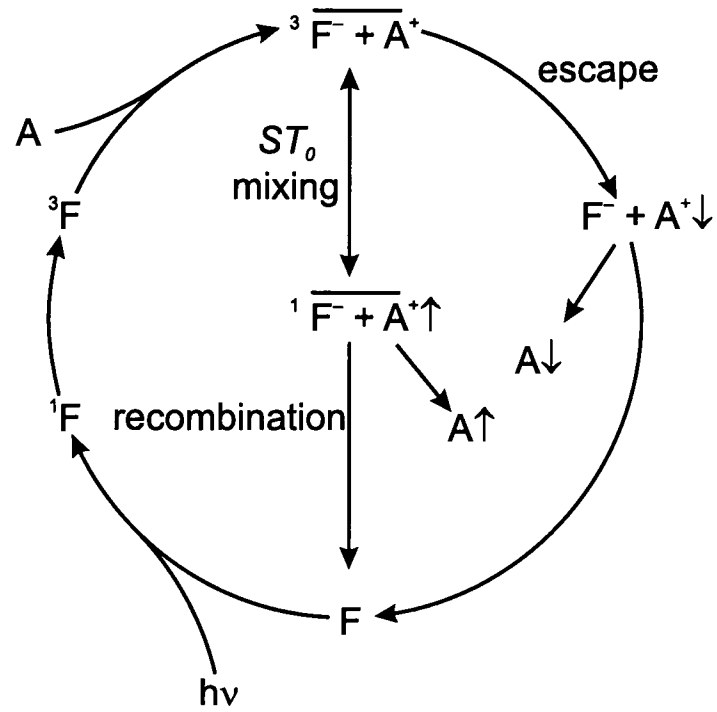


Figure 1.5: The cyclic reaction mechanism between a flavin molecule (denoted F) and an aromatic amino acid side chain (A). The symbols ↑ and ↓ denote an excess of α and β nuclear spins respectively.

potentially polarisable amino acid residues are histidine, tyrosine and tryptophan. Methionine also shows weak polarisation at the H^ϵ position next to the thioether, however, it is not of interest in this thesis.

The generation of nuclear polarisation in the aromatic amino acid side chain begins with the photochemical excitation of the flavin molecule. This excited singlet state rapidly undergoes intersystem crossing to an excited triplet state molecule. This triplet state flavin then reacts with the histidine, tyrosine or tryptophan side chain to form a triplet spin correlated radical pair. From this point the reaction scheme differs slightly from that presented in Figure 1.4, in that the idealised reaction is cyclic. This means that the polarisation produced is detected in the original amino acid (whether free in solution, or as part of a protein), not some chemically modified species. The cyclic reaction scheme is shown in Figure 1.5. Since the reaction scheme is cyclic, the products of the escape and recombination channels are identical. Therefore it would be expected that since the nuclear polarisation produced in the

recombination and escape products is equal and opposite in phase, there would be cancellation of the polarisation and no enhancements to the NMR signal would be observed. However, this is not the case. The escaped radicals, having a lifetime longer than the recombining radicals (10^{-4} s and 10^{-7} s respectively), undergo rapid nuclear spin relaxation via dipolar coupling to the unpaired electron, with a nuclear spin-lattice relaxation time, $T_1^{\text{radical}} \approx 10^{-7}$ s. This results in "leakage" of the nuclear spin polarisation in the escape products, hence incomplete cancellation of the total nuclear spin polarisation, and is known as *recombination cancellation.*

The escaped radicals are also able to transfer polarisation to the recombination products via degenerate electron exchange. Consider a spin-polarised radical, $A^{\bullet}\downarrow$, formed from the escape channel. It can undergo an electron exchange with its diamagnetic counterpart in solution:

$$A^{\bullet}\downarrow + A \rightarrow A\downarrow + A^{\bullet} \tag{1.24}$$

thereby transferring the escape polarisation to the recombination products. Since this escape polarisation is opposite in phase to that produced via the recombination channel (Figures 1.4 and 1.5), this results in a decrease in the recombination product polarisation, and is termed *exchange cancellation.* The signal intensity of a given nucleus undergoing exchange cancellation is given by the following equation [58]:

$$I_{\text{A}} = \mathcal{P}_{\text{A}}^{\text{R}} \frac{R_1^{\text{radical}}}{k_0 + k_{\text{A}}^{\text{ex}}[\text{A}] + R_1^{\text{radical}}} \tag{1.25}$$

where $\mathcal{P}_{\text{A}}^{\text{R}}$ is the polarisation generated in the absence of other competing processes, k_0 is the rate of decay of the triplet flavin either by fluorescence or by quenching with molecular oxygen, k_{A}^{ex} is the electron exchange rate constant, and R_1^{radical} is the nuclear spin lattice relaxation rate of the radical ($= 1/T_1^{\text{radical}}$).

As proteins are large macromolecules, they undergo relatively slow translational

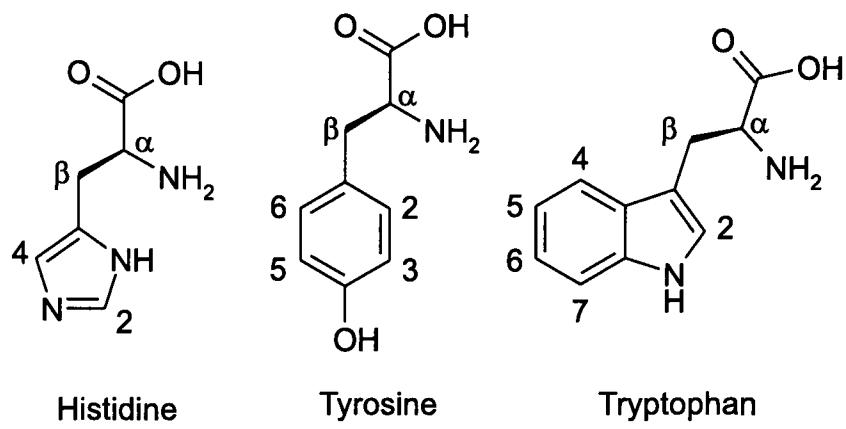


Figure 1.6: The structures of the three polarisable amino acids. The numbering scheme indicates that used in this thesis.

diffusion in solution, and hence the recombination and exchange cancellation effects described here play only a minor rôle in determining the observed CIDNP intensity. The effects are more pronounced in solutions of the free amino acids [55].

1.3.2.1 Photo-CIDNP of the Amino Acids

The structures of the three polarisable amino acids are shown in Figure 1.6. The mechanism of formation of the radical pair is different for the three amino acids. Histidine reacts with the triplet flavin via hydrogen atom abstraction [59]:

$$^3\mathrm{FMN} + \mathrm{HisH} \rightarrow {}^3\overline{\mathrm{FMNH}^\bullet + \mathrm{His}^\bullet} \tag{1.26}$$

where the over-bar denotes the spin-correlation. The rate constant for the reaction of the triplet flavin with histidine is given by the following expression [59]:

$$k_\mathrm{q} = \frac{K_{\mathrm{q}1}(k_{\mathrm{q}1}[\mathrm{H}^+] + k_{\mathrm{q}2}K_{\mathrm{q}2})}{(K_{\mathrm{q}1} + [\mathrm{H}^+])(K_{\mathrm{q}2} + [\mathrm{H}^+])} + \frac{k_0}{[\mathrm{HisH}]} \tag{1.27}$$

where $K_{\mathrm{q}1}$ and $K_{\mathrm{q}2}$ are the ionisation constants of $^3\mathrm{FMNH}^+$ and HisH_2^+, $k_{\mathrm{q}1}$ and $k_{\mathrm{q}2}$ are the second order rate constants for the reaction of $^3\mathrm{FMNH}$ with HisH_2^+ and HisH respectively. k_0 is the rate of decay of the triplet flavin in the absence

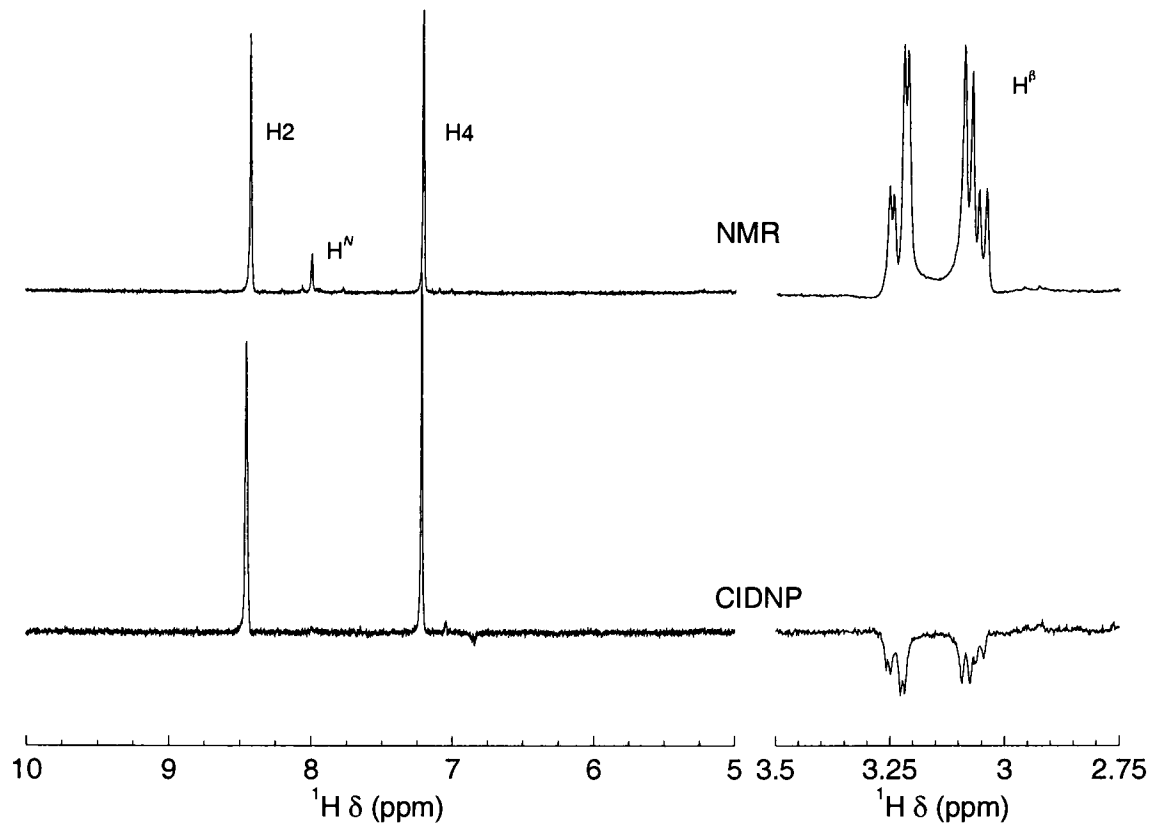


Figure 1.7: 500 MHz ^{1}H NMR and CIDNP spectra of 2 mM histidine, using 0.2 mM FMN, in H_2O at pH 7. The NMR spectrum is the average of 64 scans, while the CIDNP spectrum is the average of 32 light/dark subtraction pairs and was recorded using 50 ms laser pulses at 4 W.

of histidine. This expression is consistent with the observed pH dependence of the histidine polarisation, and based on the reactivity of the triplet flavin with the neutral and protonated forms of the histidine side chain. This therefore leads to a pronounced pH dependence of the observed polarisation on the histidine side chain, with no polarisation being observed below pH 4. Typical values for the rate constant for the reaction of histidine with the triplet flavin are in the range $k_q = 1 - 5 \times 10^7$ M^{-1} s^{-1} [59]. The polarisation produced at neutral pH is shown in Figure 1.7. Strong enhancement of the H2 and H4 protons is observed, along with emissive enhancement of H^β.

The mechanism of radical pair formation in the case of tyrosine has been the subject of some debate. It was initially thought to be hydrogen atom abstraction,

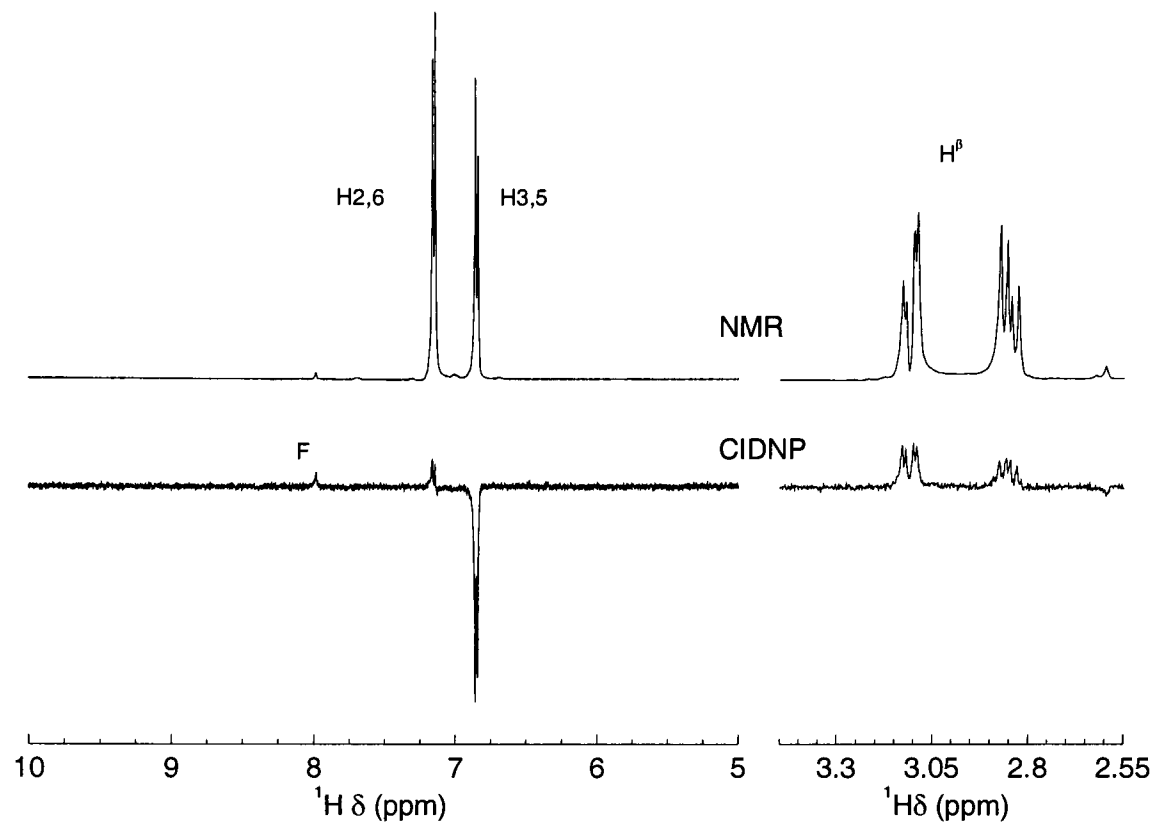


Figure 1.8: 500 MHz ^{1}H NMR and CIDNP spectra of 2 mM tyrosine, using 0.2 mM FMN, in H_2O at pH 7. The NMR spectrum is the average of 64 scans, while the CIDNP spectrum is the average of 32 light/dark subtraction pairs and was recorded using 50 ms laser pulses at 4 W. F denotes a flavin signal.

from the phenolic OH [60], but, has now been shown to be electron transfer [59]:

$$^3\mathrm{FMN} + \mathrm{TyrOH} \rightarrow {}^3\overline{\mathrm{FMN}^{-\bullet} + \mathrm{TyrOH}^{+\bullet}} \rightarrow {}^3\overline{\mathrm{FMNH}^{\bullet}\ \mathrm{TyrO}^{\bullet}} \tag{1.28}$$

This electron transfer is followed by swift deprotonation of the tyrosine radical cation to give the neutral radical being involved in the radical pair. The rate of quenching of the triplet flavin by the tyrosine is $k_q = 1 \times 10^9\ \mathrm{M}^{-1}\ \mathrm{s}^{-1}$ and is independent of pH [59], indicative of electron transfer reactions as it is at the diffusion controlled limit.

The major feature of the polarisation observed in tyrosine is the emissive polarisation of the protons at positions H3 and H5, shown clearly in Figure 1.8. This is due to the sign of the hyperfine coupling at this position, relative to the other positions around the ring. This feature does allow the easy identification of exposed

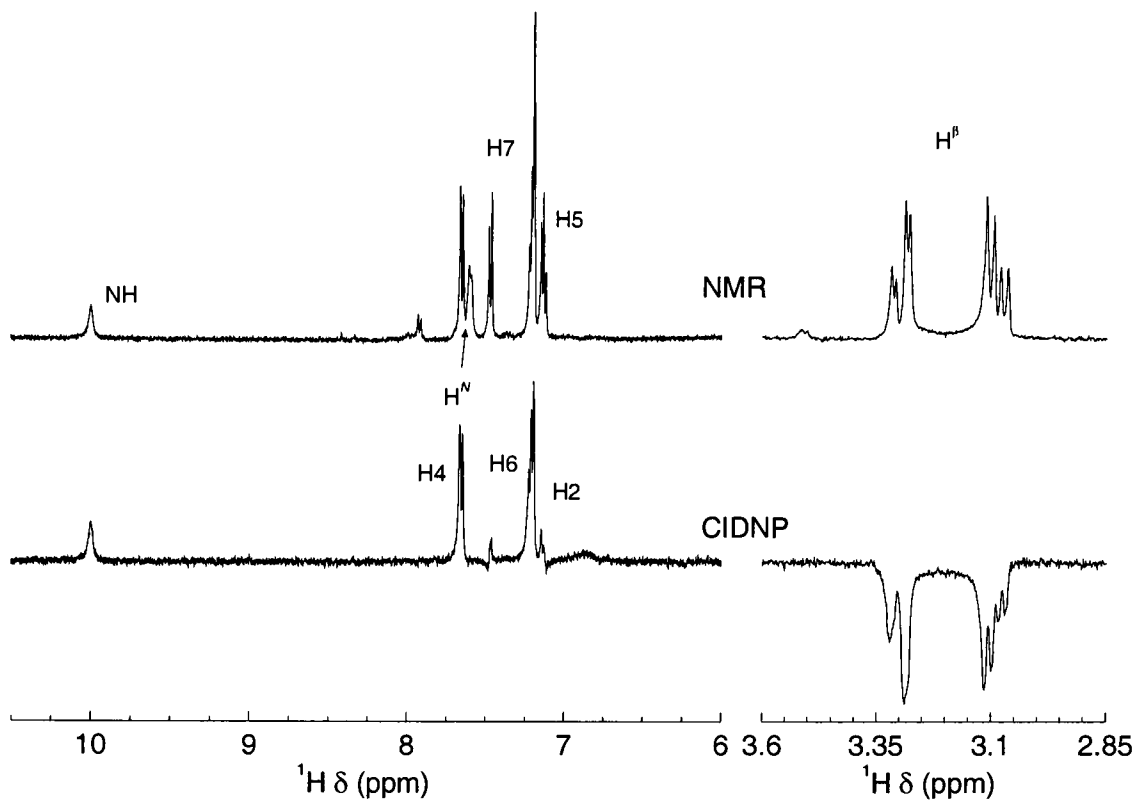


Figure 1.9: 500 MHz ^{1}H NMR and CIDNP spectra of 2 mM tryptophan, using 0.2 mM FMN, in H_2O at pH 7. The NMR spectrum is the average of 64 scans, while the CIDNP spectrum is the average of 32 light/dark subtraction pairs and was recorded using 50 ms laser pulses at 4 W.

tyrosine residues in the CIDNP spectrum of a protein. The H2 and H6 protons, along with the H^{β} show weakly enhanced absorptive polarisation.

The reaction for tryptophan also proceeds via electron transfer, followed by deprotonation of the tryptophan radical cation [59]:

$$^{3}\mathrm{FMN} + \mathrm{TrpH} \rightarrow {}^{3}\overline{\mathrm{FMN}^{-\bullet} + \mathrm{TrpH}^{+\bullet}} \tag{1.29}$$

The triplet quenching rate for tryptophan is also near the diffusion controlled limit, $k_q = 2 - 3 \times 10^9$ M^{-1} s^{-1} and is only slightly dependent on pH [59]. The polarisation produced is shown in Figure 1.9. The alternation of polarisation around the ring is due to the alternating nature of spin density in the radical's singly occupied molecular orbital. The strong polarisation of the indole NH proton is useful as an identifier for exposed tryptophan residues in proteins as there are very few other

signals in this region of the NMR spectrum.

The situation becomes more complex when binary (or ternary) mixtures of the amino acids are considered. Given that the three amino acids all have differing rate constants for quenching the triplet flavin, this results in a competition for available flavin triplets when the amino acids are present as mixtures. The signal intensity then becomes [61]:

$$I_{\mathrm{A}} = \mathcal{P}_{\mathrm{A}}^{\mathrm{R}} \frac{k_{\mathrm{qA}}[\mathrm{A}]}{k_0 + k_{\mathrm{qA}}[\mathrm{A}] + k_{\mathrm{qB}}[\mathrm{B}]} \tag{1.30}$$

where $\mathcal{P}_{\mathrm{A}}^{\mathrm{R}}$ is the polarisation generated from the radical pair reaction, k_0 is the rate of decay of the triplet flavin, either via fluorescence or quenching by molecular oxygen, and $k_{\mathrm{q}i}$ is the rate of rate of reaction for the amino acid i with the triplet flavin. At high amino acid concentrations this equation should be modified to include the effects of degenerate electron exchange by multiplying by $R_1^{\mathrm{radical}}/(k_{\mathrm{A}}^{\mathrm{ex}}[\mathrm{A}] + R_1^{\mathrm{radical}})$.

From Equation 1.30, and the quenching rate constants given above, it can be seen that the order of reactivity for the three amino acids is roughly Trp $\geq$ Tyr $\gg$ His. This leads to the conclusion that in a protein, although a histidine residue may be sufficiently exposed to undergo the hydrogen atom abstraction required to generate CIDNP, in the presence of at least partially exposed tyrosine or tryptophan it will be less able to compete effectively for the triplet flavin, and hence show decreased polarisation, or no polarisation at all.

1.3.2.2 Photo-CIDNP of Proteins

The generation of CIDNP when amino acids are held in a molecular framework such as a peptide or protein provides an interesting and powerful tool for studying protein structure and folding. The only requirement is that the residue side chain be accessible to the triplet flavin. This means that the amino acid must be on the surface of the protein and exposed to the solvent. For example, Figure 1.10 shows the aromatic region of the NMR and CIDNP spectrum of the protein bovine

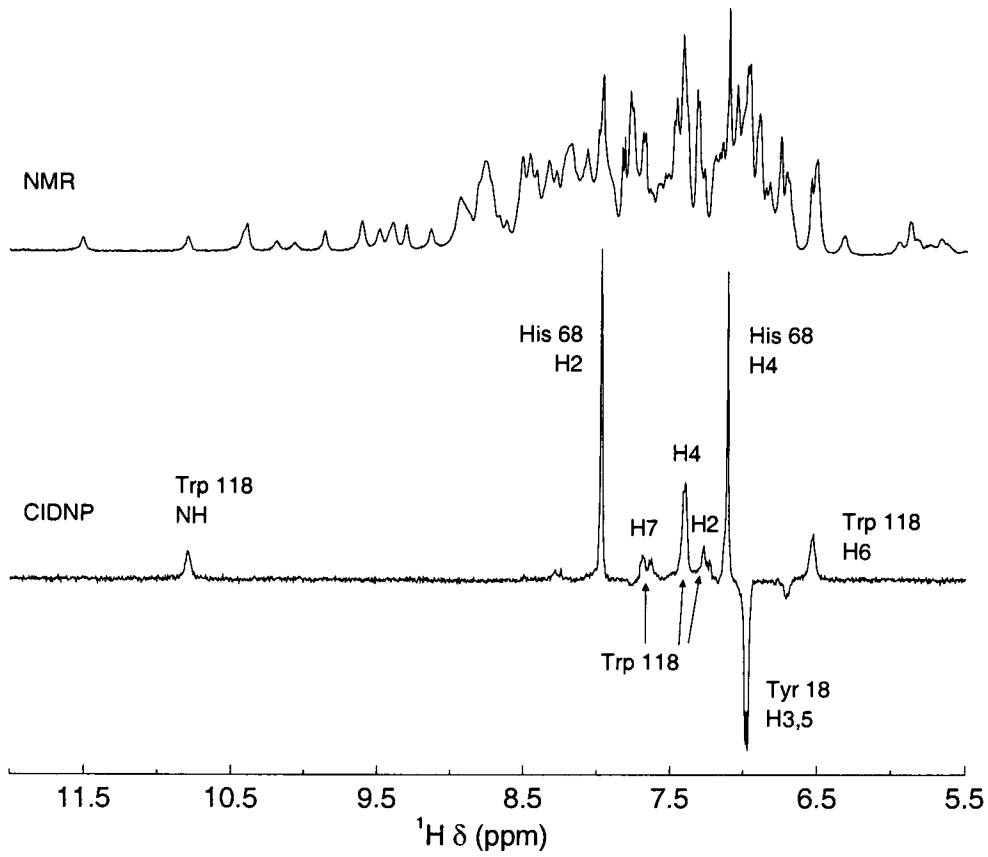


Figure 1.10: 500 MHz ^{1}H NMR and CIDNP spectra of a 2 mM bovine α-lactalbumin solution, using 0.2 mM FMN, in H_2O at pH 7. The NMR spectrum is the average of 32 scans, while the CIDNP spectrum is the average of 16 light/dark subtraction pairs and was recorded using 50 ms laser pulses at 4 W. The assignments are taken from [62].

α-lactalbumin (BLA). It is immediately clear that there is vast simplification in the CIDNP spectrum, since only signals arising from polarised amino acid residues are detected. In the native state of BLA only three of the aromatic residues are exposed, and hence polarisable. These are Tyr 18, His 68 and Trp 118, and clearly identifiable in Figure 1.10. The observation of CIDNP signals from these three residues correlates with that expected based on the protein crystal structure, shown in Chapter 4, Figure 4.23(a). This spectral simplification is one benefit of using CIDNP enhancements for the study of proteins.

1.3.2.3 Cross Polarisation

Since the generation of CIDNP produces large, non-equilibrium spin populations on certain nuclei in the amino acids, these polarisations can be transferred by the

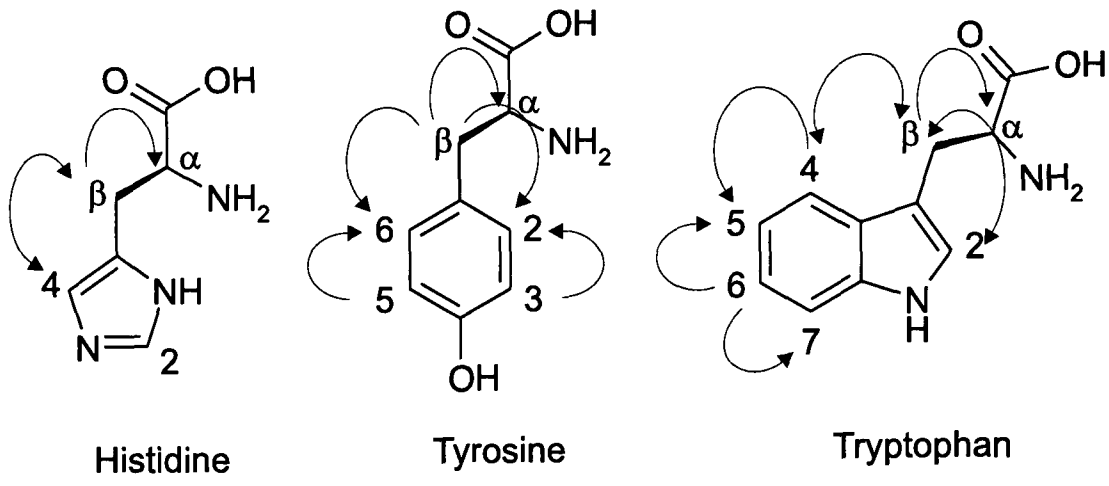
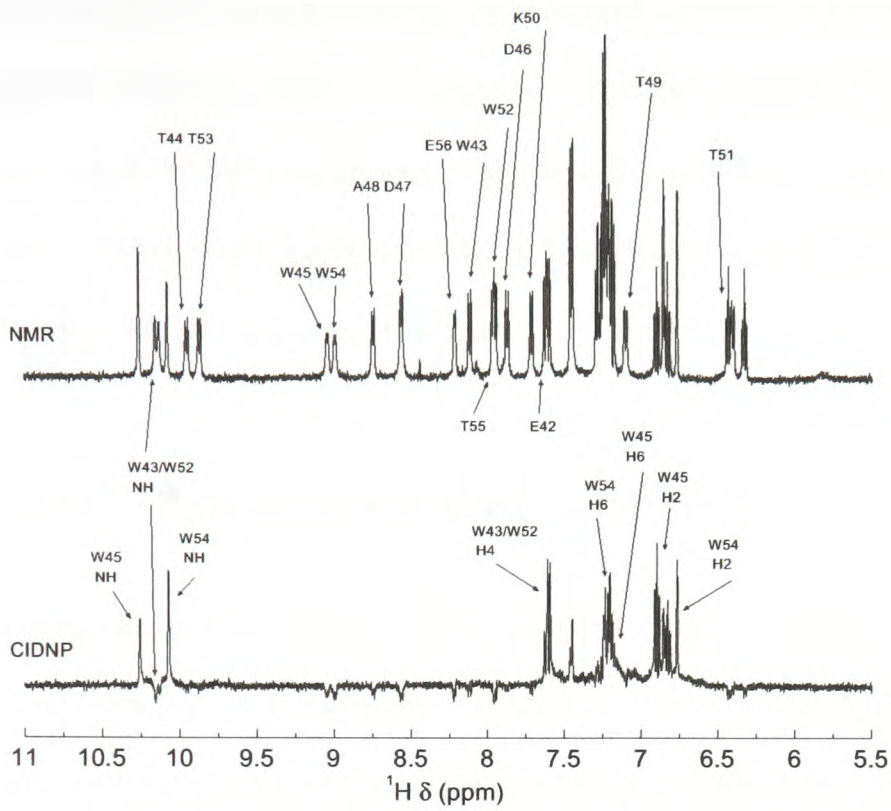


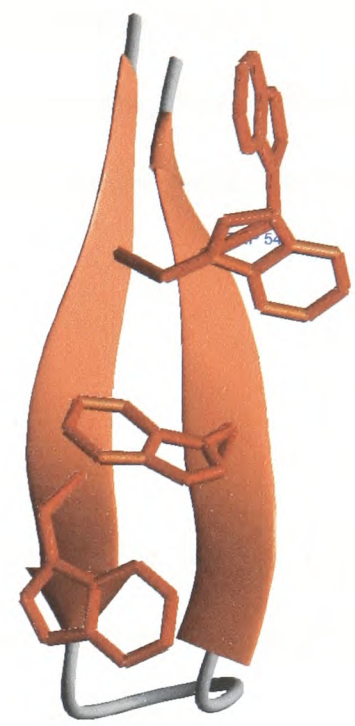
Figure 1.11: The cross-polarisation pathways for histidine, tyrosine and tryptophan.

Nuclear Overhauser Effect (NOE) to neighbouring spins. This so-called cross polarisation enables nuclei which are not directly polarised by the radical pair reaction to becomes weakly enhanced [63, 64]. Figure 1.11 shows the transfer of polarisation around the amino acid side chains. The sign of the transferred polarisation depends on which cross relaxation pathway is most efficient for a given system. This, in turn, depends on the rotational correlation time. For small molecules, such as the free amino acids in non-viscous solutions, the $W_2^{I_1 I_2}$ pathway ($\Delta m_I = \pm 2$) is most efficient, hence the transferred polarisation has the opposite phase to the directly generated enhancements. The converse is true in larger, hence slower tumbling, molecules such as proteins, where the $W_0^{I_1 I_2}$ pathway ($\Delta m_I = 0$) prevails, and transfer is with retention of phase. Cross polarisation in ^{19}F-labelled amino acids (and proteins) leads to interesting multiplet intensity patterns arising from the interaction of various relaxation pathways such as dipole-dipole interactions and relaxation arising from the chemical shift anisotropy of the fluorine [65].

An example of cross polarisation in a real system is shown in the small 16-residue peptide trpzip4. This is a triple mutant (Y45W, F52W, V54W) of the gb1 peptide, which corresponds to residues 41 to 56 of the B1 IgG-binding domain of protein G. This peptide forms a β-hairpin motif in solution, without the need for stabilising



(a) NMR and CIDNP spectra of trpzip4



(b) Structure of trpzip4 (PDB File: 1LE3 [66])

Figure 1.12: 600 MHz ^{1}H NMR and CIDNP spectra of a 2 mM solution the trpzip4 peptide at pH 6. The CIDNP spectrum was acquired using 100 ms laser pulses at 5 W. The assignments and structure are taken from [66].

counter ions. It achieves this through the π, π stacking of two pairs of tryptophan residues [66]. Figure 1.12(a) shows the NMR and CIDNP spectra of this peptide. Of the four tryptophan residues only two are directly polarised, namely Trp 45 and Trp 54. From the structure shown in Figure 1.12(b) these two residues have their side chains oriented in such as a way as to enable electron transfer with the triplet flavin, that is, the faces of the aromatic rings are protruding out into the solution, thereby allowing orbital overlap with the flavin, and formation of the radical pair. The two other tryptophan rings are held such that their edges are pointing into solution. Hence it is not possible to attain the required overlap between the tryptophan and the flavin to enable the electron transfer.

The interesting feature of the CIDNP spectrum is that several (at least half) of

the backbone amides appear as emissive signals. As it is not possible to polarise these signals directly, this polarisation probably arises via NOE cross-polarisation from the tryptophan sidechains. Since this is a relatively small system, the rotational correlation time gives rise to a negative NOE, therefore the cross-polarisation is carried with inversion of phase.

1.4 Outline of the Thesis

The final section of this introductory chapter outlines the remainder of the thesis.

Chapter 2 continues the introduction, detailing the experimental equipment and methodology used to obtained the results described in this thesis.

Chapter 3 presents results of photo-CIDNP experiments performed on a small cytochrome protein, cytochrome c_{552} from *Hydrogenobactor thermophilus* and a double alanine mutant of the same protein. The problems of performing CIDNP experiments on proteins which possess a chromophore are described and insights into the structural changes between the *apo* and *holo* mutant proteins and the wild-type are obtained.

Chapter 4 first describes the design and development of a rapid mixing device, then explains experiments performed to characterise the mixing properties of the device. The second half of this chapter describes experiments using the rapid mixing device which probe the surface structure of a series of molten globule states of the bovine and human α-lactalbumins. The information obtained is not readily available by other methods. The chapter concludes with a second use for the injection system is presented, allowing the real-time refolding of hen egg white lysozyme to be followed.

Chapter 5 continues on from the results presented in the previous chapter, applying the rapid mixing device to the folding of bovine pancreatic ribonuclease A.

The effects of sample heating by the laser illumination during a real-time CIDNP experiment are also briefly discussed.

Chapter 6 describes a different use for illumination during an NMR experiment, namely that of photochemically induced conformational change. The experiments contained in this chapter are concerned with structural changes which occur during the reversible photocycle of the LOV2 domain of phototropin. These experiments include probes of the conformational change using 2-dimensional NMR methods and photo-CIDNP experiments. Finally, a real-time NMR method is described which monitors the kinetics of the protein's return to the ground state following the activation of the photocycle.

Chapter 7 uses calculated hyperfine coupling constants and a radical pair diffusion model from the literature to simulate the nuclear polarisation obtained for the amino acid tryptophan. Comparisons are made between theory and experiment.

Chapter 8 changes direction and describes the NMR characterisation of a series of homologous peptides. These peptides have been designed to be subsequently spin-labelled and used in EPR studies. The work described in this chapter is concerned with the characterisation of the secondary structure adopted by these peptides under similar conditions to those used in the EPR measurements.

Chapter 2

Experimental Methods

In this chapter an overview of the experimental equipment and methodology used is given. Details of the NMR spectrometer hardware, pulse sequences and sequence elements used, and laser configurations needed to acquire the data presented in the remainder of the thesis are outlined.

2.1 NMR Spectrometer and Laser Systems

In order to generate and detect CIDNP enhancements in the NMR spectrum, a laser system must be coupled to an NMR spectrometer. In this thesis three such installations were used, each detailed below.

2.1.1 Oxford Centre for Molecular Sciences

The equipment at the Oxford Centre for Molecular Sciences (OCMS) consisted of two home-built spectrometers operating at ^{1}H frequencies of 500.1 and 600.1 MHz. Both spectrometers were fitted with a pair of Helmholtz coils, double tuned to ^{1}H, ^{2}H and ^{13}C, ^{15}N, suitable for triple resonance inverse detection. Each spectrometer was also equipped with a triple axis pulsed field gradient accessory, producing field gradients of up to 65 G cm^{-1}. The sample temperature was regulated within $\pm$0.1 °C of the

required value in the range 5 to 65 °C, using a Peltier thermoelectric device [67].

The laser system used with these spectrometers was a Spectraphysics Stabilite 2016-05 argon ion laser, which outputs principally at wavelengths of 488 and 514 nm. When operating in broadband mode, the laser produces 5 W of power at the laser head. The beam was focused on to the end of an optical fibre (F-MMC, Newport Optics) using a microscope ocular (F-915T Launcher Optic, Newport Optics), and was gated into 50 to 500 ms pulses using a model 200 FNC mechanical shutter from NM Laser Products, controlled by the spectrometer. Approximately 60% of the light was transmitted to the sample.

This set-up was principally used in the experiments described in Chapters 3, 4 and 6.

2.1.2 Physical and Theoretical Chemistry Laboratory

The NMR spectrometer in the Physical and Theoretical Chemistry Laboratory (PTCL) was a Varian $^{\text{Unity}}$*INOVA* 600 spectrometer operating at a ^{1}H frequency of 599.8 MHz, principally used with a 5 mm ^{1}H{^{13}C,^{15}N} triple resonance probe, fitted with a z-axis pulsed field gradient coil, capable of producing 60 G cm^{-1}.

Laser irradiation was provided by a Spectraphysics BeamLok 2085-25S argon ion laser, emitting at the same principal wavelengths as above, but also capable of emission between 300 and 370 nm (the near-UV) using suitable optics. The maximum power output of this laser, operating in broadband mode, is 25 W in the visible and 7 W in the UV region of the spectrum. Output from this laser is chopped into pulses using an LS200 mechanical shutter (NM Laser Products) and focused onto the end of an F-MBE optical fibre using a F-915 Launcher Optic (both from Newport Optics). Again, 60% coupling efficiency is obtained with this system.

This configuration was used for the experiments described in Chapters 5, 6 and 8.

2.1.3 University of Texas, Southwestern Medical Center at Dallas

The spectra shown in Section 6.2.2 of Chapter 6 were recorded at the University of Texas, Southwestern Medical Center at Dallas, using either a Varian ^{Unity}*INOVA* 500 spectrometer operating at a ^{1}H frequency of 499.8 MHz and equipped with a 5 mm ^{1}H{^{13}C,^{15}N} triple resonance probe, fitted with a z-axis pulsed field gradient coil, capable of producing 60 G cm^{-1}, or a Varian ^{Unity}*INOVA* 600 spectrometer operating at a ^{1}H frequency of 599.4 MHz and equipped with a cryogenically-cooled 5 mm ^{1}H{^{13}C,^{15}N} triple resonance probe and receiver preamplifier. This spectrometer was fitted with a z-axis pulsed field gradient coil, capable of producing 70 G cm^{-1}

Laser illumination for both spectrometers was provided using a Coherent Innova 90C-A5 argon ion laser with similar specification to that in Section 2.1.1. The light was coupled to an optical fibre using the same method as in Section 2.1.1. The light was gated into short pulses using a spectrometer-controlled Uniblitz Electronic LS 322 mechanical shutter.

2.2 Sample Illumination

Coupling the laser output to the NMR sample, via the use of fibre optics was achieved using a coaxial glass insert (Wilmad, Model WGS-5BL), originally designed for external reference compounds, placed in the NMR tube (either a Wilmad 528-PP or Shigemi PS-001). This method of illumination has been described by Scheffler *et al.* [68]. The tip of the insert was placed 4 mm above the top of the RF coil so as to minimise any field inhomogeneity caused while maintaining reasonable CIDNP intensity [31]. The arrangement of the sample tube and insert is shown in Figure 2.1. Using this set-up precludes the use of sample spinning due to the presence of the optical fibre, however, this is not a major problem with modern spectrometer hard-

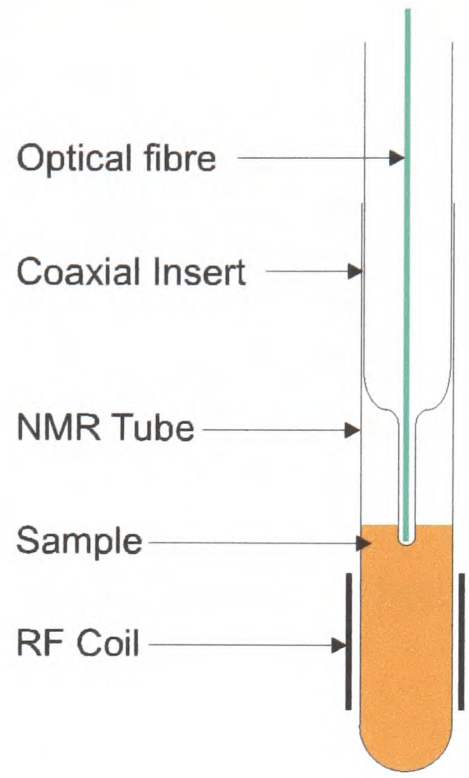


Figure 2.1: Schematic diagram showing the Coaxial insert inside a standard NMR tube.

ware. Recently, a method for the illumination of optically dense samples has been presented [69].

2.3 Photosensitiser

As described in the Chapter 1, the nuclear polarisation in these experiments is generated by an in-magnet photochemical reaction. Usually the photoactive dye used is a flavin derivative, principally flavin mononucleotide (FMN) [55, 60]. Its structure, based on a tricyclic isoalloxazine core, is shown in Figure 2.2. The chemical, biochemical and photochemical properties of these compounds are well characterised as flavins are found as cofactors in a variety of protein systems throughout biology [70, 71].

Figure 2.3 shows the UV/visible absorption spectrum of FMN. There are broad absorption bands at approximately 375 and 450 nm. These occur in the near-UV and blue-green regions of the electromagnetic spectrum. As described above the

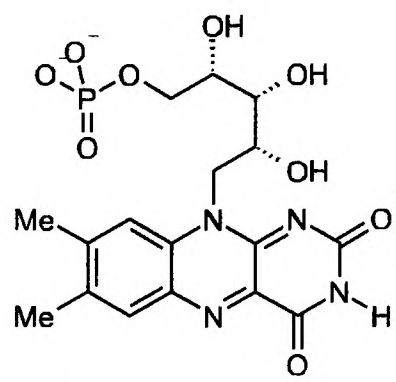


Figure 2.2: The chemical structure of flavin mononucleotide.

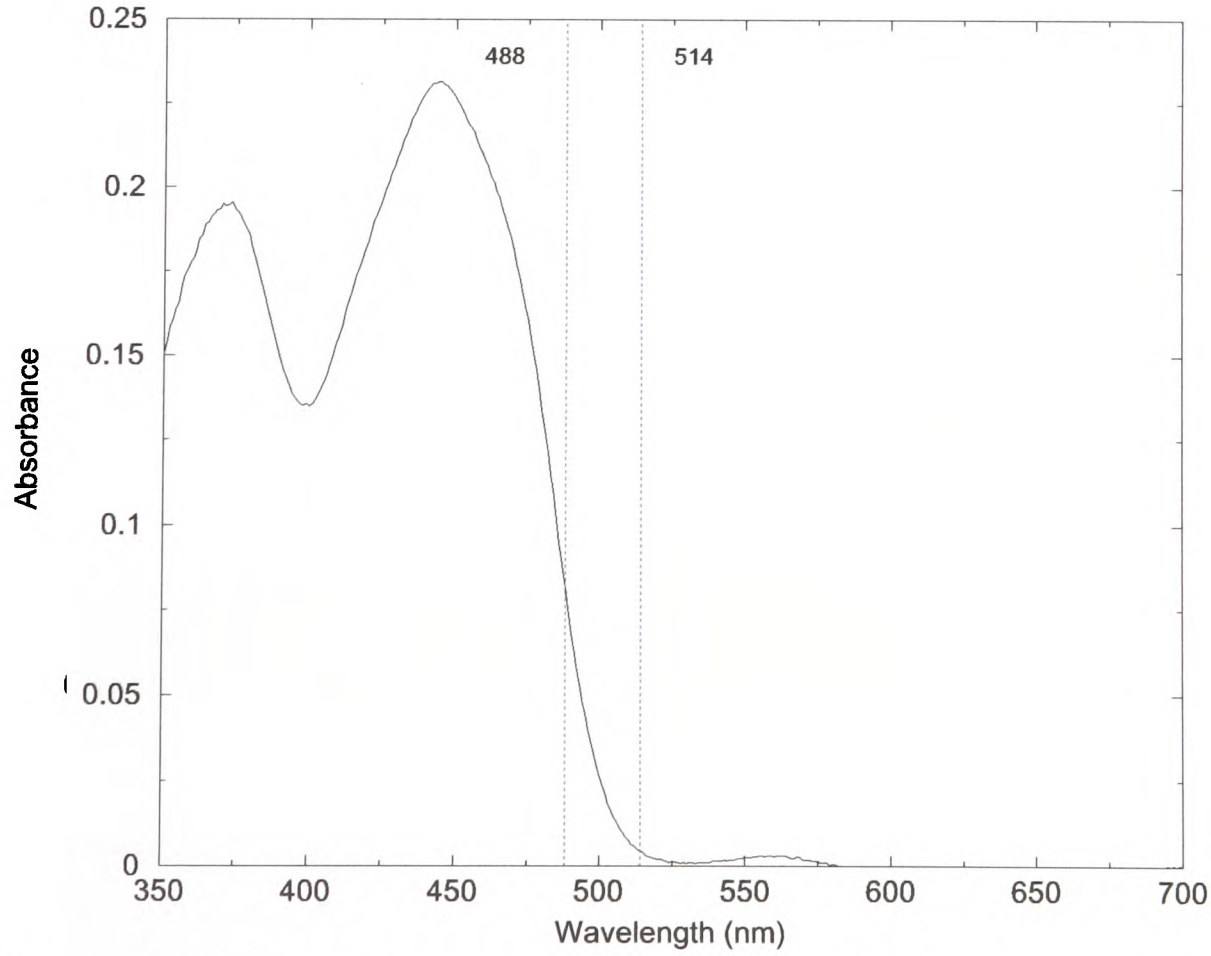


Figure 2.3: UV/visible absorption spectrum of 20 μM FMN solution at neutral pH. The blue lines represent the principal wavelengths emitted by the laser.

predominant wavelengths emitted by the laser in the visible region, are 488 and 514 nm. The absorption spectrum shows that FMN only weakly absorbs at 514 nm, but has an extinction coefficient $\epsilon > 4000$ $dm^3mol^{-1}cm^{-1}$ for the 488 nm line. It is the absorption around this wavelength which produces the excited triplet state flavin molecules which react with the aromatic amino acid sidechains, giving rise to the observed nuclear polarisation.

2.4 NMR Pulse Sequences

Nuclear spins can be manipulated using suitable combinations of radio-frequency (RF) pulses, pulsed field gradients and delays. Outlined below are the major building blocks and standard pulse sequences used to perform the experiments described in this thesis.

2.4.1 CIDNP Spectroscopy

The photo-CIDNP experiment requires that a laser pulse is inserted into the NMR pulse sequence at an appropriate position. The timing diagram for the basic CIDNP experiment is shown in Figure 2.4. This can be broken down into three stages fol-

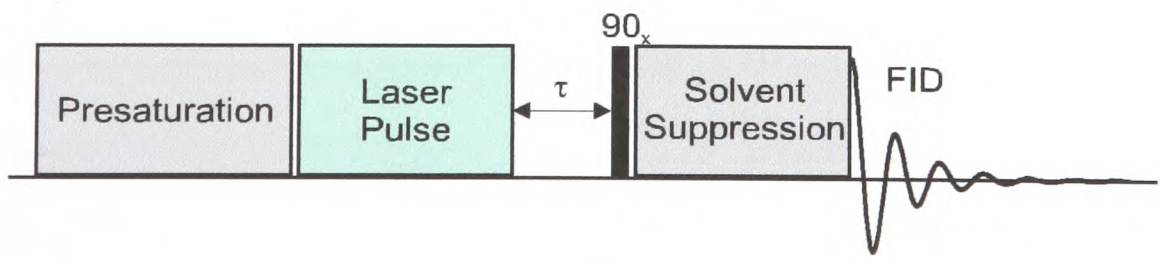
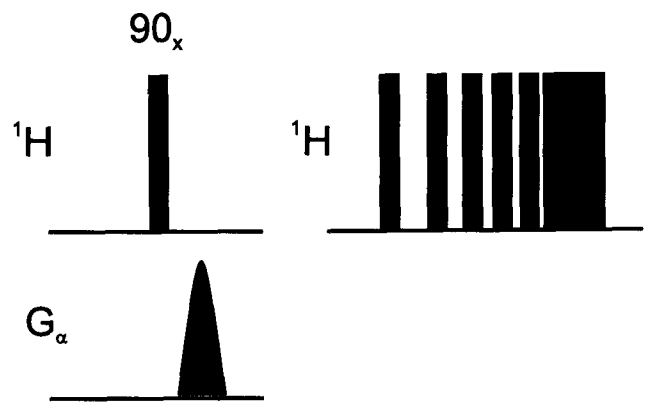


Figure 2.4: Timing diagram for the CIDNP NMR pulse sequence.

lowed by the acquisition of the free induction decay. These three stages: presaturation, the laser pulse and the solvent suppression are explained in greater detail below.



(a) The crush pulse

(b) The multiple pulse train

Figure 2.5: Presaturation techniques. Appropriate hardware permitting, the gradient in the crush pulse is applied along the magic angle. The phase of the pulses in the multiple train are $\{x, y, \bar{x}, \bar{y}\}$.

2.4.1.1 Presaturation

The generation of CIDNP results in nuclear spin systems with non-Boltzmann populations. This polarisation is detected as enhancements in the NMR spectrum, either as signals of increased intensity, opposite phase, or both. In order to observe these enhancements more clearly, presaturation is applied before the laser pulse. This minimises the signal observed from any unpolarised nuclei. Two forms of presaturation are commonly used, the pulse sequence elements for which are shown schematically in Figure 2.5. The crush pulse consists of a 90° pulse followed by a strong pulsed field gradient to dephase the coherent magnetisation (Figure 2.5(a)), while the multiple pulse train consists of a multiple of four of 90° pulses, phased as $\{x, y, \bar{x}, \bar{y}\}$, with the inter-pulse delay given by a converging geometric series of the form:

$$T_i = (m - i) \times 100 \text{ ms} \qquad \text{for} \qquad i = 1, 2, \cdots, m \tag{2.1}$$

where m is the number of pulses comprising the pulse train (Figure 2.5(a)).

The use of presaturation prior to the generation of the nuclear polarisation is

usually very effective for amino acid studies in solution as nuclear spin relaxation times are generally long compared to the duration of the laser pulse. This is not the case, however, when using proteins in which slower tumbling increases the efficiency of transverse relaxation. Difference spectroscopy is used to compensate for this, and described in more detail in Section 2.4.1.4.

2.4.1.2 The Laser Pulse

The laser pulse is triggered by the spectrometer using a spare TTL control line, following the end of the presaturation period. It is usually of between 50 and 500 ms duration, depending on what is required by the experiment. The laser pulse is followed by a delay, τ in Figure 2.4, usually of 5 ms duration, whose purpose is to allow the radicals produced by the photochemistry to recombine, before the acquisition of the NMR spectrum.

2.4.1.3 Solvent Suppression

Pure water has a proton concentration of 111 M, compared with typical protein concentrations used in this thesis of $\sim$2 mM. This results is a large solvent signal which can cause problems with the dynamic range of the spectrometer's analogue-to-digital converter (ADC) and make solute signals difficult to detect. Although it is possible to remove the solvent signal during subsequent data processing by deconvolution, it is desirable that the solvent signal be suppressed prior to acquisition of the FID. There are several methods available to achieve this (for a review see [72]), but the method which achieves the best level of suppression is the so-called Double Pulsed Field Gradient Spin Echo (DPFGSE) [73]. This sequence is applied immediately after the final 90° pulse which generates the detected transverse magnetisation. The pulse sequence is shown in Figure 2.6. The DPFGSE sequence begins with a strong field gradient which phase encodes this transverse magnetisation. A selective 180°

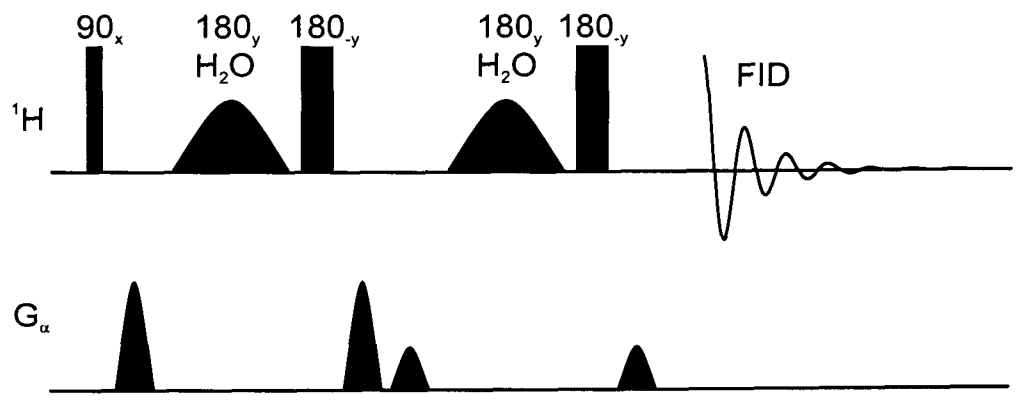


Figure 2.6: The double pulsed field gradient spin echo sequence. The selective pulses on the solvent signal are applied using Gaussian-shaped pulses, with a duration of 2 ms.

pulse is then applied to the solvent signal, followed by a hard 180° pulse applied to the whole system. The strong field gradient is then applied again. This then decodes the spins refocused by the hard 180° pulse, but further dephases the solvent signal. This unit is then repeated a second time using a different pair of gradients, taking care not to refocus the effects of the first pair of gradients.

2.4.1.4 Difference Spectroscopy

Proteins are large macromolecules and as such tumble slowly in solution, that is, they have long rotational correlation times. The major implication of this is that transverse relaxation becomes more efficient the slower a molecule tumbles. Given that the length of the laser pulse applied in the CIDNP experiment is long compared to the typical transverse relaxation time of a small protein, significant relaxation can be expected to occur during the laser pulse. This means that the effect of the presaturation applied to the spins becomes less pronounced. This can be countered by taking the difference of the NMR spectra recorded in the presence of the laser pulse and in the dark, as shown in Figure 2.7. Although it is clear that there are several polarised resonances in the spectrum of bovine α-lactalbumin, it is difficult to distinguish those resonances from presaturated signals which have relaxed during the laser pulse. The difference spectrum then only contains signals from those

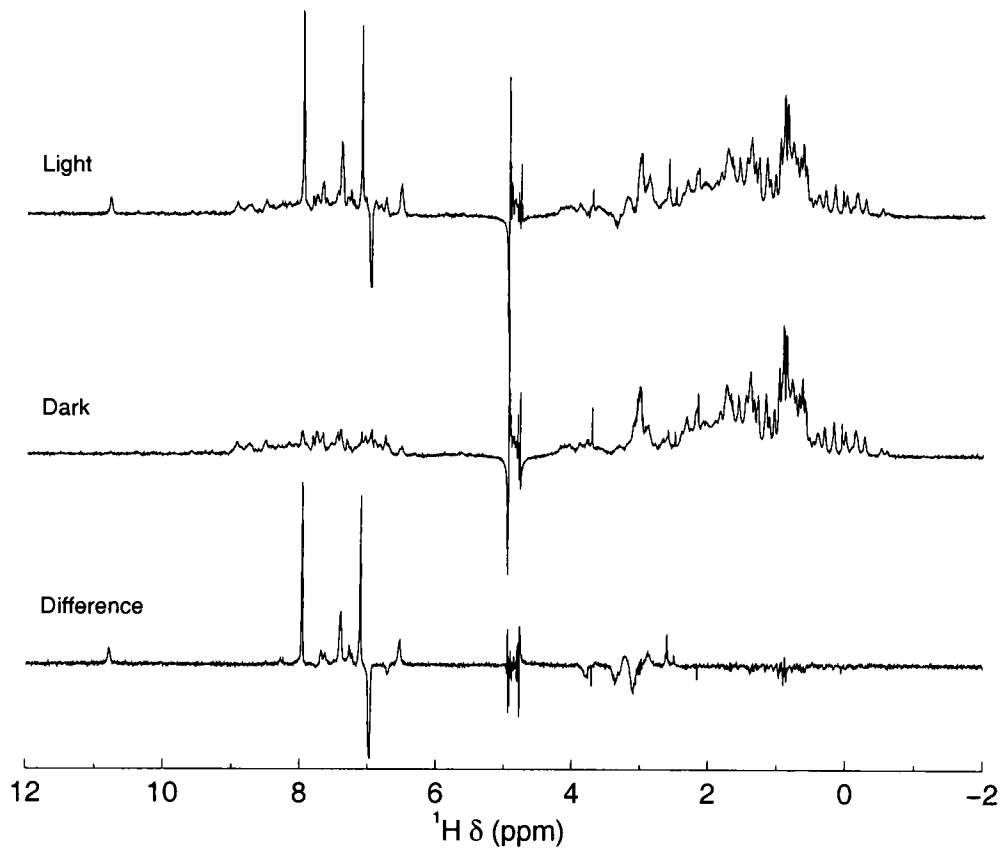


Figure 2.7: 500 MHz ^{1}H CIDNP light, dark and difference spectra of a 2 mM bovine α-lactalbumin solution at pH 7. Each spectrum is averaged over 16 scans, with a 100 ms laser pulse applied before each "light" scan.

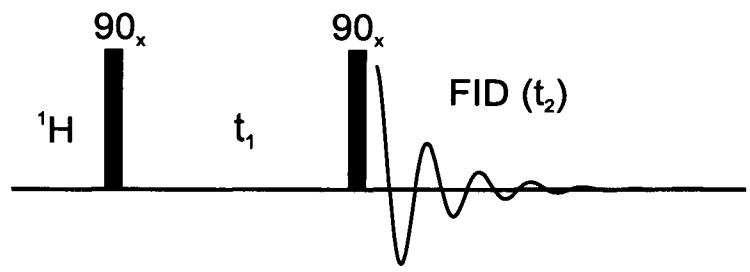


Figure 2.8: The pulse sequence timing diagram for the COSY experiment. The phases of the two RF pulses and the receiver are usually cycled in an 8-step phase cycle.

resonances which are polarised by the photochemistry. There is a signal-to-noise penalty of $\sqrt{2}$ upon subtraction, however, with the CIDNP enhancements this is usually acceptable.

2.4.2 COSY Spectroscopy

COrrelation SpectroscopY (COSY) was the first of the modern 2-dimensional NMR techniques to be developed, and provides a method for correlating mutually scalar coupled spins. The experiment consists of two RF pulses, the first of which generates transverse magnetisation, which is allowed to evolve under both chemical shift and scalar coupling Hamiltonians during the period t_1. This frequency-labelled magnetisation is then transfered from one spin to the coupling partner by the second RF pulse. The pulse sequence for the COSY experiment is shown in Figure 2.8. A 2-dimensional data set is obtained by incrementing the delay t_1. Performing a Fourier transform along t_1 and t_2 produces a 2-dimensional spectrum, in which off-diagonal cross peaks indicate the presence of a scalar coupling interaction between the two spins responsible for the associated diagonal peaks. It is also possible to extract the magnitude of the scalar coupling interaction from this experiment [74].

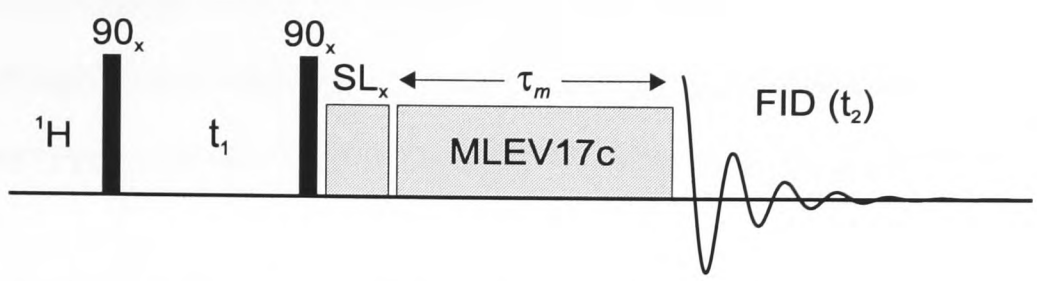


Figure 2.9: The pulse sequence timing diagram for the TOCSY experiment. Isotropic mixing is performed during the mixing period τ_m using the MLEV17c scheme [76]. SL_x represents a low power spin-lock period, usually 2 ms in duration. The phases of the three RF pulses and the receiver are usually cycled in an 16-step phase cycle.

2.4.3 TOCSY Spectroscopy

COSY spectra, while extremely useful in allowing through-bond correlations to be determined, are limited in that they only allows directly coupled spins to be correlated. This can restrict the usefulness of this technique if there is ambiguity in the spectrum. Consider for example, two spin systems AMX and $A'M'X'$. If M and M' happen to have the same chemical shift, cross peaks associated with these spins will lie at the same position. This means it is not possible to unambiguously assign the spin system topology [75]. These problems can be overcome using a relayed coherence transfer experiment [74], the most common of which is the TOCSY experiment[1]. The pulse sequence is shown in Figure 2.9. Frequency-labelled magnetisation is transferred from one spin in the system to the next using isotropic mixing under the influence of the strong coupling Hamiltonian [74, 75]. The experiments performed in this thesis use the MLEV17c mixing scheme [76].

The relayed nature of the TOCSY experiment allows the complete spin system to be mapped out. For example, consider the linear three spin system AMX, in which there is scalar coupling only between neighbouring spins, *i.e.* A to M and M to X. In a COSY spectrum cross peaks would be observed between the AM spins and between the MX spins. In the TOCSY experiment on the same spin system an

[1]TOtal Correlation SpectroscopY

additional correlation would be observed between the AX spins, despite the lack of a direct through-bond coupling. It is these additional correlations which allows the potential ambiguity of the COSY experiment to be circumvented [75].

2.4.4 HSQC Spectroscopy

The ability to correlate the proton resonances coupled to heteronuclei such as carbon-13 and nitrogen-15 provides a powerful means of gaining more information to assist in the characterisation of a protein. In the study of biological molecules the principal technique used for this is the Heteronuclear Single Quantum Correlation (HSQC) experiment. This experiment consists of three distinct parts. Firstly there is a transfer of magnetisation from the high-γ proton to the low-γ nitrogen-15, using an INEPT[2] style transfer. This magnetisation is then stored on the nitrogen nucleus as single quantum coherence and allowed to evolve under the influence of the ^{15}N chemical shift Hamiltonian for a period t_1. The final section consists of a reverse INEPT transfer of this magnetisation back to the high-γ proton nucleus for detection. Since the magnetisation begins and is detected on the proton, greater sensitivity is achieved compared with direct detection of nitrogen-15, or just a one-way transfer. Figure 2.10 shows the pulse sequence for the gradient-enhanced HSQC experiment [77]. This experiment uses the gradient pulses to select the desired coherence transfer pathway [78–81] and for suppression of the solvent signal. As drawn the experiment can be used to record ^{15}N-^{1}H HSQCs on isotopically labelled samples or ^{13}C-^{1}H HSQCs on unlabelled samples, changing the delay τ to be appropriate for the ^{13}C-^{1}H coupling constant. The pulse sequence has to be modified slightly for use on carbon-13 labelled samples, which require decoupling of neighbouring ^{13}CO and $^{13}C^{\beta}$ (and ^{15}N if the sample is doubly labelled) spins during the t_1 evolution period.

[2]Insensitive Nuclei Enhanced by Polarisation Transfer

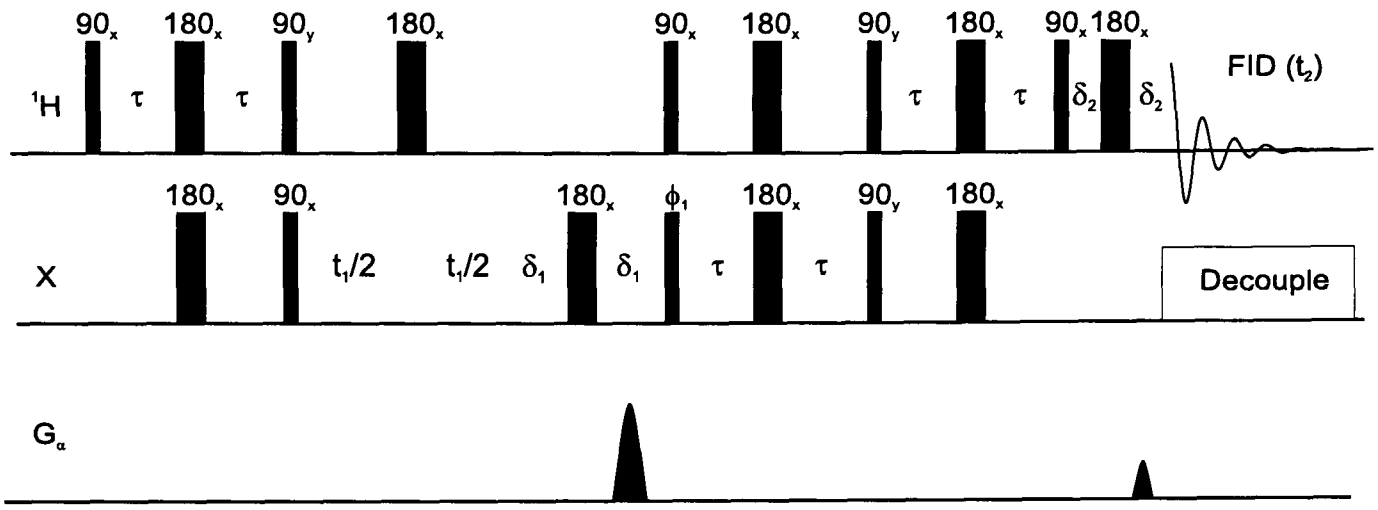


Figure 2.10: The pulse sequence for the gradient-enhanced HSQC experiment [77]. The delays are set such that $\tau = \delta_1 = 1/(4J_{\mathrm{HX}})$ and $\delta_2 = 0.3$ ms. Two transients are recorded for each t_1 increment, with the phase ϕ_1 inverted from x to $-x$ and the amplitude of the second gradient inverted also. The remaining pulses are cycled using a 16-step phase cycle.

2.4.5 NOESY Spectroscopy

Each of the NMR correlation techniques described above rely on using the through-bond scalar coupling interactions to transfer magnetisation around the spin system of interest. In the context of peptides and proteins homonuclear methods are limited to providing correlations within each amino acid residue. This is due to the extremely small H_i^{α} to H_{i+1}^{N} ^4J-coupling across the peptide bond. The so-called triple resonance pulse sequences such as HNCA, HN(CA)CO and HCA(CO)N are capable of "jumping" over this bond, however, require ^{13}C,^{15}N isotopic labelling of the protein or peptide [74]. There is an alternative and complimentary method, which uses though-space interactions rather than using through-bond couplings.

NOESY[3] makes use of the fact that there is a magnetic dipolar interaction between two spins which are close in space, and consequently do not relax independently of each other. The evolution of the z-magnetisation during the mixing period

[3]Nuclear Overhauser Effect SpectroscopY

τ_m can be described by the simplified Solomon Equations [75]:

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix}\langle I_{1z}\rangle\\ \langle I_{2z}\rangle\end{pmatrix} = \begin{pmatrix}-R_{\text{auto}} & R_{\text{cross}}\\ R_{\text{cross}} & -R_{\text{auto}}\end{pmatrix}\begin{pmatrix}\langle I_{1z}\rangle\\ \langle I_{2z}\rangle\end{pmatrix} \tag{2.2}$$

where R_{auto} is the leakage relaxation rate and R_{cross} is the cross-relaxation rate, that is, the rate of transfer of magnetisation from one spin to the other. These rates are related to the spin-lattice relaxation rate via: $R_1 = R_{\text{auto}} - R_{\text{cross}}$. For small values of τ_m the amplitudes of the cross peaks in the NOESY experiment are related to the cross-relaxation rate:

$$a_{\text{cross}}(\tau_m) \approx R_{\text{cross}}\tau_m \tag{2.3}$$

where the cross-relaxation rate is given by [75]:

$$R_{\text{cross}} = \frac{\mu_0^2\hbar^2\gamma^4}{40\pi^2 r^6}\left\{J(0;\tau_c) - 6J(2\omega_0;\tau_c)\right\} \tag{2.4}$$

where r is the distance between the spins undergoing cross relaxation and $J(\omega;\tau_c)$ is the spectral density function, which in turn depends on the rotational correlation time, τ_c:

$$J(\omega;\tau_c) = \frac{\tau_c}{1+\omega^2\tau_c^2} \tag{2.5}$$

therefore:

$$R_{\text{cross}} = \frac{\mu_0^2\hbar^2\gamma^4}{40\pi^2 r^6}\left\{\tau_c - \frac{6\tau_c}{1+4\omega_0^2\tau_c^2}\right\} \tag{2.6}$$

This equation predicts that for short rotational correlation times (fast tumbling, $|\omega_0\tau_c| \ll 1$), the NOESY cross peaks will be negative, while for long correlation times (slow tumbling, $|\omega_0\tau_c| \gg 1$), the cross peaks will be positive. It also predicts that at a critical correlation time when $\tau_c^{\text{crit}} = \left|\frac{\sqrt{5}}{2\omega_0}\right|$ the intensity of the cross peaks will vanish [75]. The rotational correlation time is of the order of 10 ns for a 14.5 kDa globular protein.

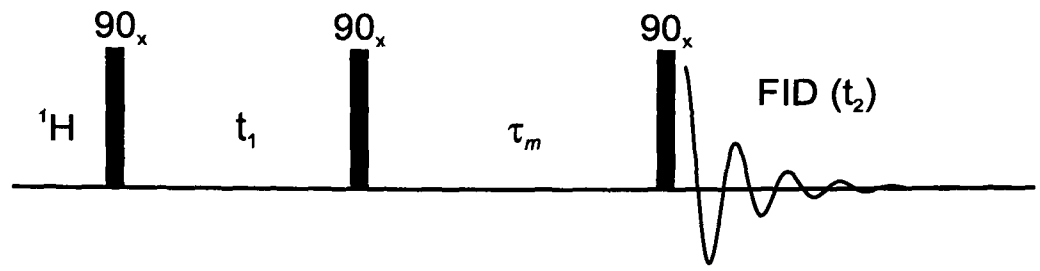


Figure 2.11: The pulse sequence timing diagram for the NOESY experiment. τ_m denotes the mixing time during which magnetisation is transferred via dipolar couplings. The phases of the three RF pulses and the receiver are usually cycled in an 16-step phase cycle.

Figure 2.11 shows the pulse sequence used for the NOESY experiment. The mixing period τ_m can be varied to allow longer range transfer, *i.e.* between spins that are distant in space. Typically, efficient transfer is achieved with $\tau_m \approx T_1$. Increasing the mixing time can allow longer range interactions to be observed, however, there is the so-called spin-diffusion limit. At this point multiple-step transfers occur, *i.e.* transfer from A to M then on to X. The second transfer reduces the intensity of the cross peak corresponding to the A to M transfer. This therefore limits the structural information available.

NOESY Spectroscopy is used mainly to obtain amino acid residue connectivities when performing sequence specific assignments, and in providing distance restraints for use in structure calculations [74].

2.5 NMR Data Processing

All NMR data in this thesis were processed and analysed using a combination of the `NMRPipe` package [82], `Felix 2.3` (MSI) and Varian's `VNMR` and `VnmrJ` software. 2-dimensional data sets were overlayed using the `NMRView` package [83].

2.6 Other Biophysical Techniques

Other biophysical techniques are also extremely valuable for use in the study of proteins and peptides. In this thesis, these techniques principally involve the use of various optical absorption spectroscopies. The equipment used for two of these techniques is described below.

2.6.1 UV/Visible Absorption Spectroscopy

The UV/visible absorption spectra shown in this thesis were recorded using a Varian Cary 3 spectrophotometer with a 1 cm cell pathlength. All measurements were performed at room temperature. The spectra shown are all baseline corrected to remove any contribution from the solvent system used.

2.6.2 Circular Dichroism

Circular dichroism (CD) uses the differential absorption of the left and right circularly polarised components of plane polarised light by chromophores in chiral molecules. This information can be used to infer details of the environment of the chromophore. In the case of proteins, far-UV CD utilises the peptide bond to provide information about the secondary structure while near-UV CD uses the asymmetric environment of the aromatic amino acid sidechains of phenylalanine, tyrosine and tryptophan to probe the tertiary structure [84].

In this thesis a JASCO J720 Spectropolarimeter was used to record the CD spectra. Quartz cells were used for all measurements, with a 1 mm cell pathlength for far-UV and 1 cm for near-UV measurements. The spectra shown are all baseline corrected to remove any contribution from the solvent system used.

2.7 Chemicals

The proteins bovine α-lactalbumin, bovine pancreatic ribonuclease A and hen egg white lysozyme used in this thesis were purchased from Sigma-Aldrich and used as obtained, as were the photosensitisers flavin mononucleotide and thionin. Human α-lactalbumin was purchased from Fluka. The sample of trpzip4 shown in the introduction (Chapter 1) was the gift of Dr. Terrence G. Oas at Duke University Medical Center. The cytochrome c_{552} protein from *Hydrogenobactor thermophilus* used in Chapter 3 was a gift from Dr. Lorna J. Smith of the Oxford Centre for Molecular Sciences. The LOV2 domain from *Avena sativa* used in Chapter 6 was from Dr. Kevin H. Gardner from the University of Texas, Southwestern Medical Center at Dallas, TX. The peptide characterised in Chapter 8 was synthesised by Ms. Janet E. Banham using standard solid-phase methods.

Pre-made 8 M guanidinium hydrochloride solution was obtained from the Pierce Chemical Company. Deuterium oxide was purchased from Apollo Scientific Ltd. Deuterated guanidine, urea, and phosphate buffers were prepared in D_2O and lyophilised three times to achieve perdeuteration. 2,2,2-trifluoroethanol-d_3 was obtained from Cambridge Isotope Laboratories. All other chemicals and buffers were purchased from Sigma-Aldrich. The sample pH was adjusted using small amounts of deuterium chloride and sodium deuteroxide, and measured using a Corning 270 pH-meter. All pH measurements are uncorrected for any deuterium isotope effect.

Chapter 3

Photo-CIDNP of a Haem Containing Protein: Cytochrome c_{552} from *Hydrogenobactor thermophilus*

Cytochromes are ubiquitous in biological systems, being one of the most evolutionary conserved protein sequences. Their primary function is that of electron transport, either in eukaryotic mitochondria, or to and from membrane bound proteins such as the photosynthetic reaction centre [85]. They are characterised by secondary structure comprised predominantly of α-helices, wrapped around a prosthetic haem group. This haem is key to the electron transport properties of the cytochrome proteins.

The study of chromophore-containing proteins using photo-CIDNP techniques has been limited due to absorption of the laser irradiation by the chromophore. This severely decreases the number of triplet state flavin molecules produced, and hence results in much smaller nuclear polarisation. There appears to be only one previous

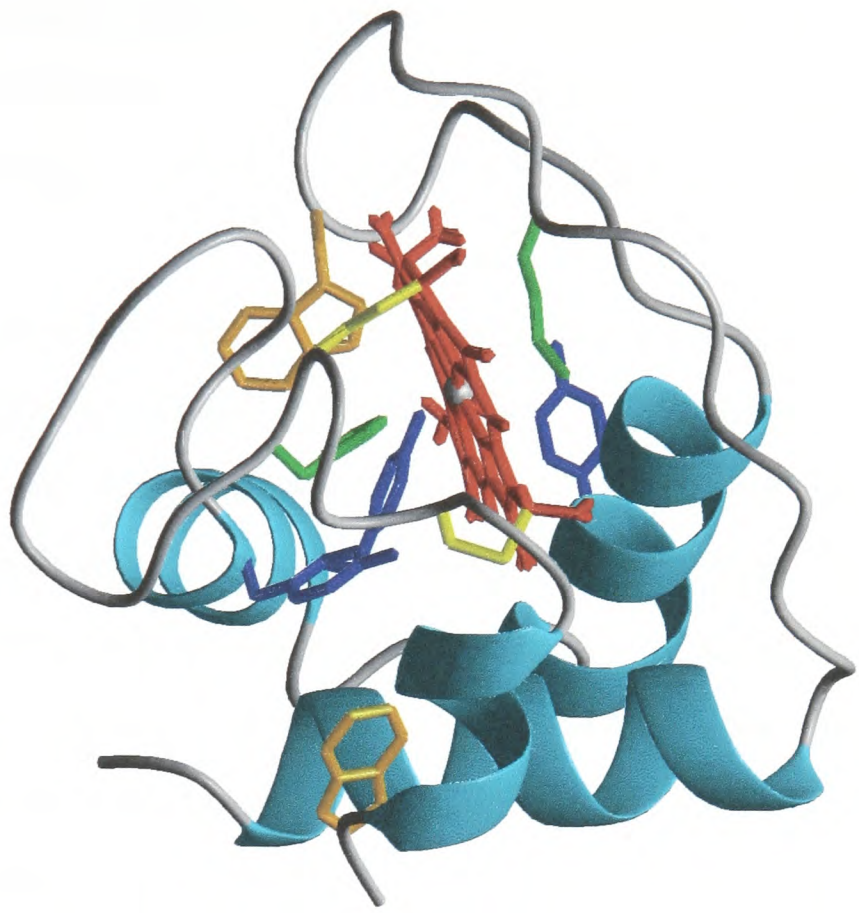


Figure 3.1: The solution structure of wild-type cytochrome c_{552} (PDB File: 1AYG [88]) The haem group is shown in red, tyrosines in blue and tryptophans in orange. The histidine and methionine iron ligands are shown in green, with the haem binding cysteines in yellow. The figure was drawn using the program `MOLMOL` [89].

example of photo-CIDNP studies of *holo* cytochromes, that of rabbit and bovine cytochrome b_5 [86].

Cytochrome c_{552} from *Hydrogenobactor thermophilus* is a small monomeric protein containing eighty amino acid residues, which covalently binds haem through two cysteine residues in a CXXCH motif [87]. The cysteine residues are at positions ten and thirteen along the amino acid sequence. Figure 3.1 shows the solution structure of wild-type cytochrome c_{552} determined using multidimensional NMR techniques [88]. The haem group is shown in red, with the potentially polarisable residues highlighted also. As can be seen, the protein consists mainly of α-helices

and loop regions, with the latter forming a lid-like structure over the haem binding site. The six potentially polarisable residues are His 14, Tyr 25, Tyr 32, Tyr 41, Trp 54 and Trp 75. Of these, in the *holo* mutant and wild-type protein histidine 14 is involved in axial ligation of the iron centre in the haem group, along with methionine 59.

Mutation of two haem-binding cysteine residues to alanines, using site-directed mutagenesis, results in a *b*-type cytochrome. The modified protein still maintains a native-like fold and binds the haem moiety non-covalently [87, 90]. It has been shown that mutation of either of the two haem-binding cysteine residues has little effect on the physicochemical properties of the protein. However, mutation of both results in reduced stability of the protein as indicated by a decrease of 40 K in the thermal denaturation temperature [91]. It has also been shown that over time, the haem group is lost by the mutant protein, and the subsequent *apo* protein undergoes amyloid fibril formation after a period of several weeks [92]. Removal of the haem moiety reveals greater changes in the *apo* protein structure. For example, the protein is found to be only partially folded by CD spectroscopy, with the total helical content reduced by 29% compared to that found in the *holo* mutant [90]. NMR diffusion measurements have shown that there is a 8.5% increase in the effective hydrodynamic radius of the *apo* state with respect to the *holo* protein [90]. Both of these observations can be attributed to the protein populating an ensemble of interconverting conformers, possibly a molten globule-like state [93]. Addition of free haem to a solution of the *apo* C10A/C13A mutant protein results in the spontaneous formation of the *b*-type *holo* species, suggesting that in the *apo* state there is a nascent haem binding pocket formed [87, 90].

Since the *holo* mutant and wild-type proteins both contain a haem group, the oxidation state of this group will have a major effect on the protein properties and consequently on the NMR spectrum. In the wild-type protein, the reduced state

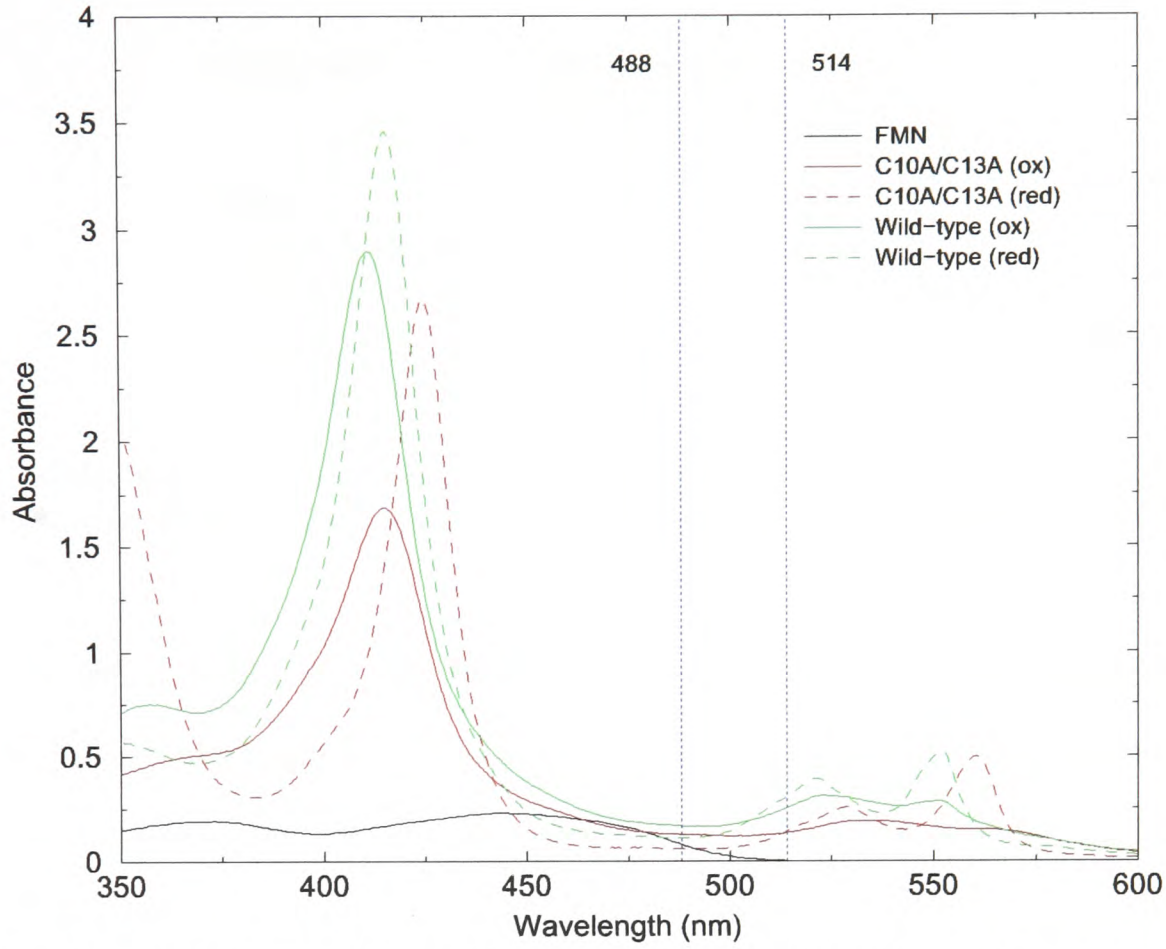


Figure 3.2: The UV/vis spectra of 20 μM solutions of the C10A/C13A and wild-type variants of cytochrome c_{552}, and a 20 μM FMN solution. The two vertical lines represent the principal wavelengths emitted by the Ar^+ laser.

of the iron in the haem group is low-spin Fe(II) which is diamagnetic, while in the air-oxidised protein it is low-spin paramagnetic Fe(III), with an electron spin angular momentum, $S = \frac{1}{2}$ [94]. There are two effects of this paramagnetism, firstly there will be a paramagnetic contribution to the chemical shift Hamiltonian, the so-called Fermi contact and dipolar shifts [95], and secondly the electron-nuclear dipolar interaction will increase nuclear spin relaxation rates in the proximity of the haem group.

The presence of the haem group results in the characteristic blood red colour of the protein solutions (with the obvious exception of the *apo* state). The UV/visible absorption spectra are shown in Figure 3.2, along with that of FMN for comparison. As can be seen from this figure, there is very strong absorption by both forms of the protein between 380 and 440 nm. As the principal wavelengths emitted by the

Compound	Extinction Coefficient	
	488 nm	514 nm
FMN	4080	236
C10A/C13A (ox)	6470	6950
C10A/C13A (red)	3075	7200
Wild-type (ox)	8600	12765
Wild-type (red)	5663	15370

Table 3.1: Molar extinction coefficients ($dm^3mol^{-1}cm^{-1}$) at 488 and 514 nm for the oxidised and reduced forms of the *holo*, wild-type proteins and for comparison, FMN.

laser are at 488 and 514 nm, the extinction coefficients at these wavelengths are of interest. These are given in Table 3.1. It is immediately clear that all forms of the protein absorb strongly, with only the reduced form of the *holo* protein having an extinction coefficient smaller than FMN at 488 nm. It will therefore be difficult to generate the nuclear polarisation as a decreased amount of light will reach the photosensitiser.

The C10A/C13A mutant and wild-type proteins were expressed in *Escherichia coli* by Ms R. Wain, using the procedure described in [87]. The *apo* form of the C10A/C13A mutant was prepared using 1 M imidazole and incubating overnight at 4°C. The resulting supernatant solution was purified using centrifugation and dialysed against 20 mM sodium phosphate at pH 7.3 [87].

3.1 Photo-CIDNP of the Air-Oxidised Proteins

Since the air-oxidised samples are the easiest to prepare, this was chosen as the best starting point for these experiments. Figure 3.3 shows the aromatic region of the NMR spectra for the *apo* and *holo* states of the C10A/C13A mutant, and the wild-type protein. It is immediately clear from this that the *apo* state of the protein has a much reduced chemical shift dispersion compared to either the *holo*

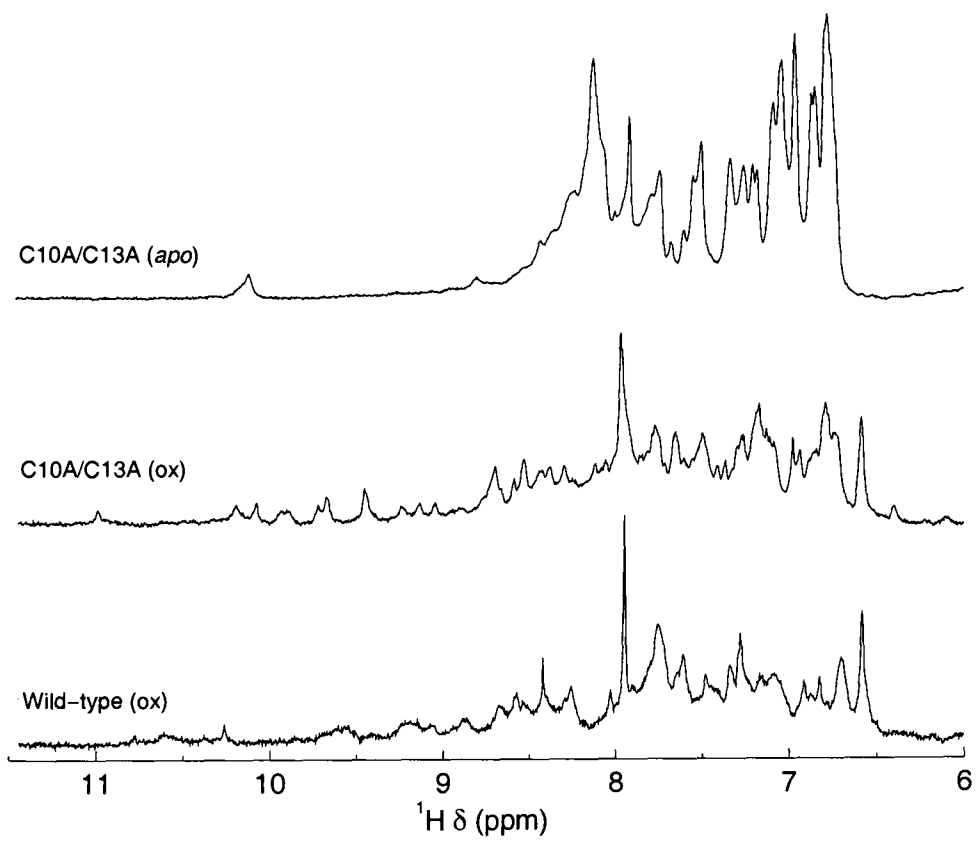


Figure 3.3: 600 MHz ^{1}H NMR spectra of a 200 μM solution of each of the proteins. The spectra are averaged over 256 scans, and processed with 2 Hz exponential line broadening prior to Fourier transformation.

state of the mutant or the wild-type protein. This is attributable to the ensemble of interconverting conformers which characterise this species, as described in the introduction to this Chapter. Both of the haem containing proteins show well resolved spectra, indicating that the proteins are folded. There are subtle differences apparent between the two spectra presumably arising from the substitution of the two haem binding cysteines for alanine in the mutant protein, causing slight structural changes in the native fold of the *holo* protein.

The generation of CIDNP in the *apo* state of the mutant protein poses no technical challenges since the protein is colourless in the visible region of the electromagnetic spectrum. This is shown in the upper spectrum of Figure 3.4. The middle trace shows that it is possible to generate nuclear polarisation in the *holo* state of the C10A/C13A mutant protein, however, with lower efficiency as indicated by the decreased signal-to-noise ratio. The bottom spectrum in Figure 3.4 is that of the

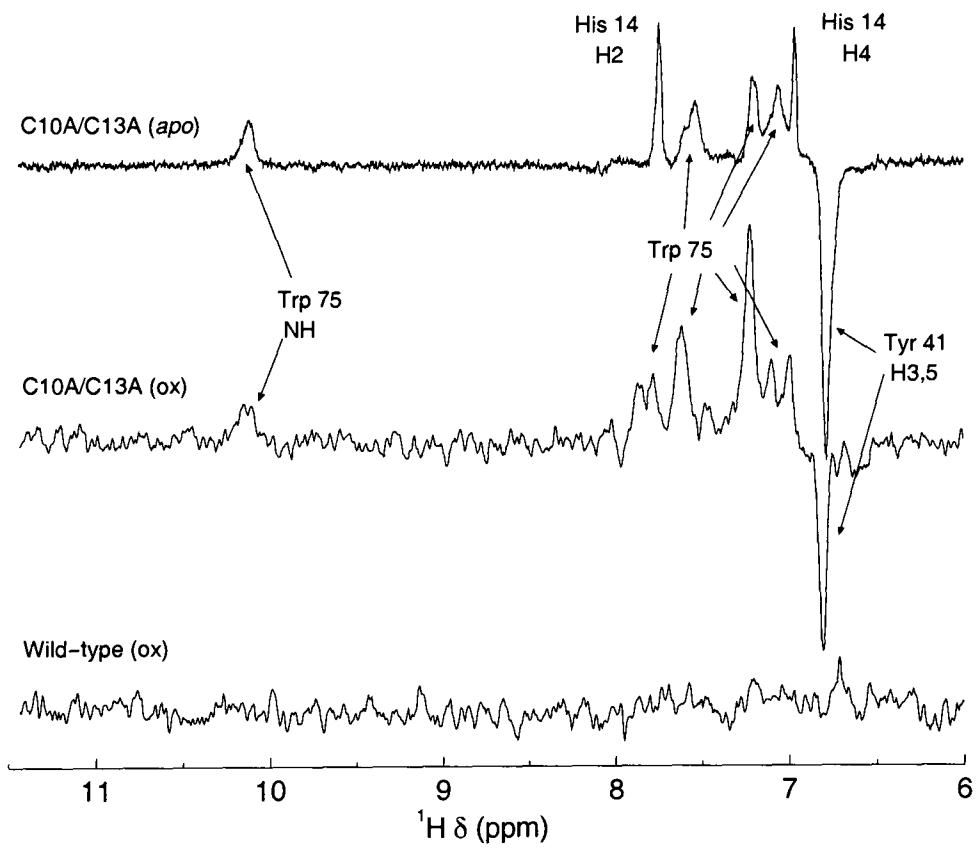


Figure 3.4: 600 MHz ^{1}H CIDNP spectra of a 200 μM solution of each of the proteins, containing 200 μM FMN. The spectra are averaged over 16 light and dark pairs, and processed with 10 Hz exponential line broadening prior to Fourier transformation, with the exception of the *apo* state which has 2 Hz line broadening applied. A laser pulse duration of 100 ms at 5 W was used in these experiments. The assignments are explained in the text.

wild-type protein, there is no evidence of any nuclear spin polarisation in this sample. The reasons for this maybe be attributable to the oxidation state of the haem group in the protein. This is considered in more detail in Section 3.2.

Since CIDNP spectra are greatly simplified compared to the corresponding NMR spectra, this opens up the possibility of assigning the observed polarised signals to specific amino acid residues in the protein. The singlets at 7.75 and 6.96 ppm are characteristic of the H2 and H4 protons of histidine. Since His 14 is the only histidine in the structure, these must correspond to the histidine in the haem binding pocket. In the *apo* state the haem binding pocket is not occupied, however, it is postulated to still be formed [90]. Therefore this allows for the idea the pocket could be occupied transiently by an excited triplet state flavin molecule producing the nuclear polarisation in the histidine residue.

The assignment of the other resonances can be made using static accessibility arguments and by comparison of chemical shift assignments obtained from 3D-TOCSY-HSQC and 3D-NOESY-HSQC experiments on ^{15}N labelled reduced-state C10A/C13A protein [96, 97]. These three-dimensional NMR experiments allow the backbone amide protons to be assigned, and they have the added benefit of allowing the indole NH proton of tryptophan residues to be assigned. Using this information, the broad singlet seen at 10.11 ppm can be assigned to Trp 75. The actual figure reported by Wain for this indole proton is 10.27 ppm, this small difference is due to the fact that these spectra were recorded using the air-oxidised species, rather than the reduced protein used by Wain *et al.* [90, 96, 97]. Only small changes in the detected chemical shift between the oxidised and reduced state of the haem containing horse heart cytochrome-*c* have been reported [98–102]. The assignment of this indole signal allows corresponding absorptive resonances in the region from 8.0 to 6.8 ppm to be assigned to the same tryptophan residue.

The accessibility of the amino acid side chain to the solvent can be calculated

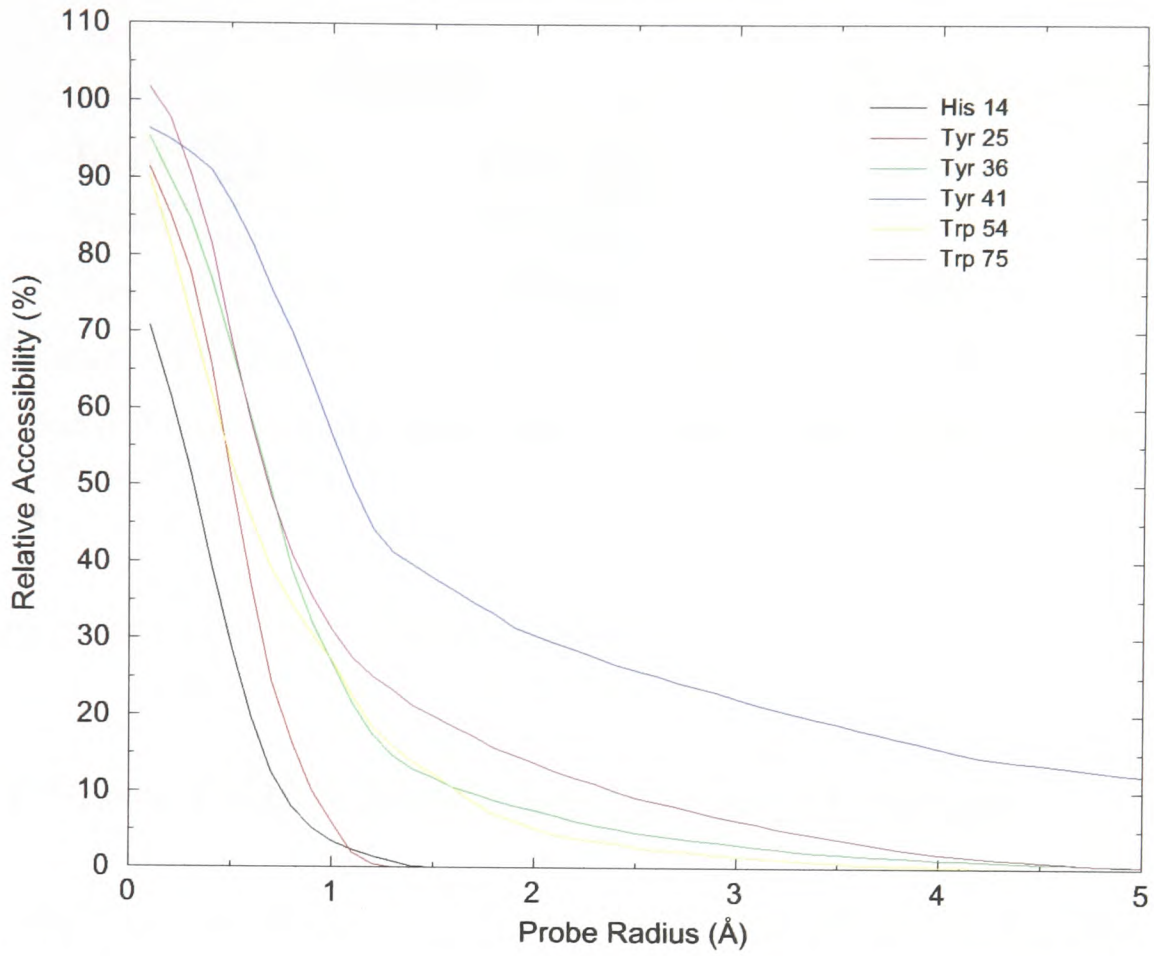


Figure 3.5: Static side chain accessibilities as a function of probe radius for the wild-type cytochrome c_{552}, calculated using PDB file: 1AYG [88].

by rolling a sphere of a given radius (the probe radius) over the surface of the protein [103]. The static accessibility is quoted as that relative to the accessibility in an extended conformation of the model tripeptide Ala-Xaa-Ala, where Xaa represents the amino acid of interest. These calculations were performed using the program `Naccess 2.1.1` [104].

Figure 3.5 shows results of these calculates for the six potentially polarisable resides as a function of the probe radius, using the NMR-derived structure of the wild-type protein. This plot allows the exposed tyrosine residue, seen at 6.81 ppm in Figure 3.4, to be assigned to the H3,5 protons of Tyr 41. This is clearly exposed, even to relatively large probe radii. The use of different size probes, and the monitoring of the change in accessibility as a function of probe radius allows the increased size of FMN compared to the standard probe (that of water), to be investigated [105]. Also, the FMN molecule is a planar tricyclic aromatic ring system, not spherical as

No.	Reaction	$E^{\ominus}$ (mV)	Reference
1	$Fe(III)^{AXXA}_{(haem)} + e^- \rightarrow Fe(II)^{AXXA}_{(haem)}$	+170	[91]
2	$Fe(III)^{WT}_{(haem)} + e^- \rightarrow Fe(II)^{WT}_{(haem)}$	+245	[91]
3	$\frac{1}{2}Fl_{(aq)} + H^+_{(aq)} + e^- \rightarrow \frac{1}{2}FlH_{2(aq)}$	−205	[106]
4	$SO^{2-}_{3(aq)} + H_2O_{(l)} + e^- \rightarrow \frac{1}{2}S_2O^{2-}_{4(aq)} + 2OH^-_{(aq)}$	−565	[107]

Table 3.2: Half-cell redox potentials the C10A/C13A mutant and wild-type proteins, along with FMN (Fl) and dithionite.

is assumed in the calculations of static accessibility.

3.2 Photo-CIDNP of the Reduced State

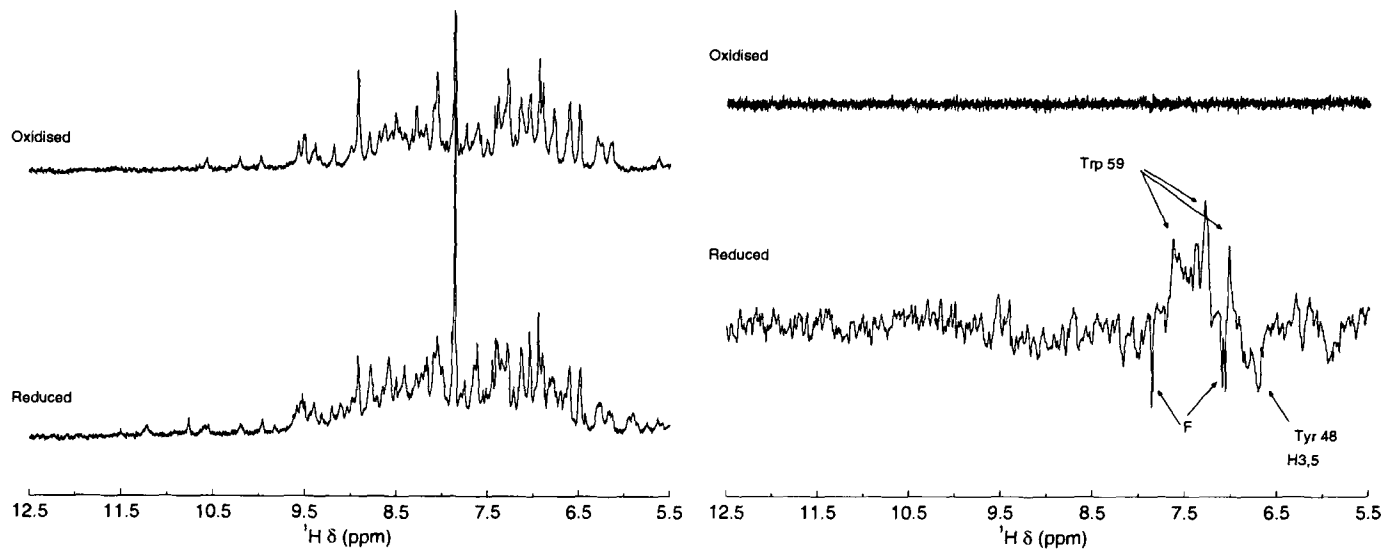
Previous NMR studies of the C10A/C13A mutant of cytochrome c_{552} have been performed on the reduced state of the protein [90, 96, 97]. This removes the complication of the paramagnetic centre in the haem group. In order to allow a comparison with previous NMR studies, it is desirable to also study the reduced state of the protein. Cytochromes are involved in electron transport chains in biological systems. These chains often also include flavin derived cofactors, so it is important to analyse the electrochemistry of the protein and FMN to determine if the reduced state experiments are feasible.

Table 3.2 shows the half cell reactions and their standard electrode potentials for the reactions of interest to this Chapter. Sodium dithionite (Eqn. 4 in Table 3.2) is the reductant of choice for these experiments as it easily reduces the protein ($E^{\ominus} = +735$ mV for the C10A/C13A mutant and $E^{\ominus} = +810$ mV for the wild-type protein). The usual protocol of placing a small about of dithionite in the bottom of the NMR tube and flame sealing [90, 97] is not practical for these CIDNP experiments since the coaxial insert must be placed in the tube, and more importantly, dithionite efficiently reduces the flavin to the colourless hydroquinone, rendering the

CIDNP experiments impossible.

In order to prepare the reduced state samples, the protein reduction must be performed under anaerobic conditions, followed by removal of the dithionite using diafiltration. Degassed flavin solution must be added to this, while maintaining an oxygen-free environment. To test the hypothesis that the oxidation state of the haem moiety influences the nuclear polarisation generated, an experiment was performed using horse-heart cytochrome *c*. This is a small globular protein of 104 residues, with four α-helices surrounding a single haem group. The iron is coordinated by a histidine and a methionine side chain, similar to cytochrome c_{552}. Two solutions were prepared, one of the air-oxidised species and one of the reduced system. The reduced system was prepared as follows: a solution of 100 μM protein was made with solvent which had been degassed with argon for two hours prior to use. This was then reduced using an excess of sodium dithionite. The reductant was subsequently removed from the solution using a YM-3 microcon diafiltration column with a molecular weight cut-off of 3 kDa. 200 μM degassed FMN solution was then added to the protein solution, which was placed in an NMR tube containing an argon atmosphere.

The results of these preliminary experiments are show in Figure 3.6. Like the wild-type cytochrome c_{552}, cytochrome *c* has two oxidation states, with the oxidised form being low-spin Fe(III) ($S = \frac{1}{2}$), and the reduced state, low-spin Fe(II) ($S = 0$). The CIDNP spectrum of the air-oxidised species does not show any evidence of the generation of nuclear polarisation, while there is weak enhancement of a tyrosine and a tryptophan residue in the reduced state. Comparison with the known chemical shift assignments for cytochrome *c* indicates that these signals arise from Tyr 48 and Trp 59 [98]. This therefore offers a possible explanation for the lack of polarisation observed in the air-oxidised, and hence paramagnetic, wild-type of cytochrome c_{552} shown in Figure 3.4. The presence of a paramagnetic centre in the system inhibits



(a) NMR Spectra

(b) CIDNP Spectra

Figure 3.6: 600 MHz ^{1}H NMR and CIDNP Spectra of 100 μM Cytochrome *c*. The CIDNP spectra were recorded using a 100 ms laser pulse at 5 W preceding each scan. F denotes a polarised flavin signal. Assignments taken from [98].

the detection of nuclear polarisation, either through quenching of the triplet flavin such that no polarisation is produced, or through efficient relaxation of the generated nuclear polarisation via dipolar coupling to the electron spin. It is, however, unclear at this time why polarisation *is* generated in the *holo* state of the C10A/C13A mutant.

3.3 Last thoughts / conclusions

This chapter shows that, under certain conditions, it is possible to obtain photo-CIDNP spectra of a haem containing protein. These spectra suffer from reduced signal-to-noise ratio due to two factors. The first is that much lower protein concentrations than usual for NMR experiments are used in an attempt to lower the optical density of the sample. This high optical density contributes to the second problem: that only weak polarisation is generated due to the decreased production of triplet state photosensitiser. A large improvement in the signal-to-noise ratio

could be obtained by the use of a different sample illumination method [69], or the use of a cryogenic probe, in which the RF coil and the pre-amplifier are cooled to around 30 K, reducing the background noise. Also, the replacement of the photo-sensitiser by a thermolytic radical initiator could by-pass the high optical density of the sample.

The fact that CIDNP was only observed in the air-oxidised form of the C10A/C13A mutant protein and not the oxidised wild-type cytochrome c_{552} or cytochrome *c* may suggest a general property of *b*-type cytochromes. The only previous examples of the observation of CIDNP in a haem-containing protein is that of bovine and rabbit cytochrome b_5, again, using the air-oxidised form of the protein [86], or of *apo* cytochromes [108].

The results of the preliminary experiments on cytochrome *c*, shown in Figure 3.6 show that is possible to efficiently reduce the protein, and generate CIDNP on these samples. Using this approach would allow the reduced state of both the C10A/C13A mutant and the wild-type cytochrome c_{552} to be studied by CIDNP, and comparisons with the air-oxidised spectra be made.

Chapter 4

Rapid Injection and CIDNP Pulse Labelling

The ability to mix reactants rapidly within the NMR magnet has many applications in protein science and chemistry in general. It allows protein folding pathways to be followed and determined [20, 22–24, 26], details of protein-substrate interactions to be studied, and permits the application of a variety of other techniques such as CIDNP pulse labelling to study the structure of folding intermediates [31, 32].

This chapter describes a simple injection set-up which has short experimental dead-times and enables rapid mixing of the injectant and buffer solutions. In order to determine the extent of the mixing, three experiments are discussed. Firstly the injection is followed on the bench-top using a digital video camera. This gives a good guide to the macroscopic mixing properties of the set-up and provides a rough guide to the time scale of the mixing process. Secondly, a pH jump experiment is performed which highlights the microscopic mixing properties, and finally NMR imaging is used to determine the distribution of the injectant within the NMR tube as a function of time after the injection event.

The main application of this mixing device is in CIDNP pulse labelling spec-

troscopy [31, 32], a technique for probing the surface structure of partially folded proteins. Results using this injection device with the proteins hen egg white lysozyme, bovine and human α-lactalbumins are presented, along with another application: that of monitoring protein folding in real-time, again using hen egg white lysozyme.

4.1 Rapid Injection

Several injection set-ups have been described in the literature [20, 26, 109–117]. While these devices are capable of mixing injectant and buffer solutions with reasonable efficiency there are problems with most of their designs, mainly long experimental dead-times, but also including poor line shapes due to B_0 inhomogeneity caused by the injector being present within the RF coil [113]. Several of these systems are constructed from modified optical stopped-flow apparatus and require modification of the NMR probe in order to incorporate the stopping syringe [109, 110, 114].

In this section the construction of the injection set-up is shown along with the results of a series of experiments designed to test and calibrate the insert for use in real-time NMR and CIDNP pulse labelling experiments.

4.1.1 The Set-up

In order to transfer protein solutions rapidly to the NMR coil for detection, an insert to the NMR tube must be constructed. This injection set-up is based on a previous design by Maeda *et al.* [117].

The insert consists of a glass capillary approximately 180 mm in length, 1.6 mm in internal diameter and an external diameter of 2.5 mm. To the end of this a narrow microlitre pipette, 25 mm in length, is fused. This pipette has an internal diameter of 0.17 mm. This produces a highly collimated jet of liquid upon injection. The

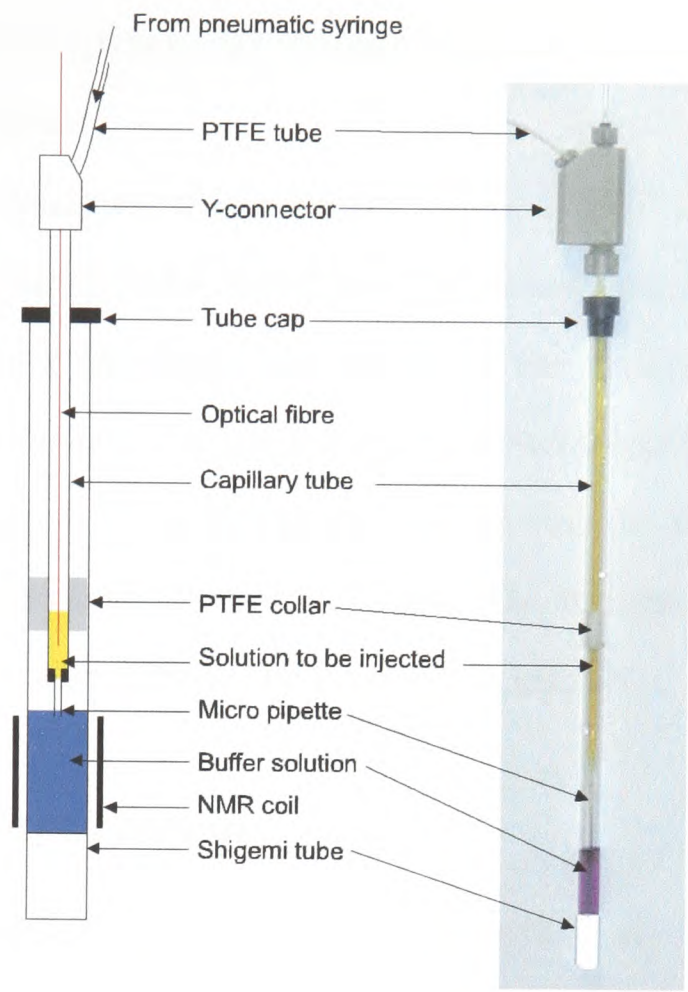


Figure 4.1: A schematic of the injection insert is shown on the left, with the photograph on the right showing the system ready for use. The photograph is enhanced by the use of 20 mM flavin mononucleotide solution loaded into the capillary tube and 1 mM methylene blue solution placed in the Shigemi tube.

glass capillary is held along the centre axis of the NMR tube by a PTFE[1] collar approximately 10 mm in length two thirds of the distance along the insert. The top of the insert is held in place by a standard NMR tube cap with a small hole drilled in the centre. The insert assembly is held 1.5 mm above the top of the RF coil in order not to disturb the field homogeneity.

At the top of the glass capillary is a Y-connector made from PEEK[2] to which a PTFE transfer line of internal diameter 0.5 mm, originating from the needle of a glass syringe (Model: SGE 500R-GT), is connected. There is also the option to introduce an optical fibre in order to perform photo-CIDNP experiments. The

[1] Poly(TetraFluoroEthylene)
[2] Poly(Ethylene Ether Ketone)

insert assembly is shown schematically in Figure 4.1, side-by-side with a photograph showing the set-up in place.

The glass syringe mentioned above is mounted in a PTFE block so as to absorb any mechanical shock caused by the injection event, therefore reducing any lineshape artifacts in the spectrum recorded. The action of the syringe is controlled by a pneumatic piston driven using nitrogen gas at 10 bar and triggered by a TTL control line from the spectrometer. Using this set-up it is possible to inject 50μL of sample using trigger pulses of less than 10 ms in duration. The syringe and piston assembly are housed on a PTFE plate outside the bore of the magnet.

4.1.2 Solvent Suppression

Modern solvent suppression sequences such as those which rely on excitation sculpting such as a DPFGSE[3] [73], or the WATERGATE[4] family of sequences [118–120] are not applicable to rapid injection experiments. The reason for this is that both of these sequences employ a gradient spin echo. The idea behind this is that the solvent signal is selectively dephased by a pair of gradients, while the signals of interest, dephased by the first gradient, are refocused by the second. This approach requires that the field experienced by the spins after the spin echo be the same as the first. This is not the case when there is bulk motion present in the sample, such as is the case after an injection event.

The intensity of NMR signal after the application of a DPFGSE as a function of diffusion coefficient and the gradient parameters is given by Equation 4.1 [121]:

$$\begin{aligned} I = \; & I_0 \;\; \exp\left(\frac{-2\tau_1}{T_2} - (\gamma\delta_1 G_1)^2 D\left(\Delta_1 - \frac{\delta_1}{3}\right)\right) \\ & \times \;\; \exp\left(\frac{-2\tau_2}{T_2} - (\gamma\delta_2 G_2)^2 D\left(\Delta_2 - \frac{\delta_2}{3}\right)\right) \end{aligned} \tag{4.1}$$

[3]Double Pulsed Field Gradient Spin Echo
[4]WATER suppression through GrAdient Tailored Excitation

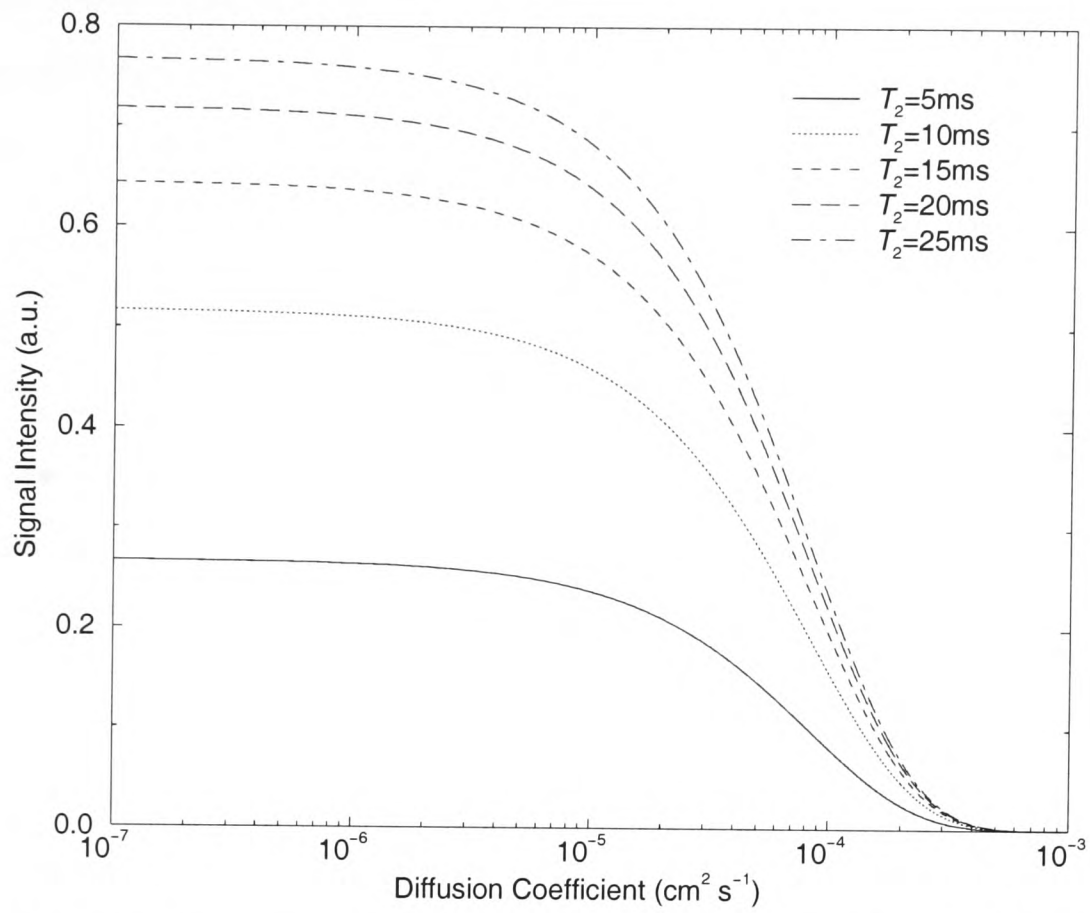


Figure 4.2: Normalised signal intensity after the application of a DPFGSE for various transverse relaxation times. $\tau_1 = 2$ ms, $G_1 = G_2 = 65$ G cm^{-1}, $\delta_1 = 1$ ms, $\Delta_1 = 4$ ms, $\tau_2 = 1.3$ ms, $\delta_2 = 0.3$ ms, $\Delta_2 = 3.3$ ms.

where I_0 is the initial signal intensity before the gradient echo, T_2 is the transverse relaxation time, γ is the gyromagnetic ratio of the nucleus undergoing the DPFGSE, D is the diffusion coefficient, and τ_i, G_i, δ_i and Δ_i are respectively the spin-echo delay, gradient strength, gradient duration and time between application of the gradients.

From this it is clear that as the diffusion coefficient increases the intensity of the signals of interest decrease rapidly. This is shown more clearly in Figure 4.2. Above diffusion coefficients of 5×10^{-5} cm^{2}s^{-1}, the signal is rapidly attenuated. For example, consider a spin with a transverse relaxation time of $T_2 = 25$ ms. This spin only has to move ~ 0.02 mm in a field gradient of 65 G cm^{-1} in order for the echo intensity to drop to 10% of that for a non-moving spin. So for this reason solvent suppression sequences of this type are not applicable to rapid injection experiments.

The problems described above are very similar to those encountered in *in vivo* NMR spectroscopy, where motion is also present. Hence *in vivo* solvent suppression sequences such as WET[5] [122], VAPOR[6] [123], DRYSTEAM[7] [124] and CHESS[8] [125] provide a good starting point for application to rapid injection experiments. Of these sequences the CHESS sequence is the simplest to understand and implement. The basic sequence consists of a selective 90° pulse on the signal to be suppressed followed by a strong field gradient to dephase the transverse magnetisation produced by the selective pulse. This pulse-gradient combination is the so-called "crush pulse". Lin *et al.* have shown that a combination of radiation damping effects and the distant dipolar field, in which long range dipolar interactions are important[9] can result in a recovery of magnetisation following the application of a crush pulse [126, 127]. Improvement of the signal suppression following a crush pulse is obtained if the field gradient is applied along the magic angle, since the dipolar interaction has a $3\cos^2\theta - 1$ angular dependence, where θ is the angle between the dipolar vector and the B_0 field vector [126–128].

For optimum results this selective crush pulse procedure is repeated three times, with the gradients applied along three orthogonal axes so that no accidental refocusing of the magnetisation occurs. There are improved versions of this sequence in which the final selective 90° pulse is replaced with a selective 180° pulse, or the delays and flip angles in the sequence are numerically optimised to compensate for T_1 and B_1 inhomogeneity effects. The optimum values are $\{\tau,\ \tau,\ 0.87\tau,\ \beta,\ 0.94\beta,\ 1.65\beta\}$, where τ represents the delay between CHESS pulses and β is the flip angle

[5]Water suppression Enhanced though T_1 effects

[6]Volume localised solvent Attenuated PrOton nmR

[7]Drastic Reduction of water signals in spectroscopY based on the STimulated Echo Acquisition Mode

[8]CHEmical Shift Selective

[9]The Distant Dipolar Field (DDF) results as follows: the dipolar interaction is a through space interaction, depending on distance as $\frac{1}{r^3}$. However, the number of molecules at a distance r, in a shell of thickness dr, contributing to the DDF, varies as $4\pi r^2 \mathrm{d}r$. Therefore the DDF decays much more slowly, as $\frac{1}{r}$.

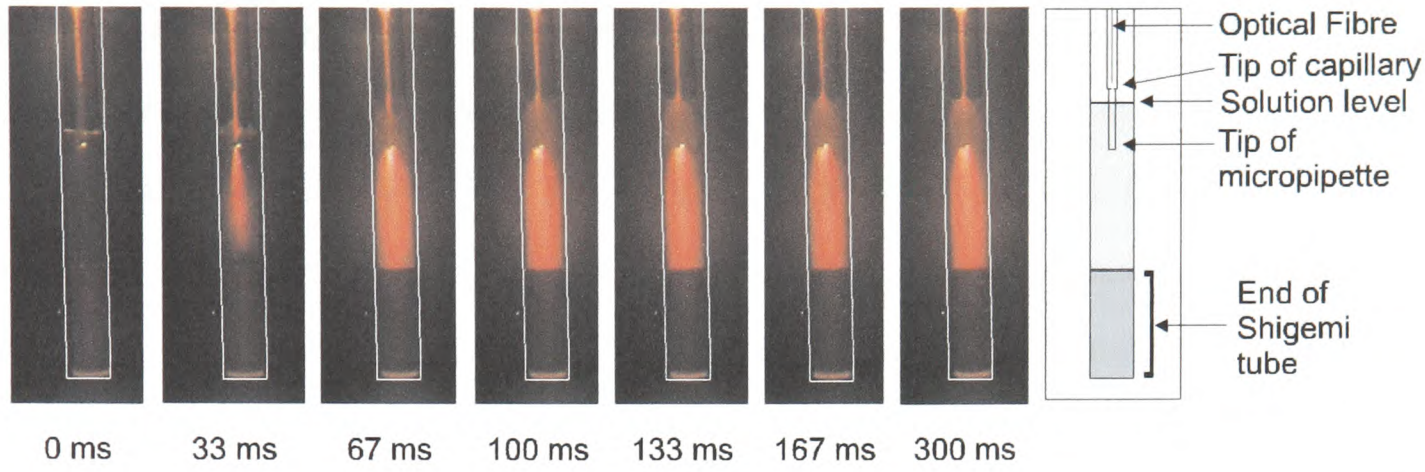


Figure 4.3: Frames captured from video footage recorded during an injection event. The time interval from the first frame is noted under each frame. The injection event occurs between the first two frames. The images appear red due to the presence of an argon ion laser filter covering the camera lens.

of the selective pulse, usually taken to be 90° [129].

For all the experiments described here the delays (2 ms) and selective flip angles (90°) were set to the same for each CHESS pulse.

4.1.3 Optical Measurements

As a first estimate of the mixing efficiency using this set-up, a series of dye injections were performed outside the NMR magnet. These involve the injection of 50 μL of methylene blue dye solution into 280 μL of H_2O in the presence of low power laser illumination (<0.5 W). Under these conditions the methylene blue fluoresces, allowing greater contrast images to be recorded. The concentration of methylene blue was set such that the optical density of the sample allowed the laser light to illuminate the entire sample tube.

Digital video footage was recorded during an injection event using a Sony DCR-TRV310 digital video camera. Still frames were captured from the footage using `iMovie` (v 2.1.1). The camera shutter speed enables a frame to be captured every 33 ms. A series of frames taken during an injection are shown in Figure 4.3. Since it was not possible to synchronise the camera shutter with the injection event, the

injection takes place at some point between the first two frames shown. From this it is clear that the macroscopic mixing is complete sometime between the 33 ms and 67 ms time points, as very little change is observed in later frames. Due to the time resolution of the camera it is impossible to gauge the mixing time more accurately, however, this does give a good guide as to the timescale of the events following injection of the dye. So as to gain a better estimate of the mixing time a different approach is needed.

4.1.4 Histidine pH Jumps

In order to measure the quality of the mixing produced by this injection device, a reaction which can be monitored by NMR and which occurs rapidly is needed. Such a reaction would have to have distinct start and end points, that is, reactants and products with differing NMR spectra. The reaction must also occur rapidly so the effects of incomplete mixing can be distinguished from any unreacted species remaining in the mixed volume. As this set-up is designed primarily for the study of protein folding reactions and protein folding intermediates, a pH jump reaction using the amino acid histidine was chosen.

The side chain of the amino acid histidine consists of an imidazole ring. Protonation and deprotonation at one of the ring nitrogens leads to changes in the chemical shifts of the two carbon bound protons (denoted H2 and H4). The ring pK_a of histidine is around 6.5, therefore a pH jump from one side of this value to the other should result in a change in the NMR signal. The experiment was carried out as follows: 280 μL of a solution of 0.15 M d-TrisDCl buffer, in D_2O, at pH 8.5 was placed in the bottom of an NMR tube. Into this 50 μL of 10 mM *N*-acetyl-histidine in D_2O was injected. This resulted in a solution with a final pH of 8.3 after injection. A series of experiments were performed using different delays after the injection, allowing a pseudo real-time data set to obtained. CHESS was used for

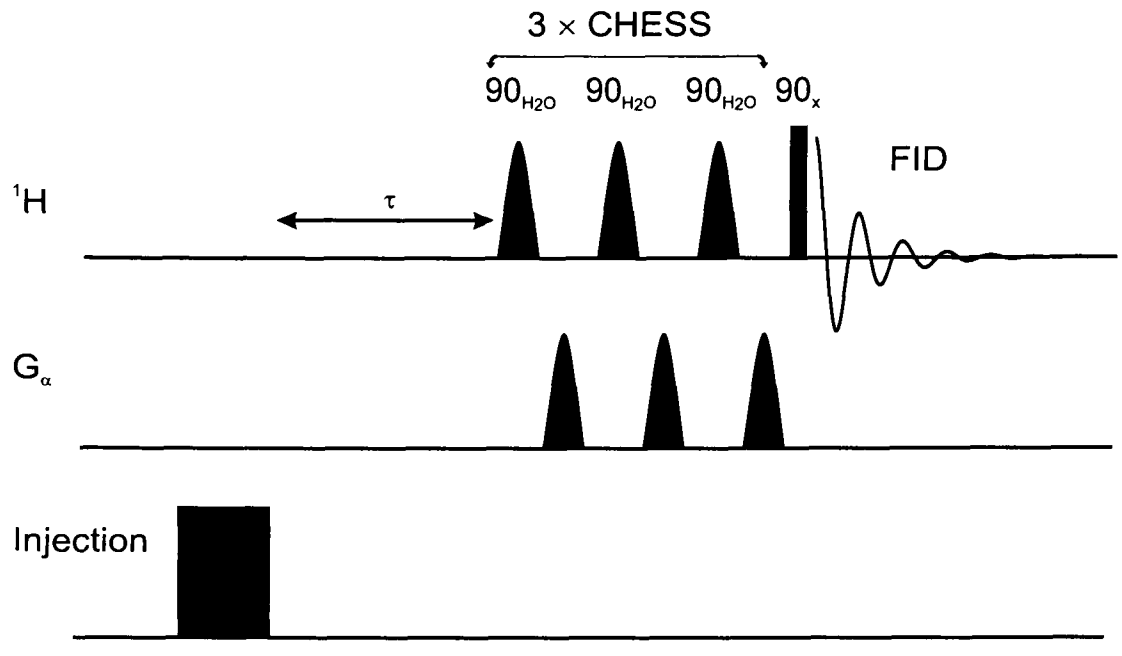


Figure 4.4: The pulse sequence timing diagram for the histidine pH jump experiment. The duration of the CHESS sequence is 12 ms. τ is the variable post-injection delay, incremented to allow construction of a pseudo real-time data set.

solvent suppression as described above, leading to a minimum delay of 12 ms. The pulse sequence timing diagram for this experiment is shown in Figure 4.4; the results for this series of injections can be seen in Figure 4.5. From this it is clear that mixing is complete within ≈40 ms of the injection event. There is very little change in the line shape or intensity of the spectra recorded with longer post-injection delays.

In the earlier spectra of Figure 4.5, specifically spectra recorded with post injection delays of between 22 to 37 ms, minor phase errors of a few degrees can be seen in the peaks that correspond to the pH 8.3 state of the system, that is, post injection. A slight broadening of the lines from the pre-injection species is also noticed. These effects can be attributed to the reaction occurring during the acquisition of the free induction decay [42, 110]. Broadening of the pre-injection species arises due to the finite lifetime of the species, that is, it is not present for the complete acquisition of the FID. The phase errors which occur in the signals from the post-injection species arise due to the transfer of transverse magnetisation from one species to another by the reaction during the acquisition of the FID. Figure 4.6(b) shows an exaggerated

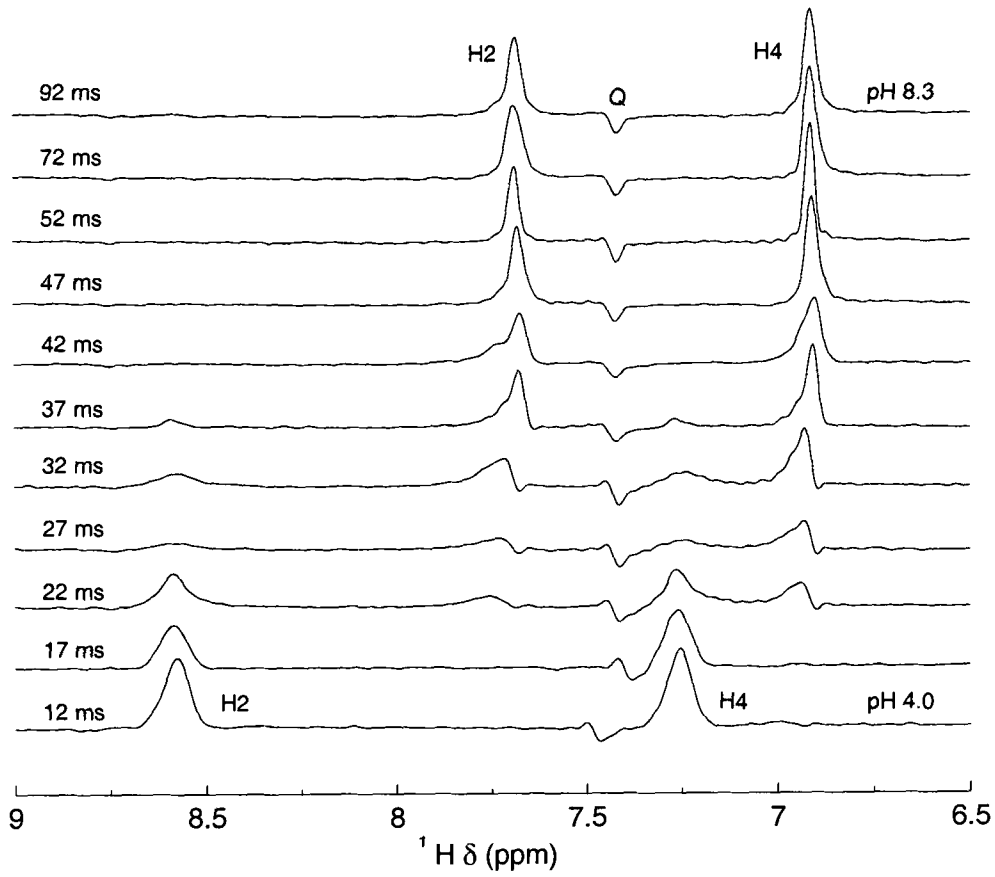
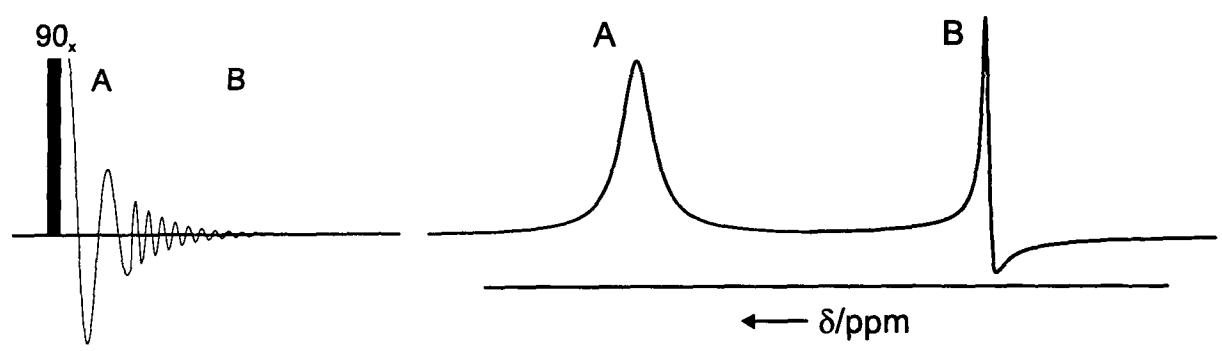


Figure 4.5: Histidine pH jump injection experiments for different post injection delays. The spectra were recored at 500 MHz, using 256 complex points and processed using forward and backward linear prediction and zero filled to 4096 complex points. A cosine-squared window function was applied prior to Fourier transformation. Q indicates the position of a quadrature image of the N-acetyl peak from the aliphatic region. As each spectrum is the result of a single scan it is not possible to apply the CYCLOPS phase cycle scheme to remove this image.



(a) The FID (b) The NMR spectrum

Figure 4.6: The schematic effect of a change in Larmor frequency during the acquisition of a free induction decay (exaggerated).

schematic NMR spectrum for a single spin-$\frac{1}{2}$ system which undergoes a change in Larmor frequency during the acquisition of the FID (Figure 4.6(a)). As can be seen from this, the NMR line from the initial component, A, is broadened, while the line arising from B has a significantly different phase due to not being present at the start of acquisition of the FID.

The pH jump experiment described above was designed to produce a complete change in the ionisation state of the histidine side chain. It is possible to monitor the pH equilibration reaction by performing a jump into the region of the side chain pK_a, in this case, around pH 6.5. This experiment was performed as for the complete ionisation jump described above, except that the final pH was chosen to be 6.5, and that no buffer was used. Using this protocol, a series of NMR spectra were recorded as a function of the post injection delay. These results are shown in Figure 4.7. This shows that the single peak for the H2 proton at pH 4 broadens and shifts as the pH in the sample increases (T = 17 to 92 ms spectra). This peak then splits into a number of overlapping peaks, resulting from a pH distribution throughout the sample volume. This then returns to a sharp peak at approximately 8.38 ppm, corresponding to the pH 6.5 state of the system.

The interpretation of these features is that a highly collimated jet of injected solution travels down the centre of the NMR tube, then after striking the flat bottom of the tube returns turbulently up the sides of the tube, producing complete mixing across the sample. This interpretation is corroborated by the video measurements.

4.1.5 NMR Imaging

The use of NMR imaging techniques to determine the spacial distribution of material is a powerful technique, widely applied in medical diagnosis [130]. In order to determine how well this injection set-up mixes the injectant and buffer solutions a series of 1-dimensional images were recored, both on steady state samples and in

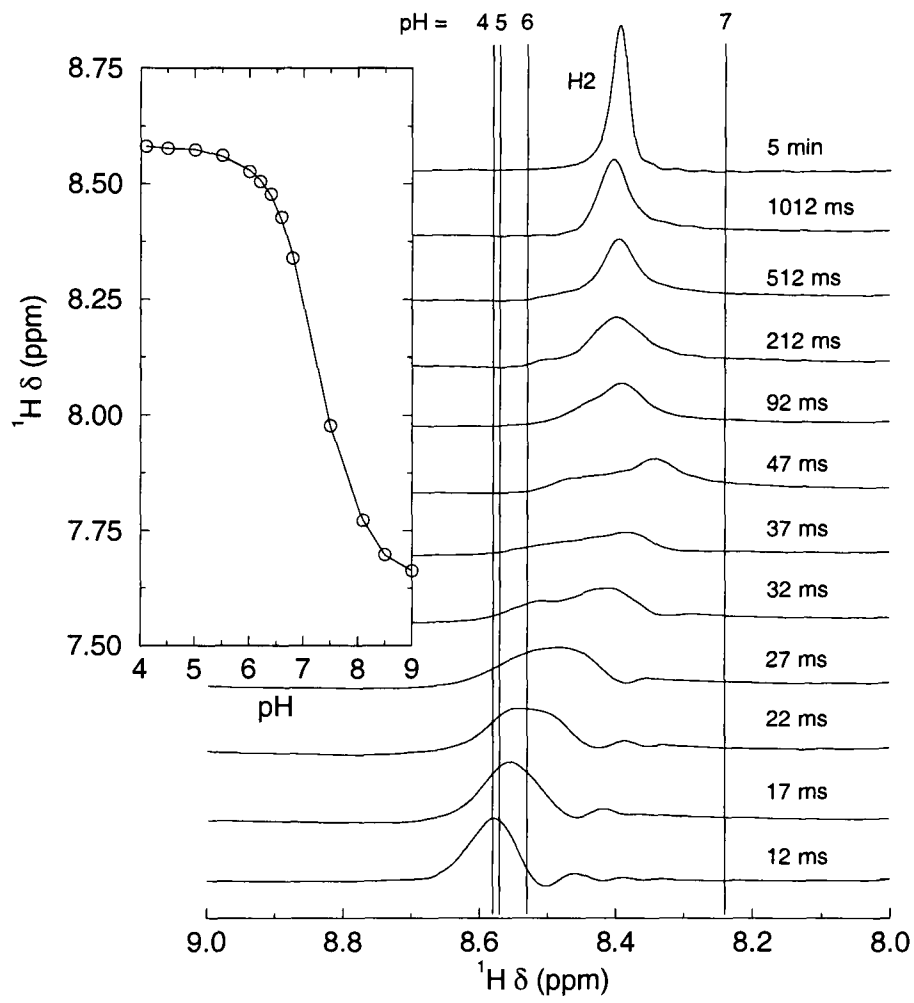


Figure 4.7: Results of a pH jump into the pK_a region of the histidine side chain. The resonance shown is that for the H2 proton. The inset shows the chemical shift titration curve for the same proton. All spectra were processed in the same manner as in Figure 4.5, except that the cosine-squared window function was replaced with 10 Hz exponential line broadening.

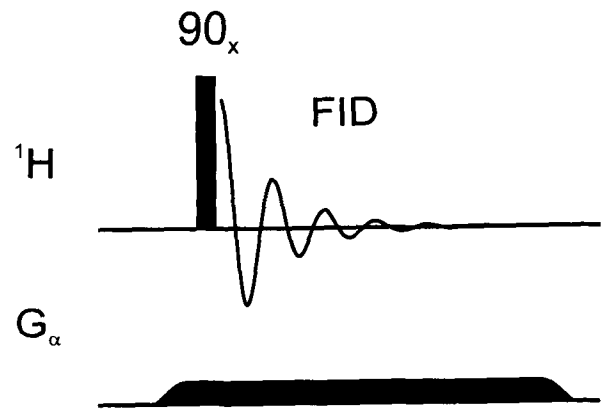
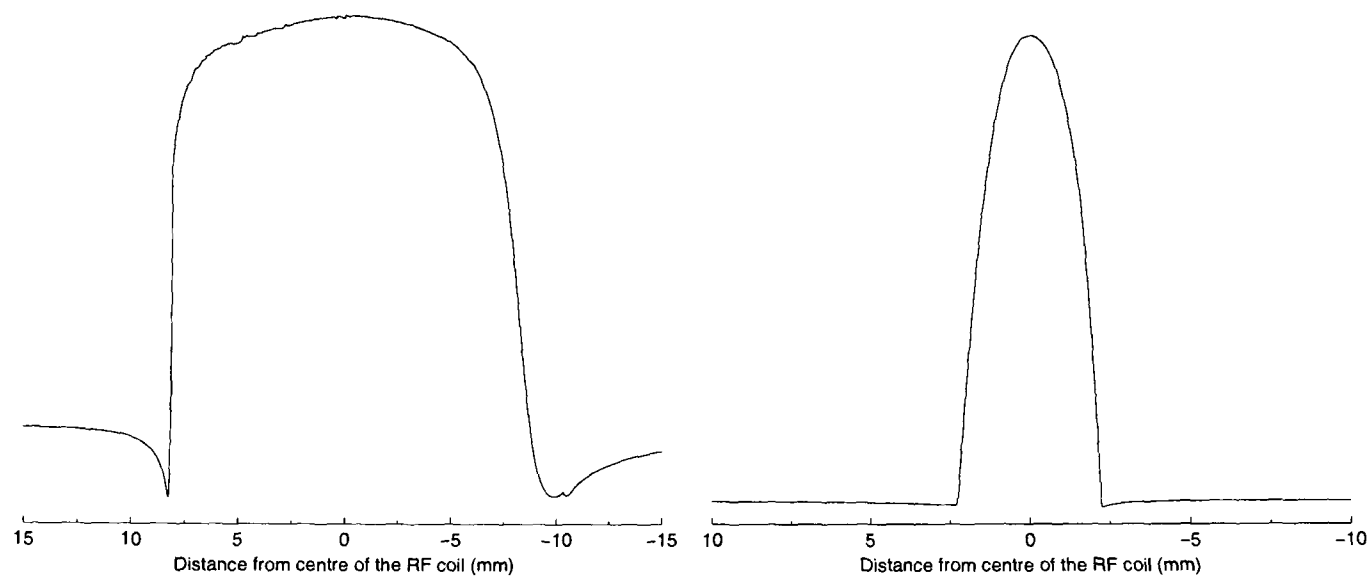


Figure 4.8: The pulse sequence for the imaging experiments. G_α corresponds to a weak field gradient applied along one of the Cartesian axes, in the laboratory frame.

real-time, as a function of time post injection.

The general approach to 1-dimensional imaging is to record the free induction decay in the presence of a field gradient. This causes the Larmor precession frequency of the spins in the sample to become a function of gradient strength and hence the position within the NMR tube. Thus the frequency recorded by the spectrometer encodes the position of the spins within the sample [130, 131]. The pulse sequence for these experiments is shown in Figure 4.8. Figure 4.9 shows images of a standard Shigemi susceptibility tube filled with a 10% H_2O / 90% D_2O mixture taken along the z- and x-axes. The right hand side of the z image is the top of the NMR tube. Images taken along the y-axis were not recorded because of the cylindrical symmetry of the system. Since, as discussed above, the Larmor frequency encodes the position within the NMR tube, the horizontal axis can be rescaled to give the distance from the centre of the RF coil.

To demonstrate the mechanical effort required to mix two solutions within the confines of an NMR tube a number of images were recorded on a layered sample. The sample was constructed as follows: approximately 230 μL of D_2O was placed in the bottom of a Shigemi tube and enough H_2O was gently layered on top so as to fill the RF coil. The tube was then carefully placed in the magnet and imaged. The tube was removed and given a gentle lateral shake and subsequently re-imaged. Finally



(a) z Image (b) x Image

Figure 4.9: 500 MHz ^{1}H NMR images taken axially and radially along a Shigemi susceptibility tube. The right hand side of the z image corresponds to the top of the NMR tube. Spectra were recorded with a flip angle of 1.54° and a spectral width of 28571.43 Hz.

the tube was inverted twice and placed back in the magnet. Figure 4.10 shows the results of these three images, all taken along the z-axis. This figure clearly shows that in order to have complete, homogeneous mixing, vigorous motion is needed, such as that provided by an injection event.

In order to gain a better understanding of the injection event, a series of 1-dimensional images were recorded in real-time during an injection event. This involved the injection of 50 μL of H_2O into 280 μL of D_2O. This results in a 6.5-fold dilution of the H_2O. Images were recored along both z and x directions, using the same parameters as for the static images. Two dummy acquisitions were recorded before the triggering of the injection device to allow the spectrometer electronics to stabilise. These have been removed from the spectra shown in Figure 4.11. From this it is clear that there is an inherent dead-time in the apparatus. The injection event took place before the first spectrum shown, however, there is no signal until the third spectrum, indicating that there is a 20 ms dead-time after the spectrom-

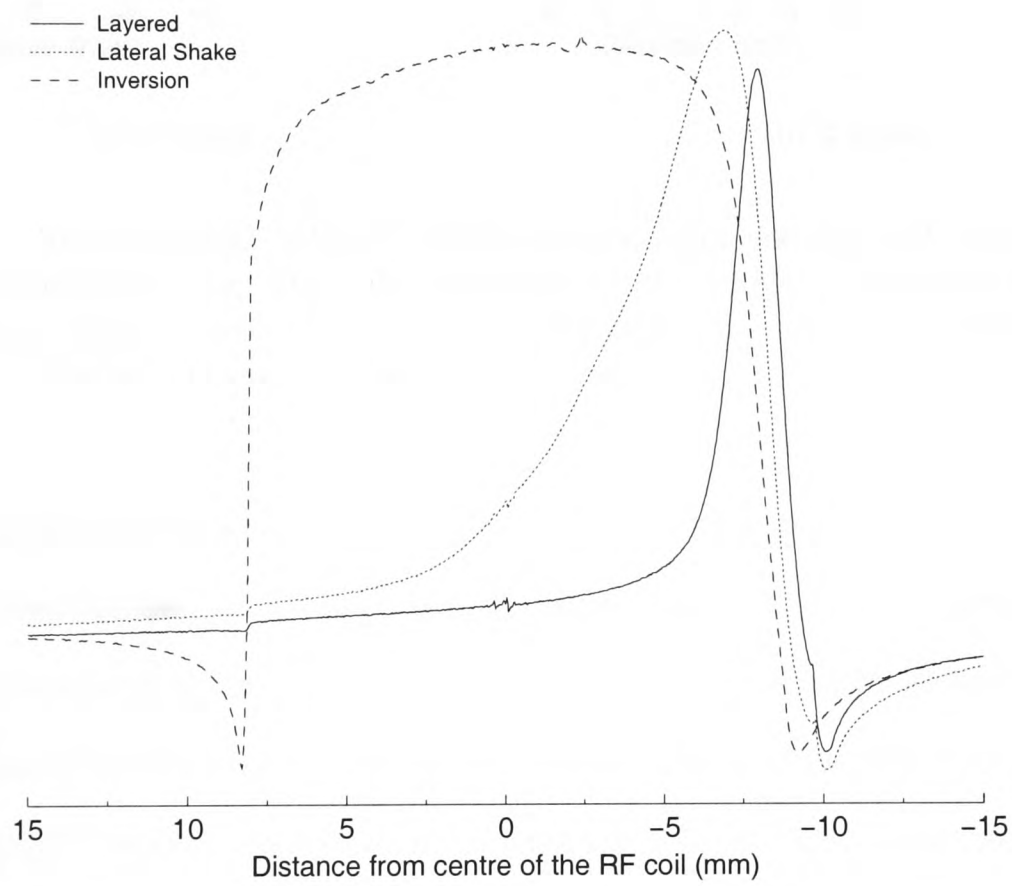
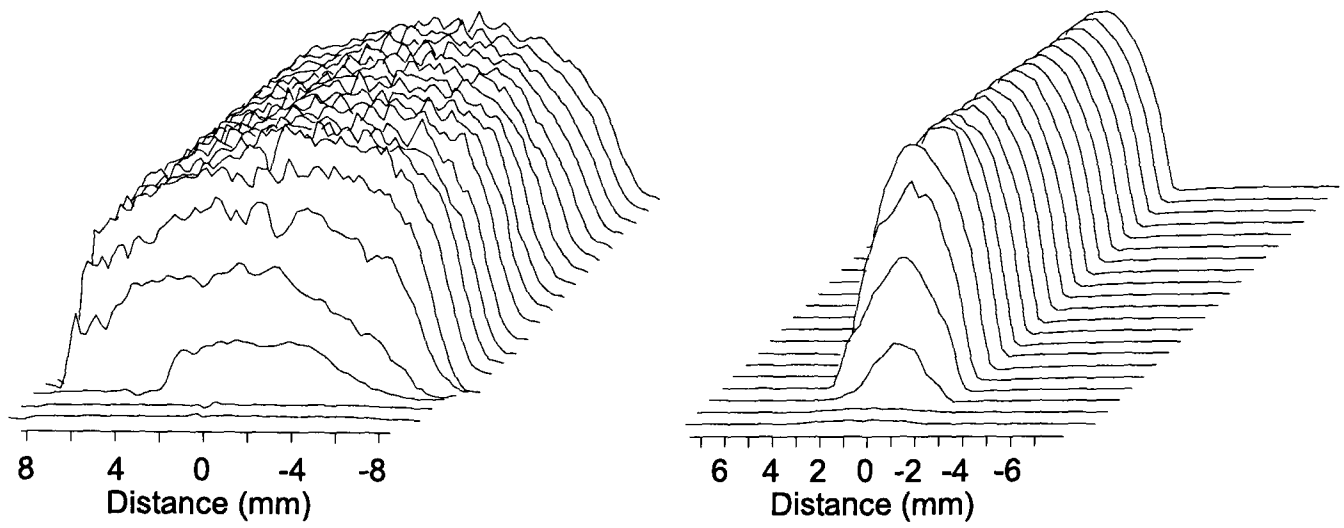


Figure 4.10: 500 MHz ^{1}H NMR images recorded using a layered sample. H_2O was placed on top of 230 μL of D_2O and imaged. Two methods of mixing are then applied to the sample, and the tube re-imaged. The right hand side of the image corresponds to the top of the NMR tube.



(a) z Image (b) x Image

Figure 4.11: Real-time 500 MHz ^{1}H NMR images taken axially and radially along a Shigemi susceptibility tube. The right hand side of the z image corresponds to the top of the NMR tube. Spectra were recorded with a flip angle of 1.54° and a spectral width of 28571.43 Hz. The spectra are recorded at 10 ms intervals.

eter triggers the injector, but before the piston is activated. Once the injector is activated, then complete mixing of the H_2O is complete within the next four scans, giving rise to a mixing time of the order of 40 ms. This is in good agreement with that determined by the other methods described above. The video footage in Figure 4.3 suggests that the injection event consists of a stream of injectant travelling down the centre of the NMR tube before striking the base of the tube. The real-time images taken along the x-axis indicate that this is the case. Measuring the width of each spectrum at half-maximum height serves as a guide to the extent of the radial mixing. The width at half-maximum height was used since, although the experiment was performed by injecting into D_2O solution, there is a small residual proton signal, which marks the edge of the NMR tube in each of the images. These widths are shown in Figure 4.12. This shows that there is indeed a narrow column of liquid travelling down the centre of the NMR tube, which spreads out radially as mixing occurs within the RF coil region. This is shown schematically in Figure 4.13.

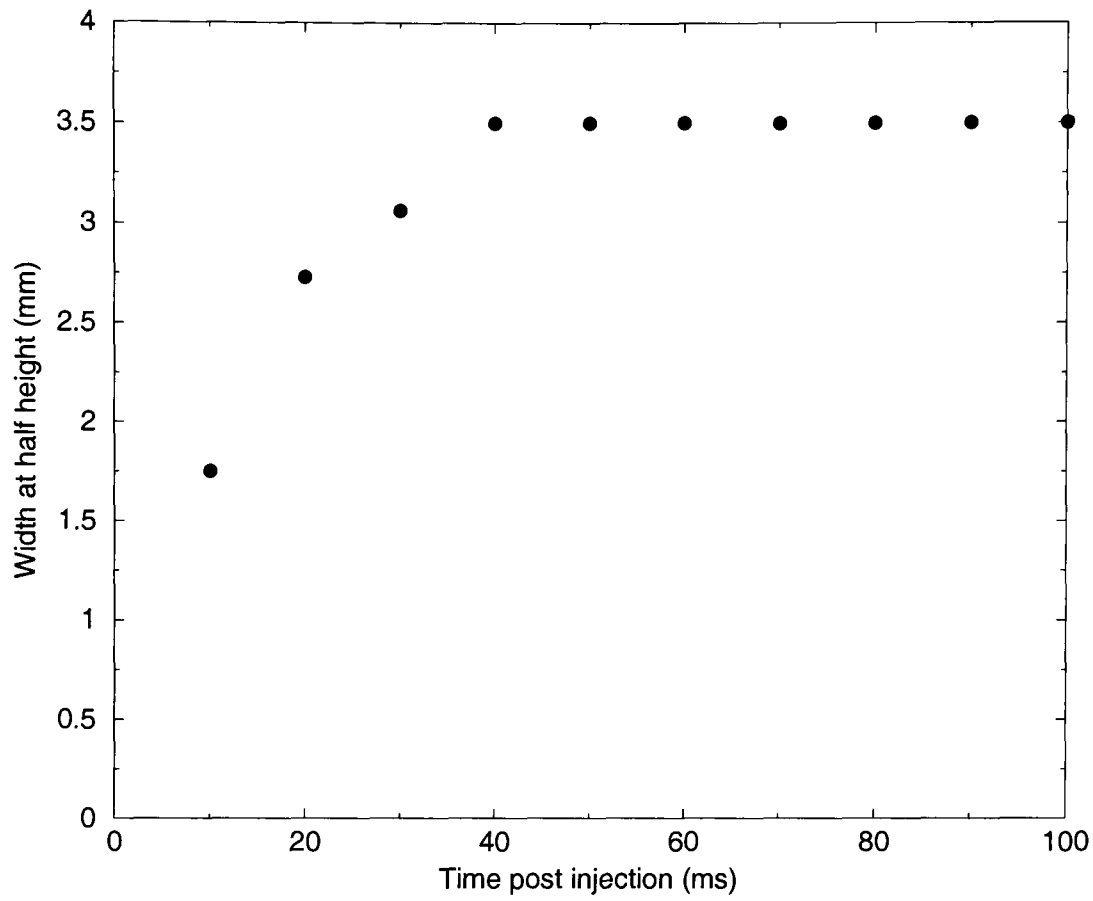


Figure 4.12: Full widths at half-maximum height for the x images in Figure 4.11(b).

To test the hypothesis that the flat bottom of the Shigemi tube aids the mixing, the real-time imaging experiments were repeated using a standard round-bottomed NMR tube. The procedure used was the same as above, however, this time injecting 50 μL of D_2O into 450 μL H_2O, as a larger sample volume is required to fill the NMR tube. The results from this series of experiments are shown in Figure 4.14. As can clearly be seen in the z image (Figure 4.14(a)) there is a significant non-uniform axial distribution of the injectant material in the NMR tube after injection. This supports the theory that the flat base of the Shigemi tube helps the mixing process. However, a point to note is that in the case of the round-bottomed tube the base of the tube is 6 mm below the RF coil, whereas it is only only 1 mm for the Shigemi tube. Therefore using the standard tube, the injectant stream must travel further in order to strike the bottom, lengthening the time before any turbulent mixing can

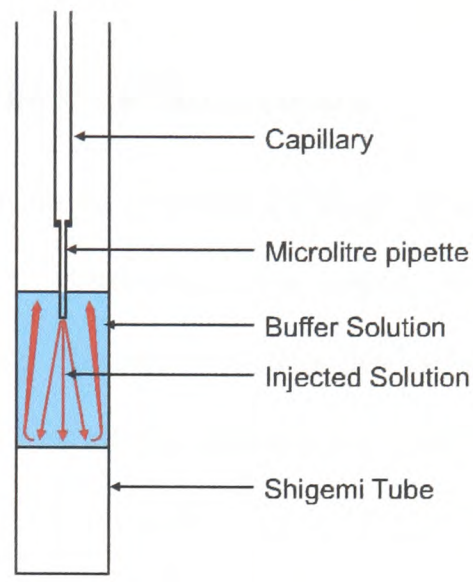
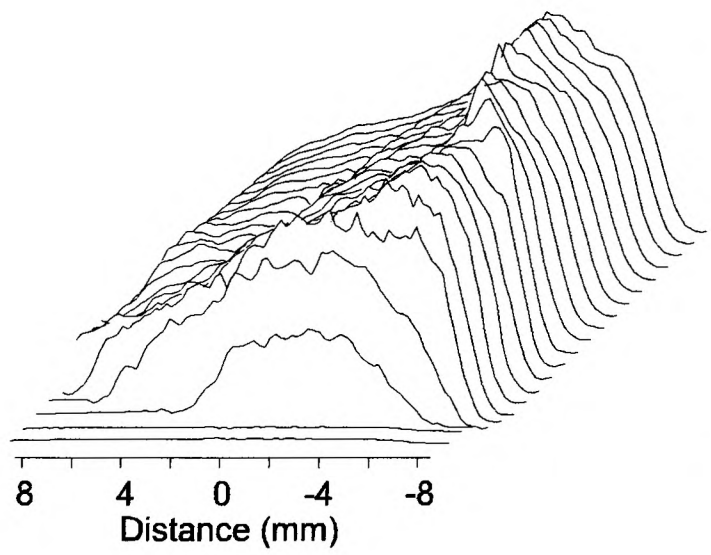
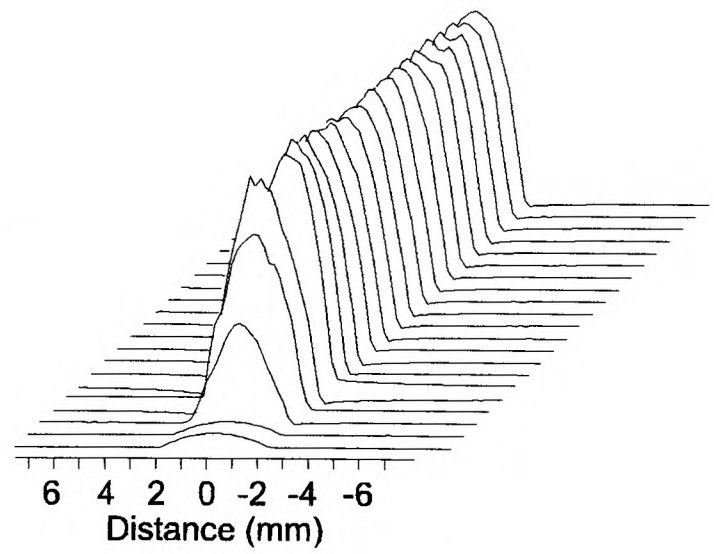


Figure 4.13: A schematic drawing of the injection event, based on evidence from the video footage, histidine pH jumps and NMR imaging experiments. The jet of liquid strikes the flat bottom of the Shigemi tube and then travels up the sides, mixing the two solutions.



(a) z Image



(b) x Image

Figure 4.14: Real-time 500 MHz ^{1}H NMR images taken axially and radially along a standard round-bottomed NMR tube. The right hand side of the z image corresponds to the top of the NMR tube. Spectra were recorded with a flip angle of 1.54° and a spectral width of 28571.43 Hz. The spectra are recorded at 10 ms intervals.

occur, also contributing to the reduced the mixing efficiency.

4.1.6 Injection of Viscous Solutions

The chemical denaturants used in protein folding experiments are usually more viscous than water. Hence to determine how much of an effect the viscosity of the solution has on the mixing efficiency of this device a further experiments were performed. Firstly, the viscosities of the denaturing solutions had to be determined and a suitable model system developed. The viscosities were measured using an Ostwald Viscometer [132], with pure water being used as the calibration standard. Using this method the viscosity of the solution is given by the following equation:

$$\eta = A\rho t \tag{4.2}$$

where η is the dynamic viscosity, A is a constant specific for the viscometer, and dependent on the temperature of measurement, and ρ is the density of the solution, determined by measuring the mass of 1 cm^3 of solution. t is the time taken for the solution meniscus to fall under gravity between two marks on the viscometer. As stated above, water was used as the calibration standard, giving the value of the constant: $A = 2.09 \times 10^{-8}$ m^2s^{-2}.

Viscosities were then measured for 8 M urea solution, 8 M guanidine hydrochloride solution, 45% (v/v) 2,2,2-trifluoroethanol (TFE), all used as chemical denaturants, and 1.5 M calcium chloride solution, the model system chosen. Each viscosity was averaged over ten measurements. The measured data are shown in Table 4.1.

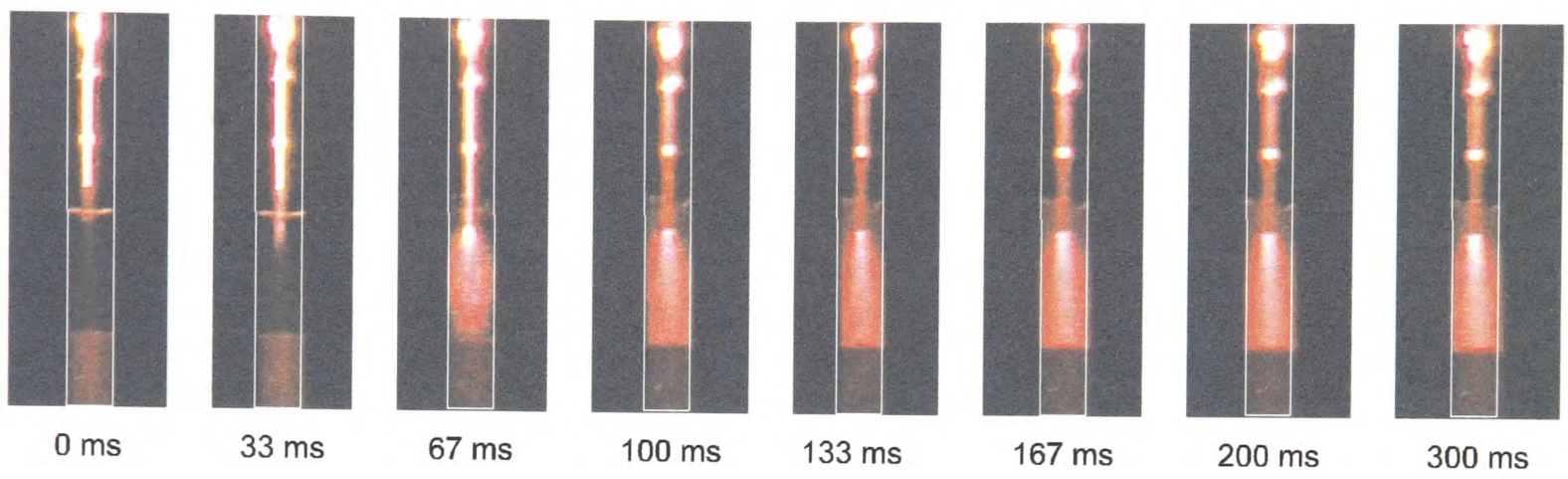
Figure 4.15 shows video footage, recorded in the same manner as that in Section 4.1.3, but injecting two different viscous solutions, 8 M guanidine hydrochloride (Figure 4.15(a)) and 45% trifluoroethanol (Figure 4.15(b)). Comparing these images with those in Figure 4.3 it is clear that a greater time is required to achieve

Compound	ρ (g cm^{-3})	η (μPa s)
H_2O	0.9849	891
8 M Urea	1.0914	1187 ± 7
8 M GdnHCl	1.1283	1241 ± 7
45% TFE (v/v)	1.1370	1429 ± 14
1.5 M $CaCl_2$	1.1235	1236 ± 8

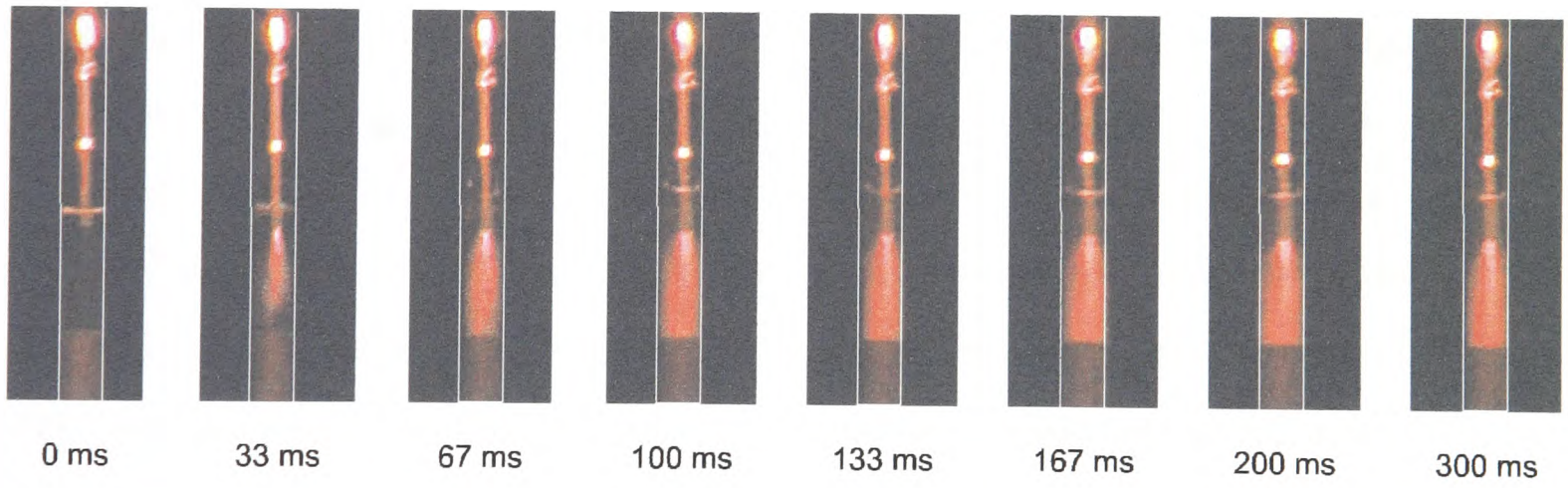
Table 4.1: Densities and dynamic viscosities for various solutions at 298 K. The viscosity of water was obtained from [133].

the mixing. In this case, the macroscopic mixing is apparently complete within the first four frames, that is, within 100 ms.

The use of urea, guanidine or TFE in the imaging experiments is complicated by the fact that this introduces a second resonance into the spectrum. When the field gradient is applied and these resonances are spread out, reflecting the spacial encoding of the nuclear Larmor frequency, these signals will overlap, complicating the interpretation of the imaging experiments. Therefore, a solution was chosen which does not provide an additional resonance in the NMR spectrum, while still increasing the solution viscosity. Figure 4.16 shows the result of a series of NMR images taken along the z-axis following the injection of 1.5 M calcium chloride solution to mimic the effect of the chemical denaturant used in protein folding experiments. Comparing this with the injection of H_2O shown in Figure 4.11(a) it is clear that the mixing takes longer. While at the top of the tube there is little change beyond 40 ms post injection, a much longer period of time, of the order of 100 ms is needed for mixing to be complete along the complete length of the tube. Images along the x- and y-axes were not recorded as mixing has been shown to be fast in these directions, coupled with the fact that radially, mixing takes place over a much smaller distance, so will be complete well before the axial mixing.



(a) 8 M GdnHCl



(b) 45% TFE (v/v)

Figure 4.15: Video Footage of the injection of two more viscous solutions into water, recorded using methylene blue florescence as before. Less than 0.5 W laser illumination was used. The injection event takes place between the first two frames in each case. The images appear red due to the presence of the argon ion filter.

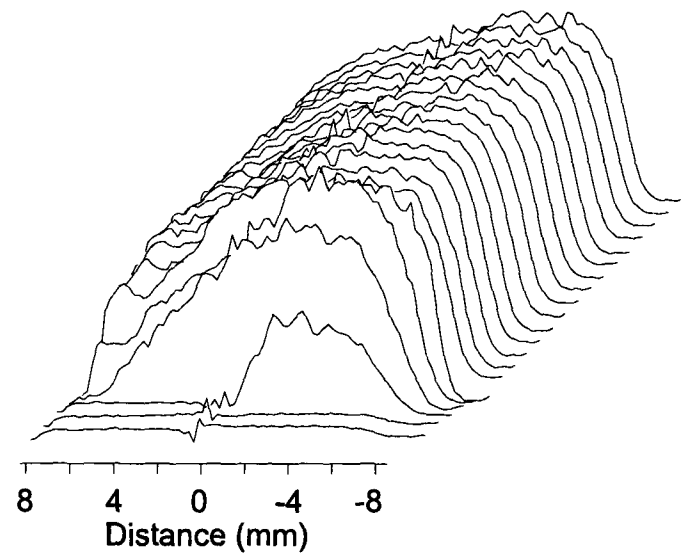


Figure 4.16: Real-time ^{1}H NMR images taken axially (along the z-axis) following the injection of a more viscous solution. The right hand side of the image corresponds to the top of the NMR tube. Spectra were recorded with a flip angle of 1.54° and a spectral width of 28571.43 Hz. The spectra are recorded at 10 ms intervals.

4.1.7 Summary

The experiments described in the preceding section provide strong evidence that this injection set-up is capable of producing complete, homogeneous mixing on a timescale of the order of 50 ms for non-viscous injectants, with a short experimental dead-time. The mixing times determined by the histidine pH jump and NMR imaging experiments agree to within a few percent. These experiments, along with the video footage of the injection event, also show that macroscopic as well as microscopic mixing of the injected solution is achieved within the NMR coil region.

4.2 CIDNP Pulse Labelling

Under certain conditions some proteins can be induced to form so-called molten globule states. These states are characterised by compact hydrophobic cores, large amounts of secondary structure, but with few specific tertiary interactions [134]. These traits are often monitored by the use of techniques such as circular dichroism (CD) or fluorescence spectroscopy. These techniques give information about the

whole structure, but provide little information on the atomic scale. NMR measurements of molten globules often given very little information also. This is because the molten globule is an ensemble of rapidly interconverting conformers, which give rise to an NMR spectrum with poor chemical shift dispersion due to a large number of overlapping resonances and the effects of millisecond conformational exchange.

Molten globules are of interest as species very similar to these are often detected as transient intermediates along the protein folding pathway. Examples include the various α-lactalbumins [62, 135, 136] and retinol binding protein [137]. CIDNP pulse labelling is a technique which enables the study of the surface structure of these molten globules.

4.2.1 Theory

CIDNP pulse labelling is based on "RF Pulse labelling" in which NOEs[10] are generated by radiofrequency irradiation of a small portion of the aromatic region of the NMR spectrum of an unfolded protein, this is then rapidly refolded and any enhancements in the aliphatic region of the spectrum due to NOE transfers are detected [138].

In CIDNP pulse labelling similar principles are used [31, 32]. Nuclear polarisation is generated in an unfolded state of the protein using the photochemical radical pair reaction between FMN and tyrosine, tryptophan and histidine residues exposed on the protein surface. This polarisation is then rapidly transferred to the native state for detection. This transfer is achieved by either a fast dilution from a high concentration of chemical denaturant, or pH jump using the injection device described above. Detection of the polarisation in the native state allows the good chemical shift dispersion and resolution to be obtained, while having the polarisation intensities from the partially folded state. This is shown schematically in

[10]Nuclear Overhauser Effects

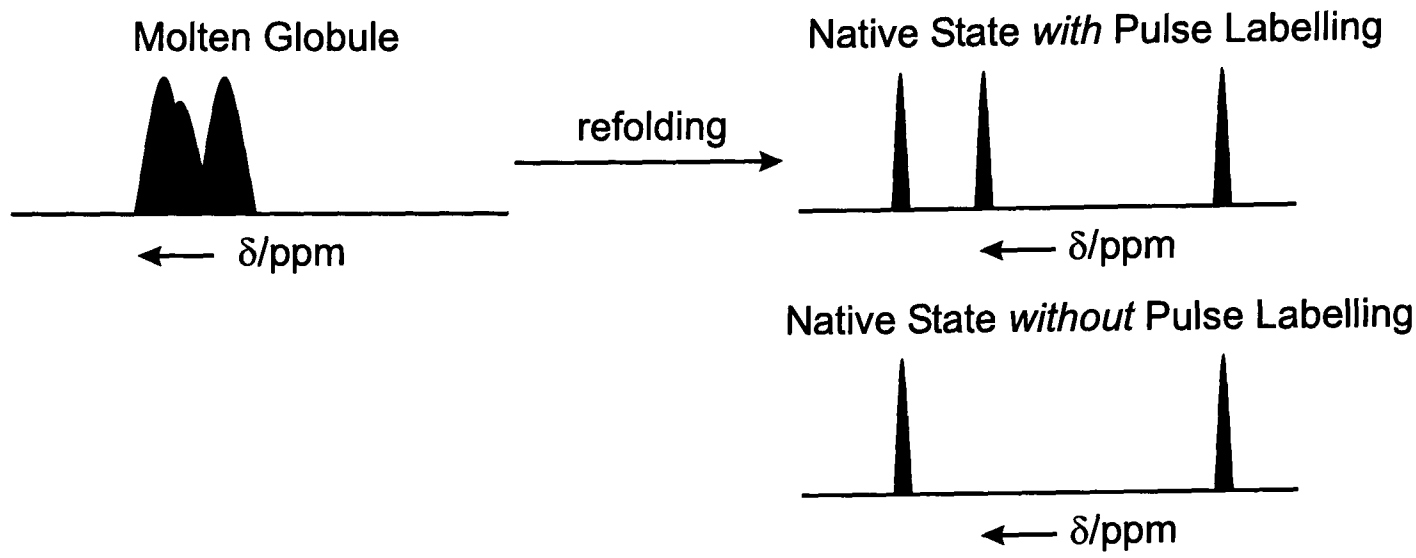


Figure 4.17: The theory of the CIDNP pulse labelling experiment. Resonances which overlap in the molten-globule state are spread out by the greater chemical shift dispersion in the native state. The corresponding native state spectrum recorded without pulse labelling is shown for comparison.

Figure 4.17. The native state spectrum, however, shows only those resonances which arise from natively accessible residues, whereas the pulse labelled spectrum includes those resonances which are additionally exposed in the molten globule state. The refolding to the native state must happen on a timescale that is fast compared to the spin-lattice relaxation time of the exposed amino acid side chain protons in order that the nuclear polarisation generated in the unfolded state still remains in the native state to be detected.

The CIDNP pulse labelling experiment has been reported [31, 32], however, there are two distinctions between this work and that performed previously. Firstly, Lyon *et al.* worked at a proton frequency of 400 MHz, whereas all the work described here was performed at 600 MHz, enabling greater resolution to be obtained. Secondly, due mainly to the lack of pulsed field gradients on the ailing 400 MHz spectrometer, the previous experiments were performed in pure D_2O solution to minimise the solvent signal. The benefit of using H_2O solution is that signals can be observed from solvent exchangeable protons such as the indole proton of tryptophan.

The pulse sequence used for CIDNP pulse labelling is shown in Figure 4.18. The

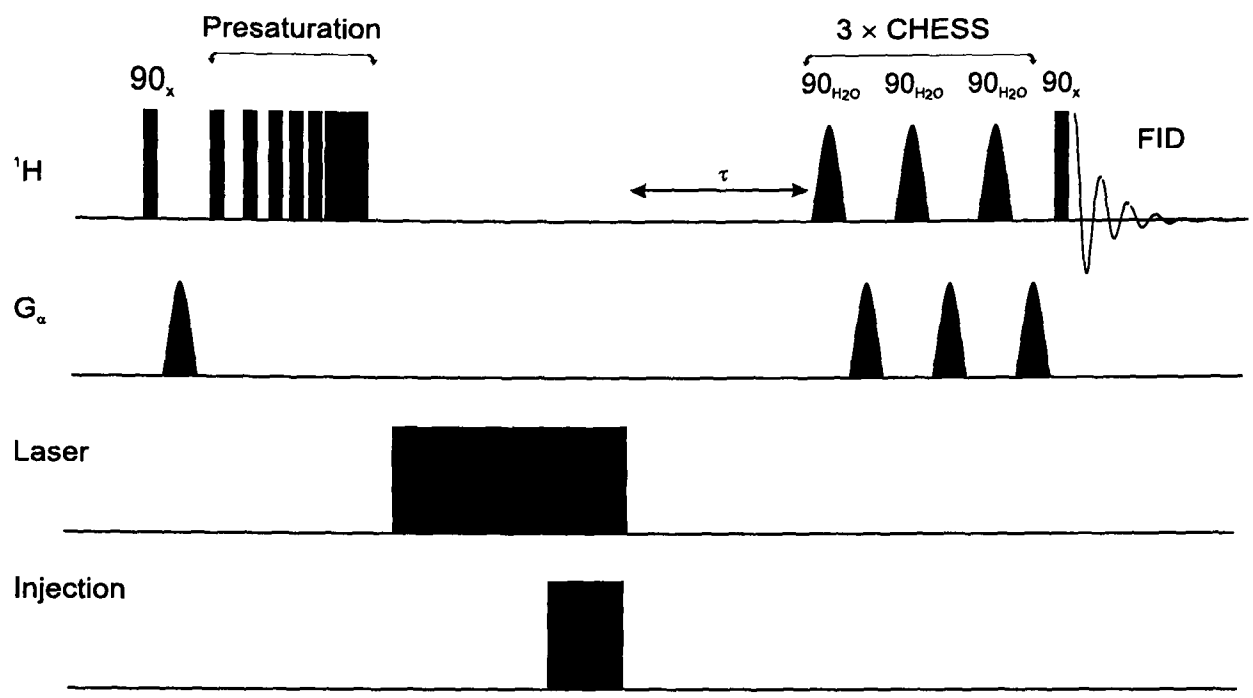


Figure 4.18: The pulse sequence used for CIDNP pulse labelling. τ is the post injection delay. G_α is the field gradient applied as a strong gradient along each of the Cartesian axis (order: z, x, y) or along three orthogonal vectors tilted at the magic angle with respect to the B_0 field direction, as explained above.

sequence consists of a pre-saturation time, which is made up of a 90° pulse followed by a crush gradient and then a train of multiple pre-saturation pulses to destroy any magnetisation in the sample. Since the protein solution it situated outside the RF coil until injection, this pre-saturation period only affects the refolding buffer. Immediately following this is the laser illumination period. This is usually between 50 and 500 ms in duration. The injection event then takes place during the final portion of the illumination. After injection there is a post-injection delay, of the order of 100 ms to allow protein folding to occur and any remaining free radicals from the illumination period to recombine. The acquisition pulse is then preceded by the application of three CHESS pulses, as described in Section 4.1.2. CIDNP pulse labelling spectra are then averaged over a series of light and dark subtraction pairs, usually four or eight, in order to improve the signal to noise ratio.

4.2.2 Hen Egg White Lysozyme

Hen egg white lysozyme is a reasonably small monomeric protein consisting of an amino acid chain of 129 residues (molecular weight 14.6 kDa). The structure consists of two distinct domains separated by the active site. The α-domain consists of four α-helices, two at the C-terminus and two at the N-terminus, along with a small 3_{10} helix. The β-domain contains a three-strand antiparallel β-sheet, another 3_{10} helix and a large loop region. The protein also contains four disulphide links, between cysteine residues 6–127, 30–115, 64–80 and 76–94. The latter acts as a second tether between the α- and β-domains, in addition to the polypeptide backbone. The disulphide bridges remain fully oxidised in all the pulse labelling experiments described here.

Hen egg white lysozyme was the first enzyme to have its structure determined by x-ray crystallography [139]. It has subsequently been the subject of a large number of studies and its structure has been characterised by NMR [140–143]. Figure 4.19 shows a representation of the solution state NMR structure of lysozyme.

The effect of various co-solvents on the structure of hen egg white lysozyme has been of great interest [145]. These co-solvents include dimethylsulphoxide (DMSO) [146–148] and 2,2,2-trifluoroethanol (TFE) [149–151]. Most notably, it has been observed that lysozyme forms a molten-globule like state in TFE concentrations above 20% (v/v) [149], the so-called TFE state of lysozyme. This state has large amounts of secondary structure, with an increase in helical content relative to the native state, but few specific tertiary interactions. Figure 4.20 shows far- and near-UV circular dichroism (CD) spectra of lysozyme in the presence of differing amounts of TFE. From the far-UV CD spectra shown in Figure 4.20(a) it is clear that there is secondary structure in all three solvent mixtures, with little difference being observed between pure water and 7% TFE (v/v). There is a noticable increase in mean residue ellipticity at about 220 nm shown in the spectrum recorded in 45%

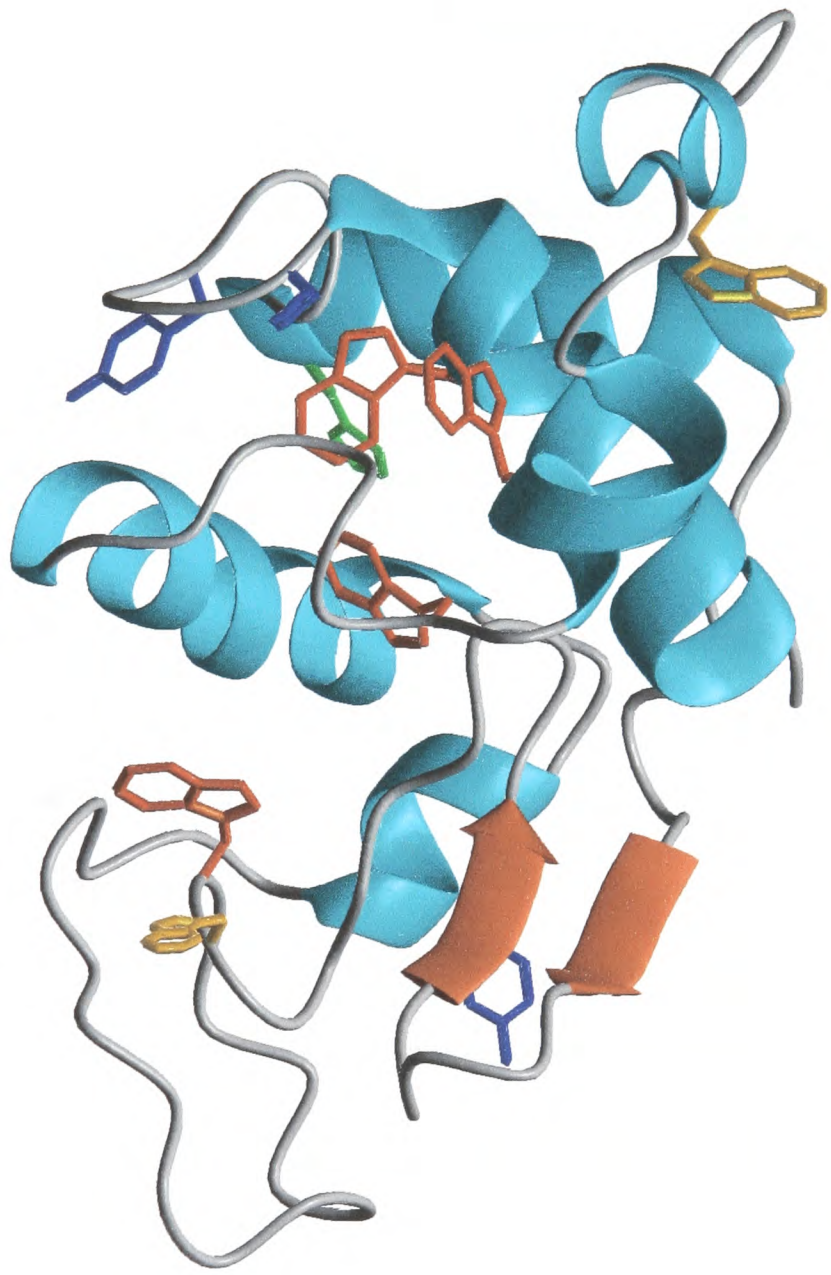
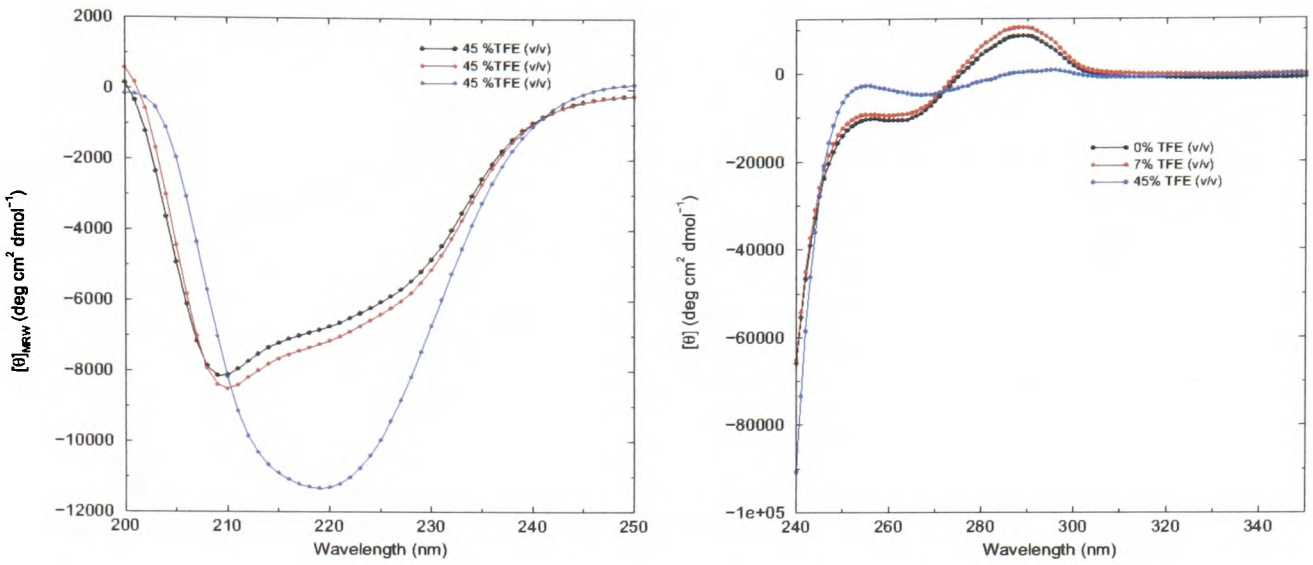


Figure 4.19: The structure of hen egg white lysozyme (PDB file: 1E8L [144]). Solvent accessible tryptophan residues are shown in orange, buried tryptophans in red. Tyrosine and histidine residues are shown in blue and green respectively. The figure was drawn using MOLMOL [89].

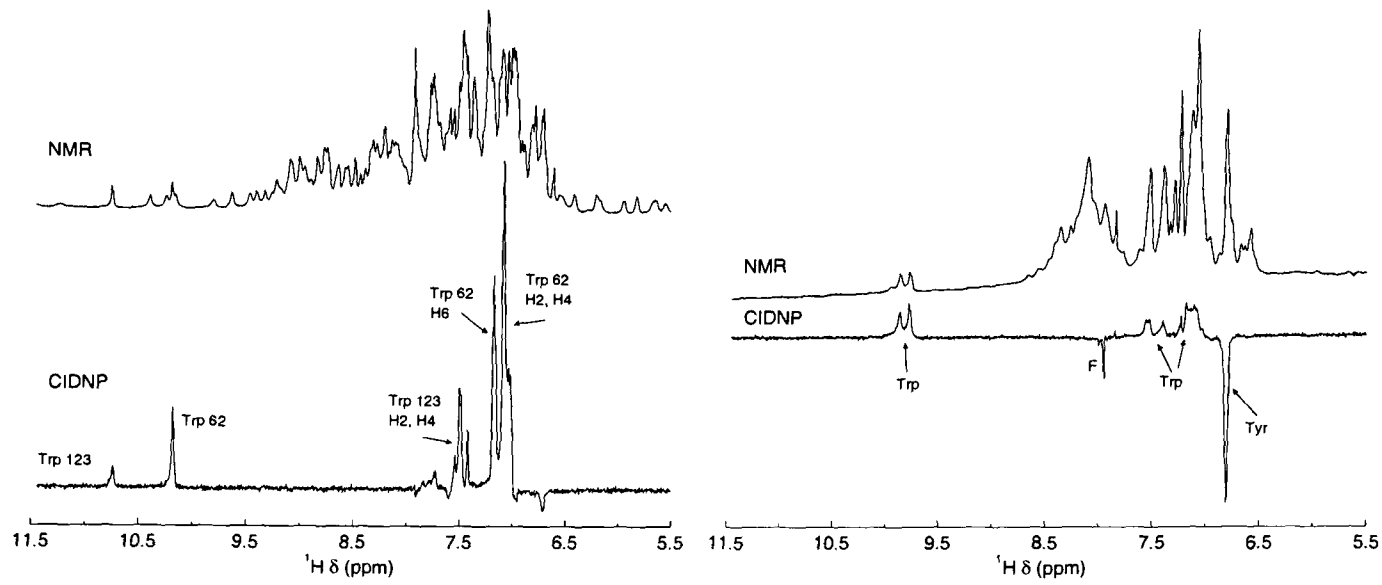


(a) Far-UV CD (b) Near-UV CD

Figure 4.20: Circular Dichroism spectra of 69 μM lysozyme at varying concentrations of TFE. The spectra were recorded at room temperature.

TFE (v/v), indicating an increase in the amount of secondary structure present in the protein [84]. Comparing these results with those obtained under identical conditions using near-UV CD to probe the environments of the aromatic amino acid residues (in this case tyrosine, tryptophan and phenylalanine), reveals that there is a significant change in the tertiary structure in protein. The near-UV CD spectra are shown in Figure 4.20(b), and again show little difference between pure water and 7% TFE (v/v). However, there are clear changes at 255 and 290 nm in the 45% TFE (v/v) spectrum, indicating a change in the tertiary structure under these conditions [84]. It is also noted by Lu *et al.* that the presence of a small amount of TFE can accelerate the refolding of lysozyme to the native state [151].

So as to have clear start and end points to the pulse labelling experiment steady-state spectra of lysozyme in 7% and 45% TFE (v/v) were recorded. These are shown in Figure 4.21. Clearly there are major structural changes occurring in the protein on increasing the TFE concentration. In 7% TFE (v/v), which closely resembles the native state, only two of the tryptophan residues are solvent accessible, namely



(a) 7% TFE (v/v) (b) 45% TFE (v/v)

Figure 4.21: 600 MHz ^{1}H NMR (128 scans) and CIDNP (32 scans) spectra of 2 mM lysozyme in H_2O. Polarisation was generated using 0.2 mM FMN dye and 4 W laser illumination for 100 ms. 50 mM sodium acetate-d_3 buffer was used at pH 7. F denotes a polarised flavin peak.

Trp 62 and Trp 123. These are highlighted by the strong enhancements in the indole region of the CIDNP spectra (Figure 4.21(a)) and corresponding enhancements in the aromatic region.

When the concentration of TFE is increased beyond a certain level, then large-scale changes in the protein structure occur, as is evident in Figure 4.21(b), both in the NMR and CIDNP spectra. Here, there is a large emissive signal at 6.9 ppm which originates from exposed tyrosine residues. Two broad tryptophan indole signals are also noticeable above 10 ppm. The signal at 8.0 ppm arises from polarised flavin molecules.

Since the TFE state of hen egg white lysozyme has a poorly resolved NMR spectrum (Figure 4.21(b)) and upon dilution will rapidly refold to the native state [151], it is an ideal choice for CIDNP pulse labelling. Two experiments were performed to probe the surface structure of the TFE state, both using a 6.5 fold dilution of hen

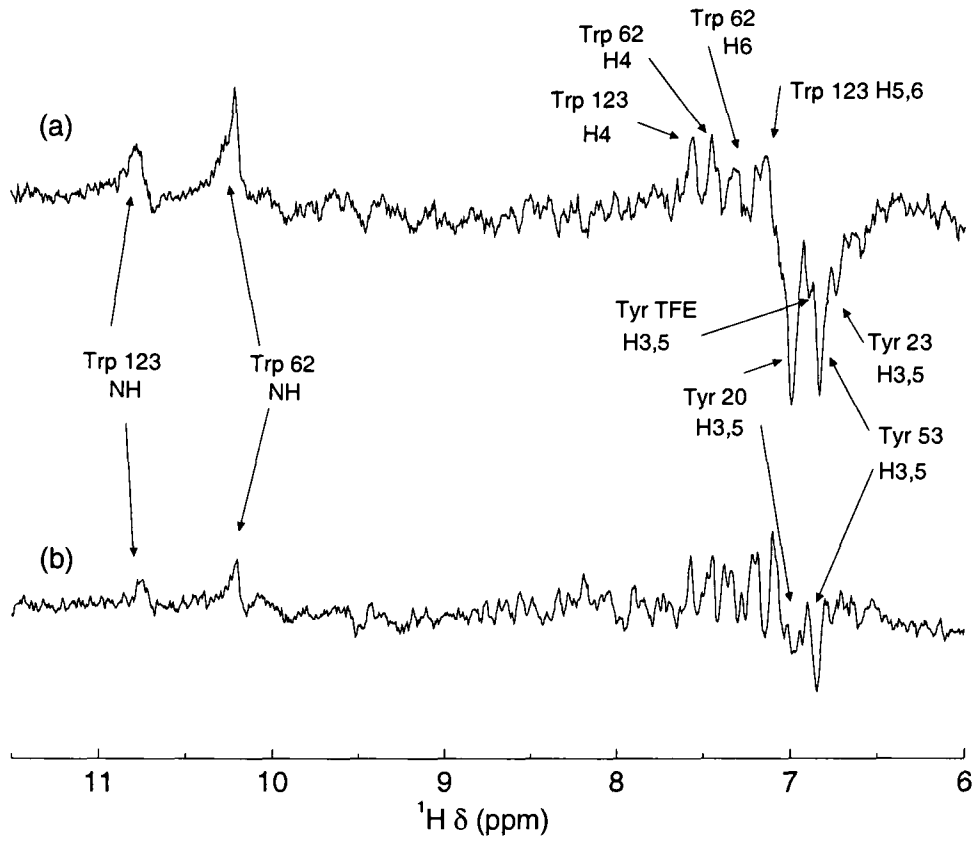


Figure 4.22: 600 MHz CIDNP pulse labelled spectra of hen egg white lysozyme upon refolding from 45% to 7% TFE (v/v) in 50 mM sodium acetate-d_3 at pH 7. Both spectra were recorded using an injection time of 50 ms and a post-injection delay of 100 ms. Spectrum (a) was recorded with a laser pulse width of 500 ms, while spectrum (b) used 100 ms. Assignments are taken from [141].

egg white lysozyme in 45 % TFE (v/v). Both experiments were recorded using an injection time of 50 ms and a post-injection delay of 100 ms. Two different laser pulse widths were used, first 500 ms, then 100 ms to minimise the effects of cross-polarisation during the laser illumination. A subtraction of four light and dark pairs results in the spectra shown in Figure 4.22.

From Figure 4.21(a) it is clear that there are two solvent-accessible tryptophan residues in the native state of hen egg white lysozyme, these are Trp 62 and Trp 123. There are no other "CIDNP-active" residues exposed. The results of these pulse labelling experiments show that there are three solvent accessible tyrosine residues as well as the two tryptophans accessible in the TFE state, remaining exposed in the native state. The spectra are shown in Figure 4.22. This agrees with what is seen in the steady-state CIDNP spectrum of the TFE state (Figure 4.21(b)): that tyrosine

residues are solvent exposed. The pulse labelled spectra provide greater detail by way of the improved chemical shift dispersion allowing the exposed residues to be assigned. The large tyrosine peak at 6.83 ppm is attributed to Tyr 53 with the other large emissive signal at 6.99 ppm corresponding to Tyr 20. Smaller signals at 6.89 and 6.73 ppm arise from a residual TFE state tyrosine resonance and Tyr 23, respectively. The chemical shift assignments are taken from [141].

It is useful to compare the results obtained here with previous CIDNP pulse labelling results obtained by Lyon *et al.* [31, 32]. Table 4.2 shows the residues detected in the two CIDNP pulse labelling experiments. Using the advantage that the

Residue	Figure 4.22 Results	Previous Study	Residue	Figure 4.22 Results	Previous Study
Trp 28	No	Yes	Trp 108	No	Yes
Trp 62	Yes	Yes	Trp 111	No	Yes
Trp 63	No	Yes	Trp 123	Yes	Yes
Tyr 20	Yes	Yes			
Tyr 23	Yes	Yes			
Tyr 53	Yes	Yes			

Table 4.2: Comparison of solvent accessible residues detected in the TFE state using CIDNP pulse labelling. Results from this study and a previous study [31, 32].

number of tryptophan residues exposed in the TFE state can easily be determined by counting the number of peaks in the indole region. It is clear that the absorptive signals in the aromatic region of Figure 4.22, between 7 and 8 ppm, must arise only from two tryptophans, rather than all six as previously thought [31, 32]. The exposure of all three tyrosine residues has been confirmed by these experiments.

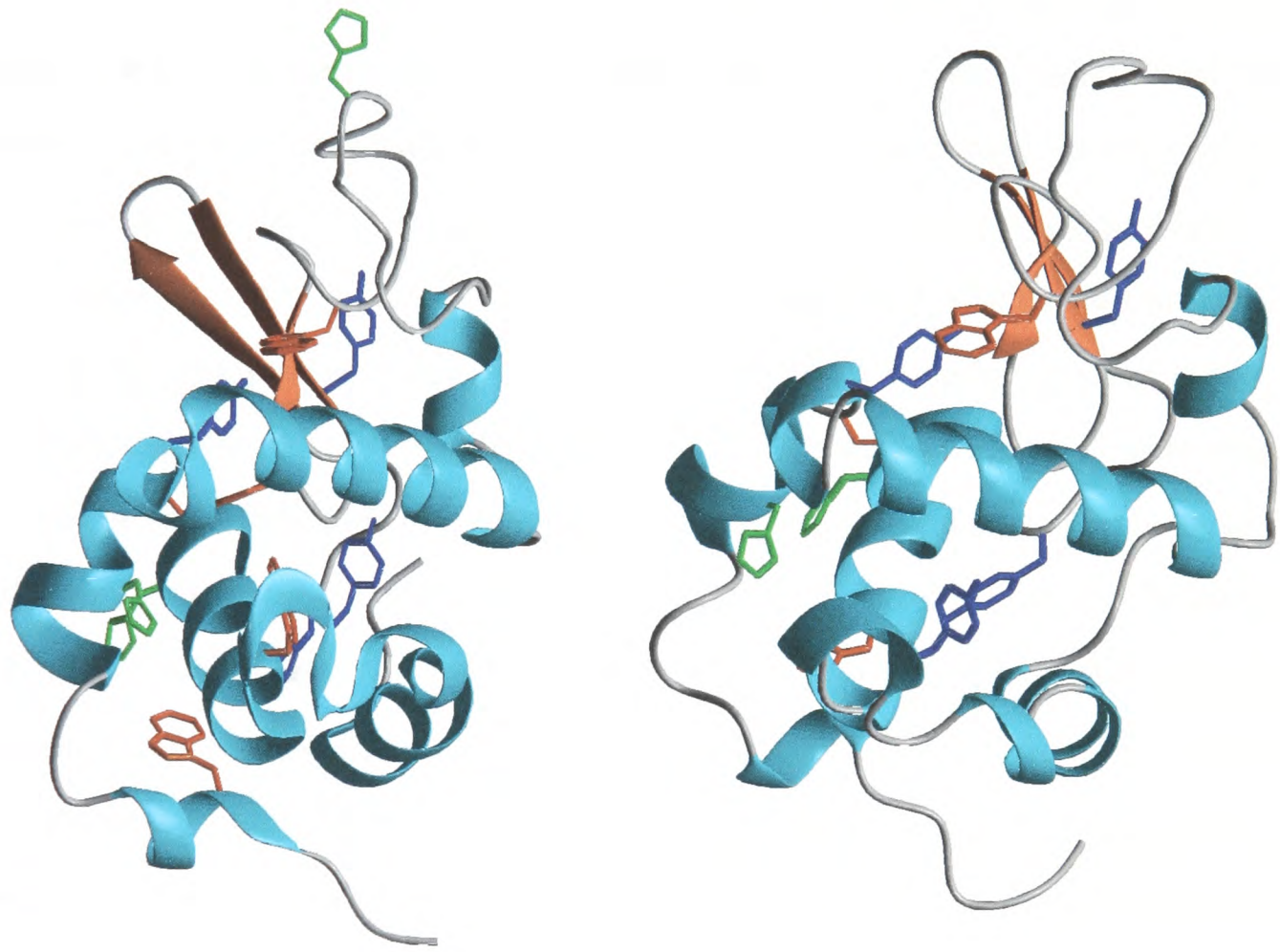
The observation that only tryptophan residues 62 and 123 are exposed in the TFE state of lysozyme indicates that tryptophan residues 28, 108 and 111 which form part of the hydrophobic core of the protein in the native state, are also buried in the TFE state. This shows that some native-like character may persist in this

unfolded state, an idea which has been postulated previously [152], and supports the premise that the formation of secondary structure in transition state regions plays an important rôle in protein folding [153, 154].

4.2.3 Bovine and Human α-Lactalbumins

The α-lactalbumins are small monomeric globular proteins consisting of 123 amino acid residues (molecular weight 14.2 kDa). These proteins are found in mammalian milk and are involved in the lactose biosynthetic pathway [155]. Their structure is somewhat similar to that of hen egg white lysozyme, in that α-lactalbumin contains two domains. The α-domain consisting of four α-helices and two shorter 3_{10} helices; the β-domain consists of a three stranded antiparallel β-sheet, another 3_{10} helix and number of loop regions. The α-lactalbumins contain four disulphide bridges, two in the α-domain, one in the β-domain and one which joins the C-terminus in the α domain to a 3_{10} helix in the β-domain. The cleft between the two domains is the binding site for a single calcium ion, which is coordinated in a distorted pentagonal bipyramid [156]. The structures of the bovine and human α-lactalbumins (BLA and HLA) are shown in Figure 4.23. The bovine and human forms of α-lactalbumin share 75% sequence identity [62], with only two changes in the "CIDNP-active" residues, namely Trp 26 in BLA becomes leucine in HLA, and His 68 becomes glutamine. Many properties of the various α-lactalbumins have been well studied by a variety of techniques [159, 160], including CIDNP [161–163] and their refolding has been well characterised [29, 30, 164, 165].

While α-lactalbumins do not adhere to the classical two-state cooperative folding model which is obeyed by hen egg white lysozyme [166, 167], they are known to form transient folding intermediates which are remarkably similar to the equilibrium molten globules. These can be formed in a number of ways, either in low pH (the A state), a moderate concentration of chemical denaturant (the P state) or in the



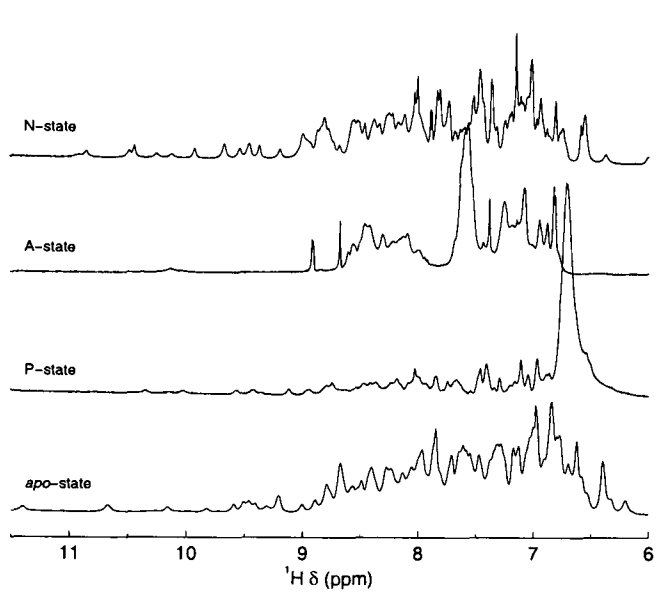
(a) Bovine (b) Human

Figure 4.23: The structures of bovine (PDB file: 1F6S [157]) and human (PDB file: 1HML [158]) α-lactalbumins . The CIDNP-active residues are colour coded as follows; tyrosines in blue, tryptophans in orange and histidines are in green. The figure was drawn using `MOLMOL` [89].

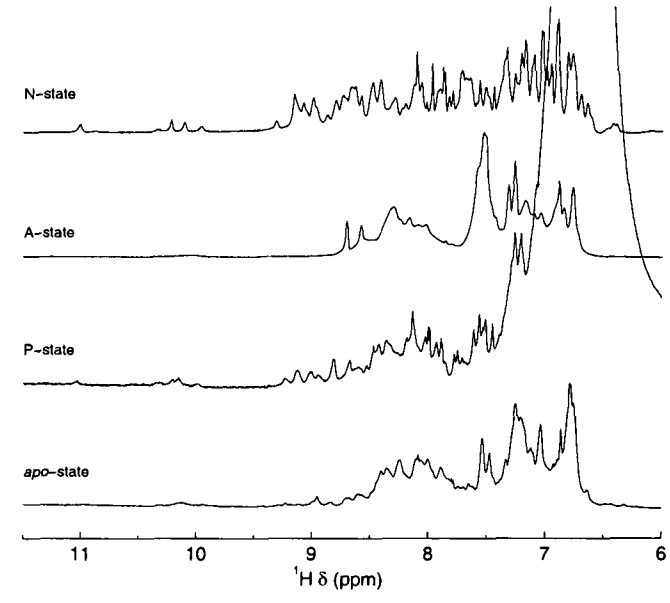
removal of a cofactor (the *apo* state) in this case, calcium ions. Based on results from a number of other biophysical techniques, such as circular dichroism (CD) and fluorescence spectroscopy, the molten globule state is known to have a large amount of secondary structure, but, with few specific tertiary interactions [136].

The question then arises, how similar are the molten globules formed under different conditions? There is, of course, no reason why they should be similar, however, CD or fluorescence spectroscopy the molten globules are virtually indistinguishable. The major drawback of these techniques is that they only provide global information about the protein, *i.e.* they do not give details about the protein on a residue-specific scale. Figure 4.24 shows NMR and CIDNP spectra of the three molten globules in question, along with the native state for comparison. The native state spectrum was recorded at pH 7, with the A state at pH 2. The P state was generated by the addition of 2 M guanidine hydrochloride (responsible for the large, broad signal at around 6.7 ppm in Figures 4.24(a) and 4.24(b)), and the removal of the bound calcium generated the *apo* state. Both the NMR and CIDNP spectra of the molten globules a show distinct reduction in the "useful information" which can be extracted. This arises from the poor chemical shift dispersion, making individual peak assignment impossible. These spectra do show, however, that there are subtle differences in these molten globules. For example, in Figure 4.24(c) there is an emissive signal, characteristic of the H3,5 ring protons of tyrosine, which appears at slightly different chemical shift: at 6.82 ppm in the A state, 6.71 ppm in the P state and 6.87 ppm in the *apo* state. Similar features are also seen in the HLA case (Figure 4.24(d)). However, in this case, there is an emissive signal which appears at a rather similar chemical shift in each of these spectra, that is at 6.78 ppm.

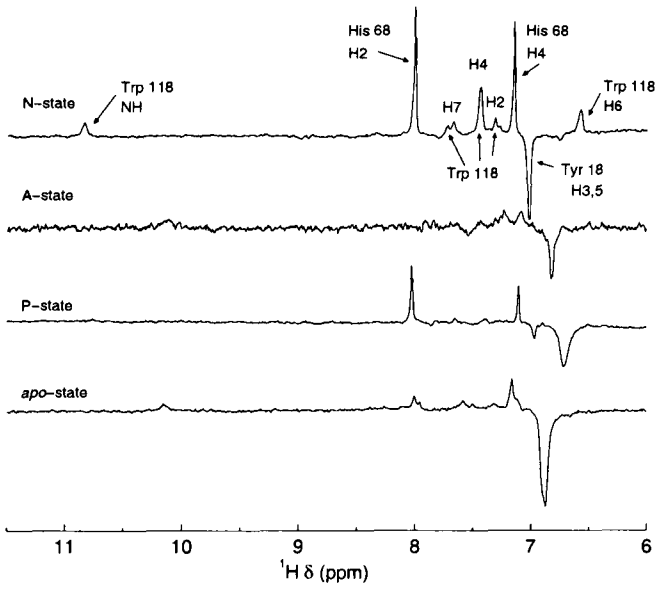
As it is not possible to obtain assignments for these spectra since they suffer from such poor resolution, and attempts to perform heteronuclear measurements on these states also provide little information as few cross peaks are detected [168, 169].



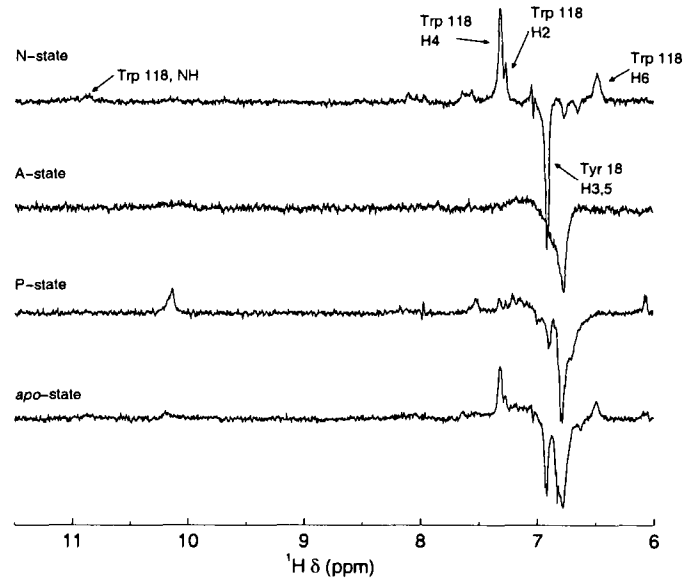
(a) Bovine — NMR



(b) Human — NMR



(c) Bovine — CIDNP



(d) Human — CIDNP

Figure 4.24: NMR and CIDNP spectra of the bovine (2 mM) and human (1 mM) α-lactalbumin molten globules. All spectra were recorded using 4096 complex points with a spectral width of 9009 Hz. NMR spectra were averaged over 64 scans, CIDNP spectra were averaged over 32 light and dark pairs, using 50 ms laser pulses. The native state assignments are taken from [62].

CIDNP pulse labelling is therefore the ideal choice to probe the structure of these partially unfolded states. Presented below are the CIDNP pulse labelling results for these three different molten globules of BLA and HLA.

4.2.3.1 The Bovine α-Lactalbumin Molten Globules

As for the experiments performed using the TFE state of lysozyme, the initial starting point for this set of pulse labelling experiments was the well-studied A state of BLA. The pulse labelling experiment was performed by injecting 50 μL of 1.5 mM protein solution at pH 2 into 280 μL of 55 mM sodium cacodylate buffer at neutral pH. This gave a final protein concentration of 0.23 mM. The spectra were recorded using 4096 complex points over a spectral width of 9009 Hz, and are the result of averaging over eight light and dark pairs, that is, sixteen injections. These results are shown in Figure 4.25. There are more features visible in these spectra compared to the steady-state spectra of the molten globules shown in Figure 4.24(c). Using the known assignments of the native state, it is immediately obvious that three of the tryptophan residues are exposed in the molten globule, shown by the presence of the characteristic indole proton resonances. These belong to residues Trp 60 (9.92 ppm), Trp 104 (11.48 ppm) and Trp 118 (10.75 ppm). The corresponding aromatic resonances from the H2, H4 and H6 protons can also be found in the region 6.0 to 8.5 ppm, and are clearly labelled in Figure 4.25(a). As for the tyrosine residues, again there are clear emissive signals from three exposed residues, namely Tyr 18, Tyr 36 and Tyr 50. In addition to this, there is a residual signal from the molten globule state at 6.81 ppm, and there is also a non-native signal observed at 7.01 ppm, denoted by Tyr NN. This could be related to the non-native signal observed by Wirmer *et al.* at 7.05 ppm [30], following the photochemical release of calcium ions to induce refolding from the *apo* state to the native state. That signal is attributed to the formation of an intermediate species along the folding pathway.

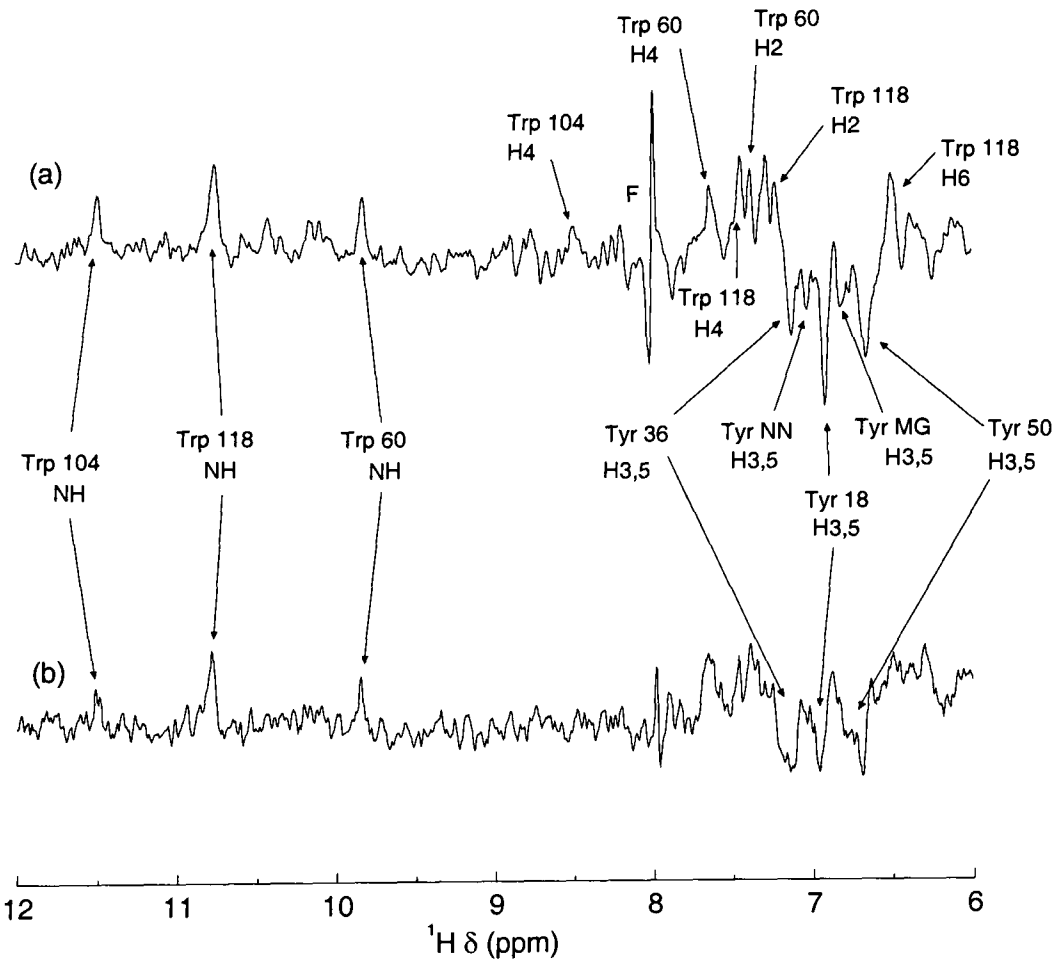


Figure 4.25: 600 MHz ^{1}H CIDNP pulse labelled spectra of bovine α-lactalbumin upon a jump from pH 2 to 7. Both spectra were recorded using an injection time of 50 ms and a post-injection delay of 100 ms. Spectrum (a) was recorded with a laser pulse width of 500 ms, while spectrum (b) used 100 ms. Assignments are taken from [62]. F denotes a polarised flavin signal.

It is possible that an intermediate similar to this is formed during the pulse labelling experiment and remains during the detection period, giving rise to this signal.

Figure 4.25(b) shows results of the same experiment, but performed using a short laser pulse in order to minimise any cross-polarisation [64]. There is, however, a problem associated with this: the signal-to-noise ratio is reduced as there is less polarisation produced during the shorter laser pulse. This mirrors observations seen in the case of lysozyme (Figure 4.22). While it is still possible to attribute different resonances to specific residues, either by using the known assignments, or by comparison with the 500 ms laser pulse experiment, this can only be done with lower confidence. For these reasons, the pulse labelling experiments performed on the other two molten globules used the longer, 500 ms, laser pulse.

As for the TFE state of lysozyme, it is informative to compare these results with those obtained previously at 400 MHz [31, 32]. Table 4.3 shows the comparison of detected residues in this work and that of Lyon *et al.* Once again, performing

Residue	Figure 4.25 Results	Previous Study	Residue	Figure 4.25 Results	Previous Study
Trp 26	No	No	Tyr 18	Yes	Yes
Trp 60	Yes	No	Tyr 36	Yes	Yes
Trp 104	Yes	No	Tyr 50	Yes	Yes
Trp 118	Yes	Yes	Tyr 103	No	No

Table 4.3: Comparison of solvent accessible residues detected in the A state of BLA using CIDNP Pulse labelling between this work and previous experiments [31, 32].

these experiments in H_2O allows the indole protons the tryptophan residues to be observed, permitting easy identification of these exposed residues. Three tryptophan residues can now be seen to be exposed in the A state molten globule: Trp 60, Trp 104 and Trp 118 as supposed previously. These new results also confirm the exposure of three of the tyrosine side chains, namely, Tyr 18, Tyr 36 and Tyr 50.

Figure 4.26 shows the results of the CIDNP pulse labelling experiment performed

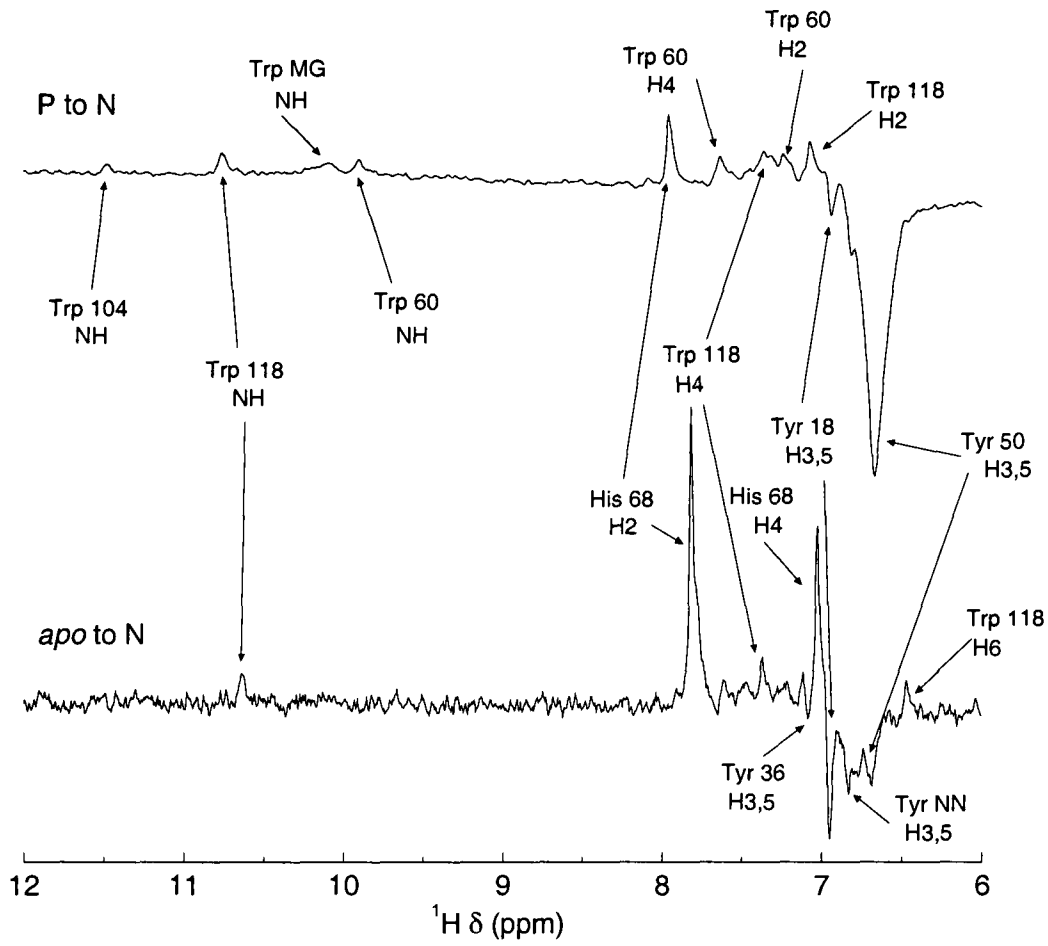


Figure 4.26: 600 MHz ^{1}H CIDNP pulse labelled spectra of bovine α-lactalbumin following jumps from the P state and *apo* state to the native state. Both spectra were recorded using an injection time of 50 ms, a laser pulse width of 500 ms and a post-injection delay of 100 ms. Assignments are taken from [62].

on the P and *apo* states. The P-to-N jump was performed by the dilution of 2.0 M guanidine hydrochloride to 0.31 M in which the protein is native-like, while the *apo* to N jump was achieved by injection of *apo* protein into 10 mM calcium chloride solution. The spectra obtained are quite different to that for the A-to-N jump, and to each other, immediately indicating that the structures of the three molten globules are different. The P state is most similar to the A state, in that the same three tryptophan residues are exposed, however, only Tyr 18 and Tyr 50 are exposed. Is is also possible to attribute the signal at 7.95 ppm to the H2 proton of histidine 68. The situation is different in the *apo*-to-native jump. In this case, the only exposed tryptophan residue is Trp 118, which is also exposed in the native state. Three tyrosine residues can be seen, these are Tyr 18, Tyr 36 and Tyr 50. One feature

unique to this experiment is the observation of both imidazole ring protons from histidine 68. Care must be taken, however, in the interpretation of the absence of CIDNP signals due to histidine residues. This cannot be easily attributed to solvent inaccessibility as the intensity of the CIDNP signal is strongly pH dependent, with the signal disappearing at extremes of low or high pH [55] and histidine also suffers from ineffective competition for flavin triplets in the presence of exposed tyrosine and tryptophan residues [61]. Therefore, if histidine polarisation is observed in the presence of exposed tyrosine and tryptophan residues, then it follows that the histidine residue must be *very* exposed. A summary of the bovine α-lactalbumin results is shown in Table 4.4.

4.2.3.2 The Human α-Lactalbumin Molten Globules

Since human α-lactalbumin also forms molten globules under the same conditions as the bovine protein (Figure 4.24), these states are also amenable to study using the CIDNP pulse labelling experiment. Figure 4.27(a) shows the result of the A-to-N pH jump experiment, recorded under the same conditions as for the bovine case in Figure 4.25(a). The first point to note is that refolding to the native state is not complete, there are large contributions from the A state molten globule still present in the spectrum. This leads to complications in the interpretation of the spectrum. It is possible to immediately assign resonances from Trp 60, Tyr 36 and Tyr 18. However, whether Tyr 50 is exposed is uncertain since it should appear at a similar chemical shift to the molten globule tyrosine resonance. There is a small shoulder on the low field side of this peak, at approximately 6.81 ppm however, this could also be the result of inhomogeneity arising from the mixing. In order to confirm the origin of this signal the experiment was repeated using a longer post injection delay 150 ms. This is shown in Figure 4.27(b). Using this longer post-injection delay, the residual molten globule peak is reduced in intensity, allowing the shoulder to

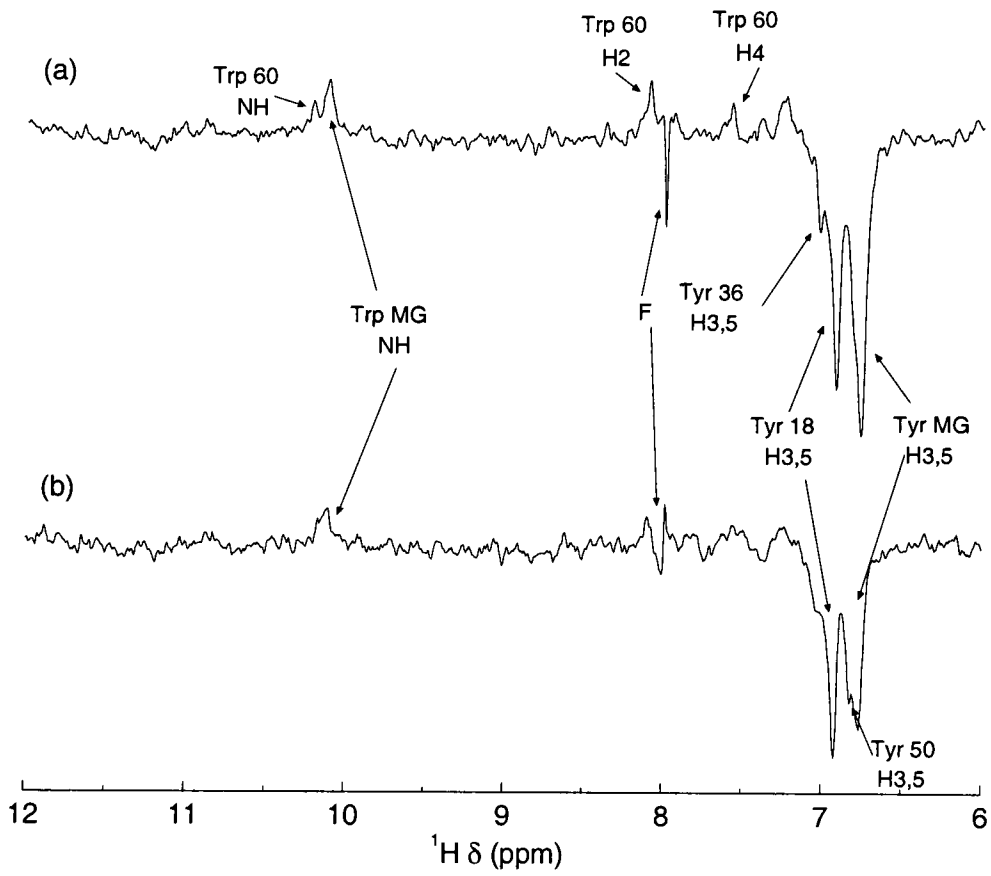


Figure 4.27: 600 MHz ^{1}H CIDNP pulse labelled spectra of human α-lactalbumin upon a jump from pH 2 to 7. The spectrum was recorded using an injection time of 50 ms and a laser pulse width of 500 ms. Spectrum (a) used a post-injection delay of 100 ms, which was increased to 150 ms in (b). Assignments are taken from [62]. F denotes a polarised flavin signal.

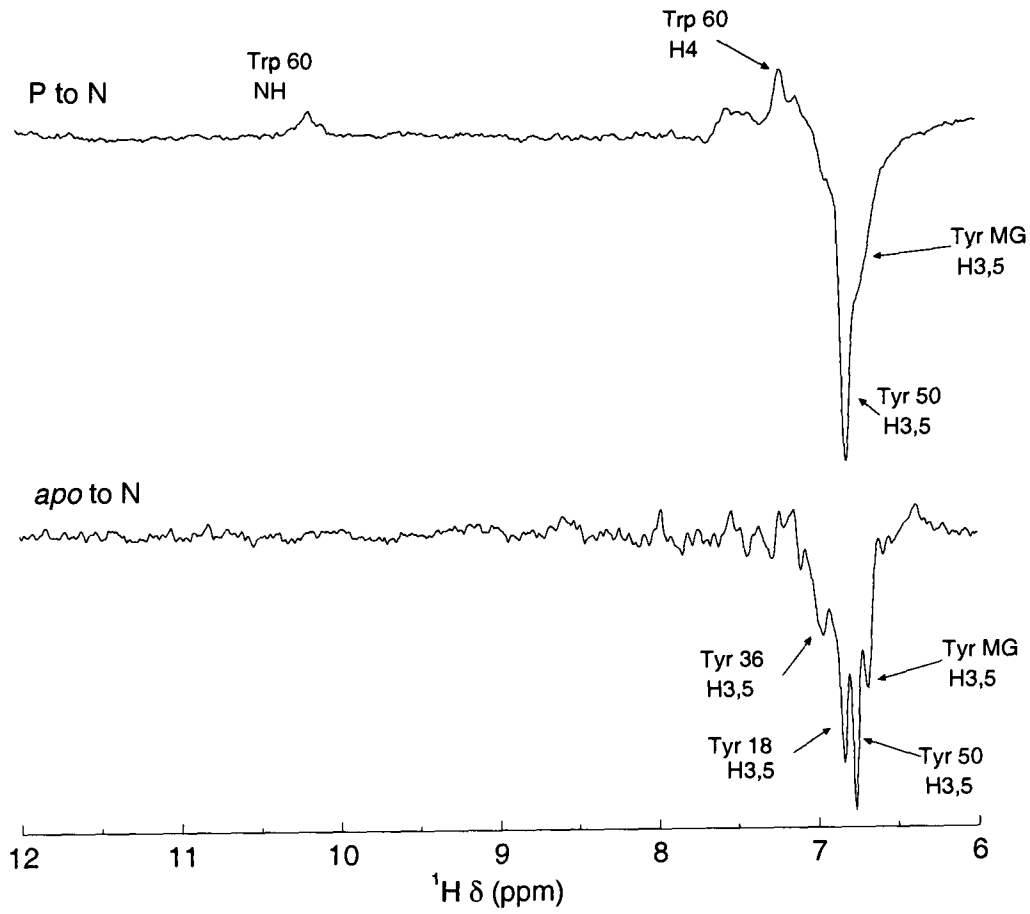


Figure 4.28: 600 MHz ^{1}H CIDNP pulse labelled spectra of human α-lactalbumin following jumps from the P state and *apo* state to the native state. Both spectra were recorded using an injection time of 50 ms, a laser pulse width of 500 ms and a post-injection delay of 100 ms. Assignments are taken from [62].

be clearly assigned to the H3,5 protons of the Tyr 50 residue, indicating that this amino acid side chain is exposed to the FMN dye in the A state molten globule.

Figures 4.28 and shows the result of experiments to study the P state to N state jump and the *apo* to *holo* state jump. These were performed under the same conditions as for the bovine cases in Figure 4.26. These spectra are somewhat less well resolved, especially in the case of the P to N state experiment. This is probably due to the presence of the denaturant affecting the folding of the protein. However, it is clear that there are several exposed residues in both cases. In the P to N state case Trp 60 and Tyr 50 are both accessible, while in the *apo* to *holo* case three out of four tyrosine residues are exposed, with Tyr 103 remaining buried in the molten globules.

As for the bovine case above, these three molten globules of human α-lactalbumin all have very different CIDNP pulse labelling spectra, indicating that the side chain accessibilities and hence the structure of the molten globules is different. The results of these experiments are summarised in Table 4.4.

4.2.3.3 Interpretation

These results, summarised in Table 4.4, are in broad agreement with those obtained by Lyon *et al.* [31, 32]. However, the ability to determine exactly which tryptophan residues are exposed does provide some points for comment. As before, there is a large amount of unfolding occurring in the β-domain of the protein, comprising residues 40 to 81. Exposure of Tyr 36, in the interface between the two domains, is also observed in each case, except in the P state molten globules.

As for the α-domain, there are some subtle differences between this study and previous results. In bovine α-lactalbumin, Trp 104 in BLA is anomalous in that it does not show the same native-like accessibilities that are shown by the other residues in the α-domain. This residue in HLA is not observed. Previous studies have shown that this domain is particularly resistant to unfolding by chemical denaturant [168] and that these amide protons undergo deuterium exchange less readily [170] that those in the other domain, or in the interface. So it is not surprising that these aromatic residues show native-like accessibilities.

These differences between the various equilibrium molten globule states highlight different degrees of hydrophobic cluster formation. In general the β-domain of the α-lactalbumins is unfolded, while the α-domain retains native-like character. The minor differences seen for the different molten globules may correspond to occupation of differing small local minima, within the deep folding pathway on the free energy landscape.

Differences between the molten globule states of the bovine and human variants

Domain	Residue	N state	A state	P state	*apo* state
		Bovine α-lactalbumin			
α	Tyr 18	Yes	Yes	Yes	Yes
α	Trp 26	No	No	No	No
α	His 32	No	n/d	n/d	n/d
α	Tyr 103	No	No	No	No
α	Trp 104	No	Yes	Yes	No
α	His 107	No	n/d	n/d	n/d
α	Trp 118	Yes	Yes	Yes	Yes
Interface	Tyr 36		Yes	No	Yes
β	Tyr 50	No	Yes	Yes	Yes
β	Trp 60	No	Yes	Yes	No
β	His 68	Yes	n/d	Yes	Yes
—	Tyr NN	No	7.10 ppm	n/d	6.85 ppm
		Human α-lactalbumin			
α	Tyr 18	Yes	Yes	No	Yes
α	(Trp 26)		Leucine		
α	His 32	No	n/d	n/d	n/d
α	Tyr 103	No	No	No	No
α	Trp 104	No	No	No	No
α	His 107	No	n/d	n/d	n/d
α	Trp 118	Yes	No	No	No
Interface	Tyr 36	No	Yes	No	Yes
β	Tyr 50	No	Yes	Yes	Yes
β	Trp 60	No	Yes	Yes	No
β	(His 68)		Glutamine		
—	Tyr NN	No	n/d	n/d	n/d

Table 4.4: Summary of the solvent accessibilities of the tryptophan, tyrosine and histidine residues in the A, P and *apo* states of bovine and human α-lactalbumin, based on CIDNP pulse labelling results. Tyr NN represents the non-native tyrosine peak observed in some of the experiments. n/d = not detected.

of the protein indicate that while these proteins maybe share a common evolutionary ancestor, there folding properties are somewhat different.

4.3 Another Application of Rapid Mixing

The use of a rapid injection mixing device has many other applications in addition to CIDNP pulse labelling. The other major use is to follow protein folding in real-time [22, 23, 117]. To demonstrate this, a real time refolding experiment was performed using the mixing device described in Section 4.1. Denatured lysozyme was produced by dissolving a 10 mM solution of lysozyme in 8 M guanidine hydrochloride at pH 7.5, containing 0.2 mM flavin mononucleotide dye, along with a small amount of 1,4-dioxane as a chemical shift reference. 50 μL of this solution was injected into 280 μL of H_2O also containing 0.2 mM FMN. This resulted in final concentrations of lysozyme and guanidine HCl of 1.5 mM and 1.2 M respectively, with a final pH of around 6.5.

A series of real-time CIDNP spectra was recorded using a laser pulse width of 100 ms. Spectra were recorded every 340 ms, using 2048 complex points and a spectral width of 9009 Hz. The data was linear-predicted out to 4096 complex points and a 10 Hz exponential line broadening window function was applied. Water suppression was provided by the use of the CHESS sequence as above [125].

Figure 4.29(a) shows the indole region of the NMR spectrum, averaged over four injections, each with laser illumination. No improvement was found by subtracting the dark spectra. In the native conformation of lysozyme only tryptophans 62 and 123 are solvent accessible. These are clearly seen at 10.18 and 10.75 ppm respectively. The broad signal at 10.04 ppm corresponds to the indole signal from the denatured protein.

In order to extract kinetic information about the folding process of lysozyme,

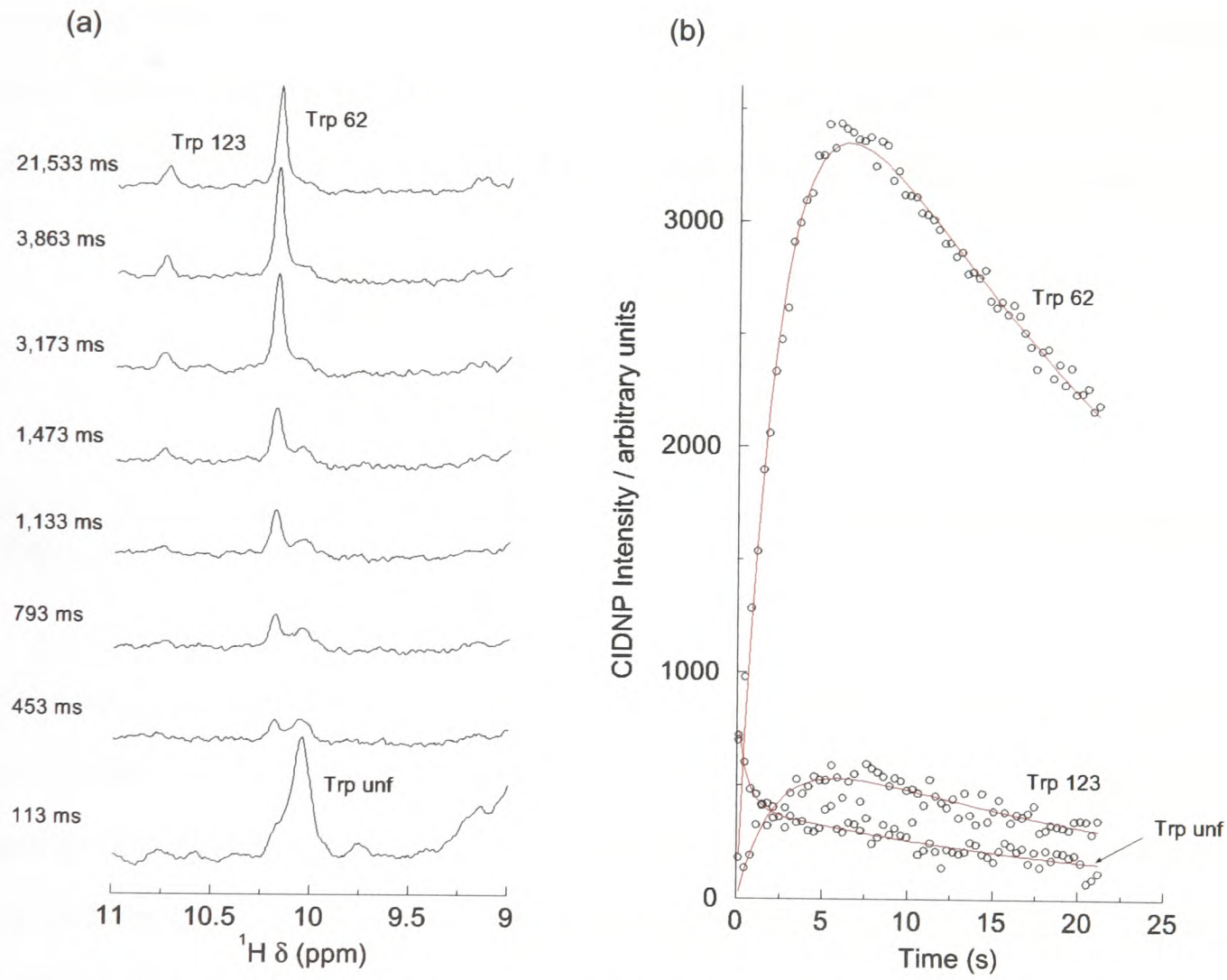


Figure 4.29: Results from the real-time refolding of hen egg white lysozyme from 8 M guanidine HCl. The time indicates the time between the injection event and the start of the acquisition of the FID. The red lines are the result of fitting the data to Equation 4.3. Assignments are taken from [141].

Figure 4.29(b) shows a plot of the peak height as a function of time following the injection event. The experimental intensity points were fitted to a function of the form:

$$I(t) = A\left\{\exp\left(-\frac{t}{T_d}\right) - \exp\left(-\frac{t}{T_r}\right)\right\} \tag{4.3}$$

where $I(t)$ is the intensity at time t after injection, T_r is the time-constant for the rise in intensity, T_d is the time-constant for the decay in signal intensity and A is a scaling factor. Table 4.5 shows the fitted time-constants and scaling factors for the Trp 62, 123 and unfolded tryptophan signals. The differences in the scaling factors arise from

Residue	A	T_r	T_d
Trp 62	4628.3	2.38 s	27.5 s
Trp 123	756.3	2.24 s	23.15 s
Trp unf	407	0.57 s	22.22 s

Table 4.5: Time-constants and scaling factors for the function fitted to the peak intensities in Figure 4.29.

differences in the solvent accessibilities of the two tryptophan residues in the native state of the protein. The time-constant for rise in signal intensity is marginally larger for Trp 62 indicating that the exposure of this residue may occur more slowly than that for Trp 123. This is in agreement with previous studies in which Trp 62 is believed to be involved in hydrophobic clustering in the unfolded state and hydrophobic interactions in the very early stages of the folding process [171]. However, comparing this data with time-constants for the refolding process obtained by fluorescence spectroscopy reveals some disparity [172]. Matagne and Dobson measured the fluorescence of the lysozyme inhibitor 4-methylumbelliferyl-N,N′-diacetyl-β-D-chitobioside (MeU-diNAG) following an eleven-fold dilution from 6 M guanidine HCl solution. From this they concluded that approximately three-quarters of the protein molecules folded with a time constant of 350 ms; the other quarter folding

with a faster time constant of 100 ms. These rates do not agree with those obtained from the real-time NMR experiment, especially for Trp 62 which is situated in the active site of the protein and would be expected to show similar behaviour to the inhibitor. However, there are several explanations for this. Firstly the actual refolding experiments performed are not directly comparable since they result in differing final conditions. In the real-time CIDNP experiment a 6.4 fold dilution from 8 M guanidine was performed. This results in a 2.5 times greater denaturant concentration at the end of the experiment. While the denaturant concentration would not be expected to affect the rate of hydrophobic collapse, which occurs on a time scale not amenable to real-time NMR, is would be expected to have a significant effect on the rate of folding from this compact hydrophobic cluster to the native state [173]. Since this refolding rate is extremely sensitive to the concentration of denaturant the different between the measured rate constants and those expected can be attributed to slight imperfections in the mixing. These imperfections may arise as a result of the increased viscosity of the 8 M guanidine HCl solution compared to that of the buffer. There is also a fast folding phase which has a time constant of the order of 100 milliseconds. It is expected to be complete by the time the data acquisition starts and so it is not detected in this experiment.

The time constants for the decay of the signal intensity are in reasonable agreement with each other and can be attributed to depletion of the photosensitiser, caused by photo-reduction of the flavin to the colourless semiquinone [117]. However, a second argument can also be put forward, based on work by Connolly and Hoch [174]. They attribute the decay in tryptophan signal intensity following a CIDNP experiment to photo-damage of the tryptophan aromatic ring. They observed extra peaks seen in both 1-dimensional NMR and 2-dimensional COSY experiments performed after the CIDNP experiments. These are attributed to photoproducts of reaction between the tryptophan and the photosensitiser, in this case

lumiflavin. Indeed there is evidence in the literature of photochemical oxidation, sensitised by flavins, for each of the CIDNP active amino acids, tyrosine, tryptophan [175] and histidine [176]. Studies have also been performed on the flavin sensitised conformational changes in both hen egg white lysozyme [177] and ribonuclease A [178] produced by photochemical oxidation. In both cases changes in the fluorescence spectrum and protein activity were noted. These have been interpreted as conformational changes in the protein rendering it inactive. Unfortunately the exact nature of the changes were not mentioned.

In order to determine whether these effects are contributing to the signal decay, a series of experiments were performed. Figure 4.30 shows NMR spectra of a tryptophan solution both before and after prolonged laser irradiation. In this case 128 laser pulses were applied, each at 4.9 W for a duration of 100 ms. There is very little difference between the spectra. This is in stark contrast to the data obtained by Connolly and Hoch, in which they saw definite evidence of additional resonances within the aromatic envelope. As a further test two COSY experiments were performed on the system, one before and a second after the application of the laser irradiation. These spectra are overlayed and shown in Figure 4.31. As with the one-dimensional spectra the results of the COSY experiments show that there is very little change occurring after the application of the laser irradiation. The spectrum recorded after laser irradiation overlay perfectly with that taken before exposure, with the exception of the diagonal peak at 8.0 ppm. This peak is attributed to the amide NH proton. It is unclear as to why this signal should be missing in the post irradiation spectrum, however, this could be the result of a number of factors include deuterium exchange.

Based on the results of these experiments it is unlikely that photo-damage to the extent reported by Connolly and Hoch has taken place. Therefore the decay in the signal intensity noted in the real-time CIDNP experiment appears to arise purely

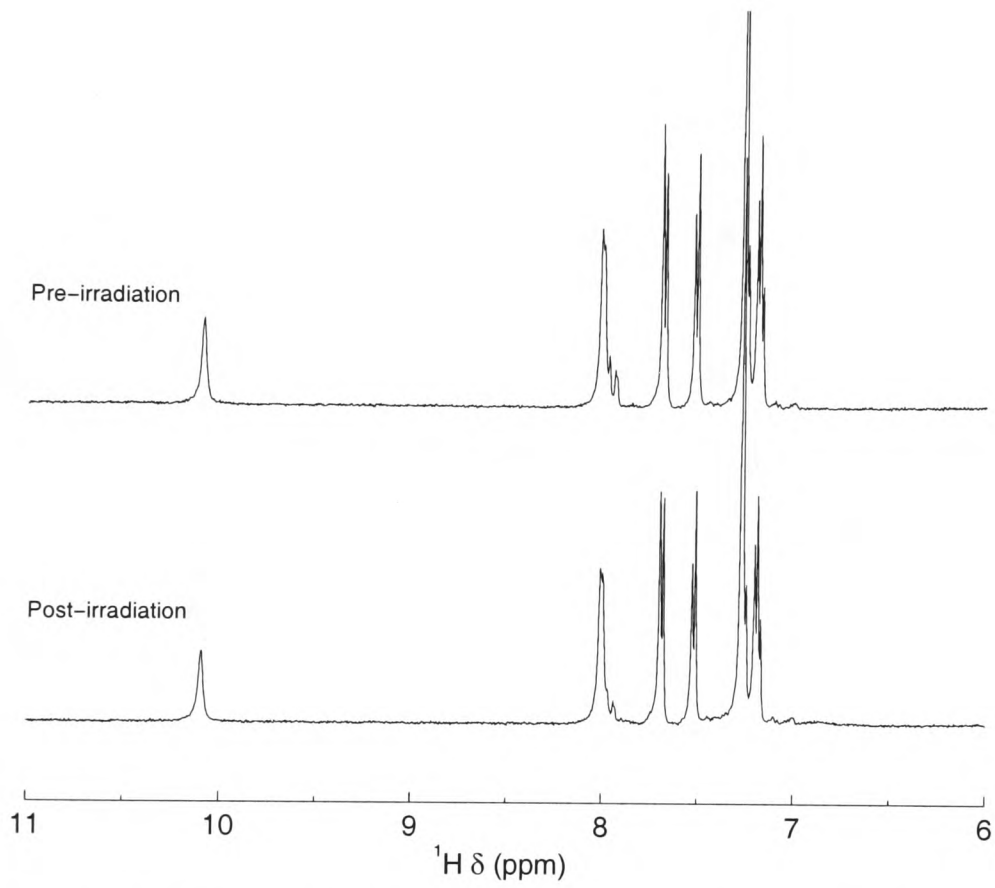


Figure 4.30: 500 MHz ^{1}H NMR spectra of a solution of 2 mM tryptophan containing 0.2 mM FMN as a photosensitiser. The spectra were recorded using 4096 complex points a spectral width of 8000 Hz and apodised with a cosine-squared window function. Laser irradiation consisted of 128 pulses of 100 ms duration at a power level of 4.9 W.

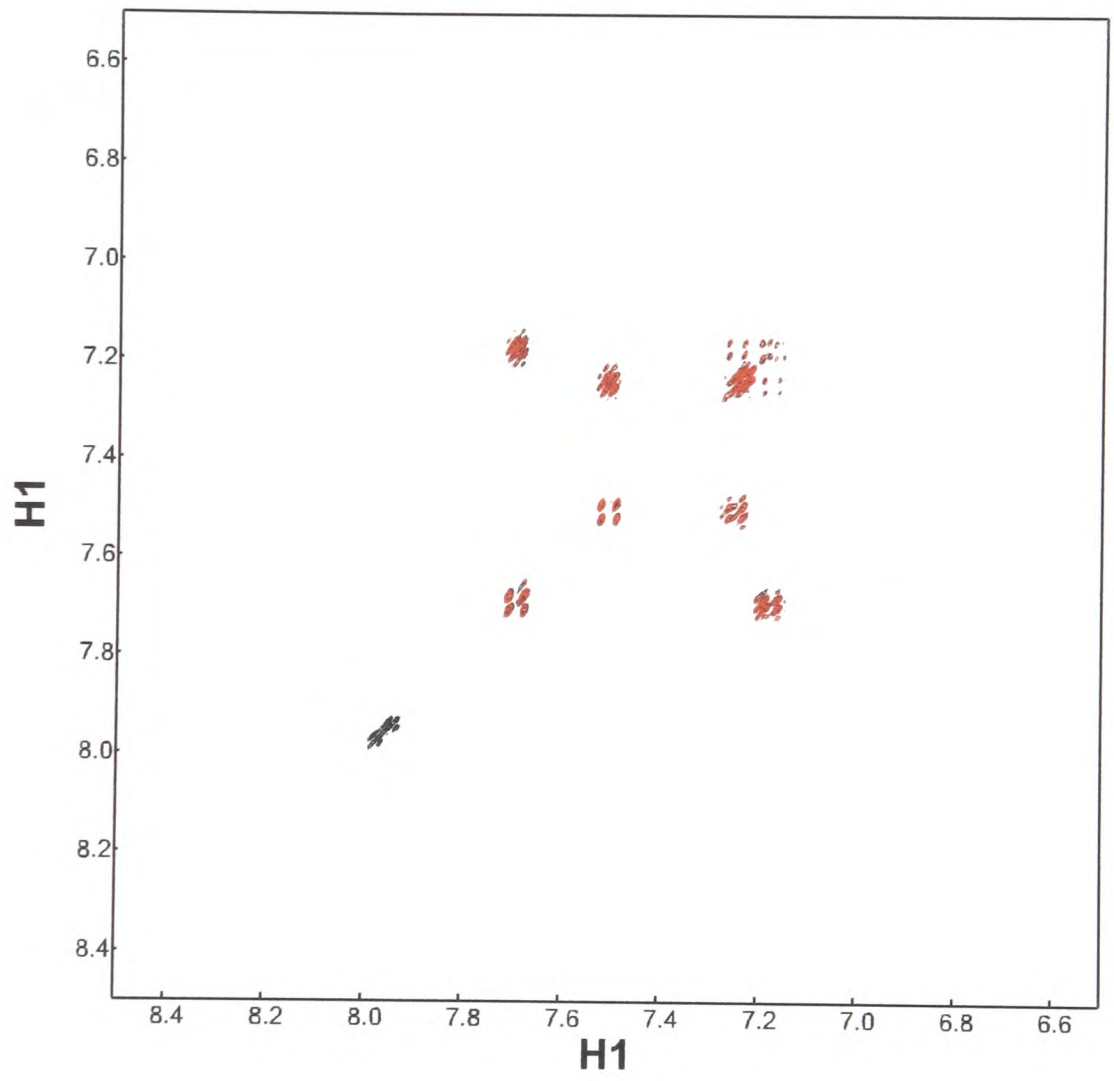


Figure 4.31: 500 MHz ^{1}H COSY spectra of a solution of 2 mM tryptophan containing 0.2 mM FMN as a photosensitiser. The black spectrum was taken before the laser irradiation with the red spectrum being taken after. The spectra were recorded using 4096 complex points in the direct dimension and 512 points in the indirect dimension. Spectral widths of 8000 Hz and sine-bell window functions were used in both dimensions. Laser irradiation consisted of 256 pulses of 100 ms duration at a power level of 5.0 W.

as a result of reduction of the flavin to the colourless, inactive form. Methods to reduce this photoreduction have been proposed by Maeda *et al.* [117].

4.4 Last thoughts / Conclusions

Described in this chapter is an injection device which is simple to use, but is capable of producing high resolution NMR spectra within 50 to 100 ms after the injection event. The design of the insert is simple and requires no modification to the spectrometer. Coupling this injection device to the CHESS solvent suppression scheme has enabled solvent exchangeable protons to be detected and information about protein structure and folding inferred.

There is, however, scope for improvement. Construction of the insert from susceptibility matched glass would enable the tip of the insert to be placed closer to the RF coil, possibly enabling greater mixing efficiency. The device could also be modified to include liquid from a second syringe, allowing ternary mixing to be performed, thus following two step processes, such as protein folding followed by ligand binding.

Previous work by Lyon *et al.* detecting nuclear polarisation transferred from ^{15}N-labelled tryptophan indoles via CIDNP-HSQC experiments [28] suggests that it may be possible to perform the CIDNP pulse labelling experiment in a similar manner. This would involve recording ^{15}N-edited one-dimensional spectra and taking advantage of the much longer spin-lattice relaxation times of nitrogen-15 as compared to protons, allowing greater time for solution mixing and protein folding between injection and acquisition, thereby giving a single signal for each exposed tryptophan residue as the NH backbone amide is too far removed from the delocalised spin system to show any nuclear polarisation.

Recent experiments following the refolding of single tryptophan mutants of the

small histidine-containing phosphocarrier protein (HPr) from *E. coli* using photo-CIDNP [33] show the potential scope for using a device which is capable of rapid mixing allowing the earliest points along the folding pathway to be observed and characterised.

Chapter 5

Real-Time Refolding of Bovine Pancreatic Ribonuclease A

There are several techniques which allow the study of protein folding in real-time, including various stopped-flow compatible techniques such as fluorescence and circular dichroism, which provide information on global changes occurring as folding is induced. More recently, real-time NMR and photo-CIDNP methods have been introduced. They have the capability to provide residue specific (both backbone and side chain) information [20, 22–24, 26, 27]. This chapter applies the rapid mixing device [27] and real-time methodology presented in Chapter 4 to a second protein system ribonuclease A from bovine pancreas, which has been extensively studied.

Bovine pancreatic ribonuclease A (RNase A) has long been the subject of study by protein biochemists. It was this protein which Anfinsen used to show that proteins can be reversibly denatured, concluding that the overall fold of the protein, its tertiary structure, is encoded purely by the amino acid sequence [179]. It was also used in the first experimental demonstration of a protein nuclear magnetic resonance spectrum in 1957 [180]. RNase A is a small globular protein constructed from 124 amino acids (molecular weight 14.2 kDa). Figure 5.1 shows the solution-state NMR

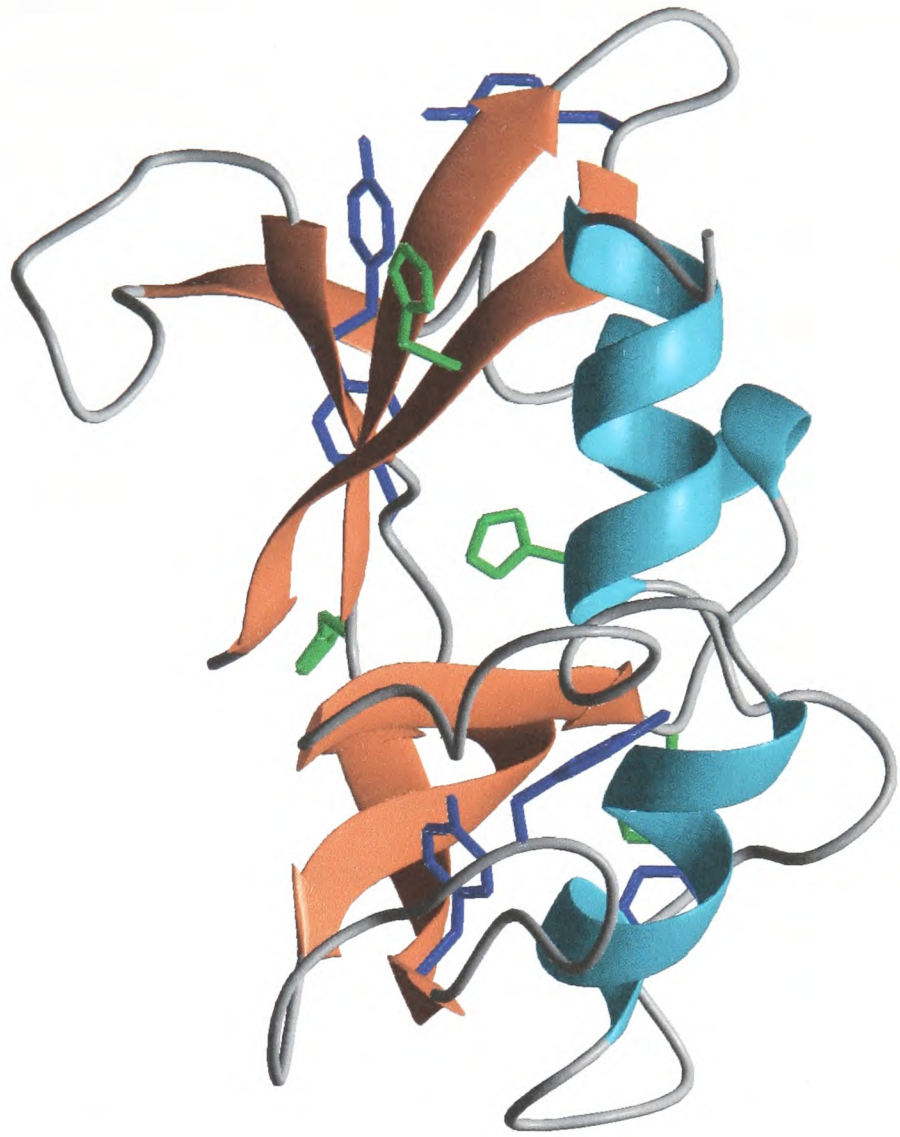


Figure 5.1: The solution structure of ribonuclease A (PDB File: 2AAS [181]). Tyrosine and histidine residues are shown in blue and green respectively. The figure was drawn using MOLMOL [89].

structure. It consists of three α-helices which serve as a scaffold for a large β-sheet region. This contains the cleft of the RNA binding site. The overall topology of the protein is stabilised by the presence of four disulphide bridges between residues 26–84, 58–110, 40–95, and 65–72. It contains ten potentially polarisable residues, six tyrosines and four histidines. Interestingly, two out of the four proline residues, namely Pro 93 and Pro 114, are in the *cis* conformation in the native state [182–184]. RNase A is involved in the hydrolysis of RNA strands in cells, via cleavage of the phosphodiester bond preceding a uracil or cytosine nucleotide. Two of the histidine residues are employed in the active site. It cannot hydrolyse DNA, because it lacks

the 2′-hydroxyl group, as this is involved in the formation of a cyclic intermediate in the hydrolysis mechanism [184].

The folding and unfolding properties of RNase A have been widely studied [185], employing a variety of techniques including NMR using equilibrium unfolding [186, 187], and temperature jump measurements [188]. The unfolding at low temperatures in the presence of cryosolvents has been monitored [189, 190]. Optical techniques have also been extensively employed using stopped-flow fluorescence [191], optical absorption [190–192], infrared spectroscopy using both temperature jumps [193] and pressure-induced folding [194] and the monitoring of 2′-cytosine monophosphate inhibitor binding and protein activity [195].

Drawing on a combination of optical spectroscopies and ^{1}H-^{3}H exchange labelling techniques, Baldwin and co-workers have proposed the following kinetic scheme for the refolding of RNase A [196, 197]:

$$\mathrm{U} \Rightarrow \begin{cases} \mathrm{U_F} & \longrightarrow & & & \mathrm{N} \\ \mathrm{U_S^I} & \longrightarrow & (\mathrm{I}') & \longrightarrow & \mathrm{N} \\ \mathrm{U_S^{II}} & \rightarrow \mathrm{I_1} & \rightarrow \mathrm{I_N} & \rightarrow & \mathrm{N} \end{cases} \tag{5.1}$$

In this scheme refolding from the unfolded state U, consists of three phases, one fast folding ($\mathrm{U_F}$), on the order of seconds, and two slower folding phases, $\mathrm{U_S^I}$ (minor contribution) and $\mathrm{U_S^{II}}$ (major contribution), folding on the order of tens to hundreds of seconds. The latter two are thought to differ in the *cis* / *trans* configuration of the four proline residues. Both these slow folding phases give rise to the formation of kinetic intermediates along the folding pathway. The intermediate $\mathrm{I_N}$ is native-like in that it possesses some enzymatic activity as monitored by the binding of 2′-cytosine monophosphate [195].

Using a series of single-jump stopped-flow studies of the refolding of RNase A following a 5-fold dilution from 4.2 M guanidine hydrochloride, Houry *et al.* obtained

the data shown in Table 5.1 [198]. In addition to the fast and slow phases given in the reaction scheme above, they observed an additional very fast folding phase (U_{VF}). The slow folding phase given is the sum of U_S^I and U_S^{II}.

Method	τ_{VF}	τ_F	τ_S $(= \tau(U_S^I) + \tau(U_S^{II}))$
absorbance	47.8 ms	3.85 s	303 s
fluorescence	71.9 ms	6.15 s	341 s

Table 5.1: Results of single-jump stopped-flow refolding of RNase A following a 5-fold dilution from 4.5 M GdnHCl [198].

The remainder of this chapter details results of equilibrium unfolding experiments and real-time refolding experiments, all monitored by the use of photo-CIDNP spectroscopy.

5.1 Equilibrium Unfolding

There is a variety of methods available to denature proteins. These include changes to the solvent system by altering the pH [137, 163], the addition of co-solvents [145, 146, 148–151, 199–201] or chemical denaturants such are urea and guanidine hydrochloride (GdnHCl) [146, 148, 164, 202, 203]. Unfolding can be induced by physical methods such as changes in temperature [199, 204], or pressure [169, 194].

Figure 5.2 shows the equilibrium unfolding of RNase A using guanidine deuterochloride as a chemical denaturant, probed using photo-CIDNP. The spectrum recorded in the absence of denaturant shows that four of the tyrosine residues are exposed to the solvent and accessible to the triplet flavin, namely Tyr 73, Tyr 76, Tyr 92 and Tyr 119. This is in agreement with inspection of the RNase A crystal structure. The H3,5 proton resonances of Tyr 76 and Tyr 92 are heavily overlapped, differing in their assigned chemical shifts by 0.02 ppm [205]. There is a noticeable

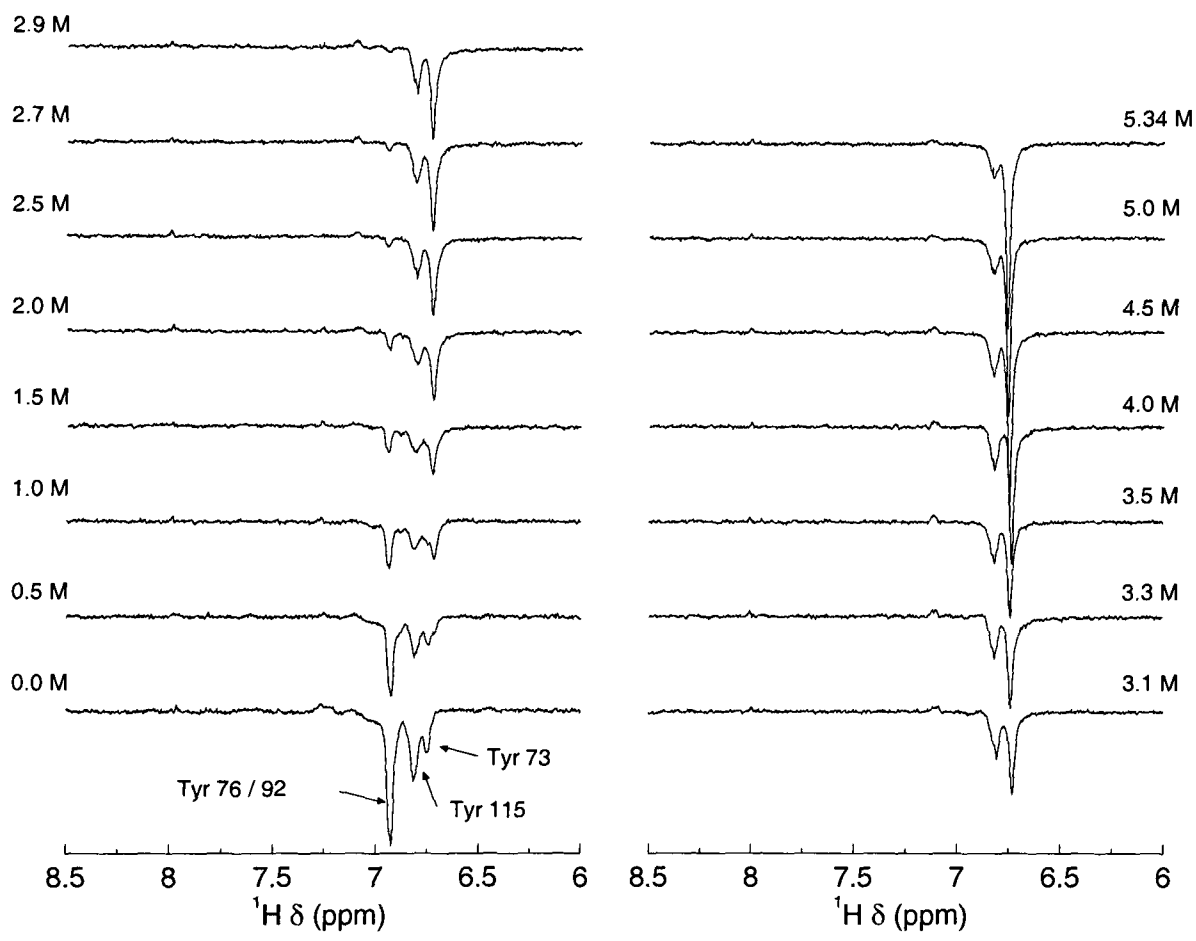


Figure 5.2: 600 MHz ^{1}H CIDNP spectra of the equilibrium unfolding of 2 mM RNase A using guanidine deuterochloride, at pH 6. The spectra were recorded using a 100 ms laser pulse at 10 W. The assignments are taken from [205].

change in the appearance of the spectra as the denaturant concentration is increased. The change in protein conformation away from the native state results in the disappearance of the signals arising from Tyr 79 and 92, and a weakening of the signal at 6.81 ppm, corresponding to Tyr 115. At high denaturant concentration, there is a large signal at 6.75 ppm arising from a number of exposed tyrosine residues in the unfolded state. From a visual inspection of these spectra it is apparent that there is a shift from the native to unfolded state at around 2.5 to 2.9 M GdnDCl. No further changes were observed beyond 5.34 M denaturant. This is in good agreement with changes made monitoring the variation in optical absorption as the protein undergoes unfolding [202].

A second unfolding experiment, using urea as the denaturant, again under equilibrium conditions, shows a similar trend. Figure 5.3 shows the CIDNP spectrum recorded at a series of different urea concentrations. Upon unfolding, there is a single, broad signal at 6.81ppm, which may contain a contribution from as many as all six of the tyrosine residues in the protein. The midpoint of the unfolding appears to occur somewhere between 3.5 and 5.0 M denaturant. This is noticeably lower than that observed by Ahmad and Bigelow, who observed the unfolding midpoint to be around 6.5 M urea using optical absorption spectroscopy [202]. This difference can be attributed to the different method employed to obtain the data. Interestingly, the H2 proton of histidine 119 appears to be unaffected by the increase in denaturant.

The denatured states attained by these two experiments, shown in Figures 5.2 and 5.3, are clearly different. This can be understood in terms of a denatured state ensemble. That is, the denatured state of a protein is not comprised of one single structure, but an ensemble of conformations which are in a dynamic equilibrium. Therefore the observed spectrum is an average over that ensemble [203, 206, 207]. The use of difference chemical denaturants to effect the unfolding shifts the relative weights of the members of the ensemble, each with differing amounts of residual

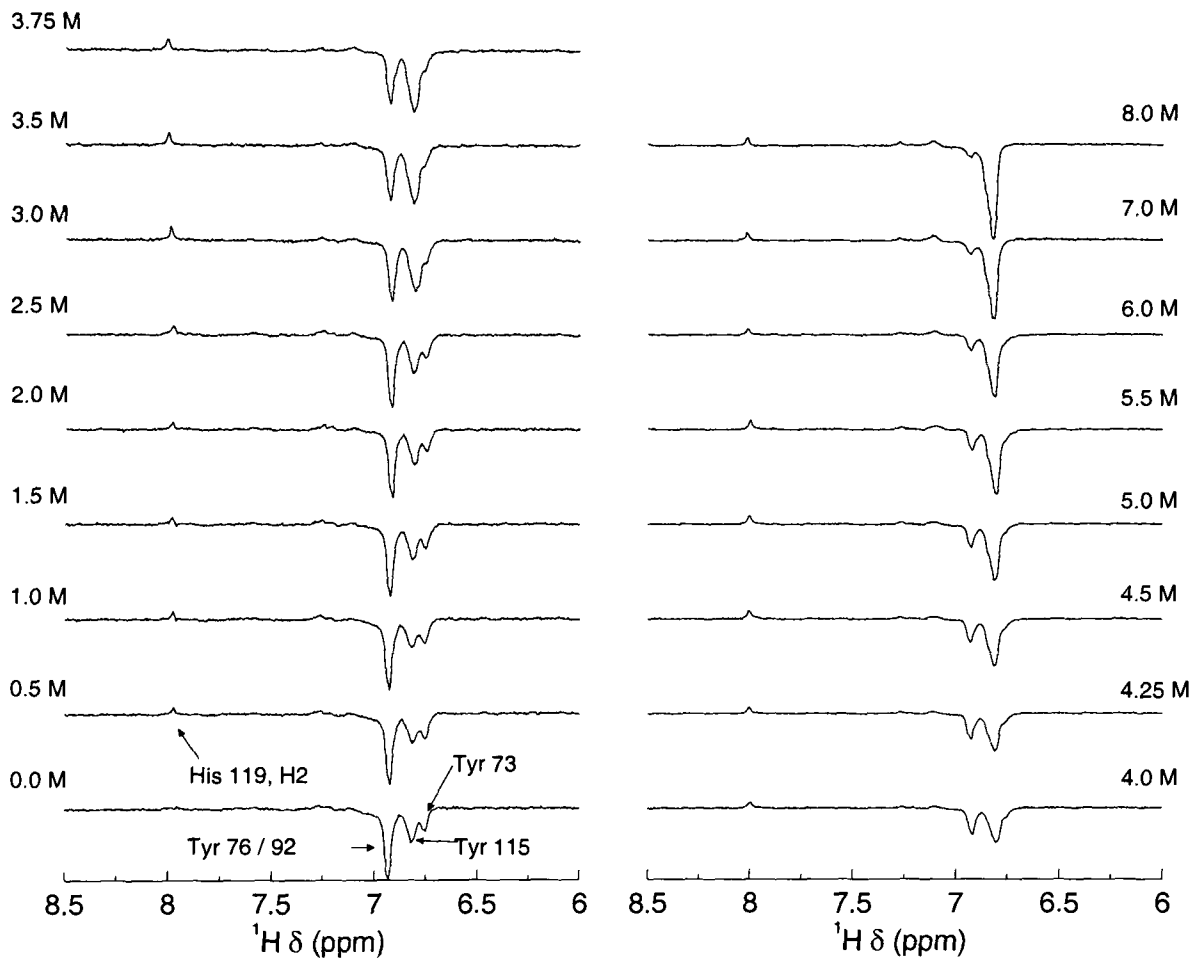


Figure 5.3: 600 MHz ^{1}H CIDNP spectra of the equilibrium unfolding of 2 mM RNase A using deuterated urea, at pH 6. The spectra were recorded using a 100 ms laser pulse at 10 W. The assignments are taken from [205].

structure, giving rise to distinct spectra. It is also interesting to note that spectra recorded at intermediate denaturant concentrations are not identical. For example, the spectrum recorded with 3.75 M urea does not have a similar appearance to any of the spectra in the guanidine unfolding series, indicating that the series of conformations being sampled in the two experiments is not the same.

5.2 Real-Time Refolding

Using the rapid mixing injection device [27] and the real-time methodology described in Chapter 4, a series of experiments to probe the real-time refolding of ribonuclease A, using photo-CIDNP, was performed. These experiments all follow a broadly similar protocol: 30 μL of 10 mM denatured protein containing 0.2 mM flavin mononucleotide in 100 mM sodium phosphate buffer at pH 6 was loaded into the injection device. A trace amount of dioxane was added to serve as the chemical shift reference. This solution was then injected into 300 μL of 100 mM sodium phosphate buffer, also containing 0.2 mM flavin. This resulted in a 10-fold dilution of the injectant solution. All solutions were prepared in deuterium oxide, with all pH readings uncorrected for any isotope effects.

A series of CIDNP spectra was then recorded as a function of time, using the pulse sequence shown in Figure 5.4. A laser pulse of 100 ms duration, typically at a power level of 10 W, was applied prior to each acquisition. Solvent suppression was provided using the CHESS sequence[1] [125]. The spectra were acquired using the following parameters: 2048 complex points were acquired over a spectral width of 7005 Hz. The data set was then extended to 4096 complex points by the application of forward-backward linear prediction, and apodised with a suitable window function prior to the Fourier transform.

[1] CHEmical Shift Selective

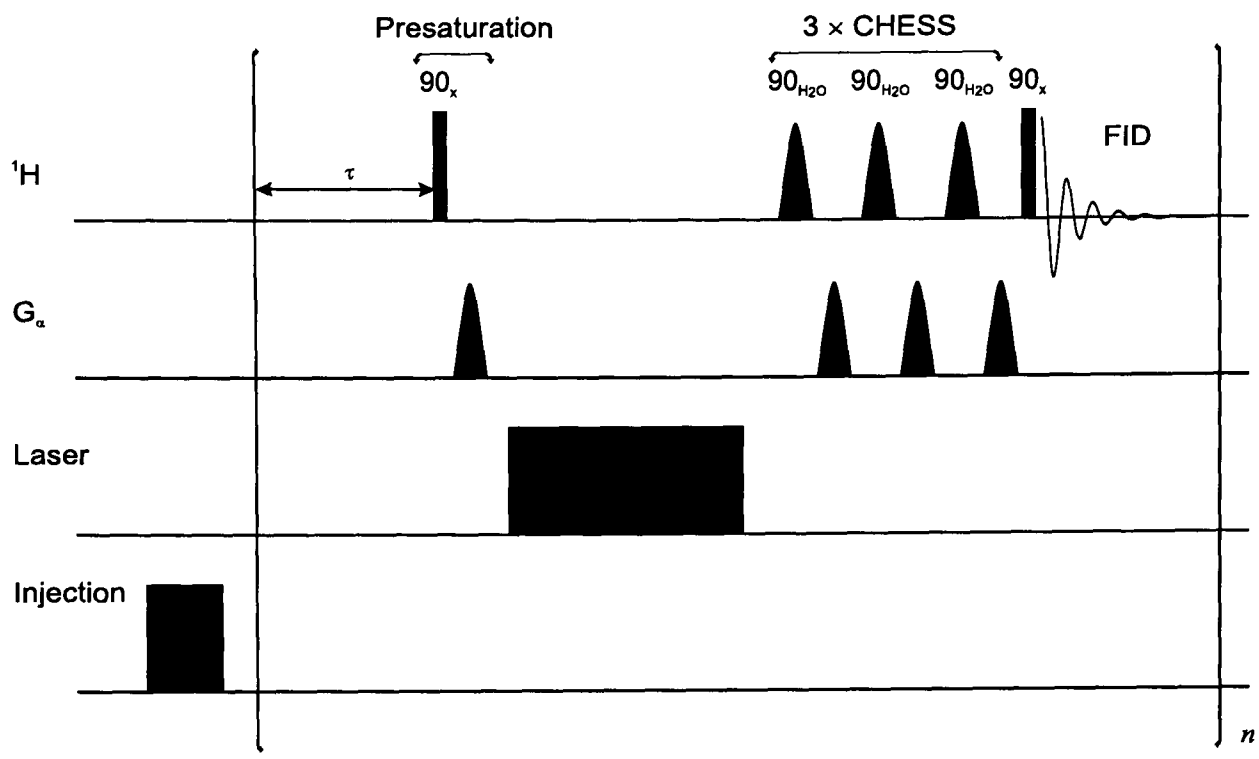


Figure 5.4: The pulse sequence timing diagram for the real-time refolding. The delay τ is set to allow the monitoring of folding over different timescales. Solvent suppression is provided using the CHESS method [125].

A series of injection experiments were then performed, using an appropriate delay τ chosen so as to cover the time period of interest. The real-time CIDNP spectra were averaged over a series of light and dark subtraction pairs to improve the signal-to-noise ratio.

As for the equilibrium unfolding experiments performed above, real-time CIDNP refolding spectra were obtained to observe the return to the native state following denaturation in either guanidine deuterochloride or urea.

The initial refolding experiment probes the refolding of RNase A from 5 M guanidine deuterochloride, giving a final denaturant concentration of 0.5 M GdnDCl, and 1 mM protein. The refolding was followed over a time period of 60 seconds following the injection at 40 time points, generating 40 CIDNP spectra. Selected time points from this experiment are shown in Figure 5.5(a). There are clearly changes occurring to the CIDNP spectra over the time course of the refolding reaction. Comparison of the final spectrum, recorded 58.9 s after the injection event, with the equilibrium un-

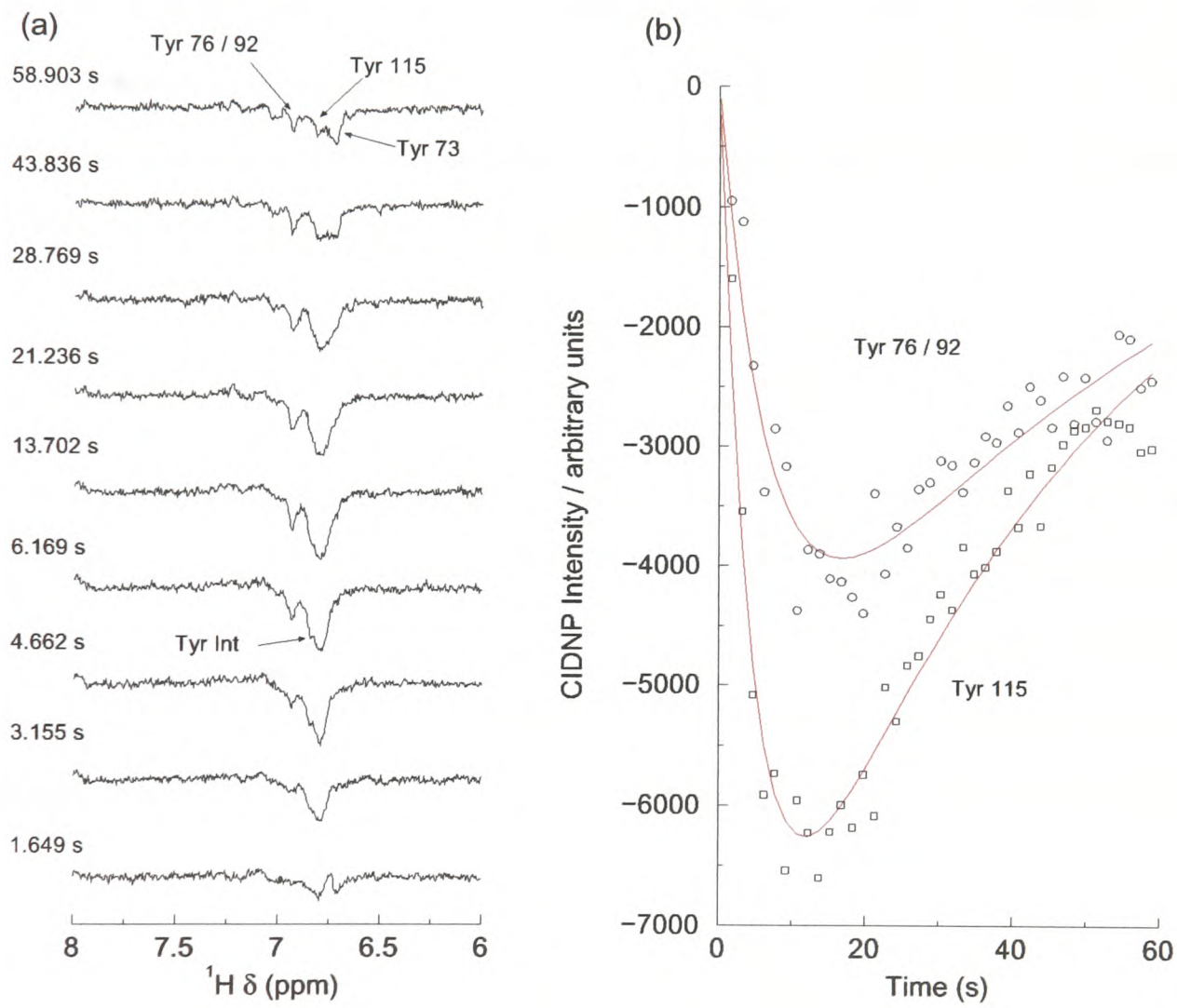


Figure 5.5: Refolding of RNase A from 5 M GdnDCl monitored over 60 s. The spectra were apodised with 2 Hz exponential line broadening. "Tyr int" denotes a signal from some folding intermediate. The time indicates the time between the injection event and the start of the acquisition of the FID. The red lines are the result of fitting the data to Equation 5.2. Assignments are taken from [205].

folding spectra in Figure 5.2 shows that, while the protein is in the process of folding it has not yet attained its native conformation. Kinetic information about the folding process can be obtained by extracting peak heights at the known chemical shifts of the native state tyrosine H3,5 protons along the course of the folding process. Figure 5.5(b) shows the peak heights for Tyr 115 and for the peak corresponding to both Tyr 76 and Tyr 92. Due to the tyrosine resonance from the unfolded state having an identical chemical shift to that of Tyr 73, it was not possible to extract useful kinetic information for this residue. The extracted intensities were then fitted to a double exponential function of the following form:

$$I(t) = A\left\{\exp\left(-\frac{t}{T_d}\right) - \exp\left(-\frac{t}{T_r}\right)\right\} \tag{5.2}$$

where T_r is the time constant for the recovery of the native state signal, and T_d is a time constant for the subsequent decay. The origin of this decay is principally photoreduction of the photosensitiser, and will be expanded upon and interpreted in Section 5.2.2. A is an arbitrary scaling factor. The parameters extracted from the fit of Equation 5.2 to Figure 5.5(b) are summarised in Table 5.2. Also noticeable in

Residue	A	T_r	T_d
Tyr 76/92	−6007	6.99 s	57.2 s
Tyr 115	−9205	4.73 s	43.7 s
Tyr 73		not fitted	

Table 5.2: Time constants and scaling factors obtained from fitting Equation 5.2 to the peak intensities in Figure 5.5(b).

this set of spectra is the presence of a signal at 6.85 ppm, denoted by 'Tyr int'. This signal does not correspond to the native chemical shift of any of the six tyrosine residues in RNase A, and is hence presumed to be a kinetic intermediate. This intermediate appears after approximately four seconds and persists as a shoulder on

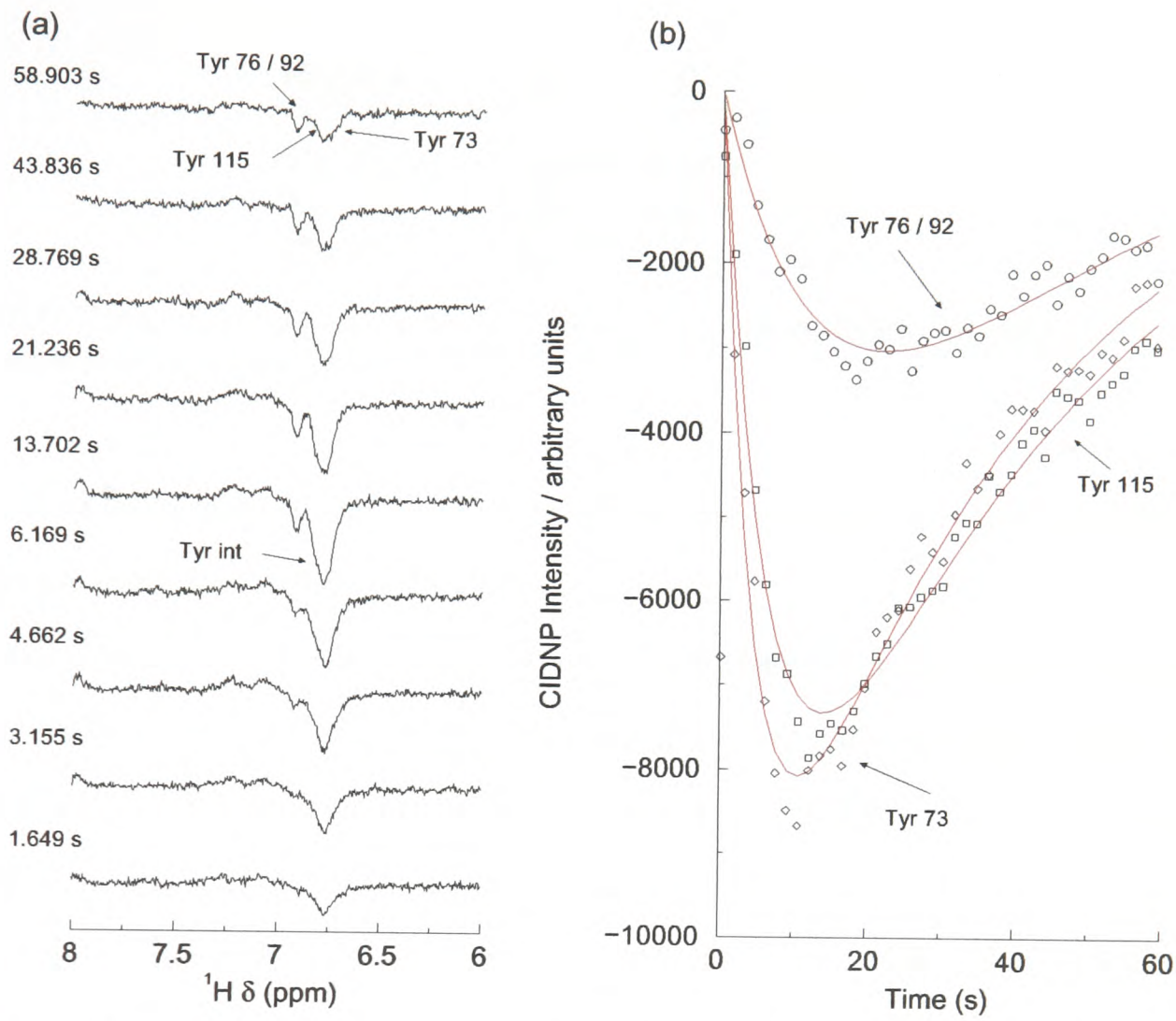


Figure 5.6: Refolding of RNase A from 10 M urea monitored over 60 s. The spectra were apodised with 2 Hz exponential line broadening. "Tyr int" denotes a signal from some folding intermediate. The time indicates the time between the injection event and the start of the acquisition of the FID. The red lines are the result of fitting the data to Equation 5.2. Assignments are taken from [205].

the Tyr 115 signal for the duration of the experiment.

A similar experiment was performed using 10 M urea as the denaturant, giving a final concentration of 1 M urea. The remaining sample conditions were unchanged. The CIDNP spectra recorded over 60 seconds are shown in Figure 5.6(a). As for the guanidine case, the native state is not fully formed by the end of this experiment. The spectra bear a striking resemblance to each other. Kinetic information can also be extracted from these data, again, fitting Equation 5.2 to the peak heights shown in Figure 5.6(b). The parameters obtained are given in Table 5.3. Since there is a small difference between the position of the emissive tyrosine signal corresponding

Residue	A	T_r	T_d
Tyr 76/92	−10688	15.45 s	34.2 s
Tyr 115	−12549	4.51 s	35.0 s
Tyr 73	−12892	6.68 s	38.0 s

Table 5.3: Time constants and scaling factors obtained from fitting Equation 5.2 to the peak intensities in Figure 5.6(b).

to the urea denatured state and Tyr 73 in this experiment, it is now possible to extract kinetic information for this residue.

The signal from a possible folding intermediate observed in the case of refolding from GdnDCl is also observed in the refolding from urea, at 6.83 ppm. Again, this intermediate is first noticed around seven seconds following the initiation of the folding.

The data obtained from these two experiments are in reasonable agreement. Since the native state is clearly not fully formed in either case, the experiments were repeated, this time monitoring the refolding at forty time points over a three hundred second period. Using guanidine deuterochloride as the denaturant, the spectra and extracted native position peak intensities shown in Figure 5.7 were obtained. It is apparent here, two, that the native state is not obtained by the end of the experiment. This is in fact due to poor subtraction of the residual guanidine deuterochloride signal, at 6.70 ppm. The contribution to this GdnDCl signal from the dark state spectra is slightly larger than in the case of the light spectrum hence upon subtraction, it appears as an emissive signal. The increased intensity of the GdnDCl signal in the dark state compared to the lit state arises from minor heating effects in the light experiment. This is not observed in the equilibrium unfolding experiments because there is a longer delay between scans. Without this complication, RNase A shows a native-like spectrum after 300 s following the initiation of refolding. Extracting the kinetic parameters for this data set, shown in Figure 5.7(b), gives the

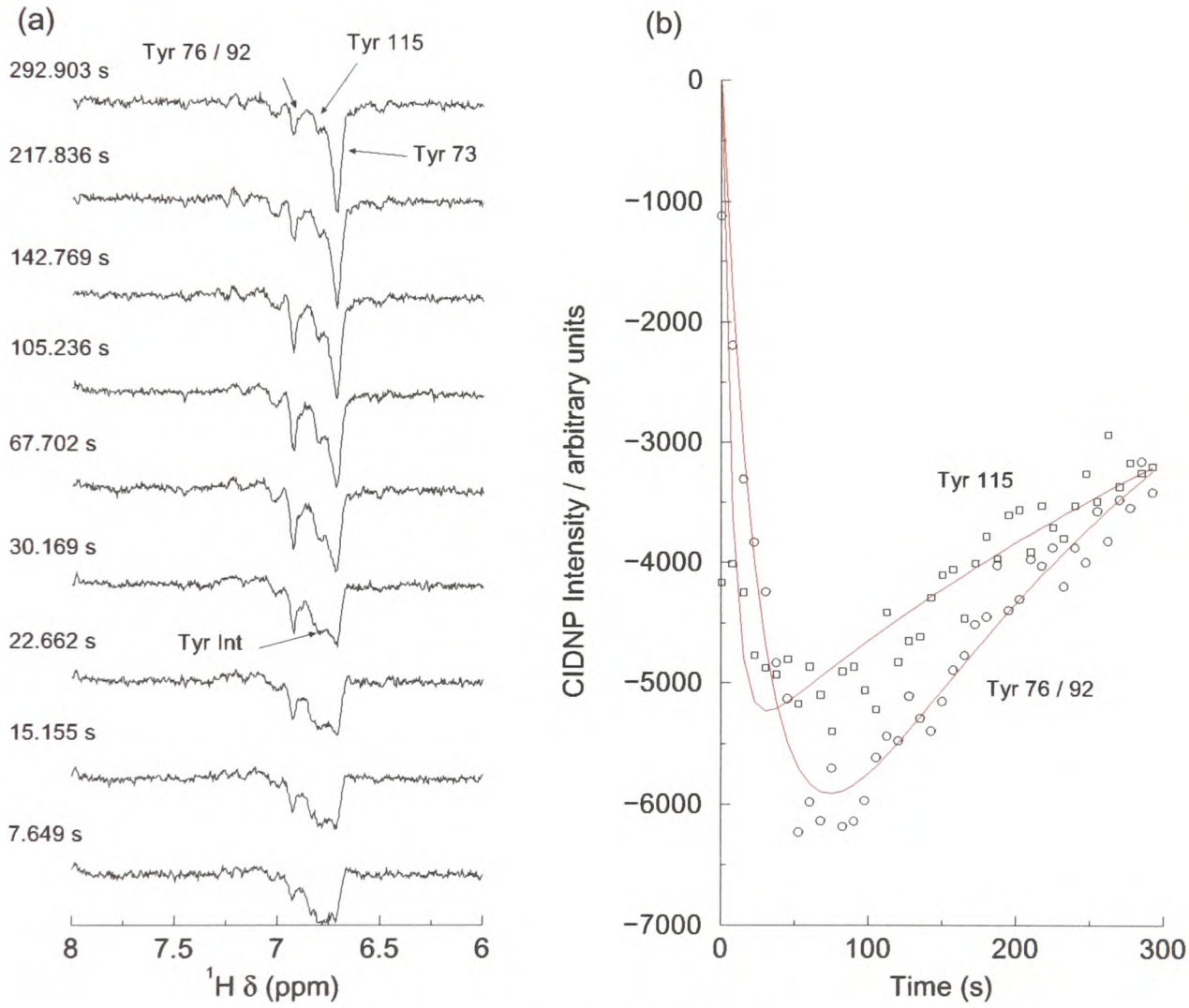


Figure 5.7: Refolding of RNase A from 5 M GdnDCl monitored over 300 s. The spectra were apodised with 2 Hz exponential line broadening. The time indicates the time between the injection event and the start of the acquisition of the FID. The red lines are the result of fitting the data to Equation 5.2. Assignments are taken from [205].

data in Table 5.4. An additional signal is also noticed to be present in this data set,

Residue	A	T_r	T_d
Tyr 76/92	−8204	27.84 s	315.0 s
Tyr 115	−5034	7.16 s	519.7 s
Tyr 73		not fitted	

Table 5.4: Time constants and scaling factors obtained from fitting Equation 5.2 to the peak intensities in Figure 5.7(b).

first observed in the spectrum recorded at 7.65 s and persisting for the next sixty seconds.

The series of experiments is completed using urea as the denaturant. The resulting spectra are shown in Figure 5.8(a). The final spectrum in this series does show native-like characteristics when compared to that from the equilibrium unfolding experiments as shown in Figure 5.3, in the presence of 1.0 M urea. The kinetic data extracted from this injection experiment are shown in Table 5.5. Again, the

Residue	A	T_r	T_d
Tyr 76/92	−8113	21.49 s	273.2 s
Tyr 115	−7280	3.86 s	229.52 s
Tyr 73	−3957	3.38 s	266.4 s

Table 5.5: Time constants and scaling factors obtained from fitting Equation 5.2 to the peak intensities in Figure 5.8(b).

presence of a non-native tyrosine signal during the refolding.

With the exception of the time constant for the recovery of Tyr 76/92 when monitored over three hundred seconds, the kinetic data sets obtained for the refolding of RNase A are reasonably consistent, across both time periods and chemical denaturants. This is in agreement with observations made by Mui *et al.* using tyrosine fluorescence measurements [203]. Moderately faster kinetics are observed in the

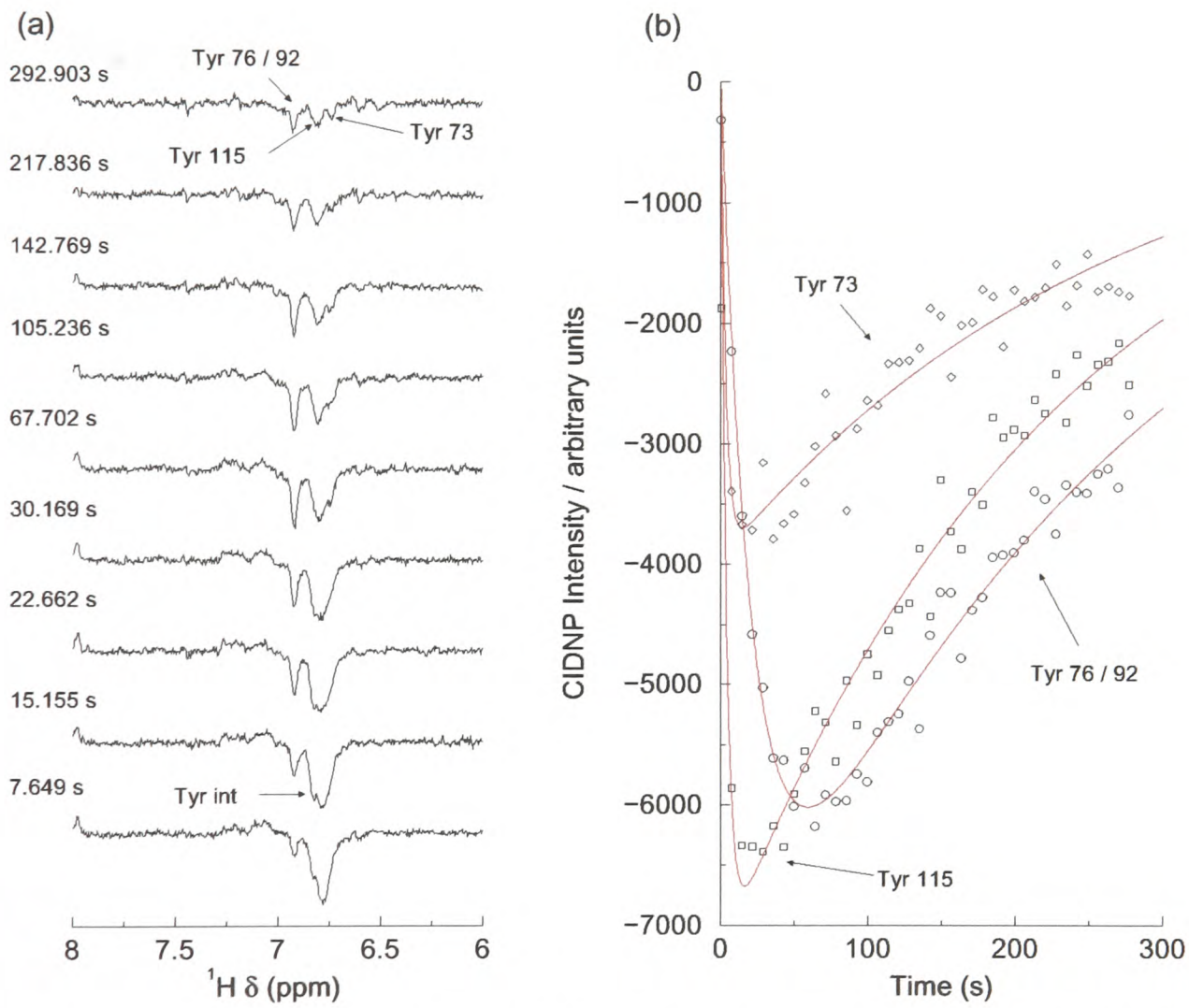


Figure 5.8: Refolding of RNase A from 10 M urea monitored over 300 s. The spectra were apodised with 2 Hz exponential line broadening. The time indicates the time between the injection event and the start of the acquisition of the FID. The red lines are the result of fitting the data to Equation 5.2. Assignments are taken from [205].

case of refolding from urea; however, considering the accuracy of these experiments, it is difficult to infer details about the folding processes occurring in the different cases. It is furthermore interesting to note that a kinetic intermediate has been observed along the folding pathway, starting in both high concentrations of urea and GdnDCl [203].

5.2.1 Interpretation of the Folding Kinetics

Ribonuclease A has been used the subject of numerous investigations in protein folding studies [185]. Comparison of the literature results shown in Table 5.1 with those obtained in the real-time photo-CIDNP experiments presented in Figures 5.5 and 5.7 shows that the principal species monitored in the folding experiment is likely to be the fast phase U_F. The very fast folding species U_{VF} cannot be observed in the photo-CIDNP detected refolding experiments since the folding of this species will be complete within the dead time of the experiment.

The slower time constants observed for the recovery of the native state signals from Tyr 76 and Tyr 92, which are difficult to distinguish given they differ in chemical shift by 0.02 ppm, can be understood in terms of proline isomerism. Usually, the peptide bond adopts the *trans* configuration, with less than 0.1% being present in the *cis* conformation [184]. Proline constitues an exception to this in that the side chain forms a five membered loop between C^{α} and N^{H}. Figure 5.9 shows the *trans* and *cis* conformations of the proline peptide bond. In free solution and small peptides the equilibrium is shifted so approximately 30% of prolines are present in the *cis* configuration. Native ribonuclease A contains four proline residues, of which two (Pro 93 and Pro 114) are found in the *cis* conformation [182–184]. Isomerism of these peptide bonds will therefore be a rate-limiting step along the folding pathway. Of these two *cis* proline residues, Pro 93 immediately follows Tyr 92 in the sequence. Thus in order for Tyr 92 to attain its correct native position, and hence give rise to

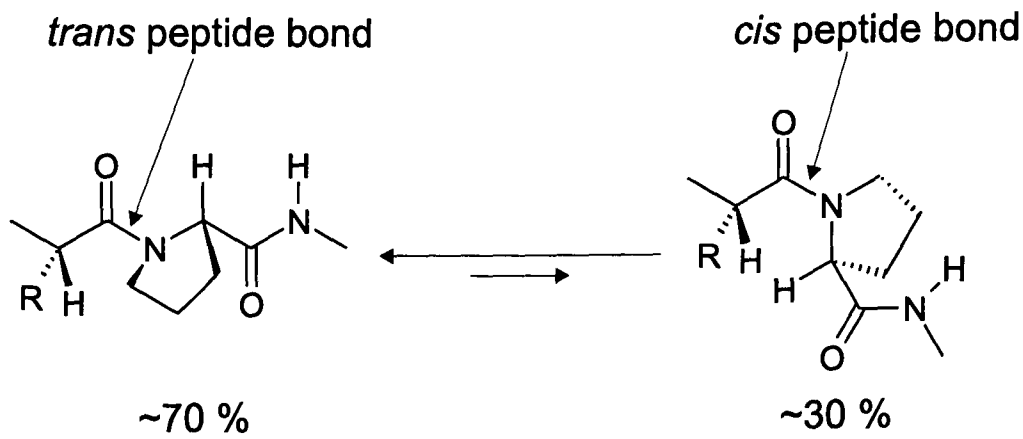


Figure 5.9: *trans* and *cis* isomers of the proline peptide bond within a protein chain.

a CIDNP signal at the correct region in the spectrum, it is dependent on the rate of proline isomerism, from *trans* to *cis*, at Pro 93.

Further evidence from Houry and Scheraga suggests that there is a fifth folding contribution, termed U_M, for medium folding phase [208]. Under similar conditions to those used for the real-time photo-CIDNP experiments time constants for the folding of the two fast phases and the medium phase suggest that these are the principal components observed since it is not possible to separate the contributions, and the slower time constants measured for Tyr 76/92 correspond to the refolding of U_S^{II} [208]. This unfolded species is known to have non-native configuration of Pro 93, *i.e.* it is in the *trans* form, and a partition between species with non-native Pro 114 or Pro 117 [209]. This non-native configuration of Pro 93 has been shown by limited proteolysis and tyrosine absorption and fluorescence spectroscopy to be a rate-limiting step in the refolding of RNase A [182, 183, 191].

Since the slower kinetics obtained for Tyr 76 / 92 are postulated to arise from proline isomerism required for the folding of U_S^{II}, it is therefore possible to envisage that the intermediate observed in the folding from both urea and guanidine deuterochloride arises from significant population of the native-like intermediate I_N. Although this intermediate has not been structurally characterised, it is known to be reasonably native-like because it binds the inhibitor 2′-cytosine monophosphate [195],

suggesting that the active site is at least partially formed.

5.2.2 Photodegradation and Sample Heating Effects

Photodegradation of the flavin is a known problem in the CIDNP experiment and several methods have been proposed to circumvent it [117]. The decay constants obtained from fitting the peak intensities in the real-time experiments presented above, especially those for the sixty second time frame, are in agreement with those obtained previously in Chapter 4 for the refolding of hen egg white lysozyme, presented in Table 4.5 [27]. As mentioned in Chapter 4, there is also some evidence in the literature of flavin-sensitised photo-oxidation of the amino acids and related compounds such as indole [174–176, 210, 211], along with photodeactivation of lysozyme [177] and ribonuclease A [178]. The photodeactivation presented in these articles occurs after illumination for tens of minutes. This, coupled with the experiments performed following the lysozyme refolding (Chapter 4), strongly suggests that the decay in the observed CIDNP intensities during the refolding of RNase A is attributable solely to decay of the flavin.

In an attempt to quantify and understand the photodegradation occurring in the real-time spectra presented in this Chapter, the following experiment was performed. The change in CIDNP intensity of the native state tyrosine residues over the same time period as that used in Figures 5.5 and 5.6, *i.e.* 60 s, was monitored, this time without an injection event. The results of this experiment are shown in Figure 5.10. The decaying intensity of the native state CIDNP signals was then fitted to the following function:

$$I(t) = A\mathrm{e}^{-t/T_d} + C \tag{5.3}$$

The parameters obtained from the fit are given in Table 5.6. It is immediately clear that the time constants for the decay of the flavin obtained from this model experiment are much shorter than those observed in the refolding experiments, which

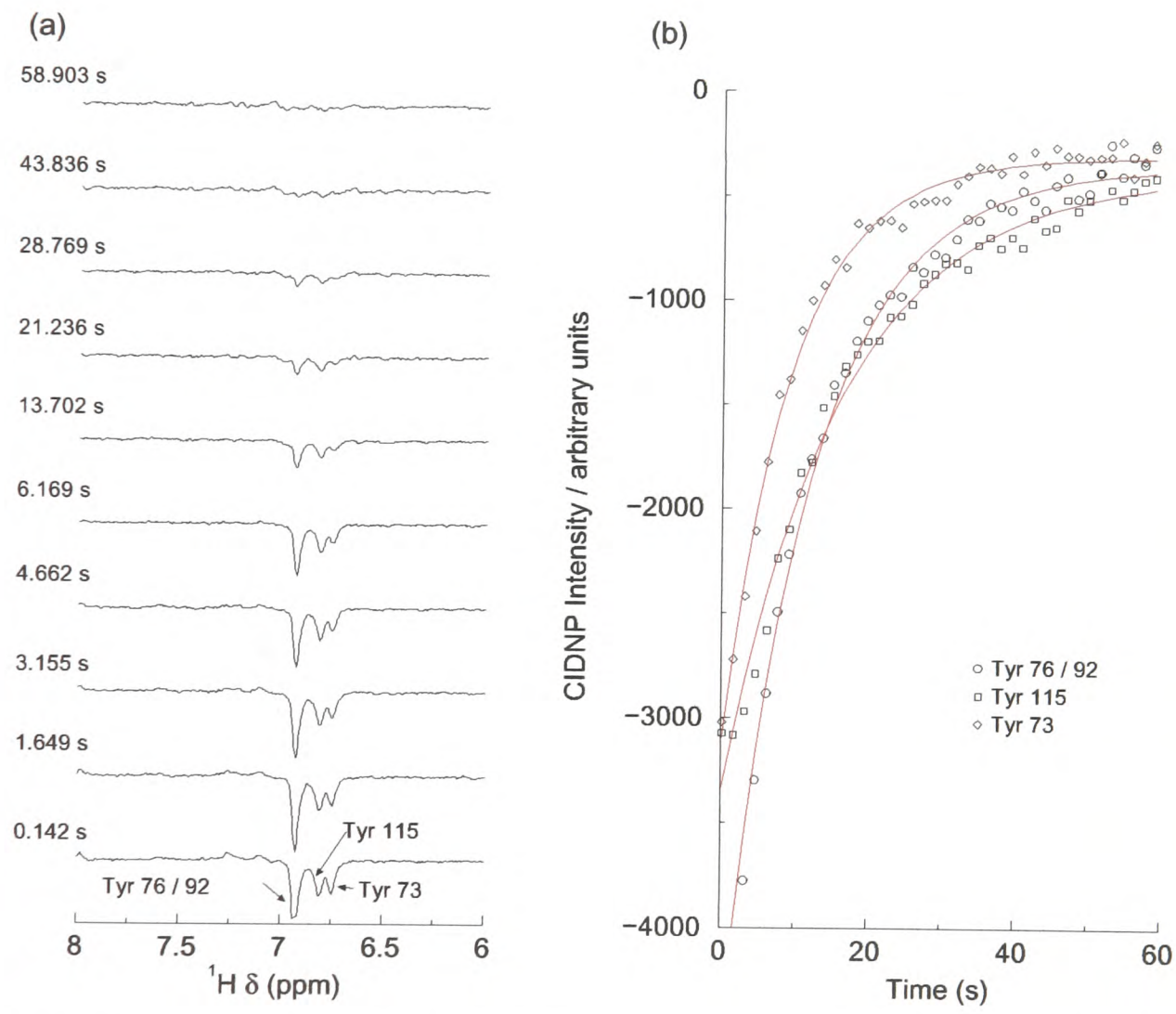


Figure 5.10: The decay of RNase A polarisation monitored over 60 s. The time indicates the time between the start of the experiment and the start of the acquisition of the FID. The red lines are the result of fitting the data to a single decaying exponential. Assignments are taken from [205].

Residue	A	T_d	C
Tyr 76/92	−4186	11.77 s	−362
Tyr 115	−2958	15.99 s	−395
Tyr 73	−2822	9.33 s	−315

Table 5.6: Time constants and scaling factors obtained from fitting Equation 5.3 to the peak intensities in Figure 5.10(b).

are of the order of $\sim$ 25 s. This indicates that reduction of the flavin is occurring more rapidly in this model experiment. This difference can be understood in terms of the shortcomings of this model: when the real-time refolding experiments are set up a small air-bubble, approximately 0.5 μL in volume, is placed at the end of the injector tip [27]. The presence of this air bubble will introduce a small amount of molecular oxygen into the sample. This oxygen will react with the colourless flavin hydroquinone to regenerate the photoactive flavin quinone. The turbulent nature of the injection event will also introduce a small amount of oxygen into the sample. This is, in part, analogous to the reoxidation via the mechanical re-injection technique proposed by Maeda *et al.* [117]. A second possible explanation is that the unfolded form of the protein causes the flavin dye to be bleached more slowly than native state RNase A would. The mechanism of this action is not yet clear.

Previous studies of protein folding monitored in real time by photo-CIDNP have employed laser pulses at a power level of approximately 4 W [26, 27, 33, 117]. The experiments presented here use a much greater power level, that is 10 W, which has the potential to complicate the interpretation of the data recorded. A quick "back of the envelope" calculation suggests that, assuming all light entering the sample is absorbed as heat[2], forty 100 ms laser pulses would introduce enough energy into the sample to raise the temperature by $\sim$ 25°. In reality, there will be a smaller temperature rise in the system since only a proportion of the light

[2]Clearly a gross over-estimate!

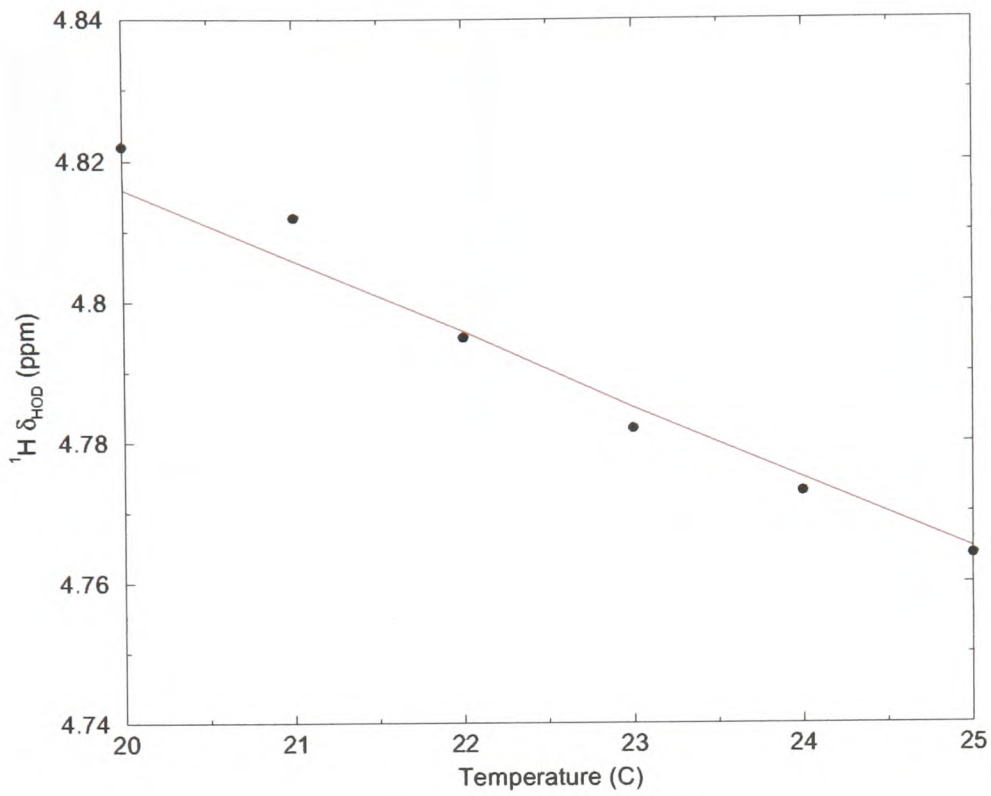


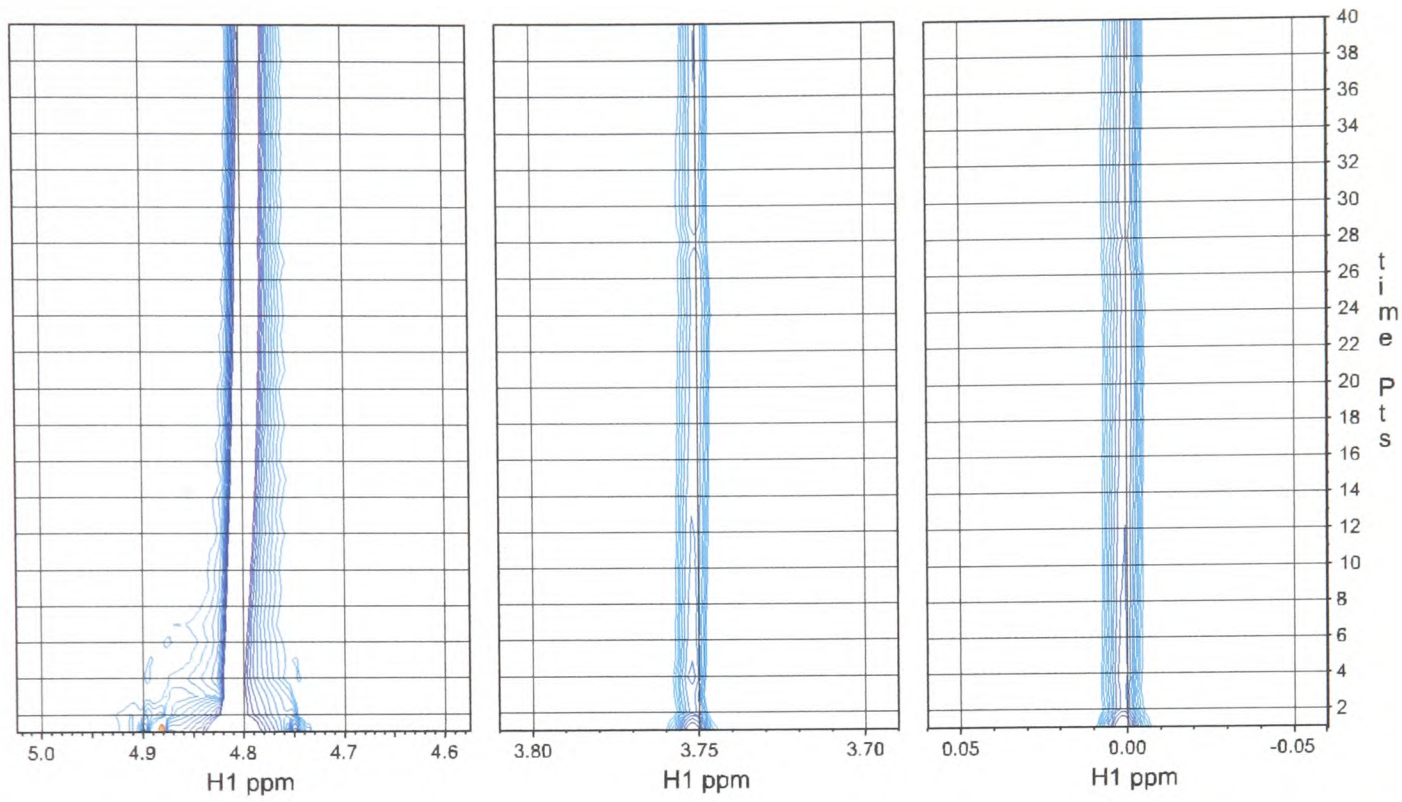
Figure 5.11: Temperature dependence of the HOD resonance. The red line is that calculated from Equation 5.4.

is absorbed and heat will be dissipated. This heating will also not be uniform, since the laser light is introduced from the top of the sample. The extent of the heating will decrease exponentially down from the tip of the injector. In order to judge what temperature change is actually occurring during the real-time experiment the following experiment was devised. The temperature dependence of the proton chemical shift of the residual HOD peak of deuterium oxide is well known, given by the equation below [74]:

$$\delta_{\mathrm{HOD}} = 7.83 - \frac{T}{96.6} \tag{5.4}$$

where T is given in Kelvin, and the solution is at pH 5.5. There is also a change in the HOD chemical shift of 0.02 per pH unit. Figure 5.11 shows the measured chemical shift of the residual HOD signal in a solution of 0.2 mM FMN containing trace amounts of dioxane and DSS[3], made up in 99.9% D_2O. The solution pH was measured to be 6.0, uncorrected for the deuterium isotope effect. Using this as

[3] 2,2-dimethyl-2-silapentane sulphonic acid



(a) HOD (b) Dioxane (c) DSS

Figure 5.12: Contour plots of the HOD, dioxane and DSS signals. The spectra were recorded every 1.1 s, and therefore span a time period of 60 s.

a baseline, a series of forty spectra were then acquired over sixty seconds using the same protocol as in Figure 5.10. The results of this experiment are shown in Figure 5.12. There is clearly a noticeable change in the HOD chemical shift, shown in Figure 5.12(a), from 4.815 ppm at the start of the experiment to 4.789 ppm by its conclusion. Similar changes are not mirrored by the dioxane or DSS signals, whose chemical shifts are constant over the complete time series and therefore shown to be independent of temperature (Figures 5.12(b) and 5.12(c)). It is possible to follow the change in the chemical shift of the HOD signal and convert this to the average temperature within the sample tube. This is demonstrated in Figure 5.13. These results indicate that the average temperature rise in the sample is $\sim 3°$, which should be insufficient to cause noticeable thermal denaturation effects under the usual experimental conditions used for the photo-CIDNP monitored refolding experiments.

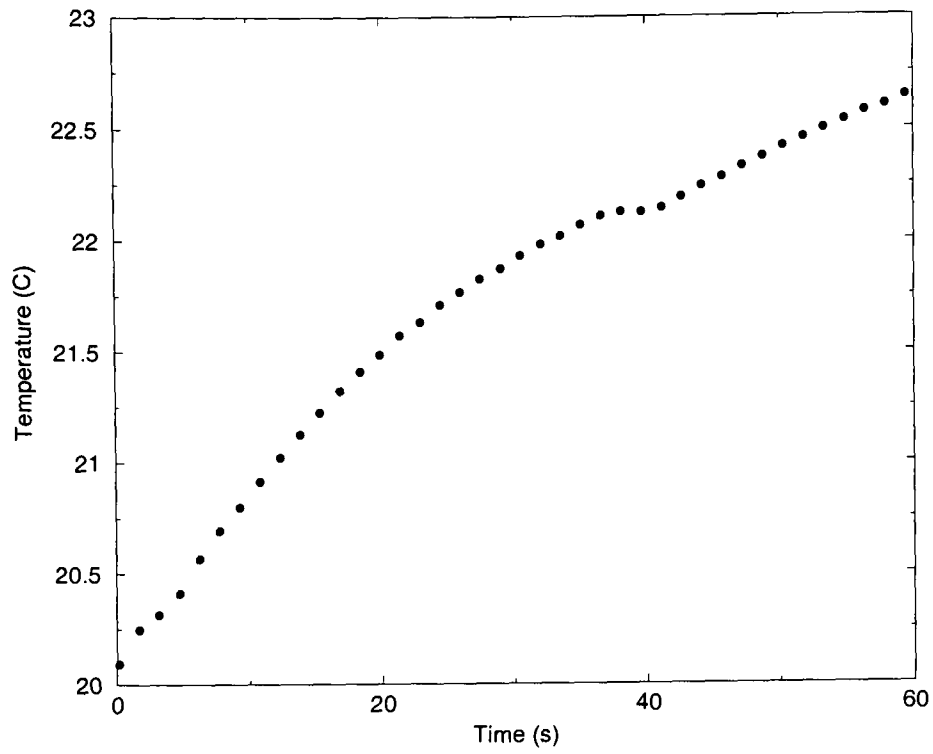


Figure 5.13: Temperature change in the HOD signal during Figure 5.12(a). The temperature conversion was performed using Equation 5.4.

5.3 Last thoughts / Conclusions

The results presented in this chapter show that photo-CIDNP is a powerful technique for monitoring the folding of proteins in real-time. It allows side chain-specific information to be obtained. This has also been clearly demonstrated in previous studies such as those on hen egg white lysozyme described in Chapter 4, and the refolding of the histidine containing phosphocarrier protein from *E. coli* [33].

In order to obtain more detailed information about the refolding of RNase A, and separate the spectral overlap in the tyrosine H3,5 region it may be possible to selectively "block" tyrosine residues, for example:

$$-\text{TyrOH} \xrightarrow{\text{MeI/K}_2\text{CO}_3} -\text{TyrOMe} \tag{5.5}$$

by forming *O*-methylated tyrosines (Equation 5.5) which are not observed in the CIDNP experiment [60]. The possibility of selectively removing signals from dif-

ferent tyrosine residues without significantly disrupting the structure of the protein would allow the relative contributions from Tyr 76 and Tyr 92 to be resolved. This approach would also allow the identity of the intermediate observed to be elucidated.

Chapter 6

Photochemically Induced Conformational Change: The LOV2 Domain of Phototropin from *Avena sativa*

Phototropins are plant serine/threonine kinases which act as photoreceptors for a wide range of processes such as phototropism, light induced chloroplast movement and stomatal opening [212]. These proteins are found in a wide variety of plants such as *Arabidopsis*, oats, corn and rice [213]. The phototropin proteins contain three folded domains, consisting of two PAS[1] domains designated LOV1 and LOV2, for Light-Oxygen-Voltage sensing [214], and a C-terminal kinase domain [215].

PAS domains are involved in a wide variety of signalling pathways including light sensing, circadian rhythm regulation and hypoxia response pathways [216] and are highly conserved across a large range of organisms [214, 217, 218]. The LOV domains of the phototropins are a subset of this family and are involved in the

[1]Per-Arnt-Sim: named for *Drosophila* PERiod, vertebrate ARyl hydrocarbon receptor Nuclear Translocator, and *Drosophila* SIngle-Minded.

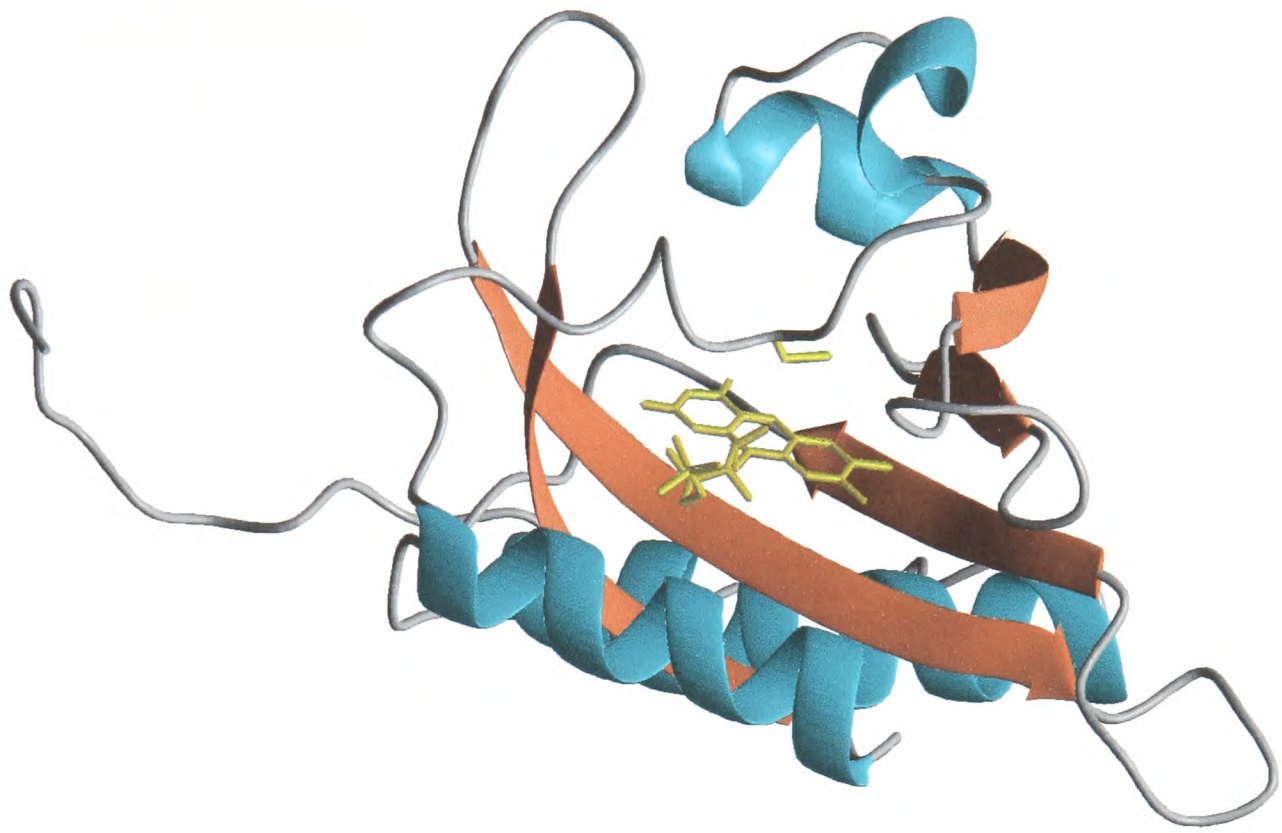


Figure 6.1: The structure of the dark state of the LOV2 domain from *Avena sativa*. The structure is based on a homology model with the LOV2 domain from *A. capillus-veneris* (PDB file:1G28 [223]). The FMN cofactor is shown in the centre of the protein. The figure was drawn using MOLMOL [89].

regulation of photoreceptor function [212] by controlling the autophosphorylation of the C-terminal kinase as a response to blue light [212, 219–222].

This chapter considers the photochemically induced changes in the second of the PAS domains, LOV2. Previous studies have shown this system to have a reversible photocycle [220] on timescales readily amenable to study by NMR. The structure of the second LOV domain in phototropin from *Avena sativa*, the LOV2 domain, is shown in Figure 6.1. This structure is based on a homology model with the LOV2 domain from *Adiantum capillus-veneris* [36]. The protein consists of 104 amino acid residues in the folded domain, spanning residues 412 to 516 of the complete phototropin sequence. The structure consists of a large five stranded β-sheet protein binding region, a series of smaller α-helices and loop regions. An α-helix, termed the Jα helix, is located outside the folded LOV domain, and lies along the β-sheet region [36]. A single molecule of flavin mononucleotide is bound non-covalently in the

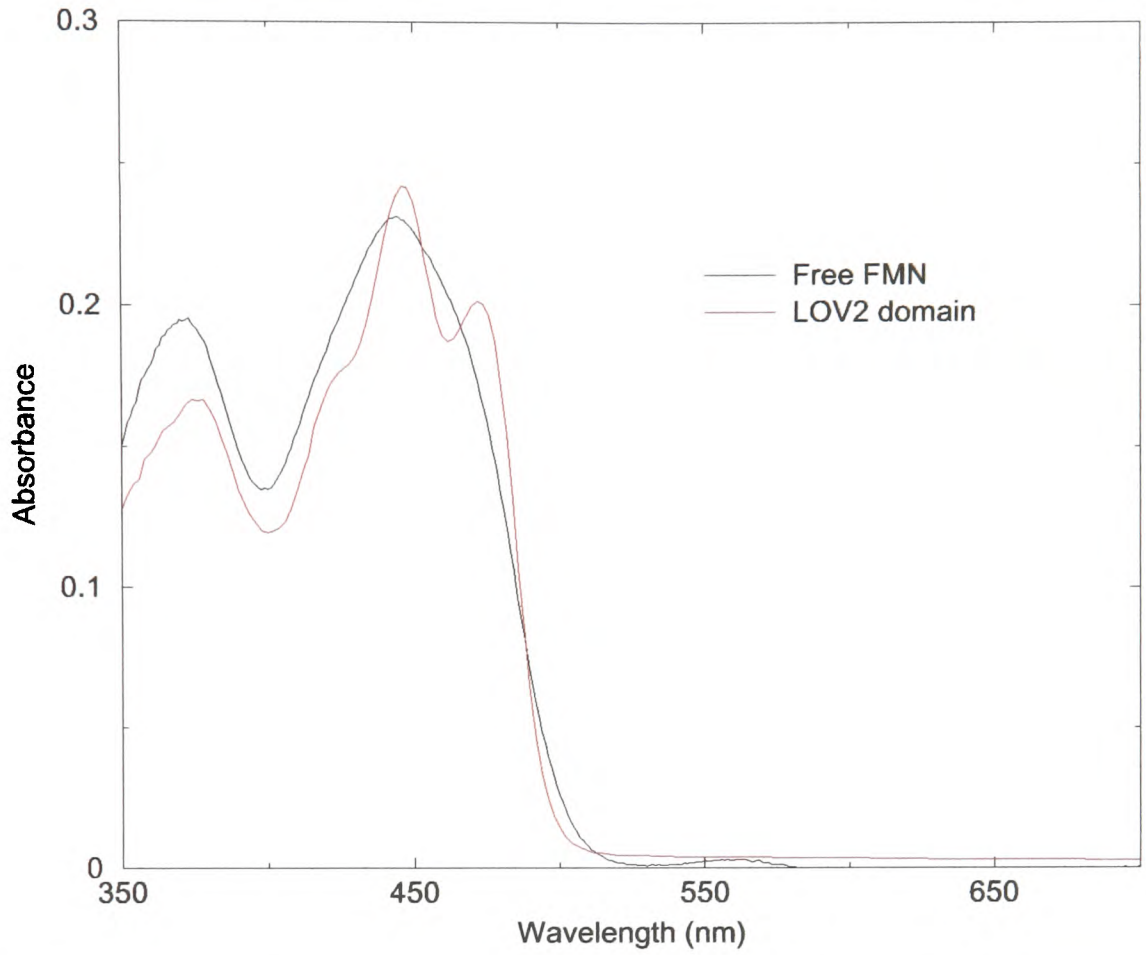


Figure 6.2: UV/visible absorption spectra of the LOV2 domain of oat phototropin and free FMN solution. Both were recorded at a concentration of 20 μM.

core of the protein [224]. Figure 6.2 shows the UV/visible absorption spectra of the LOV2 domain compared with that of free FMN in solution. It is this FMN molecule which is thought to be the source of the protein activity [36, 221, 224]. Other proteins are also known to have a reversible photocycle, for example the Photoactive Yellow Protein (PYP) from *Ectothiorhodospira halophila* uses 4-hydroxycinnamic acid as its chromophore [225, 226].

Saloman *et al.* have proposed that illumination of the protein with blue light causes the formation of a covalent link between the bound flavin and a cysteine residue in the core of the protein [221]. This cysteinyl-flavin adduct distorts the protein structure [227] and causing changes in the ^{13}C NMR spectrum of an isotopically labelled FMN cofactor [221] and in the protein NMR spectrum [36]. This

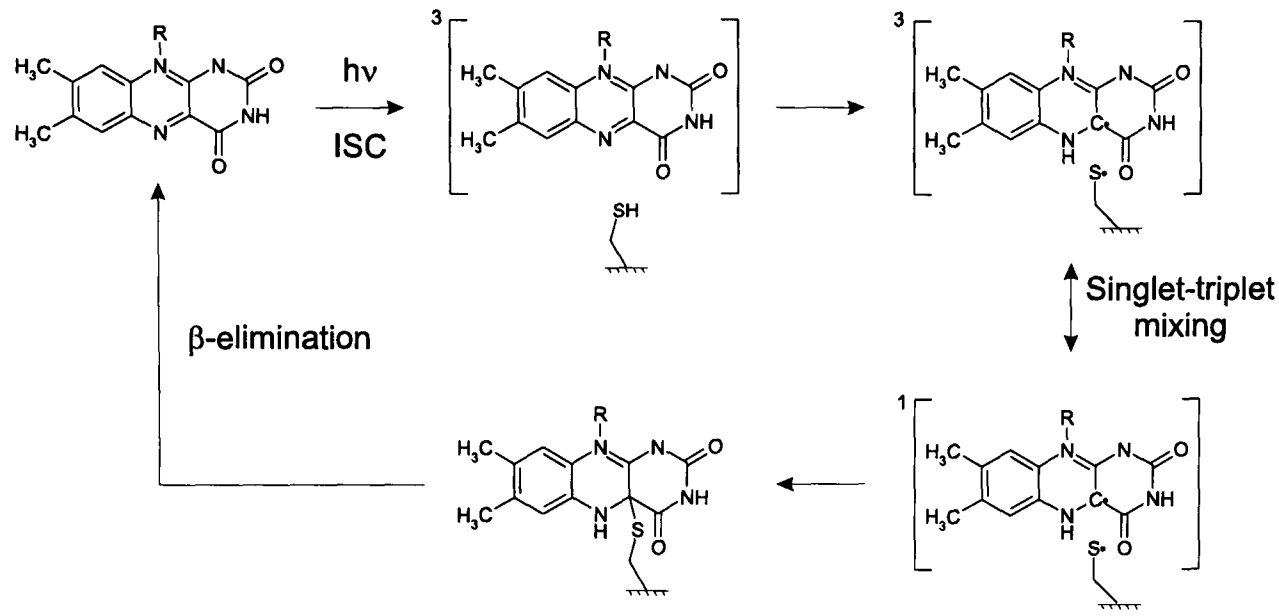


Figure 6.3: The proposed mechanism of attack of the flavin C(4a) site by cysteine-450 upon illumination of the protein with blue light [213].

mechanism has since been elucidated by EPR[2] and ENDOR[3] spectroscopy of the wild-type protein and a C450A mutant in which the cysteine residue has been replaced with alanine [213]. The proposed mechanism for the adduct formation is shown in Figure 6.3. The reaction scheme starts with absorption of a blue light photon by the FMN and as is common in flavin photochemistry, this excited state rapidly undergoes intersystem crossing to the triplet state [70]. The excited triplet flavin then extracts a hydrogen atom from the thiol of cysteine 450, and forms a triplet spin correlated radial pair. This then undergoes coherent oscillations between the triplet and singlet states, modulated by the hyperfine interactions in the radicals. Once the system is in a singlet spin correlated state it may either form the cysteinyl-flavin adduct, in the case of the wild type protein, or simply undergo a reverse hydrogen atom transfer in the case of the C450A mutant. Once this adduct is formed, the rate of return to the ground state of the protein is limited by the rate of β-elimination between the cysteine residue and the proton attached

[2]Electron Paramagnetic Resonance
[3]Electron Nuclear DOuble Resonance

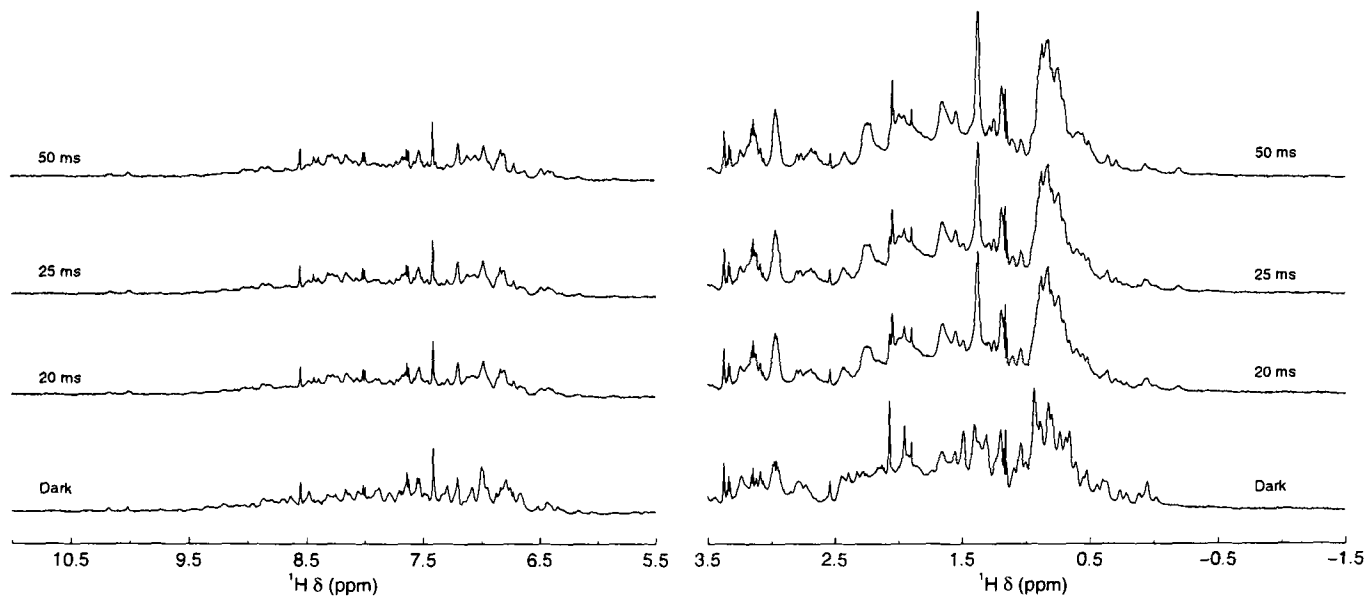
to the central ring nitrogen, N(5). An ionic mechanism has also been proposed for this reaction [228], however, this is not consistent with the EPR and ENDOR data presented by Kay *et al.* [213].

The formation of this cysteinyl-flavin adduct causes a shift in the absorption spectrum of the protein as the aromaticity of the flavin is disrupted. The return of the ground state absorption spectrum can be used to monitor the recovery of the dark state after illumination, however, this is limited to describing changes occurring around the FMN binding site. Described below are a series of experiments to determine some of the global structural changes on formation of the excited state. A new photo-CIDNP methodology is developed to study these changes and provide complimentary information to other NMR techniques. It is also of interest to study the kinetics of the decay of the adduct to the ground state.

Uniformly nitrogen-15 labelled LOV2 domain (AsLOV2) from *Avena sativa* (oats), was produced using standard bacterial expression techniques [229] by Ms. S. M. Harper of the University of Texas, Southwestern Medical Center at Dallas. The domain comprises residues 517 to 560 of the complete phototropin amino acid sequence. The protein was dissolved in 50 mM sodium phosphate and 100 mM sodium chloride and used either at pH 6 or 7 as noted when appropriate.

6.1 The Conformational Change

As outlined in the introduction to this Chapter, the LOV2 domain undergoes a reversible photocycle which has been studied using various biophysical techniques, mainly optical and EPR spectroscopies. These techniques, while giving valuable insight into the changes occurring at the photoactive centre, namely the FMN cofactor, are unable to give details on global changes occurring across the protein during the photocycle. The application of standard NMR techniques coupled with illumination



(a) The aromatic region

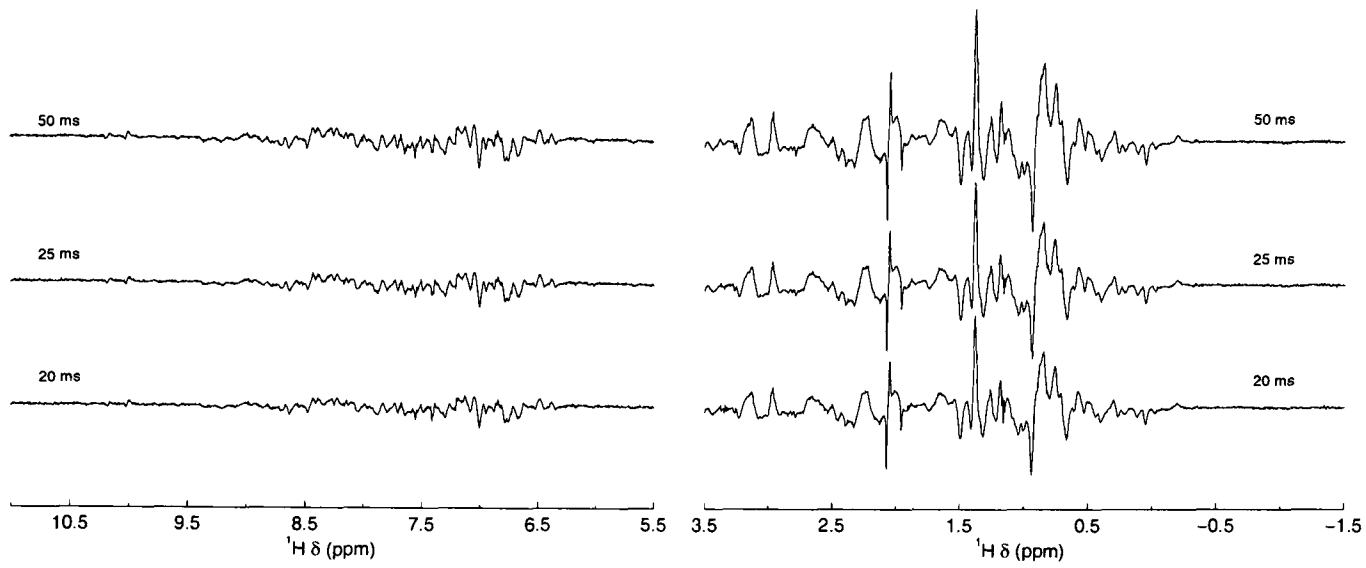
(b) The aliphatic region

Figure 6.4: 600 MHz ^{1}H NMR spectra of a 300 μM solution of AsLOV2 at pH 6. ^{15}N decoupling was applied during the acquisition of the FID.

of the protein within the NMR magnet can allow residue-specific information to be obtained. In this Section, results of initial experiments to study conformational changes in the protein structure are presented.

6.1.1 One-Dimensional Spectra

The initial experiments were performed on the LOV2 domain to determine the illumination levels requires to induce the conformational change without damage to the protein. A series of 1-dimensional NMR spectra were recorded using a low power laser pulse, 100 mW at the laser head, preceding the NMR pulse sequence. These experiments served to determine the conditions required to excite the protein and produce efficient photoconversion. This was achieved by monitoring the changes in the 1-dimensional spectra. These results are shown in Figure 6.4. It is immediately clear that the whole NMR spectrum is affected at moderate illumination times. Marked changes can be seen in the aliphatic region particularly between -0.5



(a) The aromatic region

(b) The aliphatic region

Figure 6.5: 600 MHz ^{1} NMR difference spectra of a 300 μM solution of AsLOV2 domain at pH 6. Each trace is the result of subtraction of the dark spectrum of Figure 6.4 from the corresponding illuminated spectrum.

and 2.5 ppm, and across the whole aromatic and amide region. Figure 6.5 shows difference spectra between the dark state of the protein and the spectra recorded after a short period of laser illumination. These spectra show even more clearly that there are changes across the sample. The use of short illumination times minimises the effects of sample heating due to the laser pulse.

It is obvious from these spectra that there is a large change in the protein associated with photoexcitation. Using illumination times beyond 25 ms produces little further effect. However, in order to be certain that efficient photoconversion has occurred, 50 ms was chosen as a suitable illumination time. Since the 1-dimensional spectra are crowded, more detailed information about the changes occurring can be gained by using two-dimensional ^{15}N-^{1}H correlation experiments. These are described in the following subsection.

6.1.2 ^{15}N-^{1}H Correlation Spectra

Heteronuclear single-quantum correlation (HSQC) spectroscopy is routinely used for the study of proteins [74]. It provides correlation of proton and nitrogen-15 chemical shifts of all ^{15}NH spin systems coupled through one bond in the protein.

Figure 6.6 shows the overlay of two ^{15}N-^{1}H HSQC spectra of the LOV2 domain. The black spectrum was recorded using the dark (ground) state of the protein, the second, red spectrum, recorded with a 50 ms laser pulse preceding the acquisition of each transient, hence observing the photoexcited state. Under these conditions, the lifetime of the photo-excited state is sufficient to allow the recording of the HSQC free induction decay for each transient. It is immediately clear that the changes observed in the 1-dimensional spectra are also observed in the HSQC spectra. Both the dark and light states of AsLOV2 give well resolved spectra which are not superimposable, indicating changes of the backbone H^N and side chain NH environments upon illumination. As an example, the protein contains two tryptophan residues, one from the LOV2 domain (Trp 491) and a second at the end of the linker, after the Jα helix (Trp 557). These are good choices as they are well separated from the backbone amides and can be found down field of 10 ppm in the proton dimension. The changes in chemical shift are given in Table 6.1. Similar effects are also observed for the vast majority of the backbone amides.

Trp residue	LOV2 (W491)		Linker (W557)	
	^{1}H	^{15}N	^{1}H	^{15}N
Dark	10.11	127.99	10.27	129.29
Light	10.09	128.68	10.08	128.99
Difference	−0.02	+0.69	−0.19	−0.30

Table 6.1: Changes in the tryptophan indole chemical shifts between the dark and light states of LOV2. The difference is quoted as light minus dark.

It is interesting to compare these results with those obtained from the same set

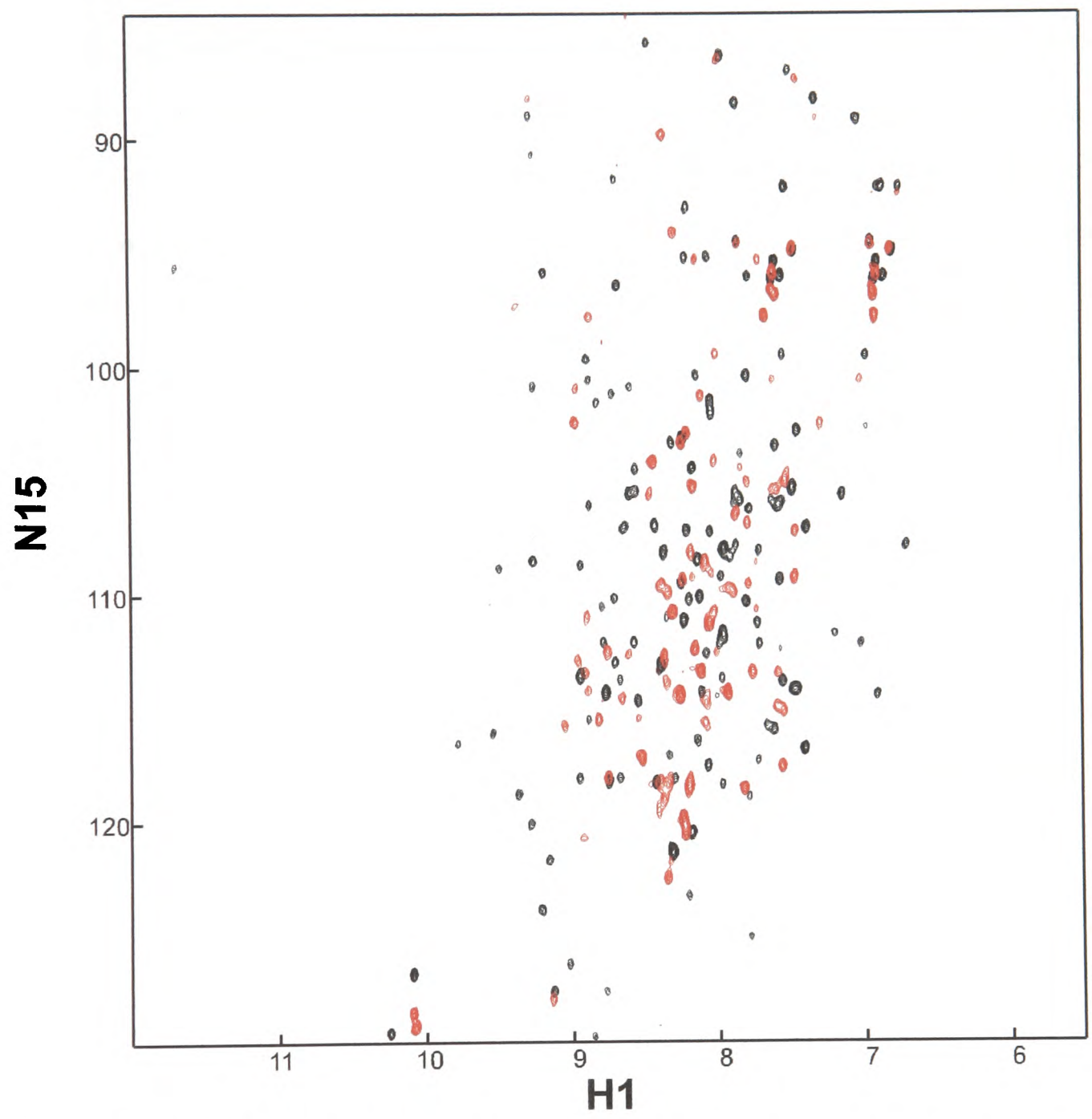


Figure 6.6: 600 MHz ^{15}N-^{1}H HSQC of AsLOV2 at pH 6. The black spectrum is the dark state and the red spectrum is the photoexcited state, recorded using 50 ms laser illumination before each transient. The spectra were recorded using a spectral width of 10 kHz, 512 complex points in the directly detected dimension (^{1}H) and a spectral width of 2777 Hz, 64 complex points in the indirectly detected dimension (^{15}N). Eight transients were recorded per t_1 increment. Shifted sine-squared window functions were used in both dimensions.

of experiments performed on a shorter construct of the LOV2 domain, termed the AsLOV2Δ construct. This protein has the Jα helix and C-terminal linker truncated, residues 522 to 560 are removed. Figure 6.7 shows overlayed light and dark HSQC spectra of this version of the protein. The differences between the ground and photo-excited states are much less striking than those seen in Figure 6.6. Only minor changes in chemical shifts are observed for a large number of the backbone H^N and side chain NH resonances. Again, looking at the now single tryptophan indole signal (Trp 491), the changes observed in the chemical shift are much smaller. These changes are of the order of a hundredth of a ppm, and are given in Table 6.2. These small scale changes observed for the backbone amides and side chain NHs

Trp residue	LOV2 (W491) ^{1}H	^{15}N
Dark	10.12	128.63
Light	10.11	128.67
Difference	−0.01	+0.04

Table 6.2: Changes in the tryptophan indole chemical shift between the dark and light states of the short construct protein. The difference is quoted as light minus dark.

indicate that, although less widespread, there are still changes in this construct upon photo-excitation.

6.1.3 Interpretation

From the results of the 1-dimensional spectra shown in Section 6.1.1 and the ^{15}N-^{1}H HSQC spectra from Section 6.1.2 it is clear that there are wide ranging changes in the amide environments, and proton environments more generally, resulting in dramatic changes in the observed chemical shifts. These changes are much more striking for the complete LOV2 domain, retaining the Jα helix, compared to the truncated construct, AsLOV2Δ. These changes can be interpreted as conformational changes

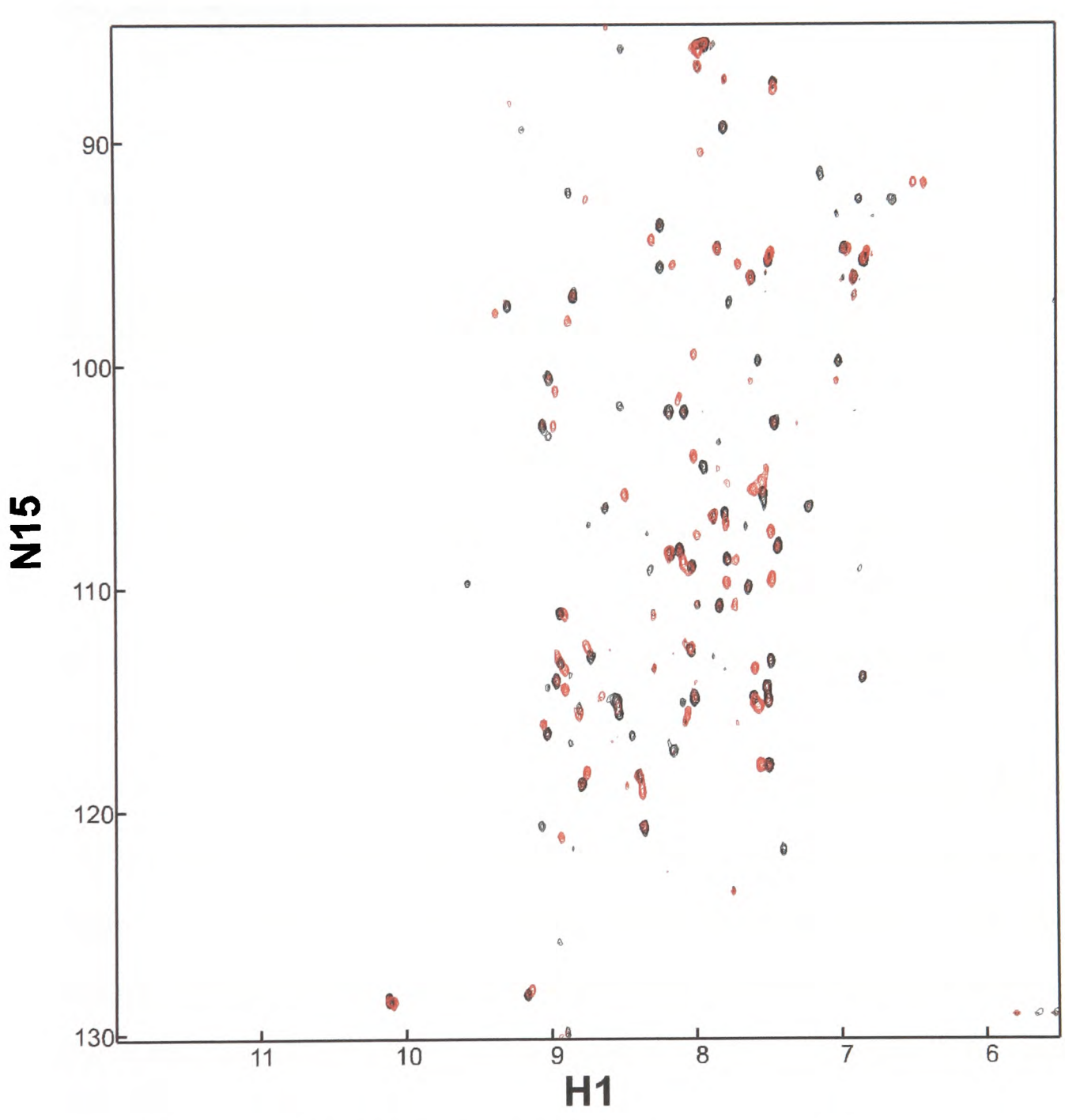


Figure 6.7: 600 MHz ^{1}H-^{15}N HSQC of AsLOV2Δ (the shorter construct) at pH 7. The black spectrum is the dark state and the red spectrum is the photoexcited state, recorded using 50 ms laser illumination before each transient. The acquisition parameters remained the same as for Figure 6.6.

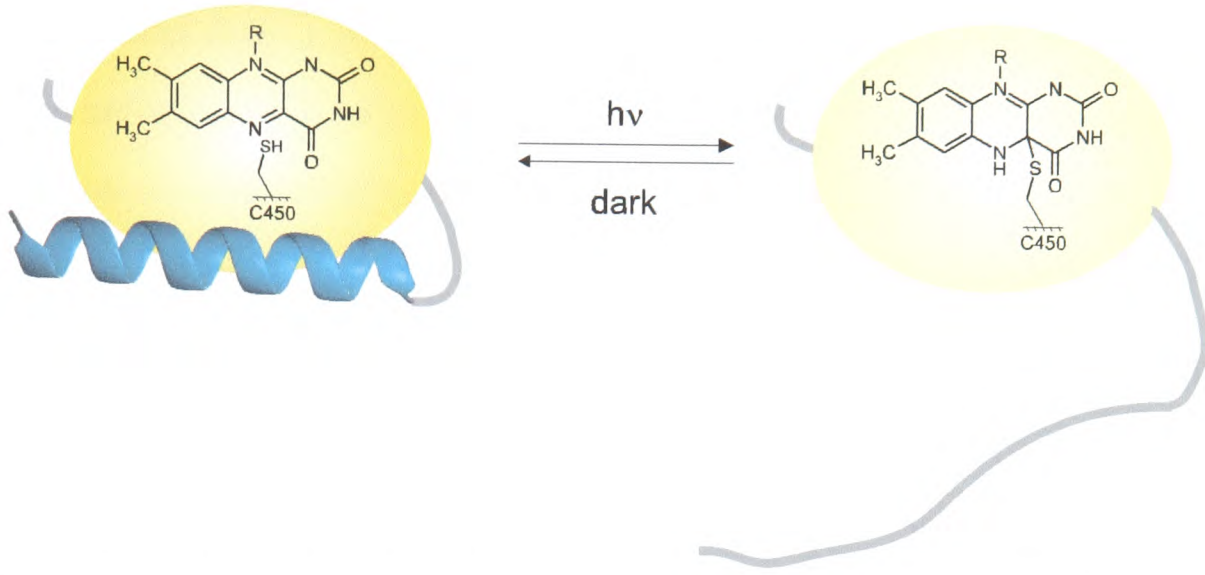


Figure 6.8: An interpretation of the structural changes occurring in the LOV2 domain upon illumination [36]. The folded LOV2 domain is represented by the yellow oval.

occurring in the protein as a result of the formation of the cysteinyl-flavin adduct upon exposure to blue light. As outlined in the introduction to this Chapter, there is a five-stranded β-sheet region of the domain which can act as a helix binding site. In the ground state of the complete construct the Jα helix, and C-terminal linker can fold back over onto this region (see Figure 6.1).

When this conformational change is triggered by the formation of the flavin adduct, the Jα helix and linker are released from the β-sheet and unfold and along with changes in the tertiary structure of the LOV2 domain, result in the changes observed in the NMR spectrum. These changes have been elucidated in more detail by Gardner and co-workers using a wide range of NMR techniques [36], and are shown schematically in Figure 6.8.

The folding of the linker and the Jα helix onto the protein binding region can also provide a tentative explanation as to the light-mediated activation of phototropins. As described above, the activity of the C-terminal kinase is controlled in response to blue light [212, 219, 221]. Therefore if, in the ground state, the kinase is held in an orientation relative to the rest of the protein in which it is unable to undergo the autophosphorylation reaction, then when the LOV2 domain (and potentially

the LOV1 domain also) is exposed to blue light and undergoes this conformational change, the unfolding of the Jα helix will release the kinase, enabling the phosphorylation reaction to occur. This would then constitute the initial part of the signalling pathways involved in phototropism, light-induced chloroplast movement and stomatal opening [212] and may provide clues as to the rôle of other PAS domains in signal transduction [217].

6.2 Photo-CIDNP Spectroscopy

Photo-CIDNP NMR spectroscopy provides a powerful method to observe solvent accessible aromatic amino acid side chains. It is particularly suited to looking at the changes in solvent accessibility following a perturbation of the sample conditions, such as in real-time refolding experiments [26, 33]. It is therefore of interest to look at changes in the solvent accessibility of the aromatic residues during the photocycle of the protein. The LOV2 domain contains a number of potentially polarisable residues of interest, there are, for example, two tryptophan residues, Trp 491 and Trp 557. The former it situated at the junction of the linker to the main folded domain, and as such may alter its environment significantly upon photoactivation of the protein. The domain also contains three tyrosine residues, Tyr 440, Tyr 483 and Tyr 508. These are highlighted in Figure 6.9 along with the FMN cofactor.

6.2.1 A New Dye System

The traditional CIDNP experiment, in which the nuclear polarisation is generated using FMN as the photosensitiser, is not readily applicable to the study of the LOV2 domain. The reason for this is that it is impossible to generate triplet state free FMN to sensitise the aromatic residues, while maintaining the protein in the ground state. Therefore, in order to perform these CIDNP experiments, a different

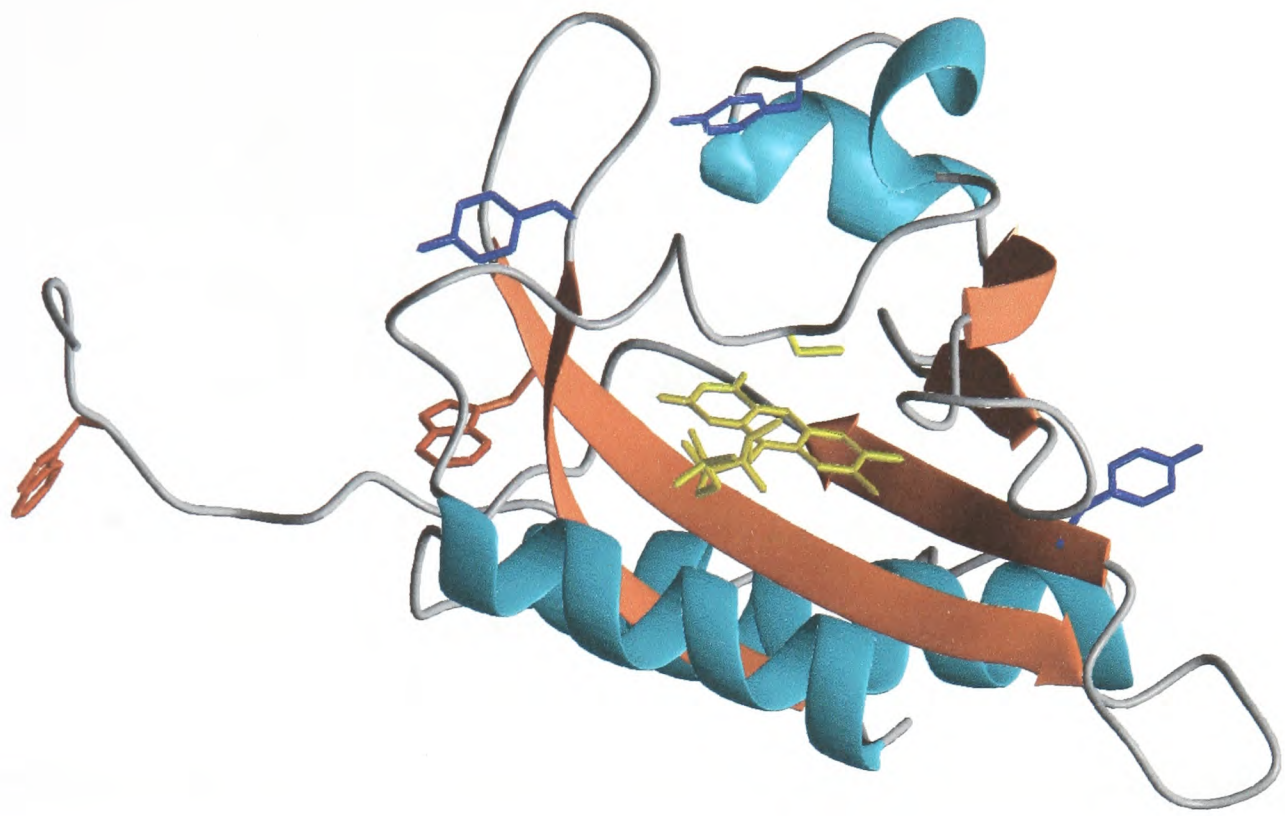


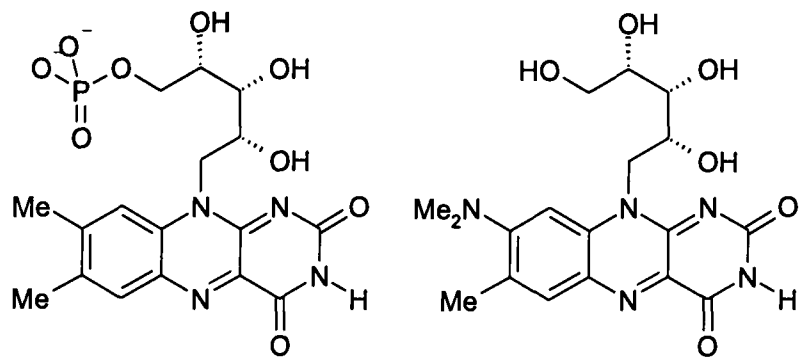
Figure 6.9: The structure of the dark state of the LOV2 domain from *Avena sativa* (Figure 6.1). The tryptophan resides are shown in red, and tyrosine in blue. The FMN cofactor is shown in the centre of the protein. The figure was drawn using MOLMOL [89].

dye system must be developed.

Several factors affect the choice of a suitable dye for photo-CIDNP experiments. These include significant solubility and a large optical absorption at the wavelength to be used to generate the nuclear polarisation. This means that only a small amount of dye is required, hence reducing the possibility of any deleterious interactions between the protein and photosensitiser. The chosen dye also needs to be capable of either an electron transfer or hydrogen atom abstraction in order to generate the radical pair, and hence produce nuclear spin polarisation. To this end, two different dyes have been studied. These are a flavin based dye, roseoflavin, and the thiazine compound thionin.

6.2.1.1 Roseoflavin

As flavins are widely used as photosensitisers in CIDNP experiments [55, 60], this was considered the best starting point. From a review of riboflavin analogues [230],



(a) FMN (b) Roseoflavin

Figure 6.10: The structures of FMN and roseoflavin

two possible compounds were identified, roseoflavin and 1-deazariboflavin. Both of these compounds have visible absorption spectra that are significantly red-shifted compared to riboflavin [230–232]. The first of these compounds was purchased from Toronto Research Chemicals, Canada, and used as obtained. 1-deazariboflavin is not commercially available and the synthetic route to this compound is non-trivial [233]. Therefore roseoflavin was considered to be the best choice.

Figure 6.10 shows a comparison of the structures of FMN and roseoflavin. As can be seen from this, the major difference is at the 8-position, where one of the methyl groups of FMN is replaced by an N,N-dimethylamine group. The presence of this electron donating group red-shifts the absorption spectrum, which is shown in Figure 6.11. The lack of the 5′-phosphate group has the effect of greatly reducing the solubility of the roseoflavin compared to FMN. The UV/visible absorption spectrum shows that roseoflavin has a broad absorption band between 450 and 550 nm, resulting in significant absorption at 514 nm and also at 488 nm, the principal wavelengths emitted by the argon ion laser. The molar extinction coefficients, for both FMN and roseoflavin at these wavelengths are tabulated in Table 6.3. At 488 nm roseoflavin has approximately twice the absorption of FMN, while at 514 nm the absorption is even larger. Since at 514 nm FMN is effectively transparent, it is this

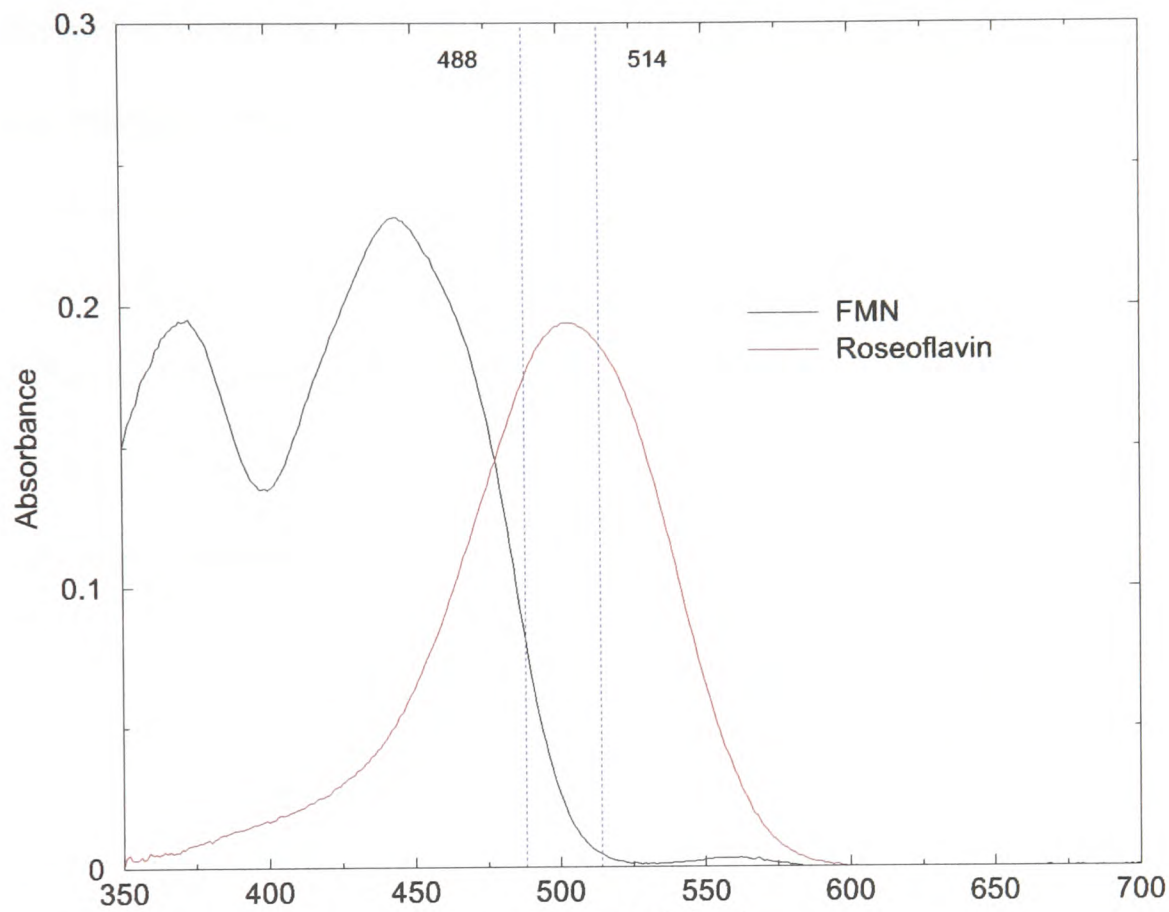
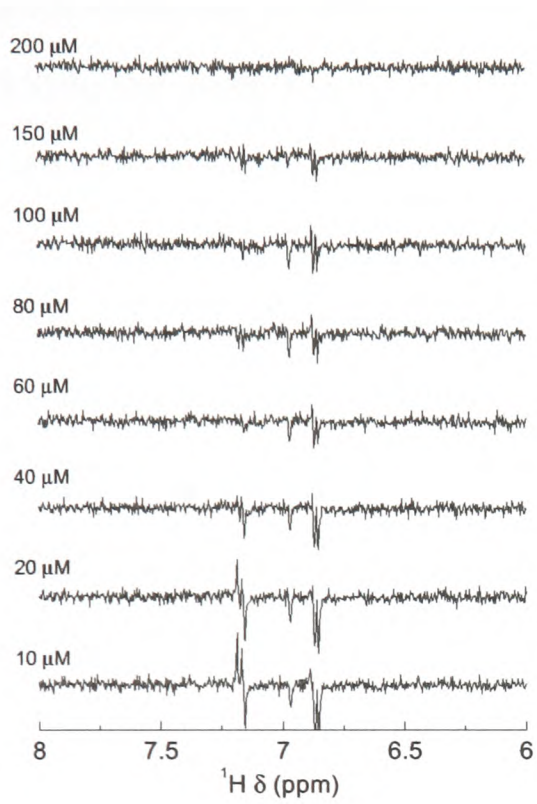


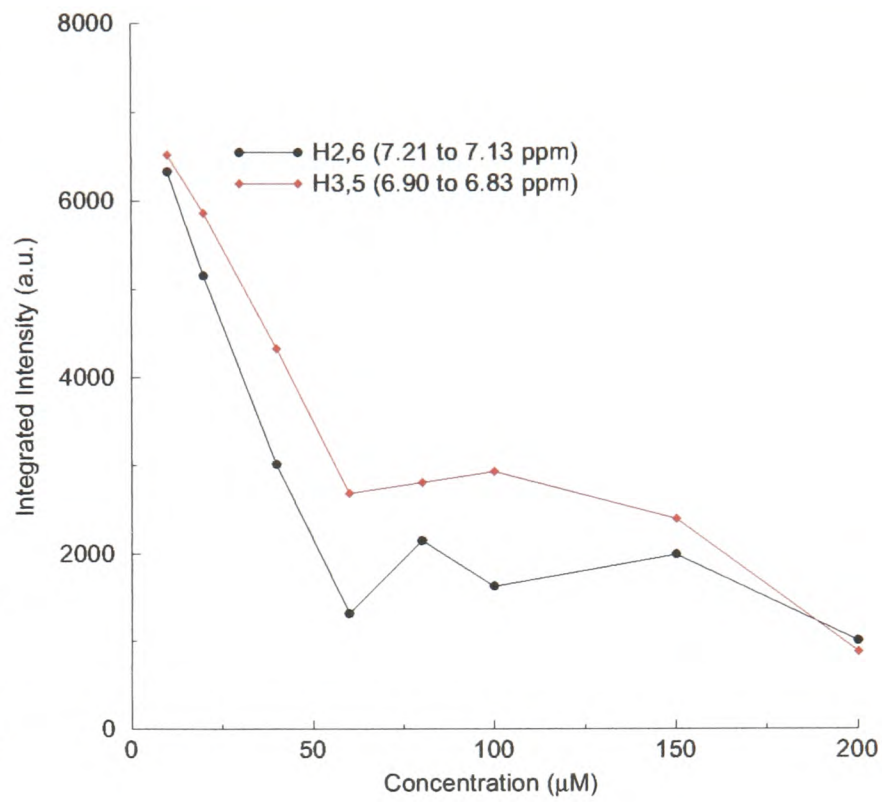
Figure 6.11: UV/visible absorption spectra of 20 μM solutions of FMN and roseoflavin in D_2O. The blue dashed lines indicate the principal wavelengths emitted by the laser.

Dye	Extinction Coefficient	
	488 nm	514 nm
FMN	4082.5	236.1
Roseoflavin	8742.5	9313.0

Table 6.3: Molar extinction coefficients ($dm^3mol^{-1}cm^{-1}$) at 488 and 514 nm for FMN and roseoflavin, calculated from the spectra in Figure 6.11.



(a) Tyr CIDNP spectra with varying dye concentration



(b) Tyr Integrated signal intensity

Figure 6.12: 500 MHz ^{1}H CIDNP spectra of 2 mM tyrosine using various concentrations of roseoflavin as a photosensitiser. The spectra were recorded using a 100 ms laser pulse and averaged over 16 scans. The laser was operated in multiline mode.

absorption which is of interest.

Figure 6.12(a) shows the CIDNP spectrum of tyrosine as a function of roseoflavin concentration. These spectra indicate that nuclear polarisation is only generated at low dye concentrations. There are two possible explanations for this observation. The first is that there is no radical pair formation at high dye concentration, resulting in no observable CIDNP. The change in the molecular structure of roseoflavin, relative to FMN or riboflavin, brings about changes in the reduction potential as well as in the absorption spectrum. The reduction potential for roseoflavin is 6 mV more negative than that for FMN [230]. It is conceivable that this difference disfavours the initial electron transfer reaction required to generate the radical pair intermediate [59]. Fritz *et al.* quote the triplet yield for roseoflavin as 0.0002, measured by

photoreduction of EDTA, with the corresponding value for FMN being 0.42 [234], which also lends support to this argument. However, the generation of polarisation at lower dye concentration indicates that the high optical density of the sample is also a contributing factor. Similar results are obtained using laser illumination only at 514 nm.

Figure 6.12(b) shows the integrated signal intensity for the H2,6 and H3,5 protons of tyrosine as a function of roseoflavin concentration. The integrated intensity was calculated using a modified trapezium rule sum, as given in Equation 6.1.

$$A = \sum_{k} \left| \frac{1}{2} \left(I_k + I_{k+1} \right) \left(\delta_k - \delta_{k+1} \right) \right| \tag{6.1}$$

where I_k is the intensity of point k, with chemical shift δ_k. This form of integration is used in order to account for the fact the the spectra consist of both absorptive and emissive signals, and provides a reasonable guide to the concentration dependence of the signal intensity.

From these curves it might be expected that the CIDNP intensity would increase further as the dye concentration is lowered. However, a problem soon arises in the fact that the number of triplet state flavins being produced is lower, hence the resulting polarisation generated is smaller. This results in a trade-off between the optical density of the sample and the number of triplet flavins produced in the RF coil region. Therefore, in the following experiments 10 μM roseoflavin was used as a suitable photosensitiser concentration.

6.2.1.2 Thiazines

Previous work by Lopez [235] and Carter [236] highlighted the thiazine class of compounds as possible photosensitisers for the photo-CIDNP experiment. Examples of this class of compounds are the pH indicator methylene blue and a homologue thionin. The structure of thionin is shown in Figure 6.13. As with the flavins, the

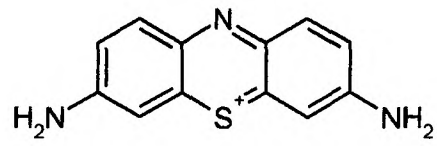


Figure 6.13: The structure of thionin, usually obtained as the acetate salt.

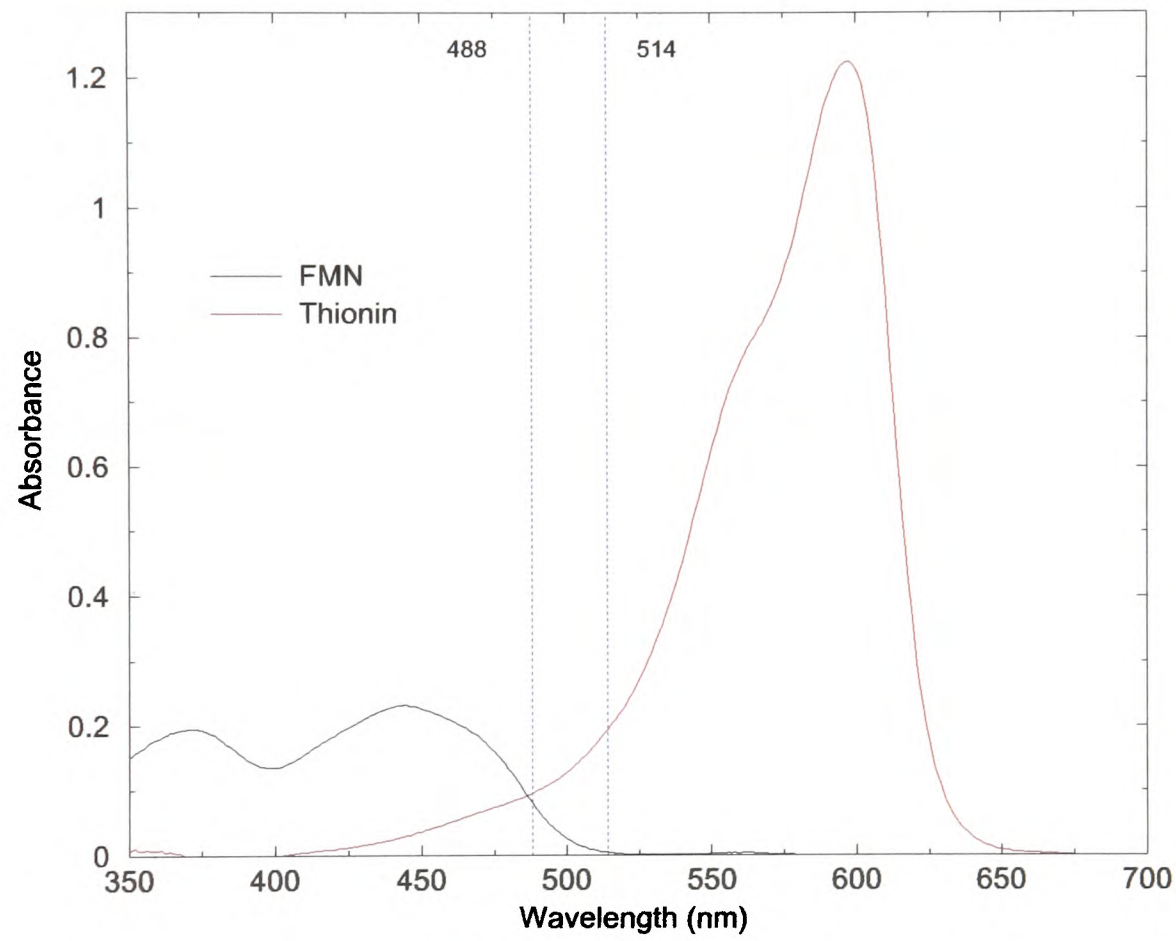


Figure 6.14: UV/visible absorption spectra of 20 μM solutions of FMN and thionin in D_2O. The blue dashed lines indicate the principal wavelengths emitted by the laser.

thiazines are based on tricyclic aromatic heterocyclic structures, resulting in these compounds being deeply coloured. The absorption spectrum is shown in Figure 6.14. At the wavelengths of interest, thionin shows strong absorption, being comparable to FMN at 488 nm and significantly increased at 514 nm. Table 6.4 shows the extinction coefficients for thionin and FMN at the wavelengths of interest.

Carter and Lopez both showed that nuclear polarisation could be generated in tyrosine and tryptophan residues using thionin. The optimum concentration of thionin was shown to be 0.05 mM [236]. No appreciable polarisation was observed with histi-

Dye	Extinction Coefficient	
	488 nm	514 nm
FMN	4082.5	236.1
Thionin	4781.0	9692.0

Table 6.4: Molar extinction coefficients ($dm^3mol^{-1}cm^{-1}$) at 488 and 514 nm for FMN and thionin, calculated from Figure 6.14.

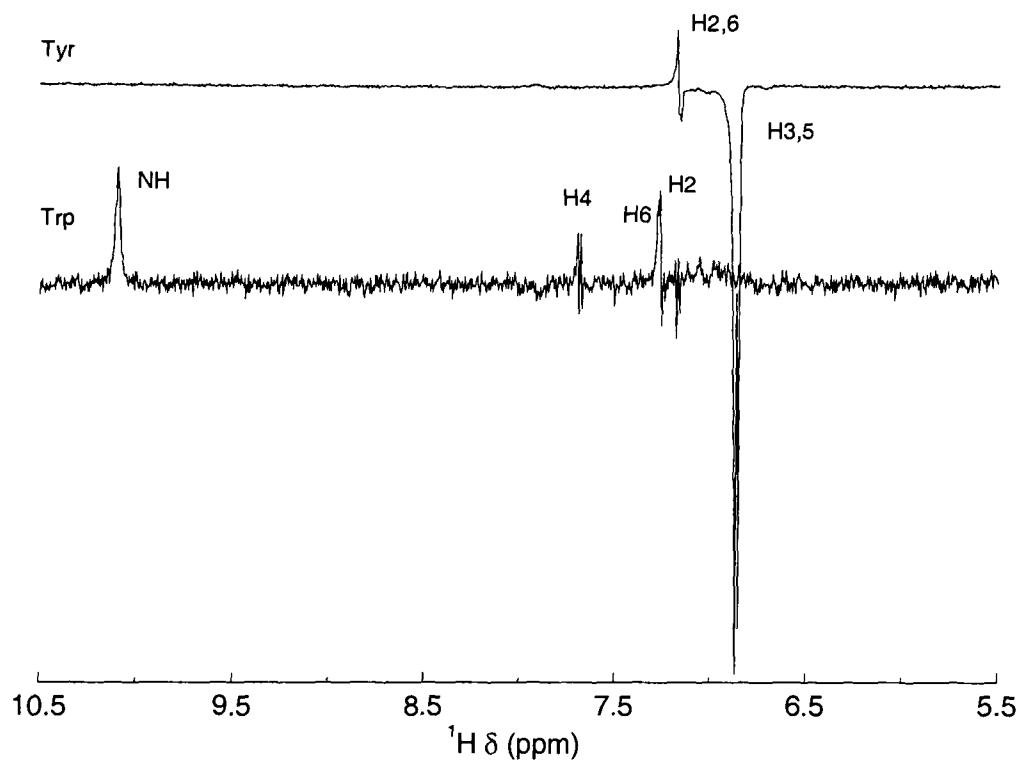
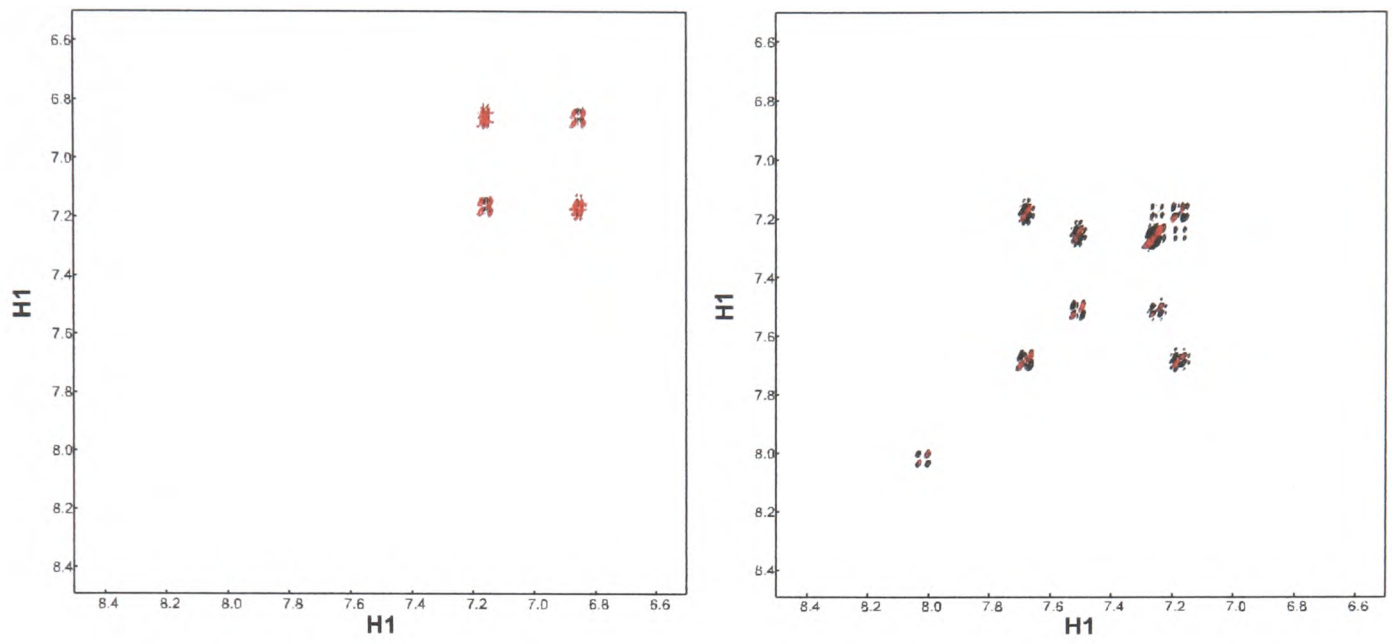


Figure 6.15: 500 MHz CIDNP spectra of 4 mM tyrosine and 2 mM tryptophan using 0.05 mM thionin. The spectra were recorded using 100 ms laser illumination and averaged over 16 acquisitions. The laser was operated in multiline mode.

dine. CIDNP spectra of tyrosine and tryptophan, sensitised using thionin, are shown in Figure 6.15. As can be seen from this, the generation of nuclear polarisation is quite successful using thionin. However, some caution must be exercised. There are a series of small signals seen in the CIDNP spectrum of tryptophan between 6.7 and 7.2 ppm. These are believed to correspond to photo-products formed by reaction of the tryptophan with the excited thionin. The principal product is thought to be kynurenine [237], formed by a complex series of oxidations and rearrangements, initiated by the formation of singlet oxygen [238].



(a) Tyrosine (b) Tryptophan

Figure 6.16: 500 MHz ^{1}H COSY spectra of 4 mM tyrosine and 2 mM tryptophan solutions containing 0.05 mM thionin. The black spectra were recorded before laser irradiation, with the red spectra taken after. All spectra were recorded using 4096 complex points in the direct dimension and 512 points in the indirect dimension. Spectral widths of 8000 Hz and sine-bell window functions were used in both dimensions. Laser irradiation consisted of 16 pulses of 100 ms duration at a power level of 5 W, in multiline mode.

In order to gain an understanding of the extent of this photo-damage, COSY spectra were recorded using tyrosine and tryptophan, both before and after the CIDNP spectra shown in Figure 6.15. The results of these experiments are shown in Figure 6.16. As can clearly be seen from this there is little change in the tyrosine spectra, shown in Figure 6.16(a), which is expected since there was no evidence of any reaction products in the one-dimensional CIDNP spectrum (Figure 6.15). However, there are obvious changes in the tryptophan spectra shown in Figure 6.16(b). There is a clear marked decrease in the intensity of both the diagonal peaks and the COSY cross-peaks, indicating that there is less amino acid present in the sample. Nakagawa *et al.* have presented a series of possible products for the photochemically sensitised oxidation of tryptophan [237]. Thus it is extremely difficult to determine what compounds are responsible for the changes in the COSY spectrum based on

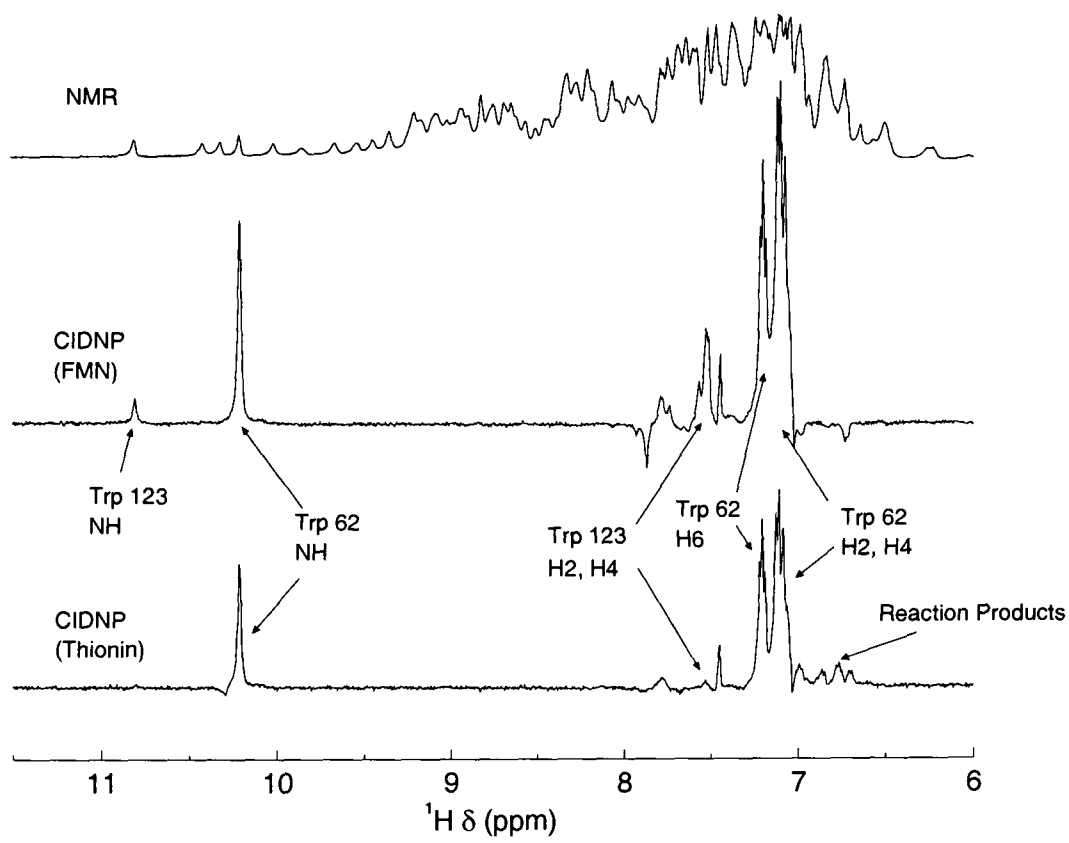


Figure 6.17: 500 MHz NMR (32 scans) and CIDNP (16 scans) spectra using 0.2 mM FMN and 0.05 mM thionin as dyes. Spectra were acquired using a laser pulse length of 100 ms, in multiline mode. The assignments were taken from [141].

this evidence alone. The point which is clear, however, is that this makes thionin a less suitable dye for the generation of CIDNP in the solvent accessible amino acid side-chains of proteins. One key requirement of the CIDNP reaction scheme is that the reaction be cyclic so as not to introduce any modification to the protein [55, 60].

As a final experiment, CIDNP spectra were recorded using the protein hen egg white lysozyme, with FMN and thionin as the photosensitisers. These spectra are shown in Figure 6.17. It is immediately obvious that there is a decrease in the signal-to-noise ratio of the CIDNP spectra recorded using thionin. Signals from Trp 123 are also missing in the thionin spectrum. This can be attributed to differing coulombic interactions between the protein and dye. As noted before, reaction products are clearly visible in the thionin CIDNP spectrum, upfield of 7.0 ppm.

From these experiments it is clear that thionin can be used to generate CIDNP with the amino acids tyrosine and tryptophan and can therefore be used to probe the

accessibility of these side chains in proteins. However, since the radical pair reaction is less ideal than in the case of FMN, side products are formed from photodegredation of the amino acid side chain. This introduces an increased level of complexity when interpreting the CIDNP spectrum. Care must be taken to distinguish peaks arising from polarised amino acids and those arising from any photoproducts.

6.2.1.3 Single-line illumination

Following on from the results shown above, which were obtained using multiline illumination, it is important to investigate the suitability of the different dyes using single line illumination at 514 nm. At this wavelength the LOV2 domain does not undergo significant photoconversion.

Figure 6.18 shows the results of CIDNP spectra obtained under similar conditions to those recorded above, but using only single line illumination at 514 nm. From this it is clear that all three dyes are capable of generating CIDNP at this wavelength. Roseoflavin generates very weak polarisation, probably due to the low triplet yield [234]. A surprising result is that FMN produces moderate polarisation in the tyrosine ring system, despite having a very low extinction coefficient at 514 nm. Thionin, however, produces extremely strong polarisation at this wavelength, of the order of four times that produced by FMN. This is in contrast to that observed under broadband illumination, where thionin gives roughly half the polarisation of FMN. As before, it is important to compare the polarisation produced in protein systems. Figure 6.19 shows CIDNP spectra of lysozyme using the three dyes of interest, illuminated at 514 nm. Clearly, the lysozyme CIDNP spectra show that there is significant polarisation in the Trp 62 residue, with very weak additional enhancement of Trp 123. The lack of Trp 123 in the case of sensitisation with thionin in Figure 6.19 can be attributed to differing coulombic interactions between the positively charge thionin and the protein as compared to the negatively charge

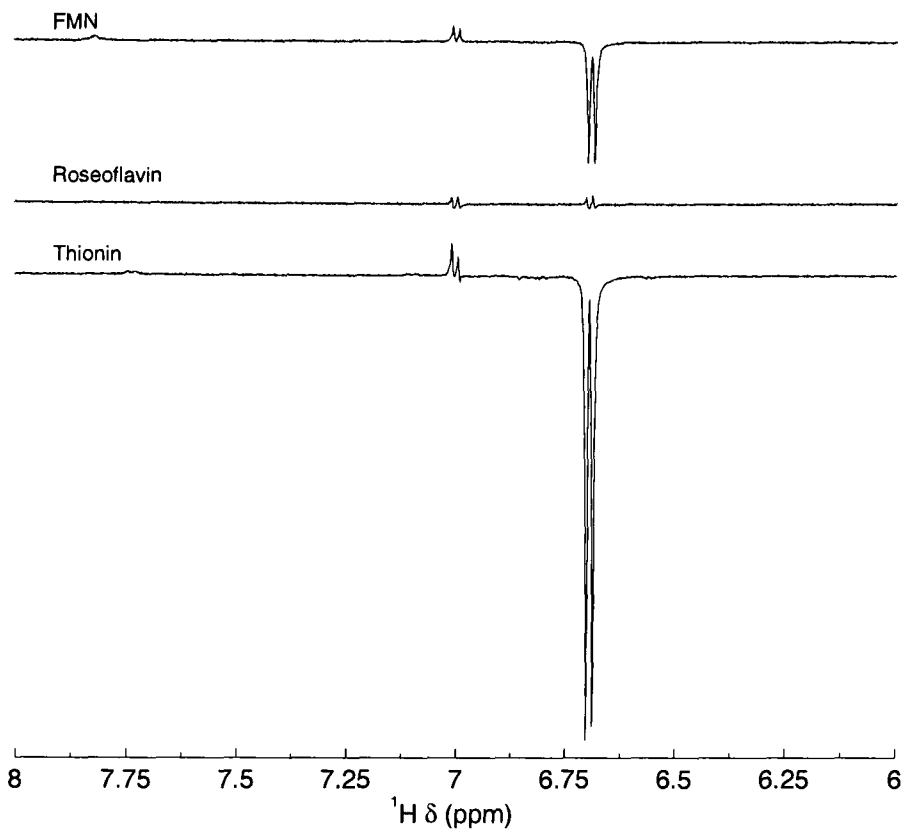


Figure 6.18: 600 MHz ^{1}H CIDNP spectra of 4 mM tyrosine using 0.2 mM FMN, 0.01 mM roseoflavin and 0.05 mM thionin. The spectra were recorded using 100 ms laser illumination at 514 nm and averaged over 16 acquisitions.

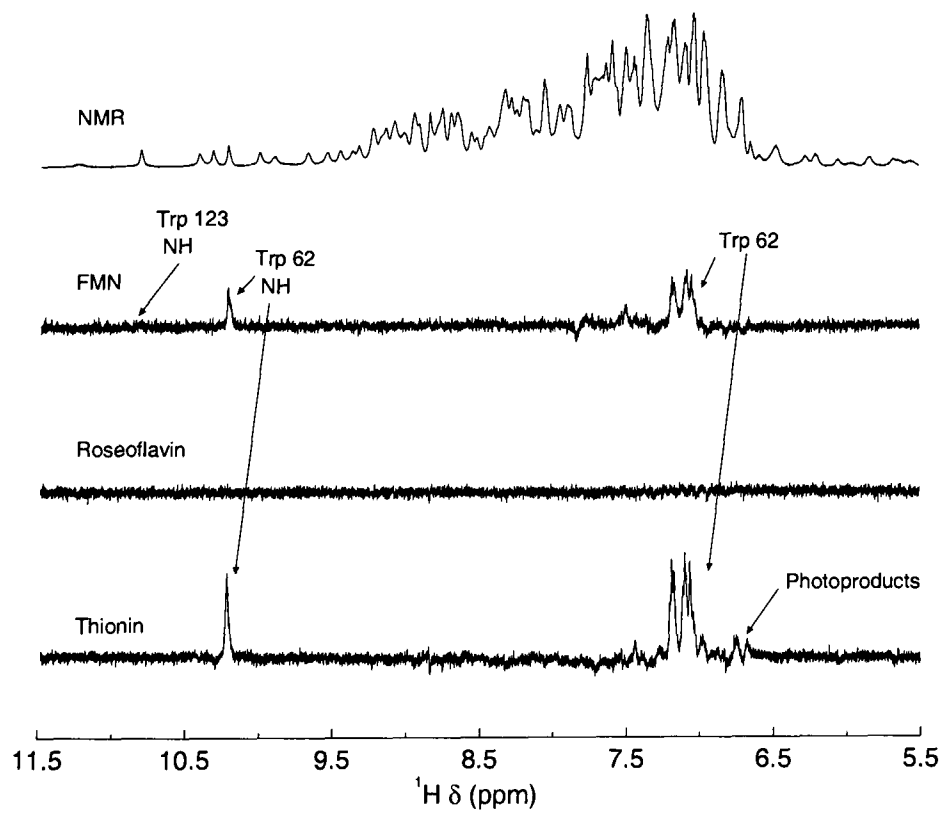


Figure 6.19: 600 MHz ^{1}H NMR (32 scans) and CIDNP (16 scans) spectra using 0.2 mM FMN, 0.01 mM roseoflavin and 0.05 mM thionin as dyes. The assignments were taken from [141].

FMN.

Based on the results of the experiments given above, it is clear that thionin is the preferred choice as photosensitiser for use in the CIDNP experiments on the LOV2 domain. The signal-to-noise ratio is greatly increased compared to the roseoflavin. While there is significant production of photoproducts, these in general occur on the high field side of the tryptophan aromatic resonances and are usually easy to identify.

6.2.2 Photo-CIDNP Spectroscopy of the LOV2 Domain

For the CIDNP experiments, two mutants of the LOV2 domain were produced, along with the wild type protein. These were a C450S mutant in which the cysteine residue which attacks the flavin upon illumination was replaced with serine. This mutant is an analogue of the dark state of the protein as it is incapable of undergoing the photocycle. The second mutant was I539E, in which isoleucine 539 was replaced with glutamate. This replacement destabilises the interaction of the Jα helix with the LOV2 domain, and so produces a protein which mimics the light state of the wild type protein, in both the appearance of its ^{15}N-^{1}H HSQC spectra and in kinase activity assays [239]. This mutant, however, is still capable of forming the cysteinyl-flavin adduct.

6.2.2.1 One-dimensional Spectra

Figure 6.20 shows CIDNP spectra of each of the wild type LOV2 domain and two mutants, sensitised with thionin and illuminated at 514 nm. This should provide a probe of tyrosine and tryptophan residues in the ground state of the protein. Firstly, it is important to note that nuclear polarisation is generated. This is evident from the enhanced tryptophan indole signals and various signals also seen in the aromatic region. However, there is also evidence of some formation of the cysteinyl-

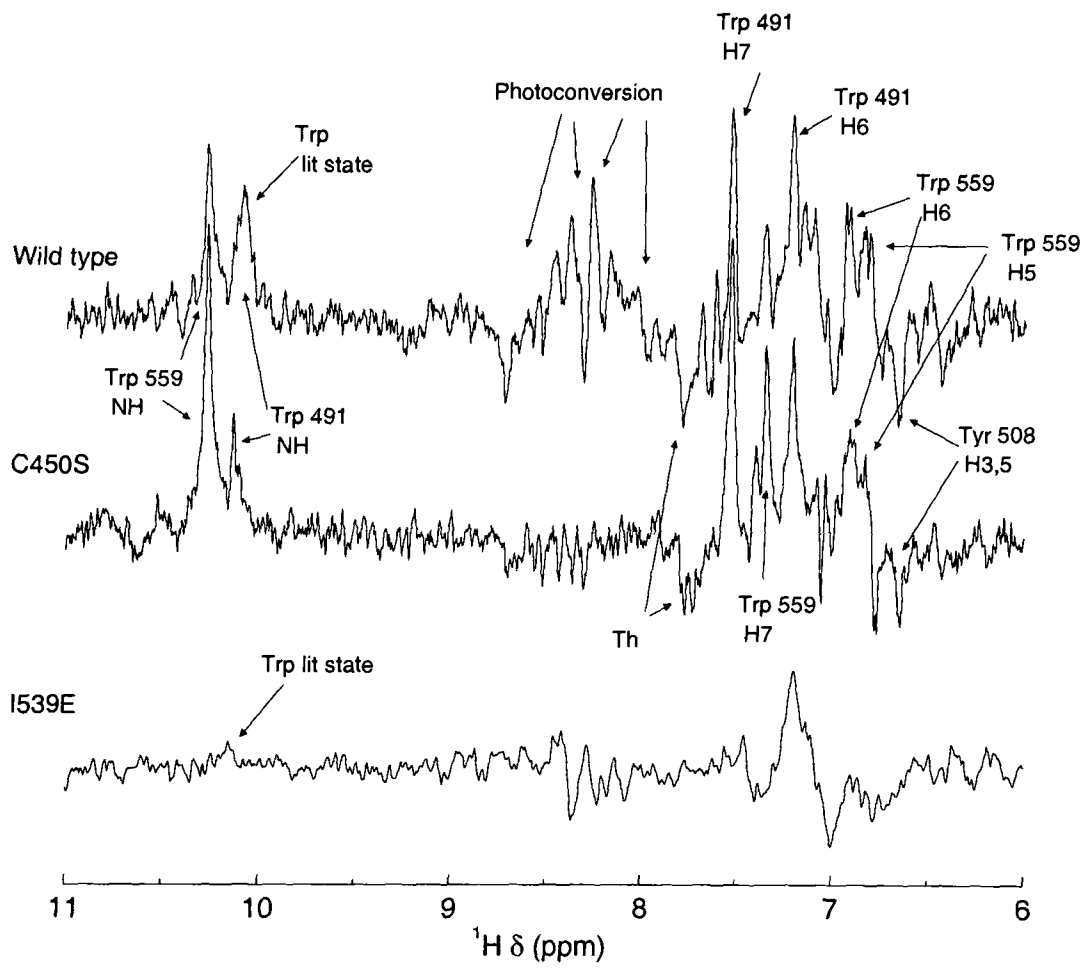


Figure 6.20: 600 MHz ^{1}H CIDNP spectra of the LOV2 domain and two mutants at a concentration of 750 μM. 50 μM thionin was used as the photosensitiser. Laser illumination was applied at 514 nm for 100 ms before each light scan. The spectra are averages of 16 light and dark pairs. The assignments are taken from the known chemical shifts of the ground state of the protein [239]. Th represents a polarised thionin signal.

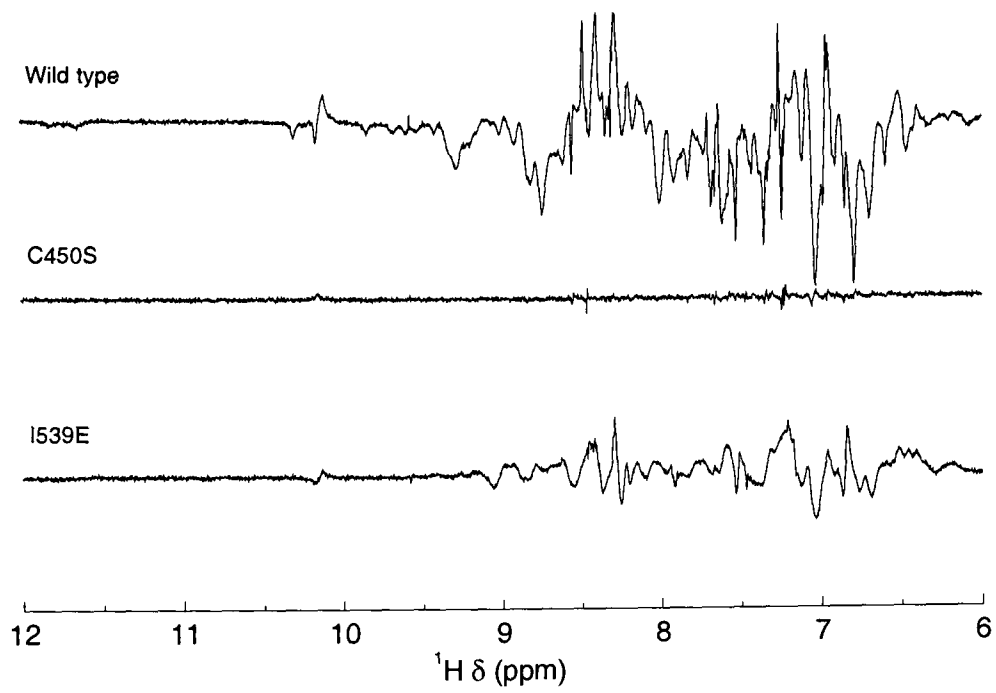


Figure 6.21: 600 MHz ^{1}H NMR difference spectra of the LOV2 domain and two mutants at a concentration of 750 μM. Each sample was illuminated at 488 nm for 200 ms before each scan.

flavin adduct, primarily from the broad signals seen between 8.0 and 8.75 ppm in the spectrum of the wild-type protein. Since the C450S mutant is incapable of forming this adduct, all of the signals in this spectrum can be assumed to arise as a result of the nuclear polarisation generated. To gain a better understanding of the photoconversion taking place in these three mutants, NMR spectra were recorded in the presence and absence of illumination (as in Figure 6.4), and the difference plotted. This is shown in Figure 6.21. There are major changes in the wild-type as expected, and also in the lit state mimic, the I539E mutant. This is due to the fact that, even though the Jα helix is no longer "docked" along the β-sheet region and it is unfolded, cysteine 450 still reacts with the FMN cofactor to form the adduct, and causes some distortion of the protein structure. The middle trace of this Figure, pertaining to the C450S mutant, confirms that there is no photoconversion taking place upon illumination. This adds support to the argument that all the signals seen in the middle trace of Figure 6.20 are truly from CIDNP. Given the fact that illumination at 514 nm does produce some formation of the cysteinyl-flavin adduct

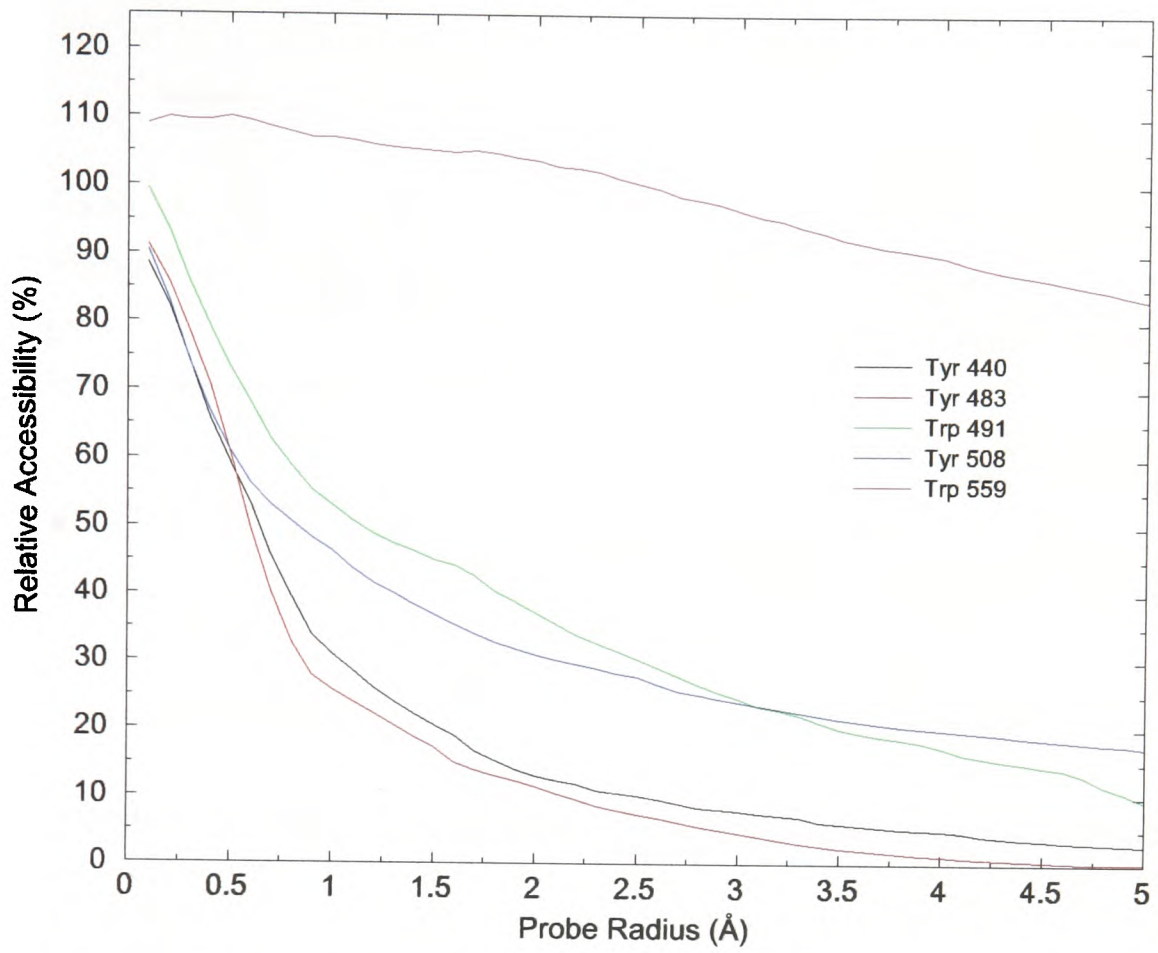
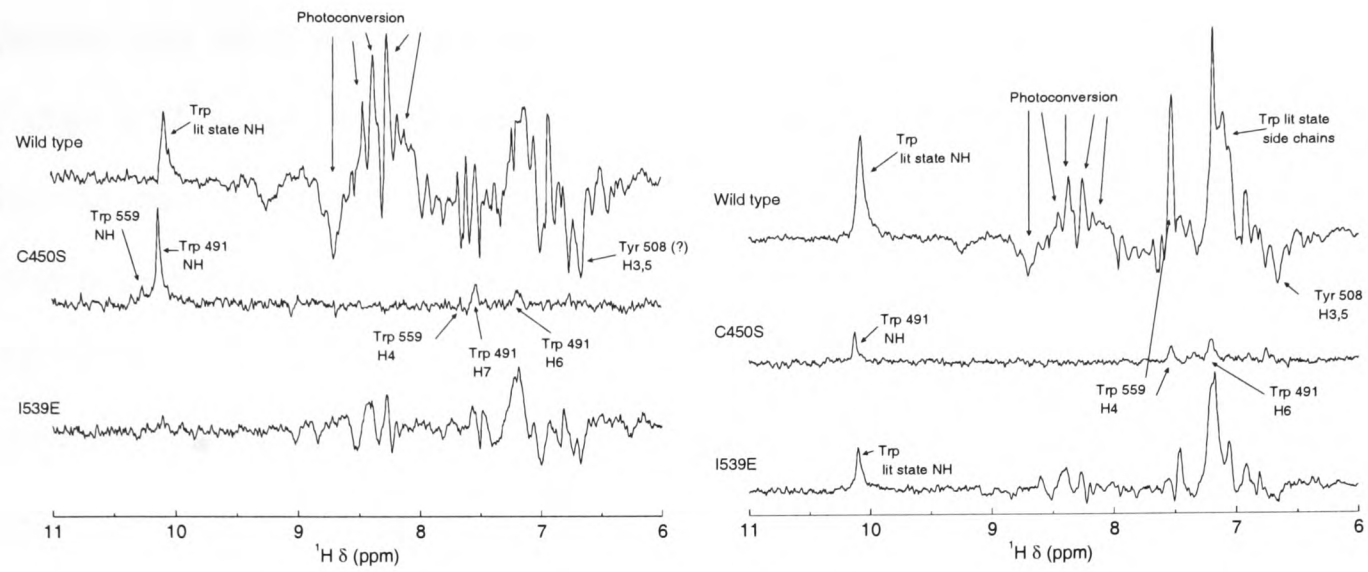


Figure 6.22: Static side chain accessibilities as a function of probe radius for AsLOV2, shown in Figures 6.1 and 6.9. The calculations were performed using the program `Naccess 2.1.1` [104].

in the wild-type protein and I539E mutant and hence associated structural changes occur, this increases the difficulty of interpreting the spectra shown in Figure 6.20.

The assignment of CIDNP spectra is aided by comparison with static solvent accessibility calculations. Figure 6.22 shows the results of these calculations performed using a structure of the dark state of AsLOV2 derived from a homology model with the LOV2 domain from *Adiantum*, as shown in Figures 6.1 and 6.9. The calculations were performed for a series of probe radii in consideration of the non-spherical shape of the flavin and thionin photosensitisers. These calculations are in agreement with the results shown in Figure 6.20 in that they predict that both tryptophan residues are solvent exposed. Trp 559 is clearly very exposed, as a result of being at the end of the unfolded linker region. Tyrosine 508 has a very similar solvent accessibility to Trp 491, and a corresponding emissive signal is observed at 6.65 ppm. The other



(a) 50 μM thionin

(b) 200 μM FMN

Figure 6.23: 500 MHz ^{1}H CIDNP spectra of the LOV2 domain and two mutants at 750 μM. Laser illumination was applied at 488 nm for 100 ms before each light scan. The spectra are averages of 16 light and dark pairs. The assignments of the C450S are taken from the known chemical shifts of the ground state of the protein [239].

tyrosine residues have very low solvent accessibilities for physically reasonable probe sizes and are not observed.

It is interesting to compare these results to those obtained when the illumination is performed at 488 nm, *i.e.* when driving the photo-excitation of the protein in addition the generation of CIDNP. Figure 6.23(a) shows this using thionin as the photosensitiser. There is no evidence of photoconversion for the C450S mutant, as expected. However, in the case of the wild type protein there are extensive changes. The tryptophan indole proton is visible at 10.09 ppm, therefore a proportion of the signals between 6.5 and 7.5 ppm can be assumed to originate from the side chain of this amino acid residue. The signal at 6.65 ppm can be tentatively assigned to the H3,5 ring protons of Tyr 508 since this signal appears at the ground state chemical shift of Tyr 508. Also, from inspection of the structure (Figure 6.9) this residue is unlikely to be greatly affected by the unfolding of the Jα helix upon photoconversion.

Since thionin and FMN possess differing electric charges, it is interesting to

investigate any effects of the different coulombic interactions between the dye and protein and what effect this has on the degree of nuclear polarisation generated. Figure 6.17 shows the differences between the spectra sensitised with the two dyes for the case of hen egg white lysozyme. The comparison for the LOV2 domain is shown in Figure 6.23. There is little difference between the two sets of spectra, except for a slight increase in intensity for the case of FMN. This is in agreement with previous experiments using thionin, which have been shown to have a decreased signal-to-noise ratio compared to the use of FMN [235].

There is a curious feature of the C450S mutant spectra shown in Figure 6.23 compared to the spectra recorded with illumination at 514 nm. Concentrating on the tryptophan indole proteins, Figure 6.20 shows both residues Trp 491 and Trp 559, with the latter having greater intensity. This can be loosely attributed to the much larger solvent accessibility of this residue, as shown in Figure 6.22. However, in the spectra recorded with illumination applied at 488 nm, Trp 559 is much weaker in Figure 6.23(a), when sensitised using thionin, and absent in Figure 6.23(b) with FMN. The precise reasons for these differences are unclear. A possible explanation is that there is an energy transfer from the internal flavin of the LOV2 domain to the photosensitiser when exciting at 488 nm. This would require the photosensitiser to be close to the internal FMN, hence could give rise to increased polarisation of Trp 491 since this lies closer to the internal cofactor than Trp 559 at the end of the linker.

There is also the possibility of the formation of a radical pair from the internal flavin and Trp 491. However, there is no evidence for the generation of CIDNP in flavoproteins when the radical pair consists of the internal chromophore and an electron donor. This is due to the need for the radical pair to separate to a distance at which the average hyperfine coupling, and the difference in radical g-values are greater that the electron-electron exchange interaction, thus allowing ST_0

mixing [60].

6.2.2.2 CIDNP-HSQC Spectra

The CIDNP-HSQC experiment was developed by Lyon *et al.* to gain greater chemical shift resolution when studying the tryptophan indole region of unfolded proteins [28, 31]. This experiment differs from the normal HSQC experiment in that there is no initial INEPT transfer of magnetisation from protons to nitrogen. Greater sensitivity is obtained by detecting the chemically pumped nitrogen magnetisation, via a reverse refocused INEPT transfer to the attached proton. Using this method there is no benefit in subtracting a spectrum in the absence of the laser pulse [28, 31].

Figure 6.24 shows an overlay of CIDNP-HSQC spectra of the C450S mutant and wild type LOV2 domain. The C450S mutant clearly shows identical features to the indole region of Figure 6.6. Both Trp 491 and Trp 559 are exposed, and visible at the dark state chemical shifts. The wild-type protein, by comparison, shows that some photoconversion is taking place during the laser pulse. The resonance corresponding to Trp 491 is no longer visible at the dark state chemical shifts, indicating that structural changes in the protein are occurring. A new signal at 10.05 ppm in the proton dimension, corresponding to the light state of the protein is seen. Trp 559 is clearly present in both spectra.

6.2.3 Interpretation

The spectra presented in this section show that it is possible to generate and detect CIDNP in the LOV2 domain using different photosensitisers and illumination wavelengths, despite the absorption of the protein. The original purpose of using thionin and exciting at 514 nm was to avoid the protein photocycle and hence probe the ground (non-signalling) state. The fact that the protein does undergo the photocycle at this wavelength, despite the low extinction coefficient complicated the

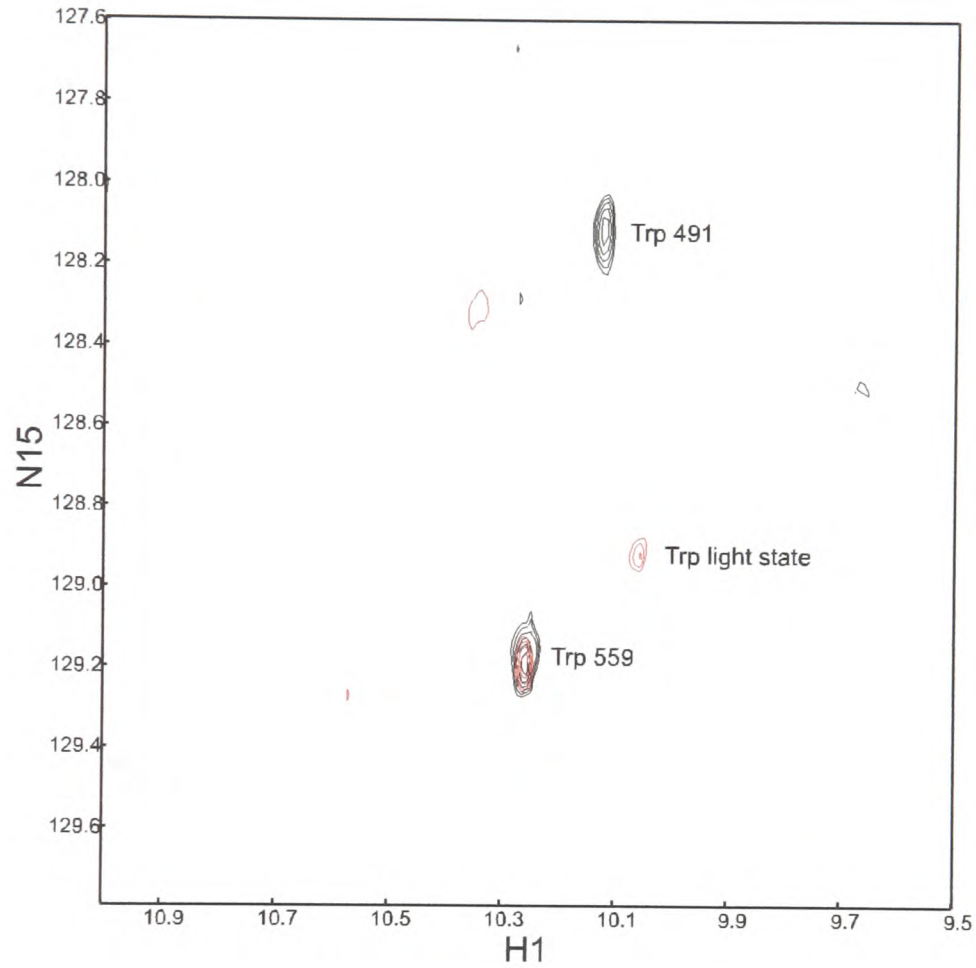


Figure 6.24: 600 MHz ^{15}N-CIDNP-HSQC spectra of C450S mutant (black spectrum) and wild type LOV2 domain (red spectrum), using thionin as the photosensitiser, showing the tryptophan indole region. A 100 ms laser pulse at 514 nm preceded each scan. The spectra were recorded using a spectral width of 9.5 kHz, 610 complex points in the directly detected dimension (^{1}H) and a spectral width of 133 Hz, 8 complex points in the indirectly detected dimension (^{15}N). The ^{15}N carrier was placed in the centre of the indole region. Four transients were recorded per increment (to achieve sensitivity enhancement [77–79] and quadrature detection in t_1). Shifted sine-squared window functions were used in both dimensions.

interpretation of the results. Of the five residues in the LOV2 domain which are amenable to study in these experiments, there appears to be little change in their solvent accessibility as the protein undergoes the reversible photocycle.

A number of improvements could be envisaged for the experiments described in this Section. The principal improvement would be to remove the complicating effect of the signals from the photo-excited state being superimposed on the CIDNP spectra of the protein. This could be achieved by recording the CIDNP "dark" spectrum in a slightly different manner. The "dark" spectrum could be recorded using the same illumination scheme as for the "light" spectrum, but without the photosensitiser present. This would therefore generate a "dark" spectrum which contained no nuclear polarisation, but had signals at the same chemical shifts as they appear in the CIDNP "light" spectrum. Subtraction of this spectrum of this from the "light" spectrum would result in only the polarised signals from the light state being present. There is, however, a problem with this approach. When recording the CIDNP spectrum a significant portion of the light entering the sample is absorbed by the photosensitiser. If the dye is not present this intense illumination would cause significant and irreparable damage to the protein. In fact, initial experiments performed prior to those shown in Section 6.1.1 using >1 W illumination caused photo-induced precipitation of the protein.

6.3 Kinetics of the Dark State Recovery

The ability to follow protein folding in real time by NMR has been well documented [21–24, 35, 240, 241]. These approaches generally follow the renaturation of the protein by monitoring the folding following a rapid pH jump, or dilution from a high concentration of chemical denaturant, such as 8 M urea. Recently, photochemically initiated refolding techniques have been performed to follow the refolding based

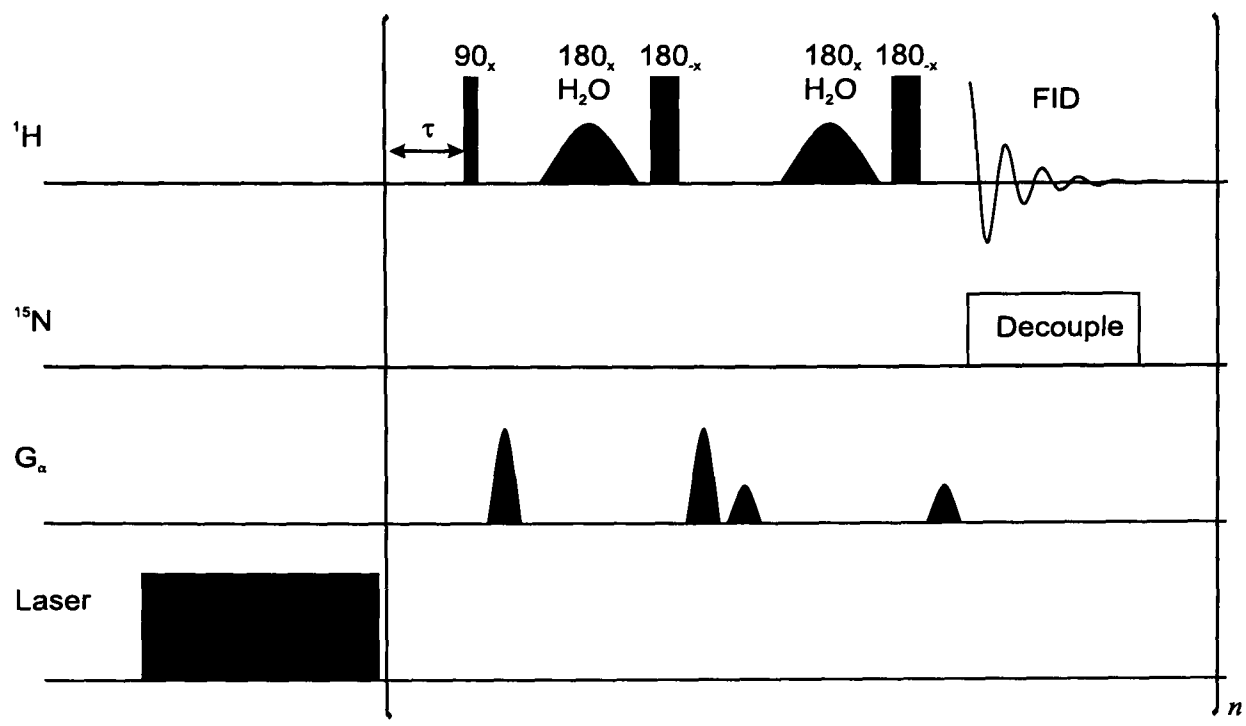


Figure 6.25: The pulse sequence for the dark state recovery experiment. The laser illumination for a period suitable to produce almost complete conversion to the light state of the domain. Water suppression is provided by the use of a DPFGSE [73].

on the release of metal ions necessary for a stable tertiary fold to be maintained [30].

Using the principles behind these experiments a method to follow the recovery of the dark state of the LOV2 domain from the light state was developed. Thi method uses changes in the one-dimensional NMR spectrum as a function of time following illumination. The methodology and results are shown below.

6.3.1 The Pulse Sequence and Methodology

The pulse sequence used for this recovery experiment is a simple sequence consisting of an illumination period, to generate the light state of the protein. This is then followed by acquisition of a number free induction decays, each preceded by a general water suppression scheme (the DPFGSE [73]), as a function of time following the end of the illumination. Figure 6.25 shows the timing diagram for this experiment. Acquisition of the NMR data using this method results in a pseudo two-dimensional data set of the form $s(t_1, t_2)$, where t_1 is the time elapsed following the illumination,

and is related to the delay τ in the pulse sequence by the following expression:

$$t_1 = n \times (\tau + t_{\text{ps}} + t_{\text{acq}}) \tag{6.2}$$

where t_{ps} is the time required to execute the RF pulses and pulsed field gradients, and t_{acq} is the acquisition time. t_2 is the usual FID acquisition period. After Fourier transformation with respect to the t_2 dimension, this yields a mixed time-frequency domain spectrum:

$$s(t_1, t_2) \xrightarrow{\mathcal{F}_2} s(t_1, \omega_2) \tag{6.3}$$

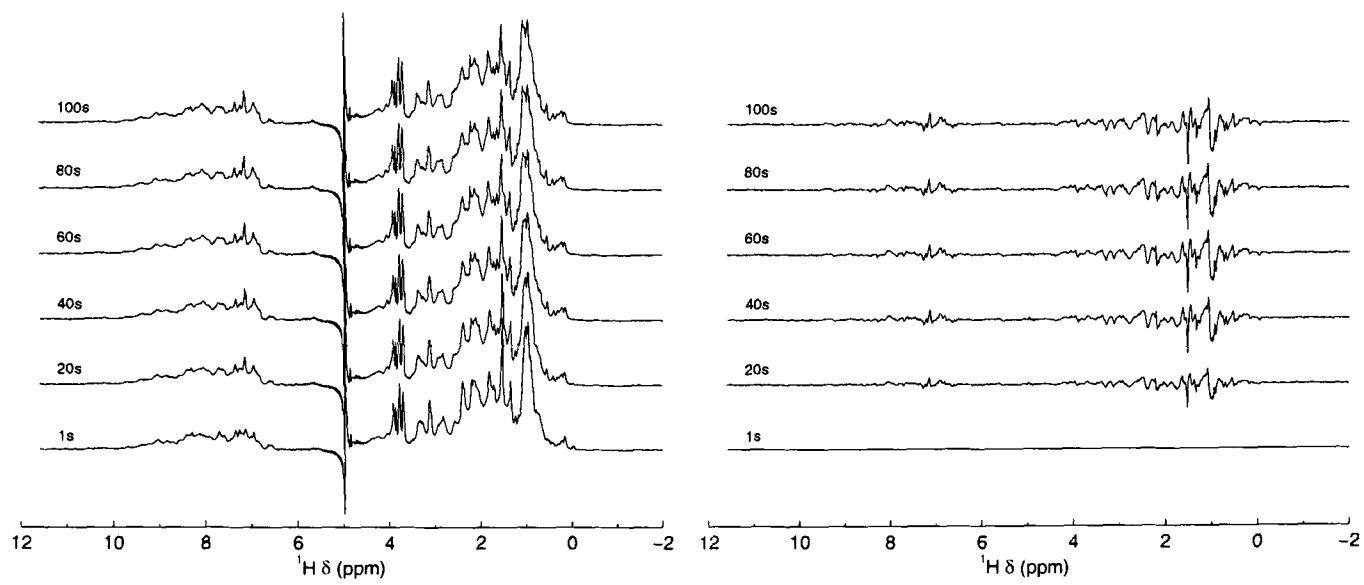
where $\mathcal{F}_2$ represents Fourier transformation in the t_2 dimension.

The protein's reversible photocycle allows repetition of the experiment a number of times and summation of the resulting data sets, thereby increasing the signal-to-noise ratio.

6.3.2 Results and Interpretation

The real-time recovery experiment described above was then applied to the LOV2 domain. A 300 μM sample was used at pH 7. The temperature was raised to 37 °C in order to increase the rate of return to the dark state. Spectra were recorded at one second intervals for two minutes following illumination, using 512 complex data points and a spectral width of 6800 Hz. The experiment was averaged over 64 pseudo 2D planes. Selected 1-dimensional spectra from the complete series are shown in Figure 6.26(a). These spectra clearly show that at later times (*i.e.* after ~60 s) there is little change in the spectrum. In fact, it is difficult to see changes even in the early spectra

The changes occurring in these real-time spectra during the recovery of the ground state can be most easily followed by subtracting the first spectrum in the series, that is, the first spectrum after illumination, from the remaining spectra. In



(a) The real-time spectra

(b) The difference spectra

Figure 6.26: Selected spectra at the times indicated from the dark state recovery experiment. Spectra were recorded at 500 MHz, pH 7, and 37°C. Nitrogen-15 decoupling was applied before acquisition of each spectrum.

this case, signals from the returning dark state will be shown as positive, while those from the light state will be shown as negative (note that this is the *reverse* of that seen in Figure 6.5). This method of presentation is shown in Figure 6.26(b).

In order to obtain quantitative information about the kinetics of the dark state recovery, the intensity at fifteen positions across both the aromatic and aliphatic regions of the difference spectra presented in Figure 6.26(b) were extracted. These intensities are plotted as a function of time after the laser irradiation in Figure 6.27. These points were fitted with a decaying single exponential function of the form:

$$I(t_1) - I(0) = A(1 - e^{-t_1/T_r}) \tag{6.4}$$

where t_1 is the time after the laser illumination, T_r is the time constant for the return to the ground state, A is a scaling factor, and $I(0)$ is the intensity of the first spectrum recorded. Table 6.5 shows these data for the fifteen points chosen. These data show that there is a distribution of lifetimes amongst the peaks selected,

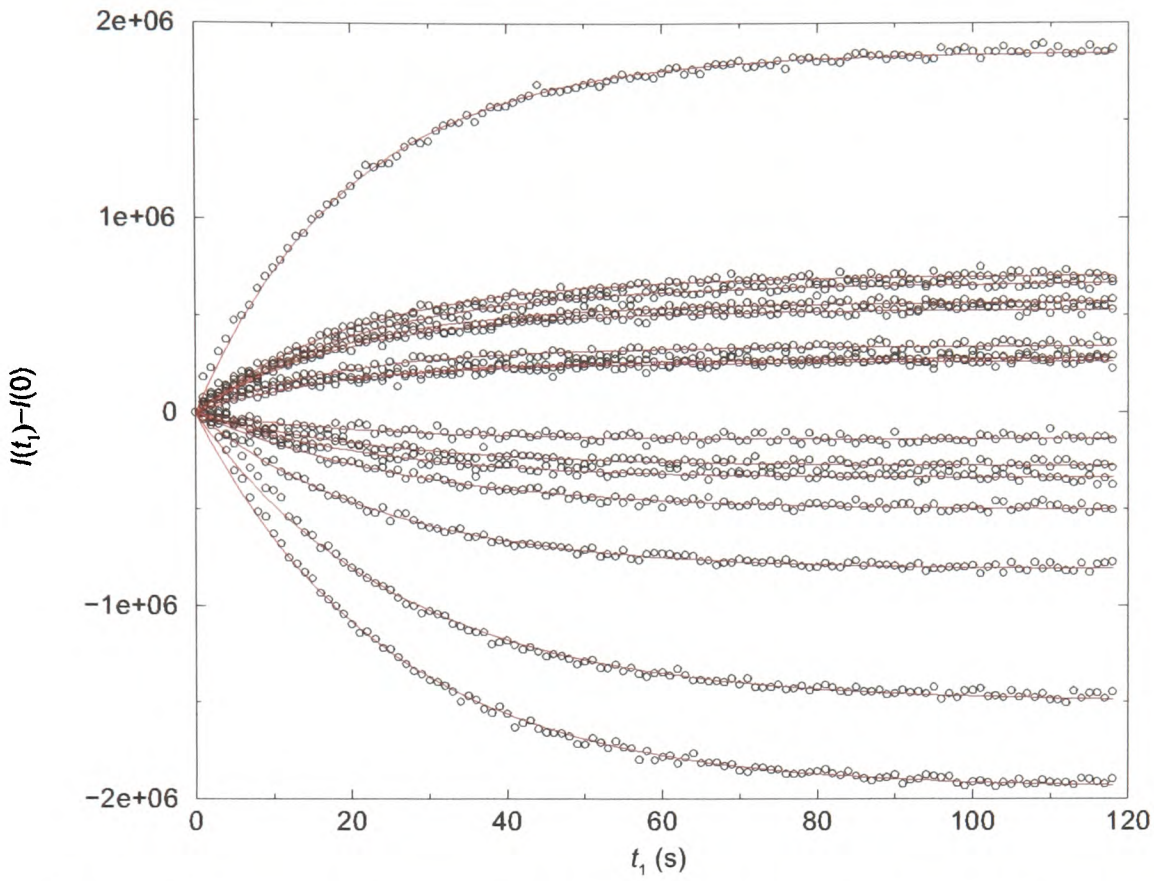


Figure 6.27: Peak heights taken from the difference spectra, plotted as a function of time after the laser illumination. The data points are fitted with a single exponential function (shown in red) as described in the text.

with an average time constant for the return of the LOV2 domain to the dark state of $\langle T_r \rangle = 22.13 \pm 2.76$ s. This distribution of lifetimes of the light state has been investigated further by Harper *et al.* using a time-resolved HSQC experiment [38]. This gives the lifetime of each of the backbone amides, and the tryptophan indole protons. The lifetimes measured here are in general agreement with those obtained by Harper *et al.* subject to two caveats. Firstly, the experiment described here was performed at a higher temperature (37°C compared to 30°C), thereby shortening the measured lifetime. Secondly, measurements taken from the one-dimensional real-time data are global measurements since it is not possible to resolve individual resonances, and hence obtain site specific information.

^{1}H δ (ppm)	A (a.u.)	T_r (s)
-0.05	-273617	24.84
0.28	283953	20.54
0.54	668885	22.22
0.81	571074	19.83
0.89	-1498800	26.29
0.97	-1945400	24.76
1.07	1852000	20.07
1.73	-811741	24.16
2.47	708199	20.55
3.88	-336638	23.59
5.26	-134115	21.55
6.89	532057	19.14
7.24	-504823	27.35
8.01	345260	18.65

Table 6.5: Parameters taken from fitting Equation 6.4 to the real-time NMR data shown in Figure 6.26(b).

6.4 Last thoughts / Conclusions

This chapter has demonstrated the application of a number of NMR techniques, including photo-CIDNP real-time methods, to the study of a domain from a plant signalling protein, phototropin. These experiments have shown that there are large-scale structural changes occurring in the protein, induced by the absorption of a light photon. In turn, these changes activate signalling pathways in the plant, initially via autophosphorylation of the C-terminal kinase, leading to the control of phototropism, light induced chloroplast movement and stomatal opening [212, 220–222].

Future experiments on this system can be envisaged along two lines. The first is to study the LOV1 domain to gain insight into the first photoreceptor of the phototropin sequence. This has similar optical characteristics to the LOV2 domain. Secondly, the techniques used here could be applied to multi-domain constructs of phototropin, for example, LOV2–Kinase and the complete phototropin sequence

LOV1–LOV2–Kinase (and potentially LOV1–Kinase), using selective labelling of the domains to combat the increased spectral complexity.

Chapter 7

Simulating Tryptophan Nuclear Polarisation

Experimental observations of the tryptophan radical are relatively recent [242–244]. In these cases tryptophan is immobilised in a protein matrix, for example in DNA photolyase [242], cytochrome *c* peroxidase [243] and in ribonucleotide reductase [244]. Identification of the tryptophan radical: whether it is the cationic or neutral radical, is usually by comparison with computed hyperfine coupling constants for various analogues and model compounds such as indole [245, 246], 3-methyl indole [247, 248] and 3-ethyl indole [248]. As well as information from EPR measurements of the tryptophan radical in a protein matrix, there are several other studies, including photodegradation during CIDNP experiments [174, 237], steady-state CIDNP [249], time-resolved CIDNP experiments [250, 251] and laser flash photolysis [59].

Knowledge of the hyperfine couplings in a radical pair makes it possible to predict both the phase and intensity of the nuclear polarisation produced in the resulting diamagnetic products. This chapter explores the case of the tryptophan radical, using computed hyperfine couplings from the literature to determine how accurately

it is possible to predict the enhancement of the NMR spectrum from the CIDNP experiment.

7.1 Theory

The probability of a radical pair recombining from a given nuclear spin state χ is given by the following equation [252, 253]:

$$P^{\chi} = \lambda \int_0^{\infty} S^{\chi}(t) f(t) \mathrm{d}t \tag{7.1}$$

where $S^{\chi}(t)$ is the probability that the radical pair is in an electronic singlet state at time t, and hence depends on the spin dynamics of the radical pair; $f(t)$ is the probability that the radical pair undergoes a reencounter within a time $t \rightarrow t + \mathrm{d}t$, following its creation; and λ is a steric factor which determines whether a given collision will be reactive.

Consider a radical pair born in a triplet spin correlated state, undergoing coherent spin evolution via the Δg mechanism. Neglecting inter-electron exchange and all anisotropic interactions, this is described by the following spin Hamiltonian:

$$\hat{H} = \sum_{j=1}^{2} g_j \hat{S}_{jz} \frac{\mu_B B_0}{\hbar} + \sum_{j=1}^{2} \sum_{i} a_{ij} \hat{I}_i \cdot \hat{S}_j \tag{7.2}$$

where j is the radical and i is the nucleus within that radical; all other symbols have their usual meaning. The probability of this radical pair being in a singlet state at time t can be determined by considering the wavefunction for the radical pair:

$$|\psi(t)\rangle = (c_S(t)\,|S\rangle + c_{T_0}(t)\,|T_0\rangle) \otimes \left|\phi_{12}^{\mathrm{N}}\right\rangle \tag{7.3}$$

where $|S\rangle$ and $|T_0\rangle$ are the electron spin wavefunctions given by:

$$|S\rangle = \frac{1}{\sqrt{2}}(|\alpha\beta\rangle - |\beta\alpha\rangle) \tag{7.4}$$

$$|T_0\rangle = \frac{1}{\sqrt{2}}(|\alpha\beta\rangle + |\beta\alpha\rangle) \tag{7.5}$$

and $\left|\phi_{12}^{\mathrm{N}}\right\rangle$ is the nuclear spin wavefunction across the whole radical pair. The coefficients in the radical pair wavefunction are given by:

$$c_S(t) = c_S(0)\cos(\omega_{ST_0}^{\chi}t) - \mathrm{i}c_{T_0}(0)\sin(\omega_{ST_0}^{\chi}t) \tag{7.6}$$

$$c_{T_0}(t) = c_{T_0}(0)\cos(\omega_{ST_0}^{\chi}t) - \mathrm{i}c_S(0)\sin(\omega_{ST_0}^{\chi}t) \tag{7.7}$$

where $\omega_{ST_0}^{\chi}$ is the rate of oscillation between the singlet $|S\rangle$ and the triplet $|T_0\rangle$ states for a given nuclear spin state χ [252]. From this the probability of being in a singlet $|S\rangle$ state at a time t of a radical pair formed in an initial triplet state, *i.e.* $c_S(0) = 0$, is therefore given by:

$$|c_S(t)|^2 = |c_{T_0}(0)|^2\sin^2(\omega_{ST_0}^{\chi}t) \tag{7.8}$$

Assuming that all three triplet sublevels are equally populated at the creation of the radical pair results in $c_{T_0}(0) = \frac{1}{\sqrt{3}}$. Then the final expression for the probability of the radical pair being found in a singlet $|S\rangle$ state at a time t following its creation as a triplet species, with the spin evolution via the Δg mechanism is:

$$S^{\chi}(t) = |c_S(t)|^2 = \frac{1}{3}\sin^2(\omega_{ST_0}^{\chi}t) \tag{7.9}$$

At high field, the $|T_{\pm}\rangle$ states are unable to interconvert with $|S\rangle$ and $|T_0\rangle$ states since they are too far removed in energy.

7.1.1 Radical Pair Diffusion Models

The radical reencounter function $f(t)$, depends purely on the kinetic (reactive) and diffusive behaviour of the radical pair. The simplest model describing this is the so-called exponential model, in which the reencounter probability is described by a single decaying exponential:

$$f(t) = k_S \exp(-k_S t) \tag{7.10}$$

where k_S is the rate constant for reaction of the radical pair via the singlet channel. Integration of Equation 7.10 multiplied by Equation 7.9 leads to the following predicted recombination probability:

$$P^{\chi} = \frac{2\lambda(\omega^{\chi}_{ST_0})^2}{3(k_S^2 + 4(\omega^{\chi}_{ST_0})^2)} \tag{7.11}$$

A more realistic treatment of diffusion has been performed by Adrian [252], which results in a reencounter function of the following form:

$$f(t) = mt^{-3/2} \exp\left(\frac{-m^2}{p^2 t}\right) \tag{7.12}$$

where:

$$m = \sqrt{\frac{3}{4\pi}} \frac{R_0 - R_\sigma}{R_0} \qquad \text{and} \qquad p = \frac{1}{\pi} \frac{R_\sigma}{R_0} \tag{7.13}$$

with R_0 being the separation of the radical pair at its creation and R_σ the critical reaction distance, *i.e.* the distance to which the radicals must diffuse in order to react, and given by $R_\sigma = \sqrt{3D\tau}$, where D is the diffusion coefficient and τ is the time between diffusive steps [252]. The recombination probability now has a more

complex functional form and is given by:

$$P^{\chi} = \frac{\lambda p\sqrt{\pi}}{6}\left\{1 - \exp\left(\frac{-2m\sqrt{|\omega^{\chi}_{ST_0}|}}{p}\right)\cos\left(\frac{2m\sqrt{|\omega^{\chi}_{ST_0}|}}{p}\right)\right\} \tag{7.14}$$

Typically $2m/p \simeq 1.5$ [252]. Adrian also argues that, on the time-scale of the ST_0 mixing responsible for the generation of CIDNP, the exponential term in Equation 7.12 tends to unity and can hence be neglected. This leads to a simplified form of the diffusion-based reencounter function:

$$f(t) = mt^{-3/2} \tag{7.15}$$

Using this simplified form the recombination probability reduces to:

$$P^{\chi} = \frac{\lambda m}{3}\sqrt{\pi|\omega^{\chi}_{ST_0}|} \tag{7.16}$$

which is equivalent to retaining just the first term in a Taylor series expansion of Equation 7.14 in powers of $\sqrt{|\omega^{\chi}_{ST_0}|}$.

7.1.2 Calculating the CIDNP Intensity

Using Equation 7.1 and the diffusion models given in the previous Section, it is possible to calculate the probability that a radical pair will recombine from a given nuclear spin state. Thence it is straightforward to calculate the nuclear polarisation at a particular nucleus since this is just the difference in the probability of recombination between the different nuclear spin states of that nucleus. For example, the difference in recombination probability between the α ($m_I = +\frac{1}{2}$) and β ($m_I = -\frac{1}{2}$)

spin states of a spin-$\frac{1}{2}$ nucleus is given by the following equation:

$$\mathcal{P}_k^{\mathrm{R}} = \frac{1}{2^{(N-1)}3^M} \sum_{\substack{i \neq k \\ j}} \left\{ P^{\chi}(\omega^{+}_{ST_0}; i, j, k) - P^{\chi}(\omega^{-}_{ST_0}; i, j, k) \right\} \tag{7.17}$$

where

$$\omega^{\pm}_{ST_0}(\ldots; i, j, k) = \Delta g \frac{\mu_B B_0}{\hbar} \pm \frac{1}{2} a_k + \sum_{m_{I_i}=-I_i}^{+I_i} m_{I_i} a_{i1} - \sum_{m_{I_j}=-I_j}^{+I_j} m_{I_j} a_{j2} \tag{7.18}$$

N is the number of spin-$\frac{1}{2}$ nuclei and M is the number of spin-1 nuclei in the calculation. Δg is defined as $g_1 - g_2$. This calculates the nuclear polarisation of a particular spin-$\frac{1}{2}$ nucleus k situated on radical one, in the presence of all other nuclei in the radical pair, in all possible configurations of nuclear spin states. Weak coupling between the nuclear spins is assumed throughout these calculations.

The program `cidnpint.f` was developed to calculate the field dependence of the CIDNP intensity using this method; the source code is listed in Appendix B.1. Figure 7.1 shows the calculated field dependence of the nuclear spin polarisation of the H2,6, H3,5 and H^{β} protons of the tyrosyl neutral radical (with an FMN neutral counter radical), using the three models for diffusion given above. The spectra were scaled to the maximum of the H^{β} curve for each model. These calculations clearly show a large difference in the predicted CIDNP field dependence. This is not surprising since the exponential model is only a very basic model. Also, Adrian's assumption that the diffusion model given in Equation 7.12 can be simplified by neglecting the exponential term to give Equation 7.15 is justified in this case, since the field dependence curves for these two models show only very minor differences at high field.

Previous experimental measurements have shown that the simplified diffusion model is in fact a good fit to experimental field dependence measurements [31,

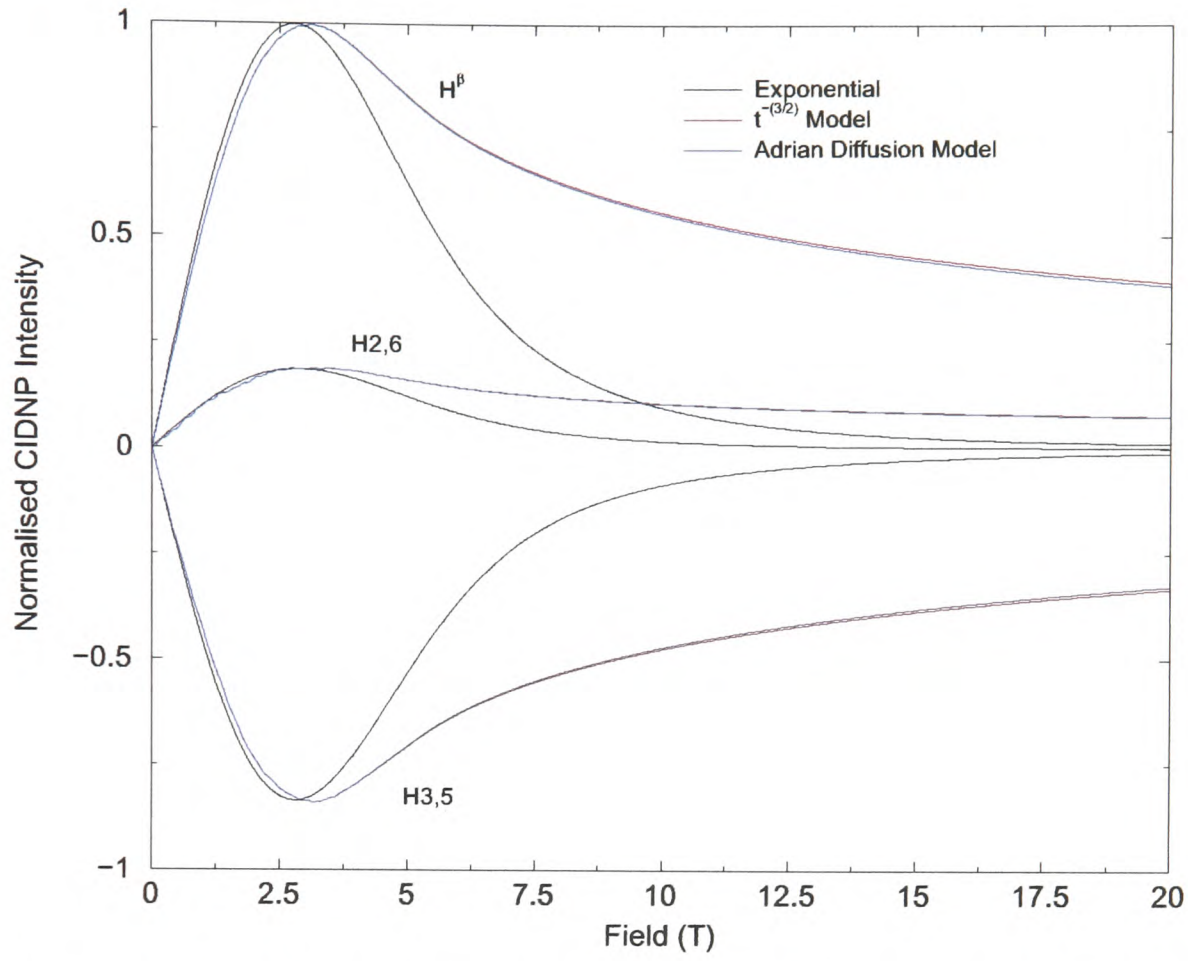


Figure 7.1: Calculated field dependence using Equation 7.17 and the three diffusion models. The hyperfine couplings were taken from Table C.1 [254, 255], and Table C.3 [256], both given in Appendix C. For the exponential model a rate constant $k_S = 5 \times 10^8$ s^{-1} was used, and for the Adrian diffusion model $2p/m = 1.5$.

257]. Calculations performed previously employed the approximation that a spin-1 nucleus could be treated as two spin-$\frac{1}{2}$ nuclei. This approximation has the effect of doubling the degeneracy of the $m_I = 0$ state. Figure 7.2 shows calculations using this approximate treatment and the explicit consideration of spin-1 nuclei described above. The calculations were performed using the same parameters as for Figure 7.1, and using the simplified diffusion model given in Equation 7.15. There is a slight shift to lower field of the maxima and minima of the field dependence curves for the case of the approximate model. There is also a slight broadening for the curves arising from the explicit inclusion of the spin-1 nuclei. Therefore the differences between the exact and approximate treatments can be seen to be minor.

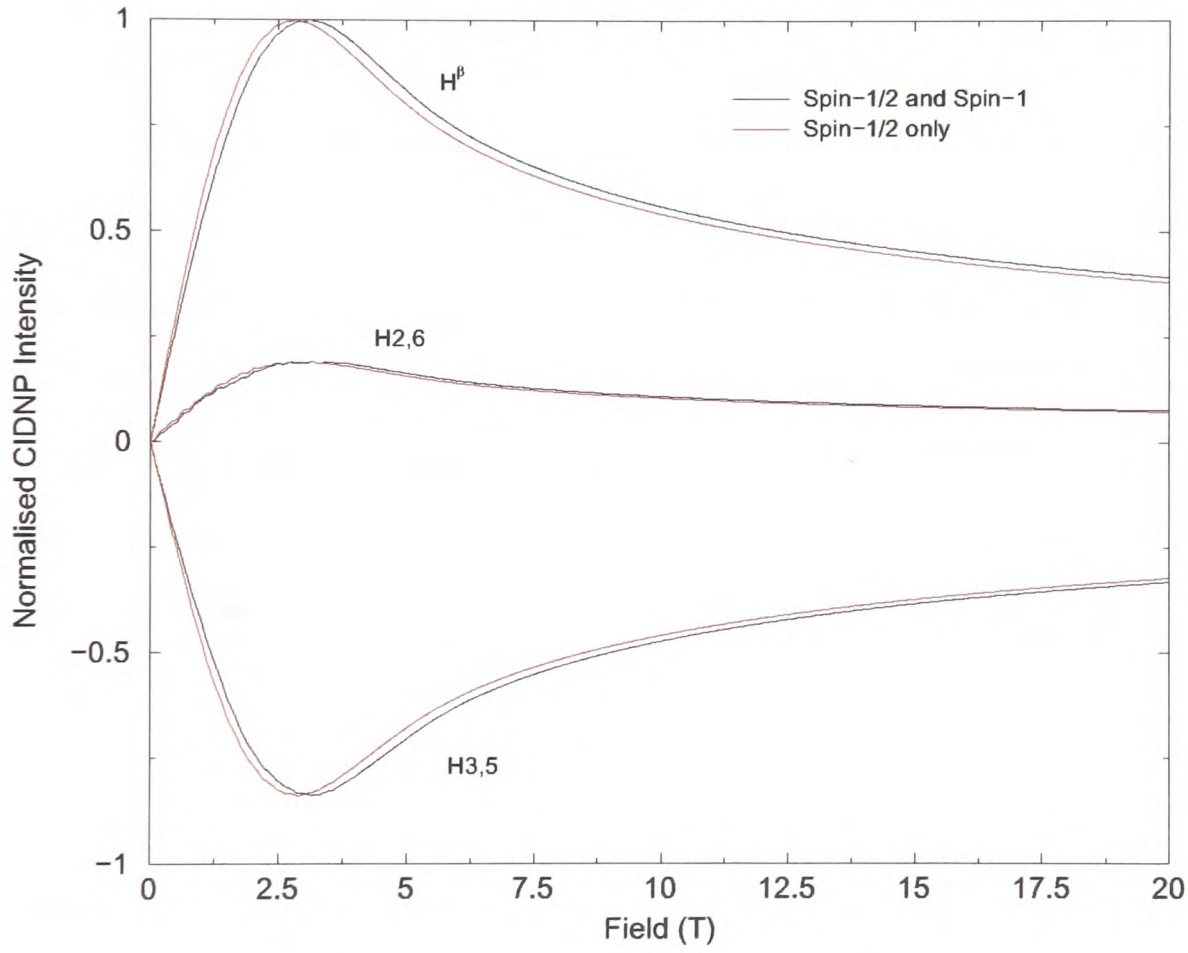


Figure 7.2: Comparison of results obtained by treating all nuclei as spin-$\frac{1}{2}$ (*i.e.* spin-1 = 2× spin-$\frac{1}{2}$) and by treating spin-1 nuclei explicitly. The hyperfine couplings were taken from Table C.1 [254, 255], and Table C.3 [256], both given in Appendix C.

7.1.3 Simulating CIDNP Spectra

Since a CIDNP spectrum is simply an NMR spectrum with enhanced intensities, it is straightforward to simulate. The starting point is the construction of a suitable time-domain signal, which is given by the following expression:

$$s(t) = \sum_k \left\{ K_k \exp(\mathrm{i}\Omega_k t) \cos^{l_k - 1}(\pi J_k t) \right\} \times \exp(-t/T_2) \tag{7.19}$$

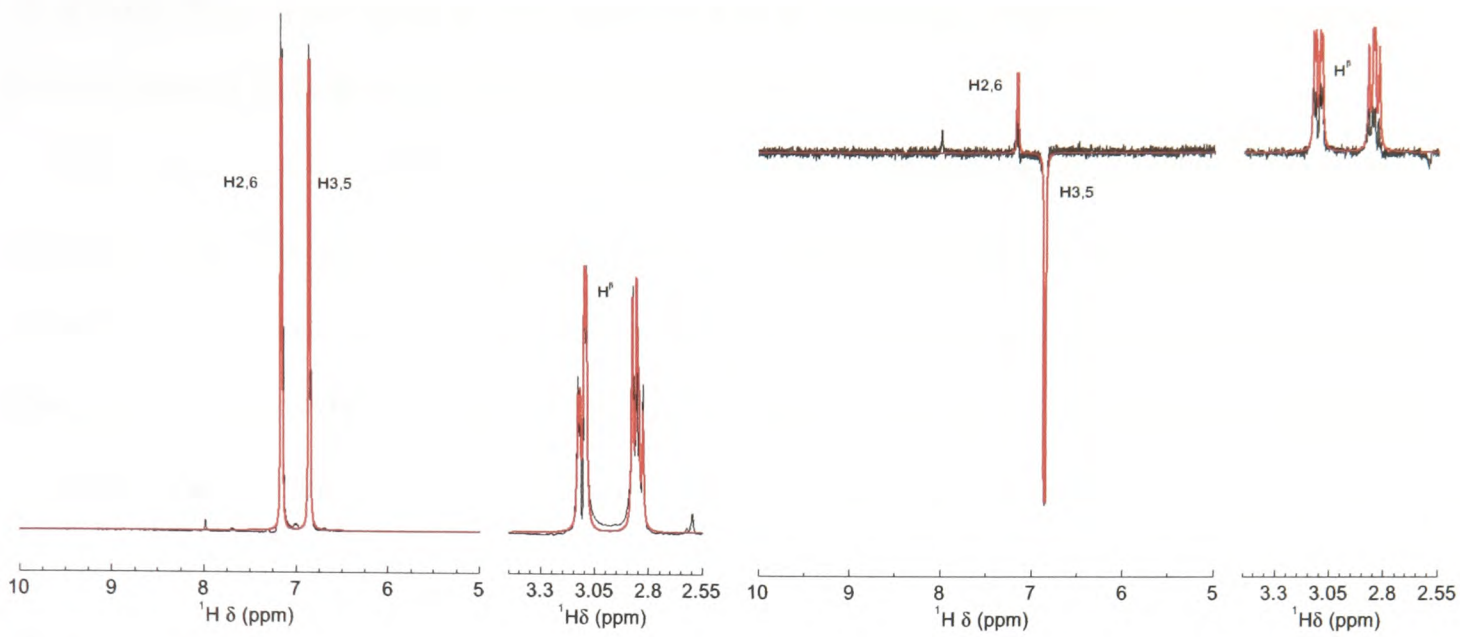
where K_k is the number of nuclear spins contributing to the k^{th} multiplet, which occurs at an offset of Ω_k, with a scalar coupling of J_k and a multiplicity of l_k. These simulations assume that the weak coupling scheme applies, which is a reasonable approximation at high field. The experimental spectra were all recorded at a proton frequency of 500 MHz, which corresponds to $B_0 = 11.75$ T. T_2 is the spin-spin relaxation time and governs the linewidth in the simulations ($\Delta\nu_{\text{FWHH}} = \frac{1}{\pi T_2}$). The same linewidth is assumed for each multiplet in the spectrum. The NMR spectrum in then given by the Fourier transform of Equation 7.19:

$$s(\omega) = \int_0^\infty s(t) \exp(\mathrm{i}\omega t) \mathrm{d}t \tag{7.20}$$

As is convention, it is the real part of Equation 7.20 which is displayed. The spectra were calculated using the program `spec-sim.f`, the source code for which is given in Appendix B.2. As the CIDNP spectra are just NMR spectra with enhanced intensities, this is accounted for by replacing the number of nuclei contributing to the multiplet, with the CIDNP intensity at 11.75 T obtained from the field dependence calculations. This is a valid approach since to a first approximation the observed CIDNP intensity is proportional to the polarisation generated.

$$I_k \propto K_k \quad \text{for the NMR spectra} \tag{7.21}$$

$$I_k \propto \mathcal{P}_k^{\mathrm{R}} \quad \text{for the CIDNP spectra} \tag{7.22}$$



(a) NMR (b) CIDNP

Figure 7.3: Experimental (black) and simulated (red) 500 MHz NMR and CIDNP spectra for tyrosine. The experimental spectrum was acquired using a 2 mM solution of tyrosine, illuminated for 50 ms at 4 W. The CIDNP intensities were taken from the simplified diffusion model curves in Figure 7.1.

An example of the simulated NMR spectrum for tyrosine, and the experimental spectrum for comparison, is shown in Figure 7.3(a). The strong coupling observed in the H^{β} proton resonances was accounted for by adjusting the relative contributions of the component multiplets in the simulation of the FID.

The CIDNP spectrum in Figure 7.3(b) was simulated using intensities obtained by applying the simplified diffusion model in Figure 7.1. This is modelling the tyrosyl neutral radical with the flavin neutral radical as the counter species. The spectrum has been scaled to the height of the H3,5 protons of the experimental spectrum. This simulation shows that reasonable agreement with the experimental spectrum is obtained using this model of diffusion and algorithm for the simulation of the CIDNP spectrum. There are, however, slight discrepancies. The most pronounced is the over-estimation of the intensity of the H2,6 ring protons. This arises since the effects of polarisation cancellation are most noticeable for nuclei with small hyperfine

couplings [55]. The same is observed for the H^{β} protons, however, the polarisation is now spread across eight lines of the multiplet.

The case of CIDNP from the tyrosyl neutral radical constitutes the "proof of concept" for this method of predicting the CIDNP intensities in the diamagnetic products of a radical pair reaction. These ideas are extended in the rest of this chapter to the study of tryptophan, and to determine how accurately is it possible to predict both the phase and intensity of the nuclear polarisation.

7.2 Results

As described in the introduction to this thesis (Chapter 1), tryptophan nuclear polarisation is generated via an electron transfer process, followed, at neutral pH, by rapid deprotonation of the tryptophan radical cation on a microsecond timescale [59]:

$$^{3}\mathrm{FMN} + \mathrm{TrpH} \rightarrow {}^{3}\overline{\mathrm{FMN}^{-\bullet} + \mathrm{TrpH}^{+\bullet}} \rightarrow {}^{3}\overline{\mathrm{FMNH}^{\bullet}\ \mathrm{Trp}^{\bullet}} \qquad (7.23)$$

where the overbar indicates spin correlation between the radicals. This mechanism therefore supposes that nuclear polarisation arises from a radical pair involving either the tryptophan radical cation, or the neutral radical.

The experimental CIDNP spectrum shown in this section was obtained using a 2 mM *N*-acetyl tryptophan solution at pH 7, containing 0.2 mM FMN. The spectra were recorded at a proton frequency of 500 MHz, using a laser pulse of 50 ms duration at a power of 4 W.

7.2.1 Tryptophan Radical Cation

Since the accepted mechanism for the generation of tryptophan CIDNP is via an electron transfer process [59], it is logical to start with the radical pair consisting of a tryptophan cation radical and an anionic flavin radical.

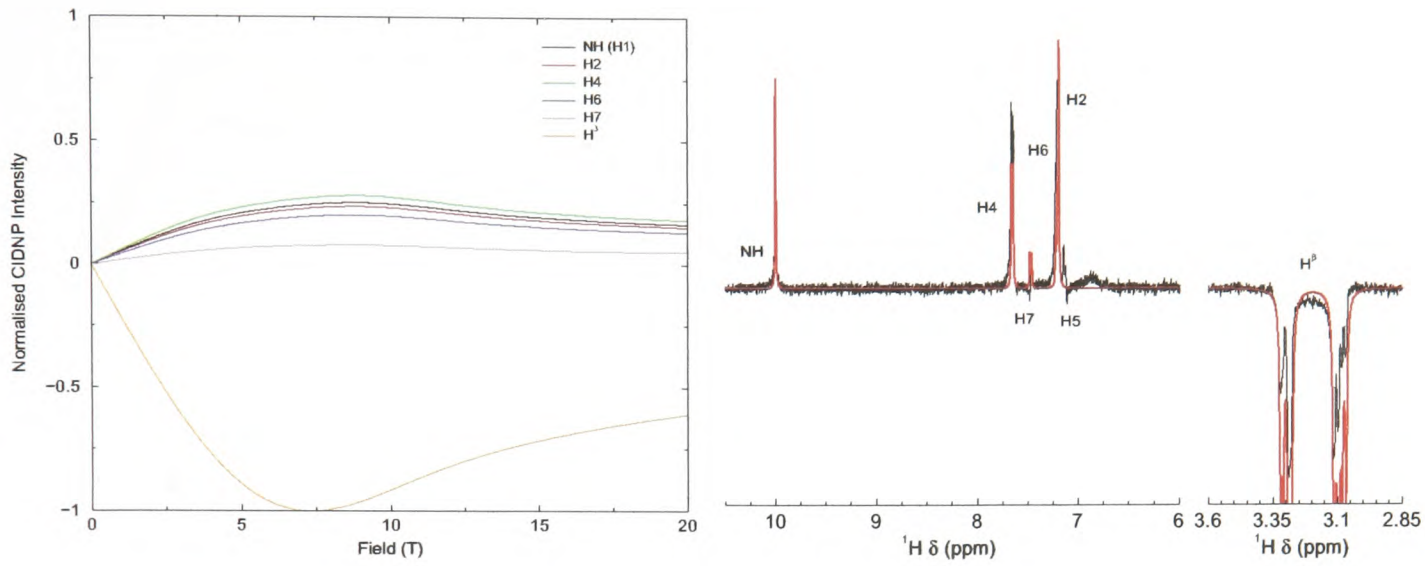
In order to calculate hyperfine couplings in the tryptophan radical on a sensible time scale it is often necessary to approximate the structure, usually to either 3-methyl indole or 3-ethyl indole. O'Malley and Ellson performed hyperfine coupling calculations on 3-methyl indole using Density Functional Theory (DFT) at the B3LYP/EPR-III level of theory [247]. These hyperfine couplings are given in Table 7.1 (and Table C.5 in Appendix C). Figure 7.4(a) shows the CIDNP field

NH (H1)	H2	H4	H5	H6	H7	$H^{\beta}1$	$H^{\beta}2$	N1
−0.52	−0.49	−0.58	-	−0.42	−0.17	2.65	1.22	0.22

Table 7.1: Isotropic hyperfine coupling constants for the cationic tryptophan radical, obtained from DFT calculations at the B3LYP/EPR-III level of theory on 3-methyl indole [247].

dependence curves obtained using these hyperfine couplings and the simplified diffusion model (Equation 7.15). These curves are dominated by the large predicted hyperfine couplings of the H^{β} protons, which are modelled as a methyl group attached to carbon-3 of the indole ring. The magnitude of the hyperfine couplings at these protons leads to a large overestimate of the CIDNP intensity, as seen in Figure 7.4(b). However, the converse is true for the other major signals. Both the resonances of H4 and H6 show smaller simulated polarisation than is observed experimentally, by factors of approximately 1/3 and 2/3, respectively. The minor signal H7 has greater predicted polarisation which can be attributed to polarisation cancellation effects. These are more noticeable given the small magnitude of the hyperfine coupling at H7. Finally, the indole proton is of interest. The observed discrepancy can be attributed to the increased linewidth of this proton resonance, arising from chemical exchange of the proton with the solvent. Increased relaxation rates, due to significant chemical shift anisotropy of the nitrogen and the fact that ^{14}N is a spin-1 nucleus.

While the model used in Figure 7.4 shows reasonable qualitative agreement with



(a) Field dependence using hyperfine couplings from [247]

(b) CIDNP spectrum using hyperfine couplings from [247]

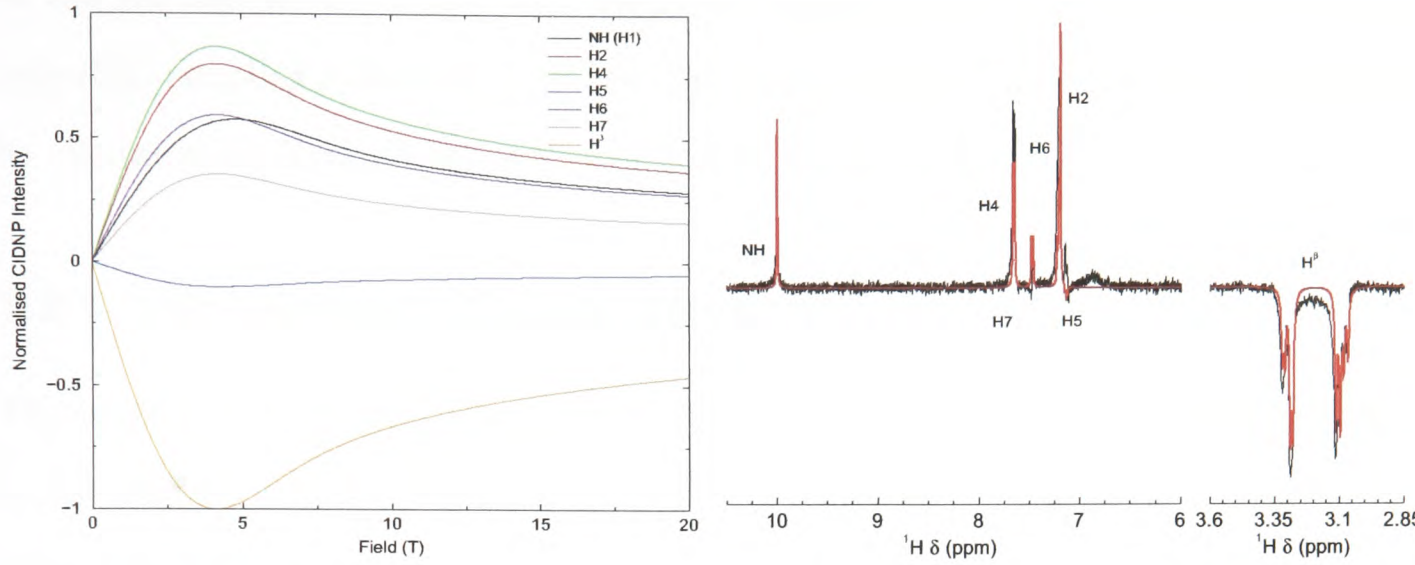
Figure 7.4: CIDNP field dependence and spectra for a radical pair consisting of a tryptophan cation radical and an anionic flavin radical. The black spectrum is the experimental spectrum and the red is the simulation, which is scaled to the height of the H2 singlet. The hyperfine couplings used are tabulated in Table C.2 [258, 259] and Tables 7.1 and C.5 [247].

the experimental spectrum, it is clearly deficient in certain respects. The extension of the alkyl chain attached at position 3 of the indole ring therefore should bring an improvement. Hyperfine couplings for 3-ethyl indole have been calculated by Himo and Eriksson, using DFT at the PWP86/ILGO-III level of theory. The hyperfine couplings used for the tryptophan radical cation are presented in Table 7.2 (and in Table C.5 in Appendix C). These hyperfine couplings give rise to the CIDNP

NH (H1)	H2	H4	H5	H6	H7	$H^{\beta}1$	$H^{\beta}2$	N1
−0.34	−0.47	−0.51	0.06	−0.35	−0.21	0.33	0.84	0.25

Table 7.2: Isotropic hyperfine coupling constants for the cationic tryptophan radical, obtained from DFT calculations at the PWP86/ILGO-III level of theory on 3-ethyl indole [248].

field dependence given in Figure 7.5(a). The simulated CIDNP spectrum obtained



(a) Field dependence using hyperfine couplings from [248]

(b) CIDNP spectrum using hyperfine couplings from [248]

Figure 7.5: CIDNP field dependence and spectra for a radical pair consisting of a tryptophan cation radical and an anionic flavin radical. The black spectrum is the experimental spectrum and the red is the simulation, which is scaled to the height of the H2 singlet. The hyperfine couplings used are tabulated in Table C.2 [258, 259] and Tables 7.2 and C.5 [248].

using the hyperfine couplings from 3-ethyl indole in Figure 7.5(b), show a better fit to the experimental spectrum than that obtained using 3-methyl indole. It is immediately clear that the polarisation of the H^{β} protons predicted is closer to the experimentally observed value. The ring protons generally have polarisation comparable to that observed for 3-methyl indole, due the fact that the hyperfine couplings are of similar magnitudes. From the simulations the polarisation of the h5 ring proton is incorrectly predicted to be of the opposite phase to that of the other ring protons. This antiphase nature may arise from polarisation transfer via *J*-coupling, in that nuclear spins which are strongly coupled may transfer polarisation, although contributions from this mechanism are known to be small [249].

Given that transfer of polarisation from NOE transfers and from strong *J*-coupling are small, there are two reasons for the incorrect estimate of the H5 and H7 polarisation: cancellation effects will be largest due to the small hyperfine coupling

at the H5 and H7 positions [55]. Secondly, inaccuracies in the calculation of the hyperfine coupling constants from the DFT calculations. These may be improved by the use of an higher level of theory in these calculations.

7.2.2 Tryptophan Neutral Radical

The reaction mechanism for tryptophan with the triplet state flavin molecule involves electron transfer. There is evidence for a proton transfer on a microsecond time scale following this transfer [59]. It is therefore worth considering attempts to simulate the tryptophan CIDNP spectrum using hyperfine couplings obtained from DFT calculations performed on the neutral radical. This species differs from the cationic radical in that is lacks the indole NH proton. There will thus be no polarisation predicted at that position, unless, when the radicals recombine, the tryptophan neutral radical extracts a proton from the N5 position of the protonated FMN radical.

The first model for the tryptophan neutral radical is the 3-methyl indole model used by O'Malley and Ellson [247]. The hyperfine couplings are given in Table 7.3 (and in Table C.4 in Appendix C). Using these hyperfine couplings the CIDNP

H2	H4	H5	H6	H7	$H^{\beta}1$	$H^{\beta}2$	N1
0.11	−0.42	–	−0.34	−0.06	2.48	2.52	0.41

Table 7.3: Isotropic hyperfine coupling constants for the neutral tryptophan radical, from DFT calculations at the B3LYP/EPR-III level of theory on 3-methyl indole [247].

spectrum shown in Figure 7.6 was obtained. As before, the simulated spectrum has been scaled to the height of the H2 singlet. It is clear that this model does not satisfactorily reproduce the experimental spectrum. The polarisation expected is over estimated at the H4 and H^{β} protons. However, a more reasonable agreement is obtained for the H6 and H7 protons. The phase of the polarisation at H5 is still

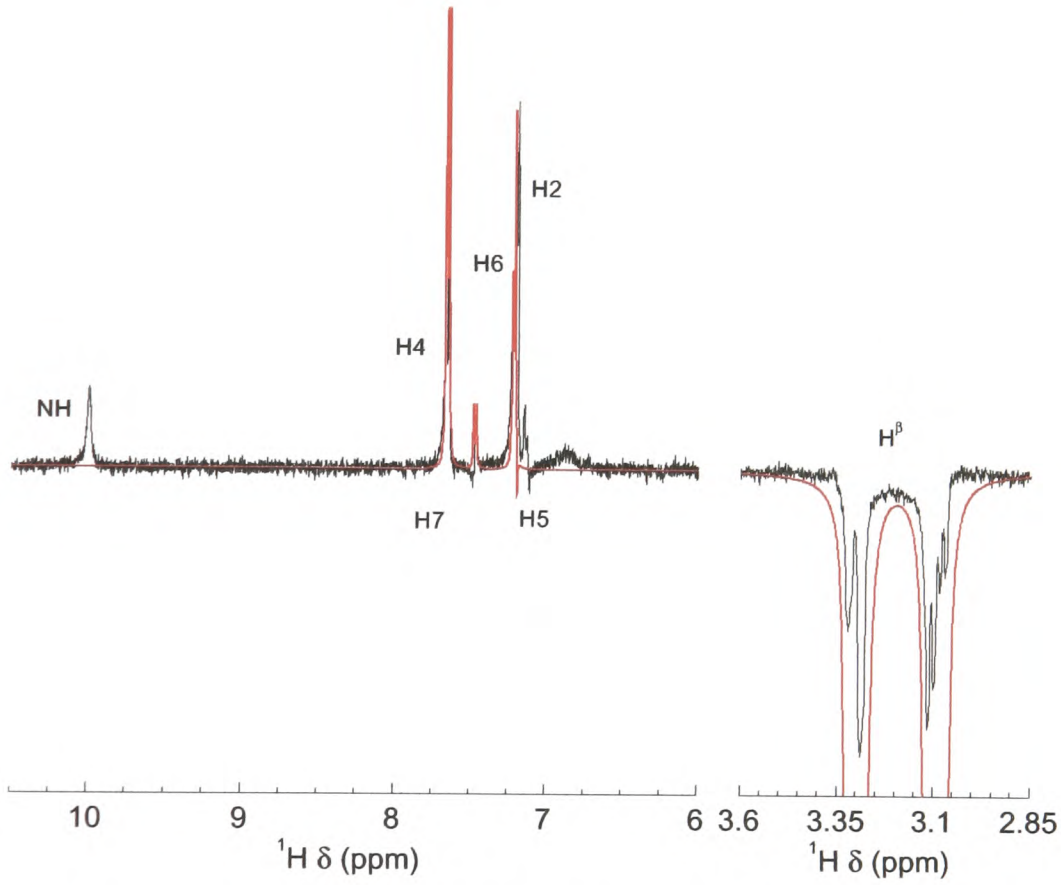


Figure 7.6: CIDNP spectra for a radical pair consisting of a tryptophan neutral radical and a neutral flavin radical. The black spectrum is the experimental spectrum and the red is the simulation, which is scaled to the height of the H2 singlet. The hyperfine couplings used are tabulated in Table C.1 [258, 259] and Tables 7.3 and C.4 [247].

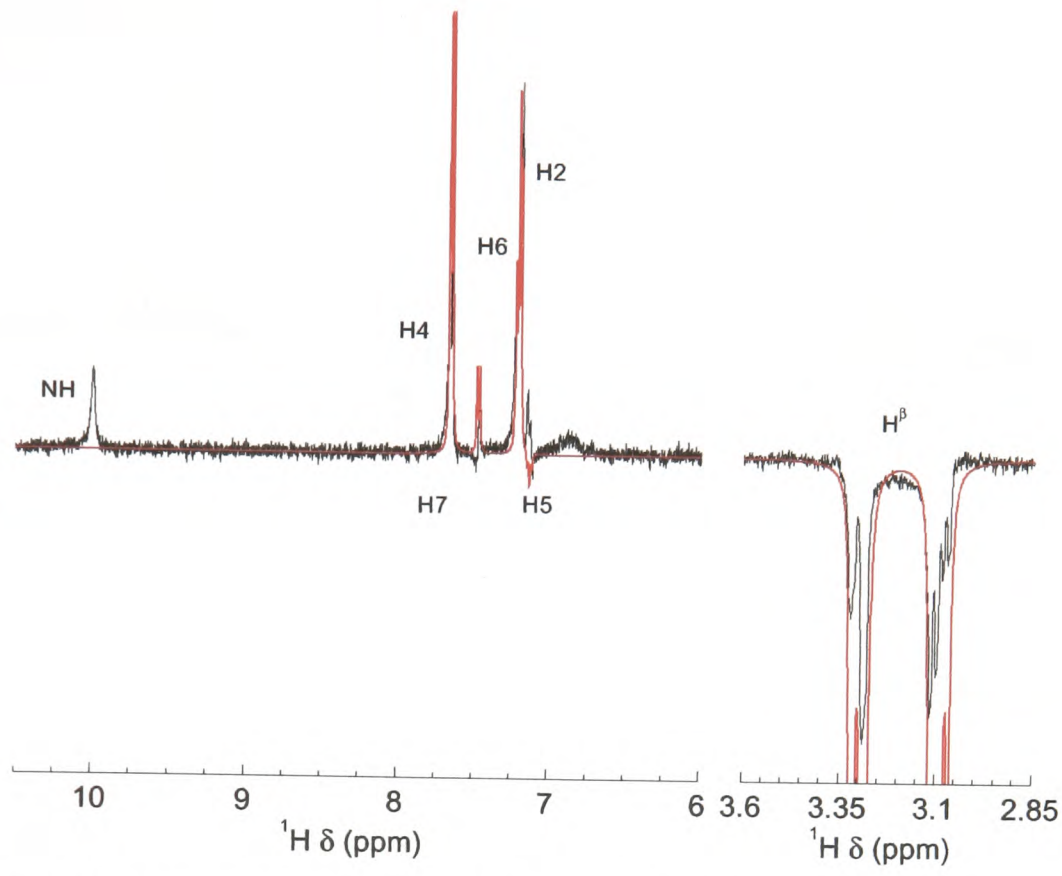


Figure 7.7: CIDNP spectra for a radical pair consisting of a tryptophan neutral radical and a neutral flavin radical. The black spectrum is the experimental spectrum and the red is the simulation, which is scaled to the height of the H2 singlet. The hyperfine couplings used are tabulated in Table C.1 [258, 259] and Tables 7.4 and C.4 [248].

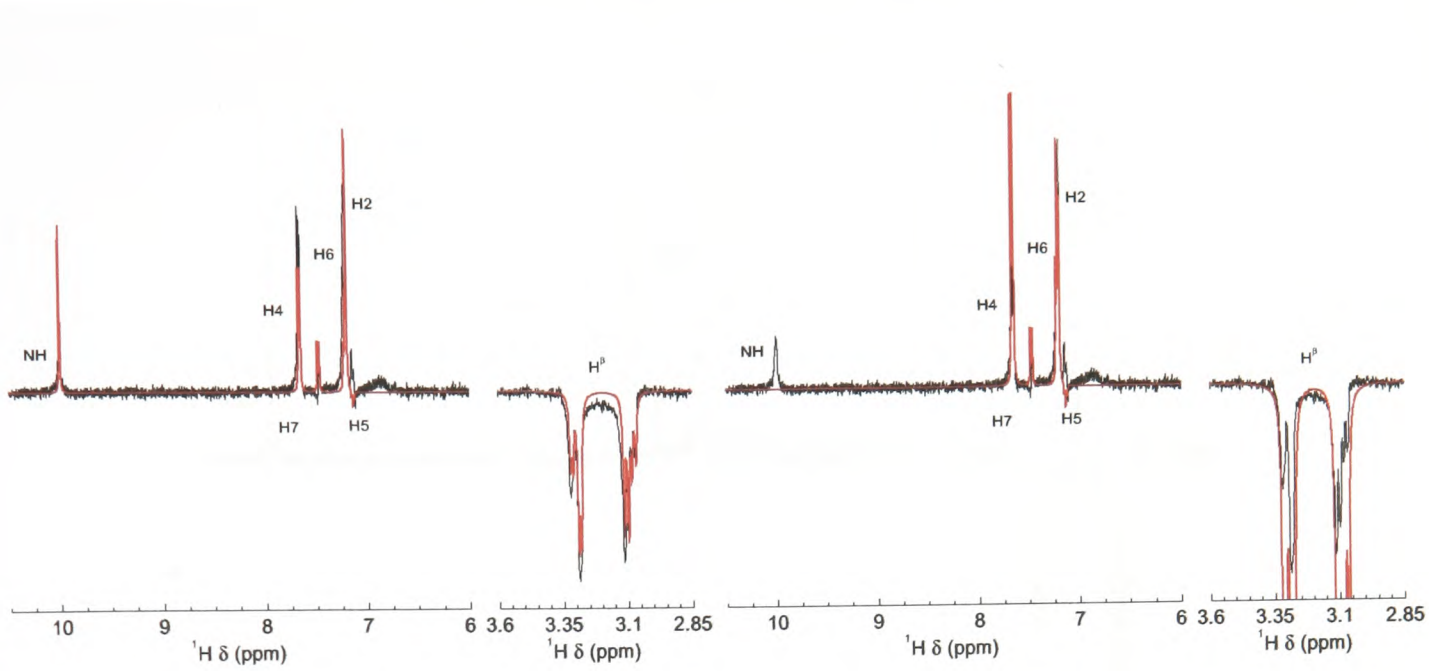
predicted incorrectly.

Using the model of Himo and Eriksson [248], that of 3-ethyl indole, does not result in any significant improvement of the simulations, as is shown in Figure 7.7. The hyperfine couplings are shown in Table 7.4 (and in Table C.4 in Appendix C). The polarisation observed at H4, H7 and H$^{\beta}$ is still overestimated. The phase of

H2	H4	H5	H6	H7	H$^{\beta}$1	H$^{\beta}$2	N1
−0.05	−0.40	0.03	−0.32	−0.08	0.29	1.08	0.25

Table 7.4: Isotropic hyperfine coupling constants for the neutral tryptophan radical, from DFT calculations at the PWP86/ILGO-III level of theory on 3-ethyl indole [248].

H5 also remains incorrect. It can be concluded from these results that the radical pair consisting of a neutral tryptophan radical and a neutral flavin radical does not



(a) Ionic radical pair

(b) Neutral radical pair

Figure 7.8: Comparison of the ionic and neutral radical pairs. The hyperfine couplings used were those for 3-ethyl indole [248].

provide the major contribution to the observed spectrum.

7.3 Discussion

From the results presented above, it is clear that a radical pair, predominantly consisting of the cationic tryptophan radical and an anionic flavin radical gives rise to the observed CIDNP spectrum. Figure 7.8 shows the experimental and simulated spectra for the ionic and neutral radical pairs side-by-side for comparison. These spectra are simulated for the case of the 3-ethyl indole [248]. Interestingly, a better fit to the experimental spectrum can be obtained by constructing a linear combination of simulated spectra: a small contribution from the neutral species is added to that from a radical pair containing the cationic tryptophan radical. Since the results above indicate that 3-ethyl indole is a better model for the tryptophan radical, the hyperfine couplings from Himo and Eriksson's calculations were used [248]. Figure 7.9 shows the best least squares fit of a linear combination of the simulations

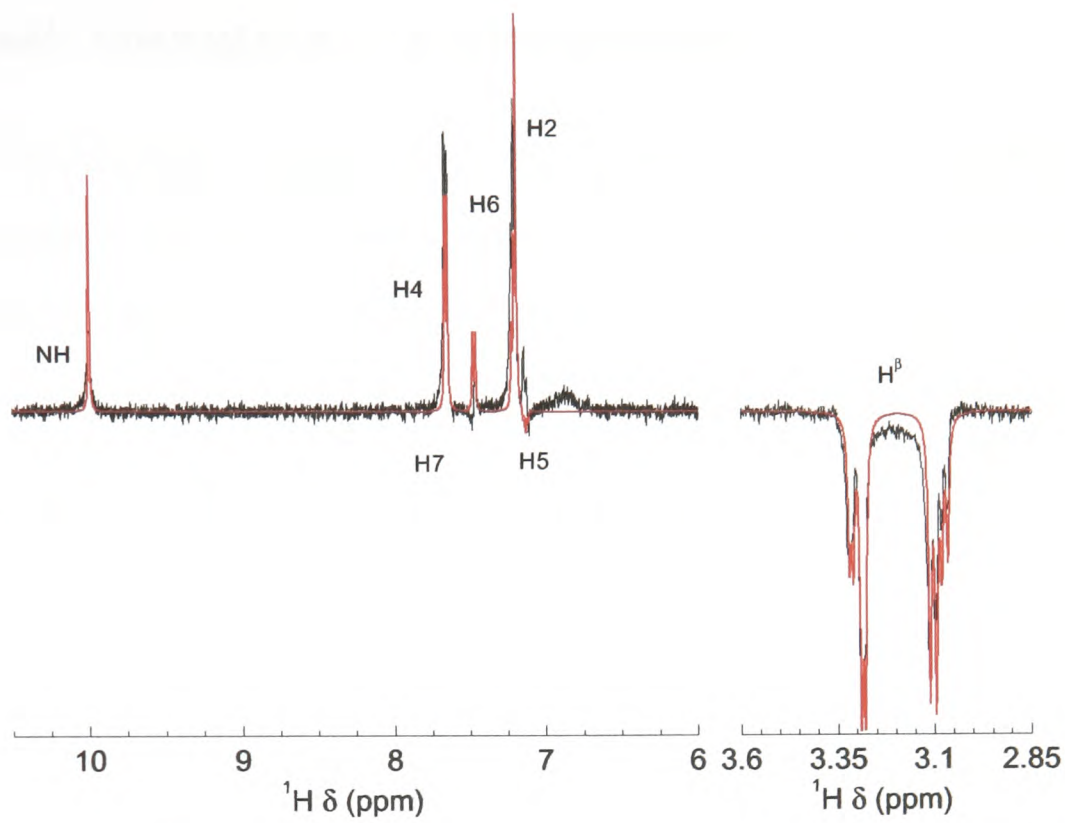


Figure 7.9: Linear combination of the simulated spectra from Figures 7.5(b) and 7.7. The determined contributions of the two radicals are 0.8661 from the cationic radical and 0.1339 from the neutral radical.

from a cationic tryptophan radical and a neutral tryptophan radical, obtained in a least-squares sense. Using this approach results in much better agreement with the experimental spectrum. There is now only a minor underestimate of the polarisation at H4 and H6, with excellent agreement for the H^{β} protons. H7 is also much closer to the experimental spectrum. Agreement for the indole proton is better, however, the linewidth of this signal in the experimental spectrum is still greater than for the simulation, for the reasons described above. This minor contribution from the neutral radical pair is to be expected: if deprotonation of the cationic tryptophan radical is slow, the observed polarisation will be small, because of the effects of degenerate electron transfer [61].

7.4 Last thoughts / Conclusions

In spite of the simplistic models used in this Chapter, reasonable agreement with the experimental results was obtained. However, more information may be obtained using a different experimental approach. In the limit of short laser pulses, so as not to accrue significant polarisation transfer via dipolar coupling, the intensity of a given signal can be expressed as:

$$I_k = \mathcal{C}_k \times t_{\text{laser}} \tag{7.24}$$

where $\mathcal{C}_k$ is the CIDNP build-up rate and t_{laser} is the duration of the laser pulse. The use of time-resolved CIDNP (tr-CIDNP) in which the polarisation is generated using a pulsed laser source (usually either an XeCl excimer laser, or a Nd:YAG laser) allows the build-up of nuclear polarisation to be followed with microsecond time resolution following the laser pulse [250, 251, 260–262]. In this regime, the effects of nuclear spin relaxation during the laser pulse can be neglected, and since the electron Zeeman interaction is much larger than the electron-nuclear hyperfine coupling, the CIDNP build-up rate is directly proportional to the hyperfine coupling.

$$\mathcal{C}_k \propto a_k \quad \Rightarrow \quad I_k \propto a_k \tag{7.25}$$

Measurement of this CIDNP build-up rate would allow the extraction of the hyperfine couplings, and given the approximations of the relevant diffusion model used, allow the identity of the participating radical to be determined. This approach also allows an indirect method for determining both the magnitude *and* sign of the hyperfine couplings in the free tryptophan radical in solution.

A more accurate approach to the simulation of the CIDNP spectrum, is to perform a full density operator calculation, which takes into account the evolution of

the correlated electron spins in the radical pair, the evolution of the nuclear spins following the recombination of the radical pair, and the subsequent RF pulse and detection. This style of calculation would still evaluate Equation 7.1, however, the spin evolution would be calculated using the Stochastic Liouville Equation [42]:

$$\frac{\mathrm{d}\vec{\sigma}(t)}{\mathrm{d}t} = \left(-\mathrm{i}\hat{\hat{L}} - \hat{\hat{\Gamma}}\right)\left(\vec{\sigma}(t) - \vec{\sigma}(0)\right) \tag{7.26}$$

where $\vec{\sigma}(t)$ is a vector in Liouville space corresponding to the density matrix and $\hat{\hat{L}}$ is the spin superoperator, calculated from the full spin Hamiltonian in the following manner:

$$\hat{\hat{L}} = \hat{H} \otimes \hat{1} - \hat{1} \otimes \hat{H} \tag{7.27}$$

where $\otimes$ is the direct product. The Hamiltonian can be constructed in such a way as to include the effects of polarisation transfer by strong J-coupling. $\hat{\hat{\Gamma}}$ is the relaxation superoperator, and describes both the electron and nuclear spin relaxation processes occurring in the system. It is in this term that processes which are responsible for cross polarisation by dipolar couplings would be included.

The probability of the radical pair being in a singlet state at time t is then given by the trace of the density vector, obtained from solving Equation 7.26 for $\vec{\sigma}(t)$, with a suitable singlet projection operator $\mathcal{S}^x$:

$$S^x(t) = \mathrm{Tr}\left[\mathcal{S}^x \cdot \vec{\sigma}(t)\right] \tag{7.28}$$

This is then combined with Equation 7.1, and can be used to calculate the observed polarisation at a given nucleus.

Chapter 8

NMR Characterisation of the Homologous 4K Peptide Series

To obtain a complete understanding of the protein folding process and its implications, it is important to know the three dimensional folds, that is the tertiary (and quaternary) structures of proteins. Classically, this has been performed by X-ray crystallography. Protein structure determination in the solution state is principally performed by fitting the known primary structure of the protein to a set of experimentally derived NMR constraints using restricted molecular dynamics simulations [74]. These constraints are typically of two types: distance constraints, obtained from nuclear Overhauser effects observed in NOESY[1] spectroscopy; and torsion angle constraints, principally obtained from scalar coupling data. Observable distances between protons are usually limited to around 5 Å since the magnitude of the NOE decays with increasing separation of the protons as r^{-6}. The use of chemically pumped ^{19}F NOEs may extend this range to approximately 8 Å [65]. Further constraints can be obtained from residual dipolar couplings by weakly aligning the protein molecules within the magnetic field of the NMR spectrometer. This weak

[1]Nuclear Overhauser Effect SpectroscopY

alignment reintroduces part of the dipolar interaction between the H^N and N^H nuclei of the amide bond and allows the orientation of the NH bond vector, relative to the alignment tensor of the protein, to be obtained [263].

Recently, the use of site-directed spin-labelling has allowed paramagnetic probes, usually stable nitroxide radicals, to be attached to proteins [264]. These probes have found a wide variety of uses, both in EPR and in NMR studies of proteins. The influence of the paramagnetic centre on the NMR spectrum allows information about a range of protein interactions to be obtained, for example protein-ligand interactions [265], domain-domain docking [266]. Spin labels have been used in the study of membrane proteins where the depth and angle of insertion of the protein into the membrane can be determined [267].

Information about the spin label environment is also available from EPR studies, via changes in the g and hyperfine tensors. Using pairs of spin labels allows the distance between the paramagnetic centres to be measured, using the so-called DEER[2] experiment. This experiment involves creating an electron spin echo in such a manner that it is modulated by the electron-electron dipolar coupling [268, 269]. Analysis of these modulations enables distance information to be obtained [270]. There has, however, been little comparison of this pulsed method with traditional continuous wave EPR spectroscopy, where the dipolar coupling is manifested as a contribution to the spectral linewidth.

A set of peptides, specially designed for the work described in this chapter will be used to compare CW-EPR and DEER methods for obtaining the distances between the spin labels. To this end, information at atomic resolution on the secondary structure adopted by the peptides is necessary. This allows comparison of the distances obtained from the EPR measurements with those expected from the peptide secondary structure [271]. This chapter describes the NMR characterisation of a

[2]Double Electron Electron Resonance

homologous series of *de novo* peptides, which will be used in CW-EPR and DEER experiments when modified using nitroxide spin labels.

8.1 The Peptides

The series of peptides under investigation in this chapter was designed around a poly(alanine) sequence spanning twenty-one residues. Every fifth residue was replaced with lysine in order to increase the helical propensity of the sequence [272], and hence results in the name "4K" to describe the series. In order to provide points to attach the spin labels, two of the alanine residues were replaced with cysteines at particular locations along the sequence. The variable separation of these cysteine residues allows different distances between the spin labels to be obtained. Table 8.1 gives the primary sequences of the six peptides which comprise this 4K series, along with individual names. The chemical structures of the three amino acids comprising the peptide series are shown in Figure 8.1. The peptides were synthesised by Ms. J. E. Banham using standard solid phase Fmoc[3] chemistry [273].

Name	Sequence				
	5	10	15	20	
4K(3,7)	AACAK	ACAAK	AAAAK	AAAAK	A
4K(3,11)	AACAK	AAAAK	CAAAK	AAAAK	A
4K(3,12)	AACAK	AAAAK	ACAAK	AAAAK	A
4K(3,13)	AACAK	AAAAK	AACAK	AAAAK	A
4K(3,14)	AACAK	AAAAK	AAACK	AAAAK	A
4K(3,18)	AACAK	AAAAK	AAAAK	AACAK	A

Table 8.1: Names and amino acid sequences for the 4K peptides. The amino acids are designated by their single letter codes (A=alanine, C=cysteine, K=lysine).

[3] 9-Fluorenylmethyloxycarbonyl

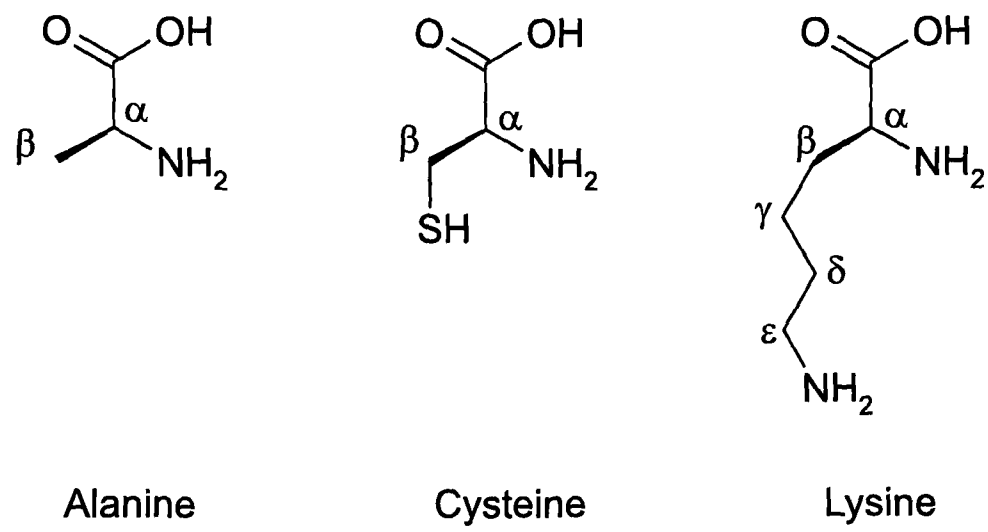


Figure 8.1: Structures of the amino acids alanine (A), cysteine (C) and lysine (K). The notation used to identify the side chain positions is shown also.

8.2 Initial Investigation

The 4K peptides were designed to have a strong propensity for the formation of helical secondary structure. It is, however, possible to further increase the helicity of an amino acid sequence by the addition of a cosolvent such as 2,2,2-trifluoroethanol (TFE). The presence of TFE stabilises the helix by interaction with the hydrogen bonding network along the peptide backbone [200]. For example, the helical content of the protein hen egg white lysozyme is increased by the presence of TFE in concentrations of 20% (v/v) or greater [149, 150], as shown in Figure 4.20, Chapter 4. For the 4K peptides an increase in the helical content is observed on increasing the TFE concentration, with a plateau observed above 40% (v/v) [271]. Therefore, all spectroscopic measurements were performed in the presence of 50% TFE (v/v), at which point no further increase in helicity was observed [271]. This concentration is also chosen to be the same as that used in the EPR experiments performed on these peptides.

As a first estimation of the secondary structure adopted by the 4K peptides, a series of far-UV circular dichroism spectra were recorded, each in 50% TFE (v/v), and at a temperature of 4°C so as to be consistent with the NMR data shown later

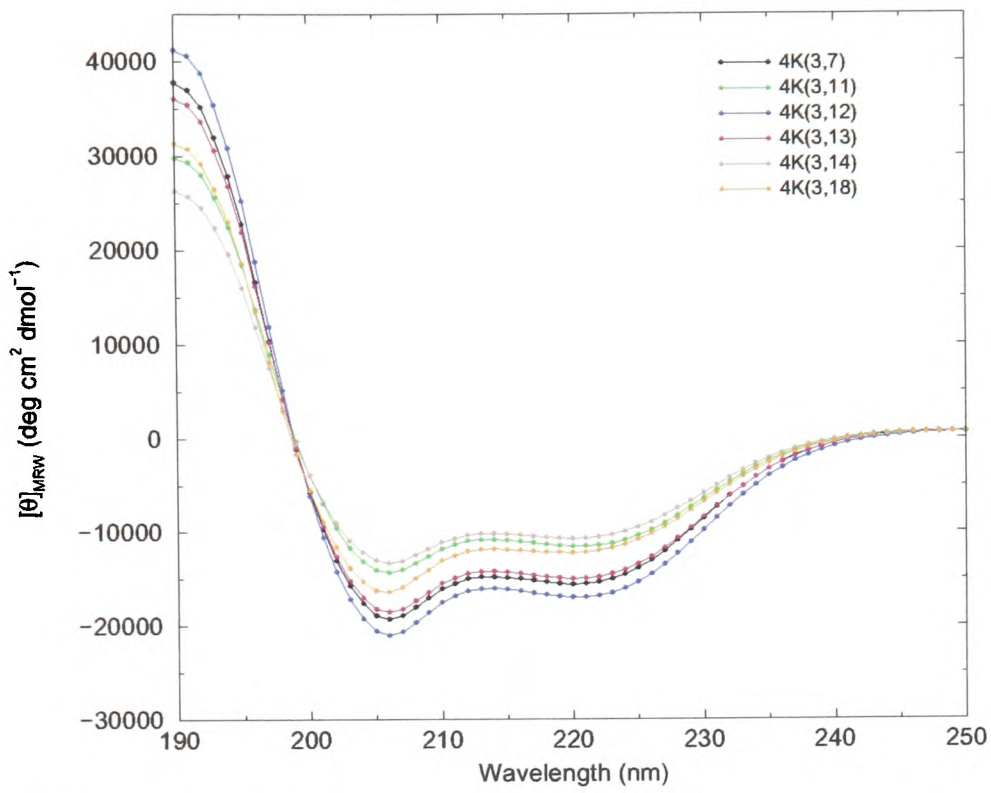


Figure 8.2: Far-UV circular dichroism spectra of 100 μM solutions of the 4K peptides in 50% TFE (v/v) at 4°C. The spectra were recorded by Ms Janet E. Banham.

in this Chapter. These spectra clearly show a large degree of helical secondary structure. This is evident by the strong negative mean residue ellipticity at 205 nm and 222 nm, both of which are characteristic of a helix [84]. While each peptide gives rise to CD spectra with similar shape, there is, however, some variation across the series. This arises principally from inaccuracies in determining the weight of peptide used to prepare the samples for circular dichroism.

Following the circular dichroism, the one dimensional NMR spectra of the peptide series were assessed. These are shown in Figure 8.3. The spectra were recorded under the same conditions as for the circular dichroism measurements, except for an increase in sample concentration. The spectra were all referenced to the chemical shift of the residual TFE signal, observed at 3.88 ppm. Despite the fact that these peptides contain only three distinct types of amino acid residue, and despite the repetitive nature of the primary sequence, there is reasonable resolution in the NMR spectra. There are also clearly subtle differences between the different peptides comprising the series, most noticeably in the amide region, between 7.5 and 9.0 ppm,

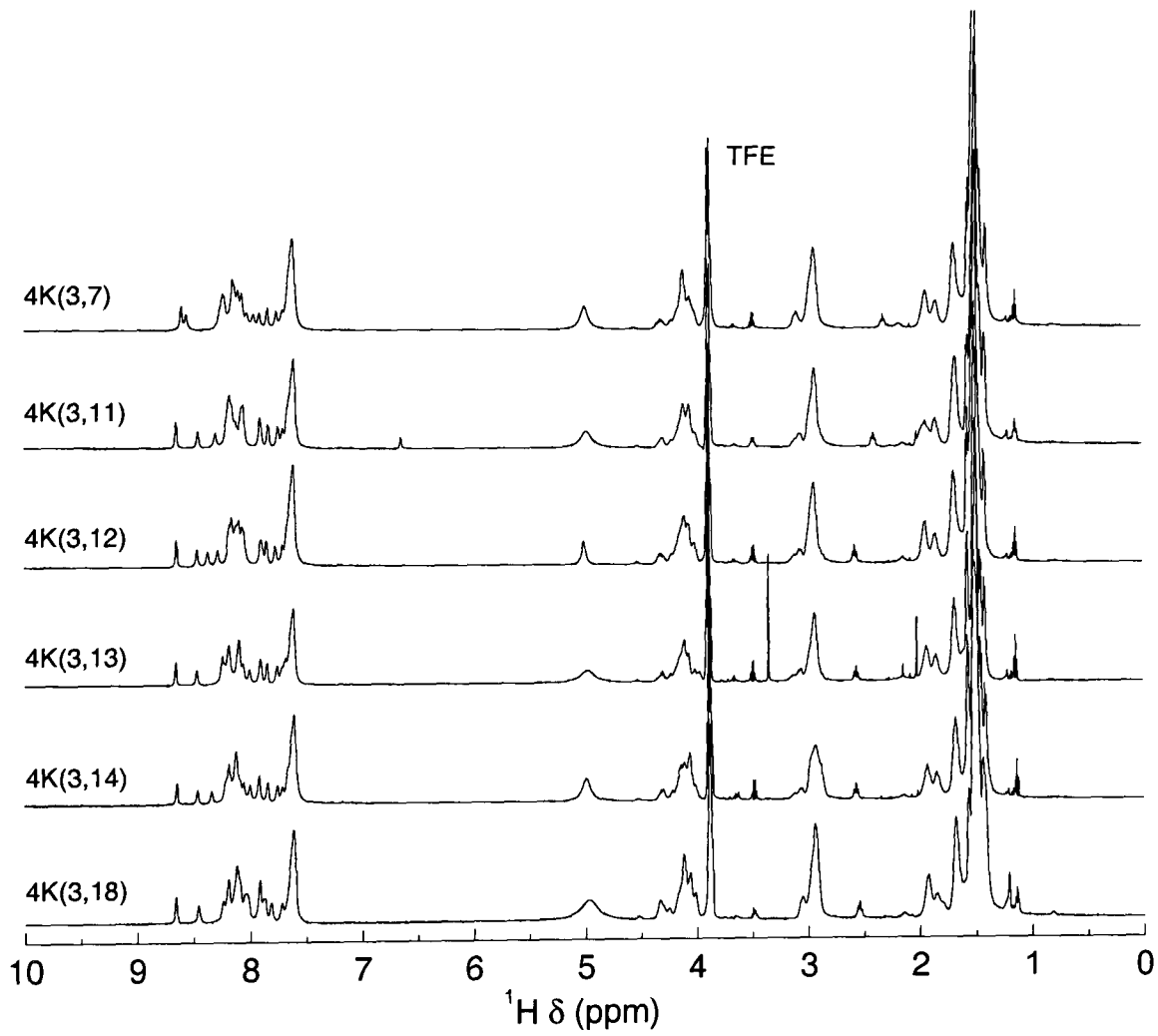


Figure 8.3: ^{1}H NMR spectra of the 4K peptides recorded at 4°C. The samples contained 5 mM peptide in 50% TFE-d_3 (v/v), 50% H_2O.

and in the H^{α} region, between 4.0 and 4.5 ppm. These observations indicate that the peptides are folded under these conditions and have well-defined secondary structure. This is in general agreement with the data obtained from CD spectroscopy, and suggests that these peptides will be amenable to further characterisation.

8.3 NMR Characterisation

Following on from the results presented above, this section describes the more detailed NMR characterisation of the 4K peptides, first outlining the methodology used to obtain the data, and then working through the data collection process, using one of the peptides as an example.

8.3.1 Methodology

There are several approaches available to obtain the sequence specific chemical shift assignment, and hence structural information, of a peptide or protein [74]. The method used in this chapter to gather the ^{1}H assignments loosely follows that proposed by Wüthrich in which homonuclear scalar coupling correlations are used in conjunction with NOE information to provide connectivities [274, 275]. These assignments can then be easily extended to the side chains, enabling complete assignment of the peptide or protein. The protocol used for the sequential assignment in this chapter is broadly as follows:

1. Identify individual amino acid spin systems using (total) correlation spectroscopy.

2. Obtain sequential connectivities on the basis of backbone NOE analysis.

3. Extend assignments to the complete spin systems, including side chains.

Once the sequential assignment has been performed, more detailed analysis of the NOESY spectra allows secondary (and potentially tertiary) structural information to be extracted. These data can then be used in further characterisation of the peptide or protein.

For the 4K peptide series investigated in this Chapter, TOCSY[4] experiments were used to identify the individual amino acid spin systems. These were usually recorded using an isotropic mixing time of 80 ms, employing the MLEV17c scheme [76]. The NOESY experiments used in both sequential assignment and structural characterisation were recorded using a mixing period of 150 ms. Both experiments employed presaturation of the solvent signal prior to execution of the pulse sequence.

All of the NMR spectra presented in this chapter were recorded at 4°C in order to increase the intensity of the NOESY cross peaks. As described in Chapter 2 the NOESY cross peak intensity is determined principally by the duration of the mixing time (τ_m) and the rate of cross relaxation between the two interacting spins, given by the following expression:

$$R_{\text{cross}} = \frac{\mu_0^2 \hbar^2 \gamma^4}{40\pi^2 r^6} \left\{ \tau_c - \frac{6\tau_c}{1 + 4\omega_0^2 \tau_c^2} \right\} \tag{8.1}$$

where τ_c is the rotational correlation time, r is the distance between the pair of protons experiencing the NOE, and ω_0 is the proton Larmor frequency. All other symbols have their usual meaning. This equation assumes a pure dipole-dipole relaxation mechanism [75]. In the slow tumbling limit, *i.e.* when $|\omega_0 \tau_c| \gg 1$, this equation simplifies to the following:

$$R_{\text{cross}} \approx \frac{\mu_0^2 \hbar^2 \gamma^4 \tau_c}{40\pi^2 r^6} \tag{8.2}$$

[4]Total COrrelation SpectroscopY

Therefore, the rate of cross relaxation, and hence the cross peak intensity, depends on the rotational correlation time of the peptide or protein under investigation.

For a sphere of hydrodynamic radius r_{H}, undergoing isotropic rotational motion in a solution of viscosity η, the rotational correlation time is given by [74]:

$$\tau_c = \frac{4\pi\eta r_{\mathrm{H}}^3}{3k_{\mathrm{B}}T} \tag{8.3}$$

hence a reduction in the temperature will result in slower tumbling (a longer correlation time), and thus in greater cross peak intensity. The spectra presented in Figure 8.2 indicate that the 4K peptides are helical, hence are rod-shaped, and do not undergo isotropic rotation in solution. This absence of rotational isotropy leads to more a complex expression for the rotational correlation time and for the cross relaxation rate. However, the general premise is still valid: that a decrease in temperature leads to an increase in the effective rotational correlation time.

In the context of peptides and proteins the proton-proton distances, and hence the NOEs, are described using the following notation:

$$d_{\mathrm{AB}}(i,j) \equiv d(\mathrm{H}_i^A, \mathrm{H}_j^B) \tag{8.4}$$

which describes the distance between the proton at position A of residue i to the proton at position B of residue j. So for example, the NOE arising from the H^α proton of residue 10 to the H^N proton of residue 13 is denoted:

$$d_{\alpha N}(10,13) \equiv d(\mathrm{H}_{10}^\alpha, \mathrm{H}_{13}^N) \tag{8.5}$$

By convention, if the residues are sequential (*i.e.* if $j = i+1$) then the indices i and j are omitted [274, 275].

8.3.2 A Worked Example — 4K(3,18)

In order to illustrate the procedures used to obtain the results shown in this Chapter, one of the peptides will used as a "worked example". Given the similarity of the peptides in the 4K series, the choice of peptide is unimportant; in this case 4K(3,18) is used.

Following on from the one dimensional spectrum shown in Figure 8.3, identification of the individual amino acid spin systems was performed using a TOCSY experiment. Figure 8.4 shows the scalar coupling correlations arising from the H^N protons along the side chains. The identification of the spin systems is straightforward given that only three types of residue are present in the 4K(3,18) peptide. The sulphydryl group of the cysteine side chain causes the H^β protons to be more deshielded compared to those of the alanine residues, and hence they tend to resonate at lower field, usually in the region 2.6 to 3.4 ppm [74]. Characteristic signals from these cysteine H^β protons are observed in region (a) of Figure 8.4. Similarly, due to the larger number of protons on the lysine side chains, these residues can be clearly identified in the highlighted region (b) of Figure 8.4. This region principally contains signals from the H^N to H^γ and H^δ correlations. Once the resonances from the cysteine and lysine residues in the spectrum have been identified, the remainder of the signals must, by a process of elimination, correspond to the remaining alanine residues. Figure 8.5 shows an enlargement of the region corresponding to the $H^\alpha \leftrightarrow H^N$ scalar coupling correlation, with the amino acid type responsible for that signal indicated. There are twenty observable cross peaks in this portion of the spectrum. The H^α to H^N correlation for the N-terminal residue is not observed as H^N undergoes rapid exchange with the solvent.

In order to determine the sequence specific assignments for 4K(3,18) the amino acid spin systems must be connected together in the correct, *i.e.* sequence, order. With ^{13}C, ^{15}N isotopically labelled samples this is best achieved using triple-

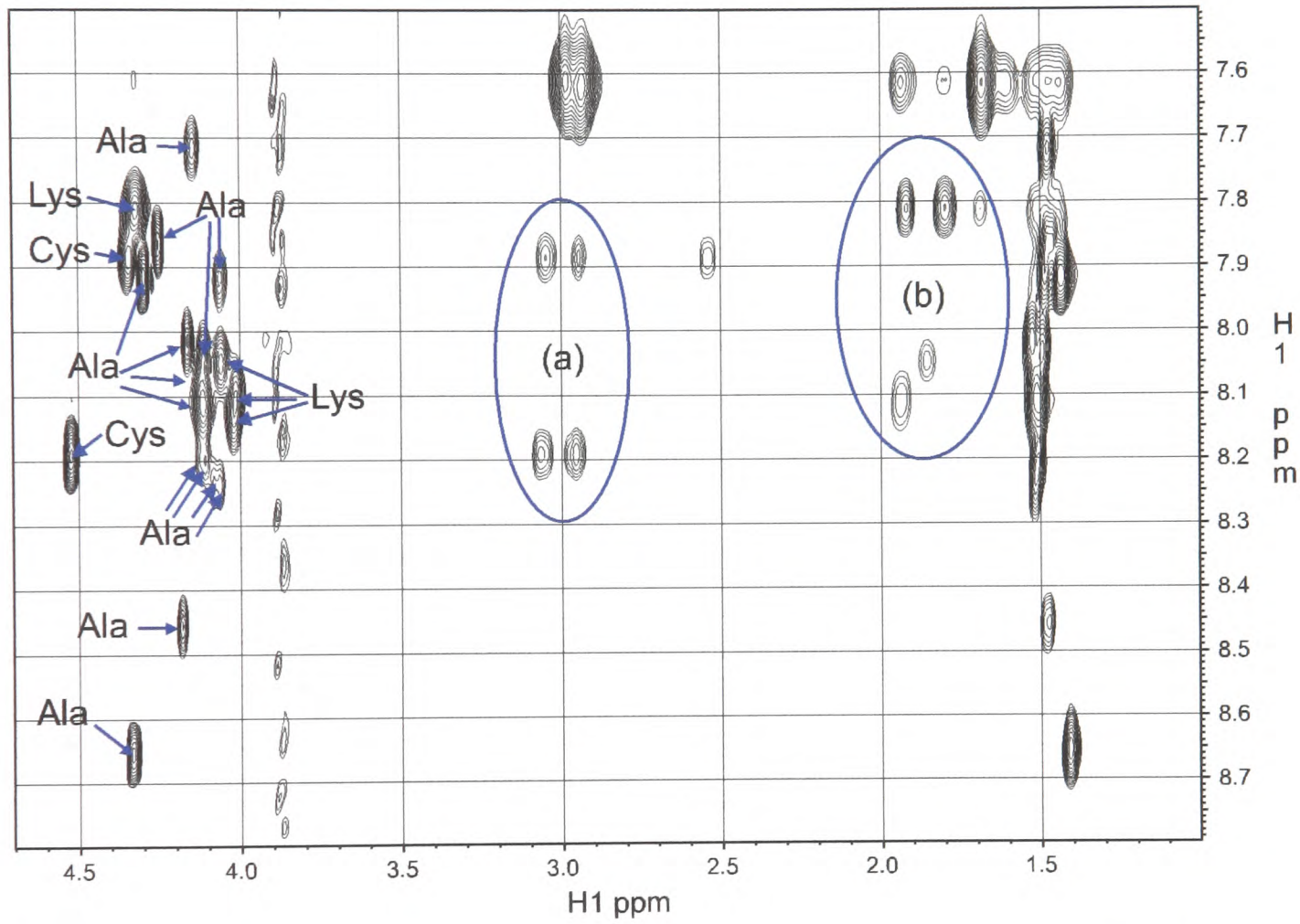


Figure 8.4: 600 MHz ^{1}H TOCSY spectrum (80 ms mixing time) showing the H^N correlations to H^α, H^β and other side chain protons for the 4K(3,18) peptide. 1024 complex points were acquired in the directly detected dimension for each of 256 increments. The spectrum was processed with a cosine-squared window function and zero-filled to 2048 points in each dimension prior to Fourier transformation. The H_2O signal was removed by deconvolution with a sine-bell extending over 32 points. The spectrum is referenced to the residual TFE signal at 3.88 ppm.

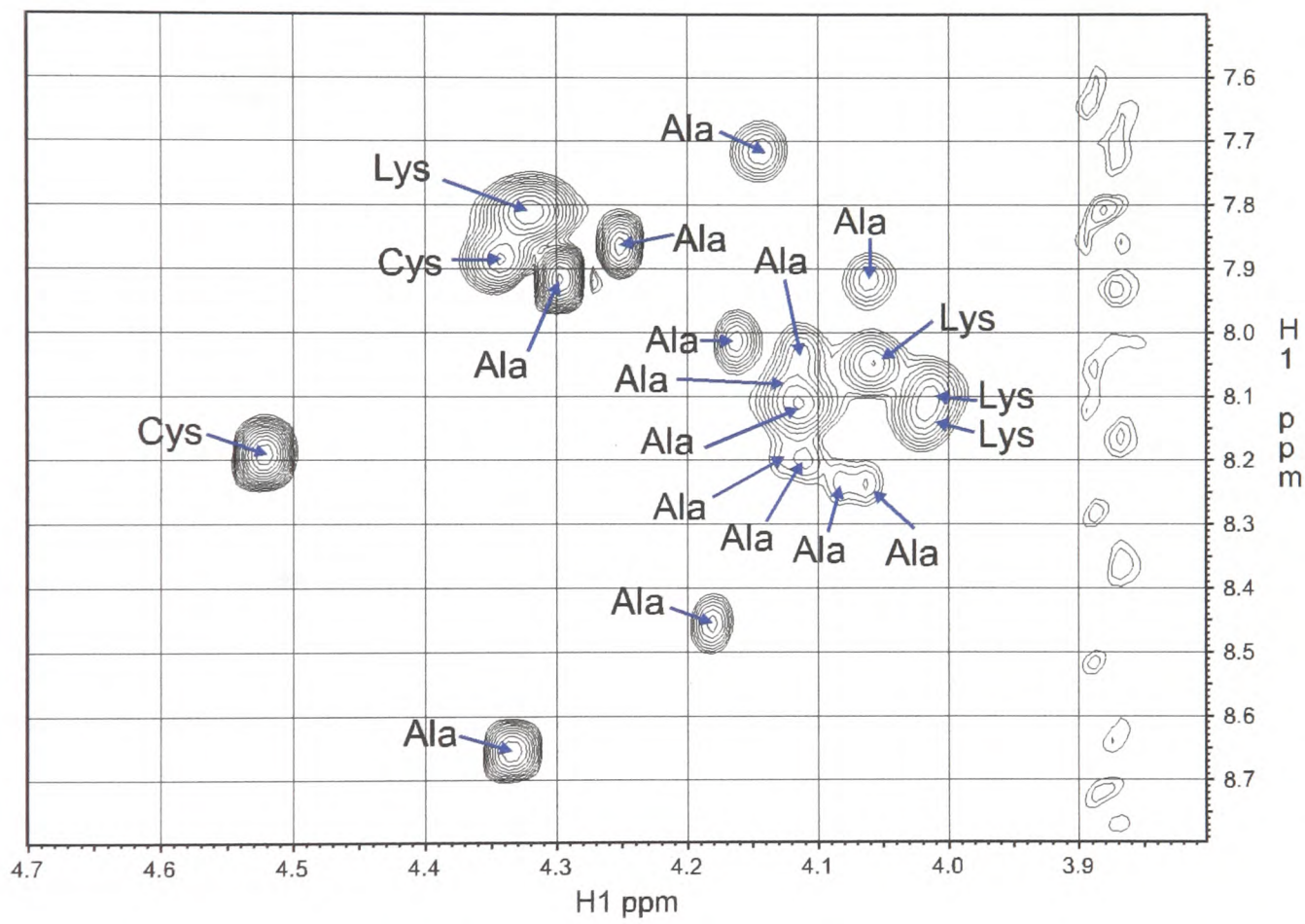


Figure 8.5: Enlargement of the $H^{\alpha} \leftrightarrow H^{N}$ region of Figure 8.4. The stripe at 3.88 ppm arises from the residual TFE signal.

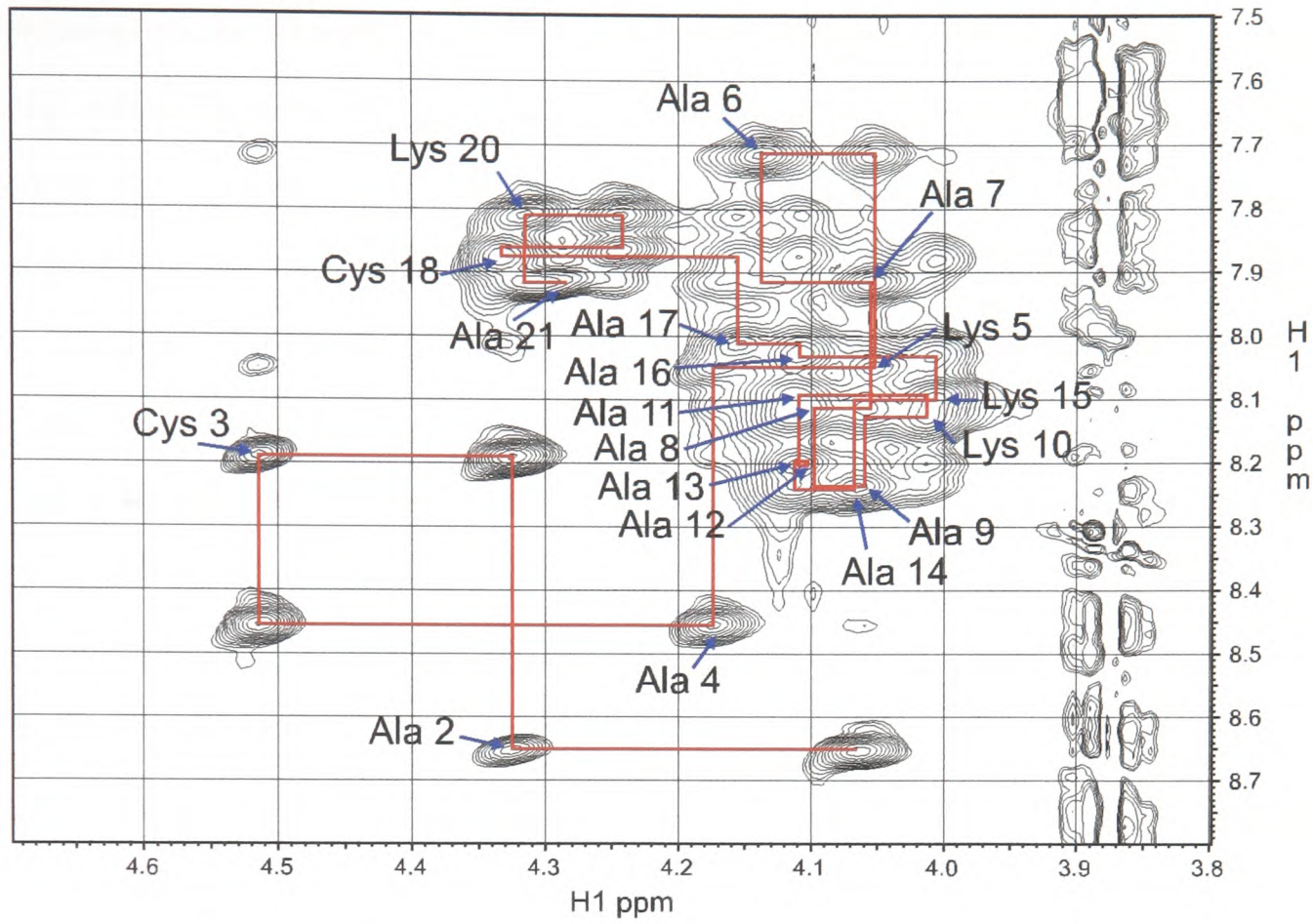


Figure 8.6: 600 MHz ^{1}H NOESY spectrum (150 ms mixing time) showing the H^{α} to H^{N} correlations for the 4K(3,18) peptide. 1024 complex points were acquired in the directly detected dimension for each of 256 increments. The spectrum was processed with a cosine window function and zero-filled to 2048 points in each dimension prior to Fourier transformation. The H_2O signal was removed by deconvolution with a sine-bell extending over 32 points. The spectrum is referenced to the residual TFE signal at 3.88 ppm. The red line indicates the "walk" along the peptide backbone.

resonance techniques such as HNCA, HN(CA)CO and HCA(CO)N. As this is a small peptide it should be possible to determine the connectivities using homonuclear, through space techniques, in this case NOESY spectroscopy. Figure 8.6 shows the H^{α} to H^{N} region of a NOESY spectrum recorded under the same conditions as for the previous TOCSY experiment, using a mixing time of 150 ms, which was found to give the most satisfactory results. Comparison of this spectrum with Figure 8.5 allows the intraresidue NOEs, that is $d_{\alpha N}(i,i)$ interactions, to be identified as these will appear at the same positions as the resonances in the TOCSY spectrum. Once this has been performed, sequential NOEs, *i.e.* $d_{\alpha N}$ must be identified. This procedure makes it possible to "walk" along the peptide backbone from the

N-terminus to the C-terminus. The route taken is shown as the red line in Figure 8.6, where each line segment connects the resonances from the $d_{\alpha N}(i,i)$ and $d_{\alpha N}$ interactions. The sequential NOEs are usually straightforward to assign as these cross peaks are of comparable intensity to those arising from the intraresidue NOEs. Once this connectivity of the amino acid residues has been obtained, the backbone assignment is complete.

Following the identification of the H^{α} and H^{N} protons, further inspection of the TOCSY spectrum allows the assignment of the H^{β} protons to be made, in this way extending the assignment out to the side chains. No attempt was made to obtain the stereospecific assignment of the cysteine H^{β} protons, neither was a complete assignment of the lysine side chains performed. The backbone and partial H^{β} chemical shift assignments for 4K(3,18) are given in Table D.6, Appendix D.

The H^{α} chemical shifts obtained from the above analysis can be used to predict the probable types of secondary structure present along the amino acid sequence. This is achieved by using the Chemical Shift Index (CSI) [276] and will be discussed in more detail in Section 8.4.

Once the sequence specific assignment has been completed it is then possible to perform a more detailed analysis of the resonances observed in the NOESY spectrum; identifying medium range NOE interactions in order to gain structural information about the peptide. To illustrate this, Figure 8.7 shows slices taken parallel to the F_1 axis from both the TOCSY experiment (Figures 8.4 and 8.5) and from the NOESY experiment (Figure 8.6). These slices are taken through the centre of the cross peak arising from the $H^{\alpha} \leftrightarrow H^{N}$ correlation for Cys 3. The TOCSY experiment shows a single resonance arising from the scalar coupling correlation between the H^{α} and H^{N} protons. The intraresidue NOE ($d_{\alpha N}(3,3)$) for this cysteine is clearly identifiable, occurring at the same position as in the TOCSY trace. Four further signals are also observed in the slice from the NOESY spectrum. The strongest of these

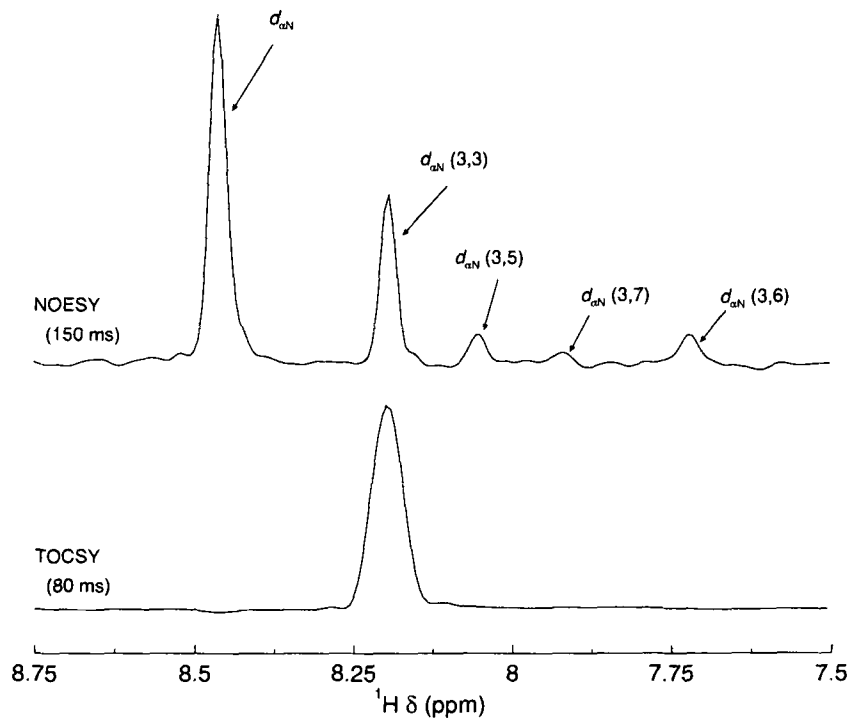
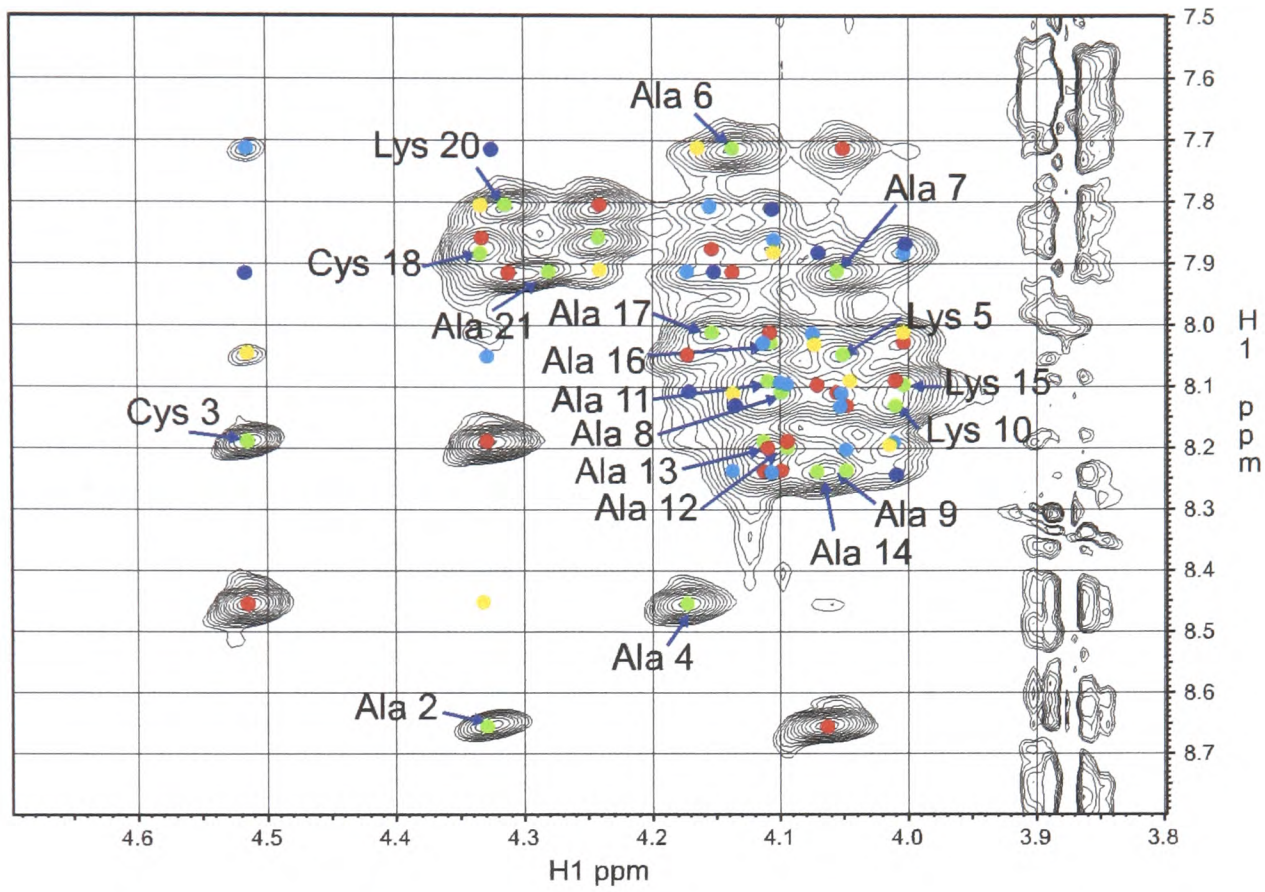


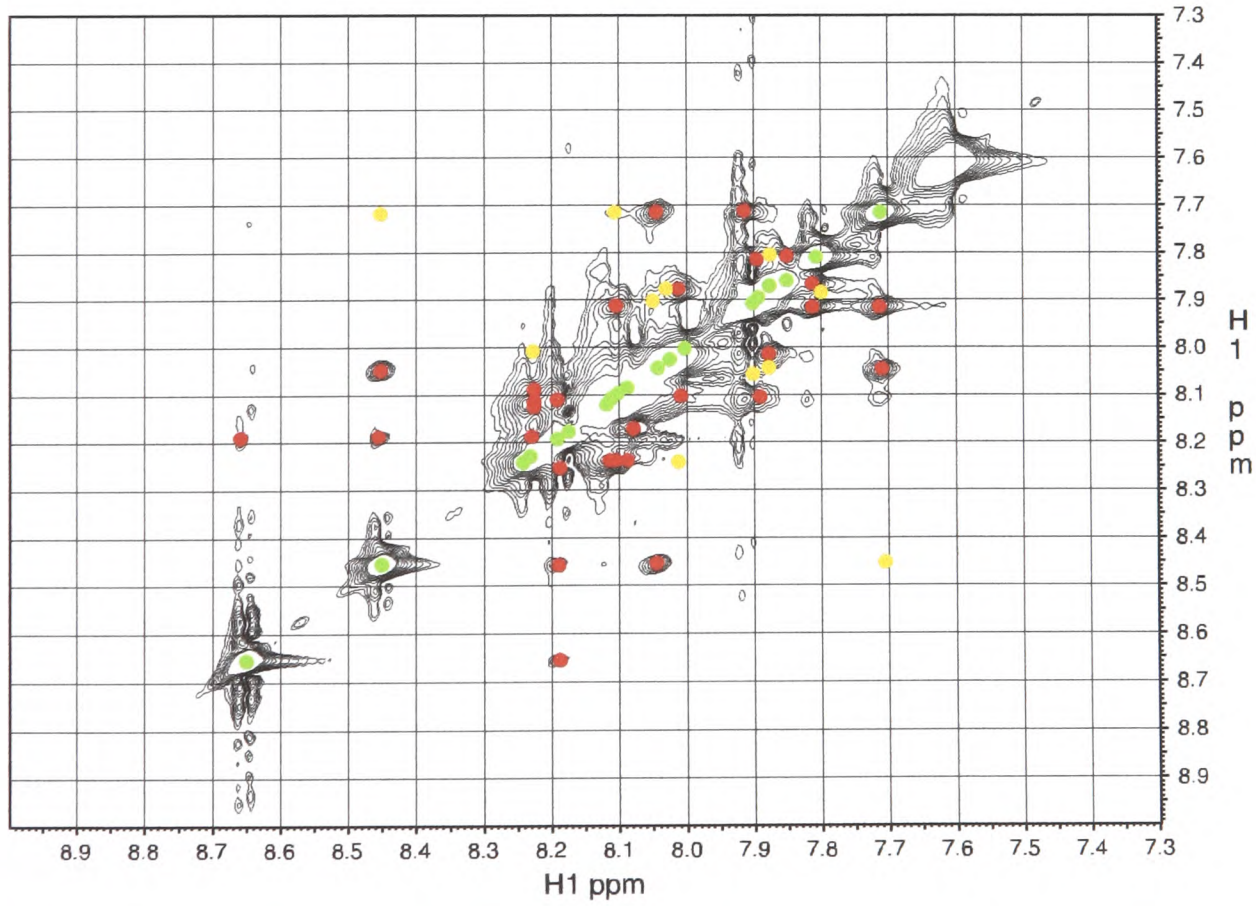
Figure 8.7: Slices taken at 4.51 ppm (the H^{α} position of Cys 3), parallel the F_1 axes of Figures 8.4 and 8.6. The $d_{\alpha N}(i,j)$ NOEs are identified on the trace from the NOESY spectrum.

corresponds to the sequential NOE $d_{\alpha N}$. The smaller signals, in order of intensity, correspond to $d_{\alpha N}(3,5)$, $d_{\alpha N}(3,6)$ and finally $d_{\alpha N}(3,7)$. That is they correspond to the NOEs arising from the $i \rightarrow i+2$, $i \rightarrow i+3$ and $i \rightarrow i+4$ NOE interactions. This pattern of observed NOEs is consistent with the cysteine residue being part of a helix [274, 275]. The same approach to the identification of the medium range NOEs is then performed for the complete amino acid sequence. Figure 8.8(a) shows the same region of the NOESY spectrum as in Figure 8.6, identifying all NOEs that can be attributed to specific interactions. The colour scheme used is that $d_{\alpha N}(i,i)$ are green, $d_{\alpha N}(i,i+1)$ are red, $d_{\alpha N}(i,i+2)$ are orange, $d_{\alpha N}(i,i+3)$ are light blue and $d_{\alpha N}(i,i+4)$ are dark blue. A total of fifty-nine NOEs arising from H^{α} to H^{N} can be seen in this region of the spectrum. These NOEs are well distributed along the amino acid sequence indicating the adoption of a well-defined secondary structure.

The second set of through-space correlations which are important for structural analysis of the peptide backbone are those arising between pairs of amide protons,



(a) $d_{\alpha N}(i,j)$



(b) $d_{NN}(i,j)$

Figure 8.8: Identification of all NOE cross peaks taken from Figure 8.6. The NOEs are identified using the following colour scheme: $d_{\alpha N}(i,i)$ are green, $d_{\alpha N}(i,i+1)$ are red, $d_{\alpha N}(i,i+2)$ are orange, $d_{\alpha N}(i,i+3)$ are light blue and $d_{\alpha N}(i,i+4)$ are dark blue. Similarly, $d_{NN}(i,i)$ are green, $d_{NN}(i,i+1)$ are red and $d_{NN}(i,i+2)$ are orange.

that is the $d_{NN}(i,j)$ NOEs [274, 275]. Figure 8.8(b) shows this amide region of the NOESY spectrum between 7.3 and 9.0 ppm, using the same colour scheme as in Figure 8.8(a). From this spectrum twenty one NOEs have been identified, comprising fifteen d_{NN} interactions and six $d_{NN}(i,i+2)$ interactions.

The results of this NOE analysis are presented in Figure 8.10(f). The identification of patterns of NOEs allows structural information to be elucidated, based on the expected proton-proton distances in the various types of regular secondary structure [274, 275]. The following section presents the results obtained using the method described in this section applied to the complete series of 4K peptides.

8.4 Results

The strong sequence identity between the peptides of the 4K series enables the results of the chemical shift assignment process to be checked, since there is expected to be little difference between the member peptides, except at the cysteine replacement sites. The analysis of the chemical shift and NOE data obtained from the NMR experiments described above is given below, in terms of Wishart's chemical shift index and interpretation of the observable NOE patterns.

8.4.1 Chemical Shift Index

Wishart and coworkers have undertaken an analysis of the chemical shift assignments of a large number of proteins and noted that the deviation of the measured H^{α} chemical shift from its random coil value is strongly correlated with the type of secondary structure at that position in the sequence [277, 278]. The CSI approach functions as follows: H^{α} protons which are shielded (upfield) compared with the random coil position tend to be found in α-helices while those residues in β-sheets tend to have H^{α} chemical shifts which are found downfield of the random coil po-

sition. This leads to the concept of the Chemical Shift Index (CSI), which can be defined as follows [276]:

$$\delta_{\mathrm{CSI}} = \delta_{\mathrm{obs}} - \delta_{\mathrm{rc}} \tag{8.6}$$

If a residue has a CSI more negative than -0.1 ppm, and is surrounded by residues with similar values of the CSI, then it is deemed to be part of a helical region of secondary structure. Similarly, a residue with a CSI more positive that $+0.1$ ppm, and surrounded by residues with similar values, is likely to be part of a β-sheet structure. Values intermediate between these two cases, *i.e.* $-0.1 < \delta_{\mathrm{CSI}} < +0.1$ ppm, tend to arise from residues lacking well-defined secondary structure.

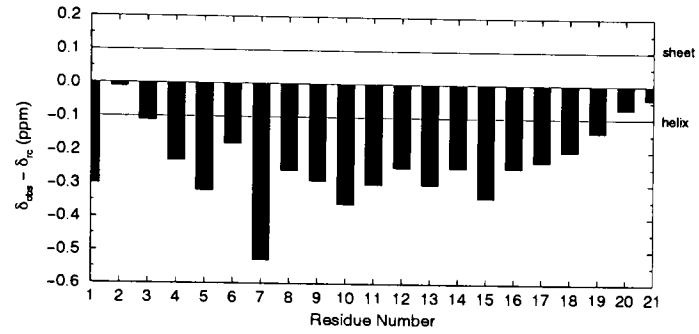
A study by Merutka *et al.* has found that the H^{α} chemical shifts are relatively unperturbed by the presence of trifluoroethanol, while the H^{N} protons are significantly perturbed [200, 279]. This therefore allows the CSI classification technique to be used for the analysis of the H^{α} chemical shift information obtained for the 4K peptides.

Figure 8.9 shows the CSI plotted as a function of the amino acid sequence for the 4K peptides. The chemical shifts for each peptide in the 4K series are given in Appendix D. The random coil chemical shifts used for the calculation of the CSI are given in Table 8.2. The chemical shift index clearly provides strong evidence

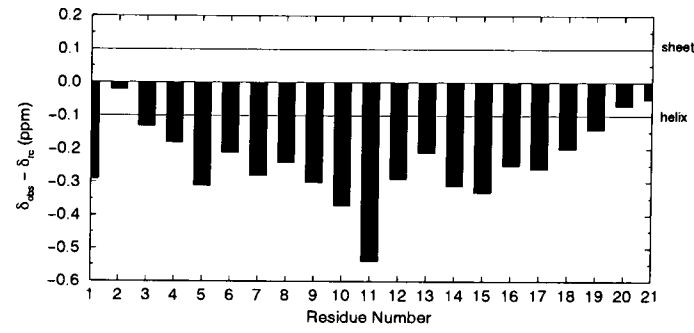
Residue	δ_{rc} (ppm)
Ala	4.35
Cys	4.65
Lys	4.36

Table 8.2: Random coil chemical shifts for the H^{α} proton of the three amino acids used in the 4K peptides [276].

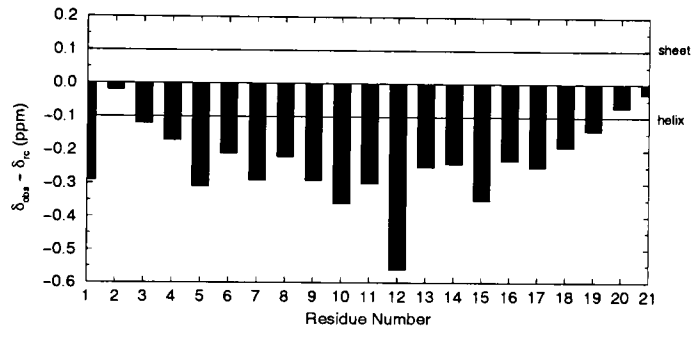
that these peptides are forming helical secondary structure. Across all members of the 4K series, residues spanning Cys 3 to Ala 19 all have chemical shift indices



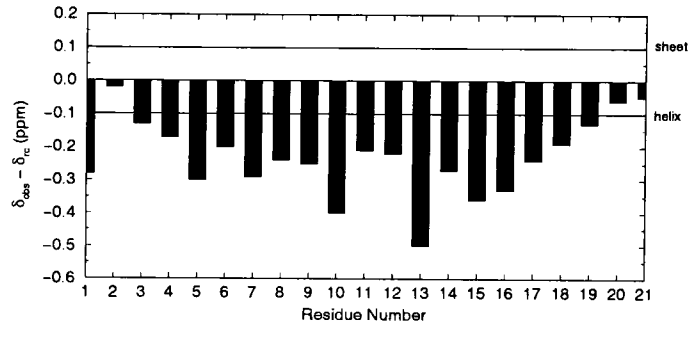
(a) 4K(3,7)



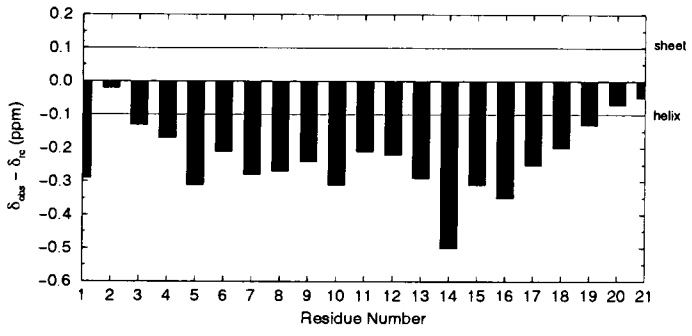
(b) 4K(3,11)



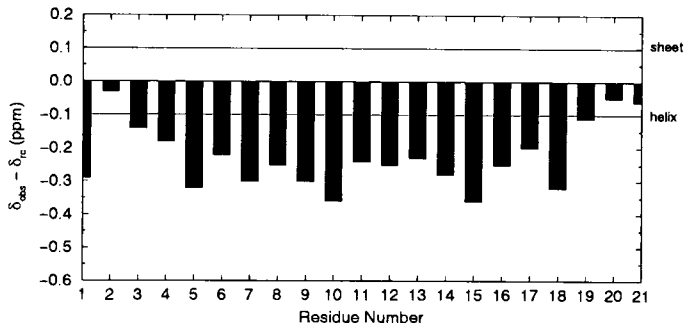
(c) 4K(3,12)



(d) 4K(3,13)



(e) 4K(3,14)



(f) 4K(3,18)

Figure 8.9: CSI as a function of amino acid sequence for the series of 4K peptides. Chemical shift assignments are given in Appendix D and the random coil H^{α} chemical shifts are given in Table 8.2.

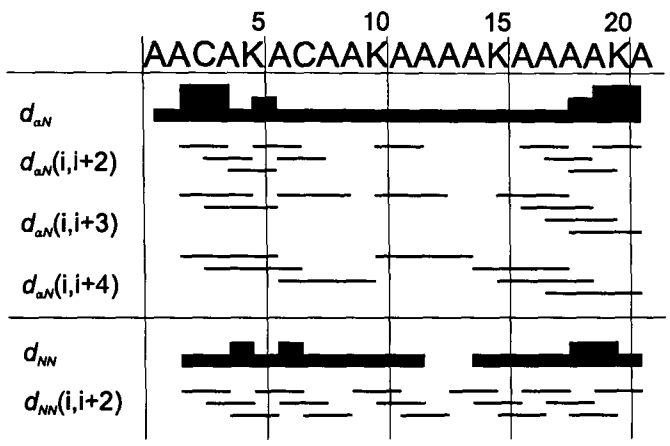
indicative of a helix, in most cases being significantly more negative than −0.1 ppm. The cysteine residue closest to the C-terminus of the peptide is clearly identifiable in all cases as having the most negative CSI, except in the case of 4K(3,18), where it is still clearly seen to have anomalous magnitude. The more neutral values of the CSI toward both the N- and C-termini of the peptides may suggest some slight "fraying" of the helix in these regions. This is expected since any secondary structure at the termini has an incomplete set of hydrogen bonds. At the N-terminus there are no hydrogen bonds to the carbonyl oxygen in the early residues. In an ideal α-helix this hydrogen bond cannot exist until the fourth residue since the helix is characterised by a $i \rightarrow i+3$ hydrogen bonding pattern. Likewise at the C-terminus the final three residues have no partner for the hydrogen bonds in an α-helix.

8.4.2 NOE Analysis

Analysis of the sequential and medium range NOE interactions allows information about the nature of the secondary structure adopted to be obtained. These NOE interactions are all obtained by analysis of NOESY spectra recorded for the complete series of 4K peptides. Figure 8.10 shows the information obtained from these experiments. The total number of unambiguously identifiable NOEs for the 4K series of peptides is summarised in Table 8.3. This corresponds to an average of

Peptide	Number of NOEs		
	$d_{\alpha N}(i,j)$	$d_{NN}(i,j)$	Total
4K(3,7)	44	33	77
4K(3,11)	58	29	87
4K(3,12)	52	34	86
4K(3,13)	54	34	88
4K(3,14)	48	26	74
4K(3,18)	59	21	80

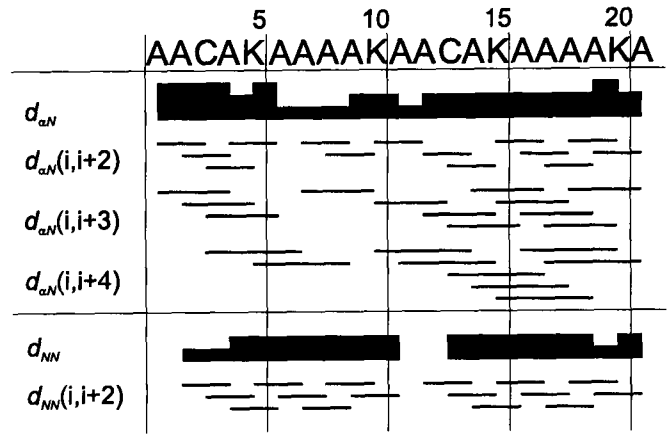
Table 8.3: Number of NOEs identified for each of the 4K peptides.



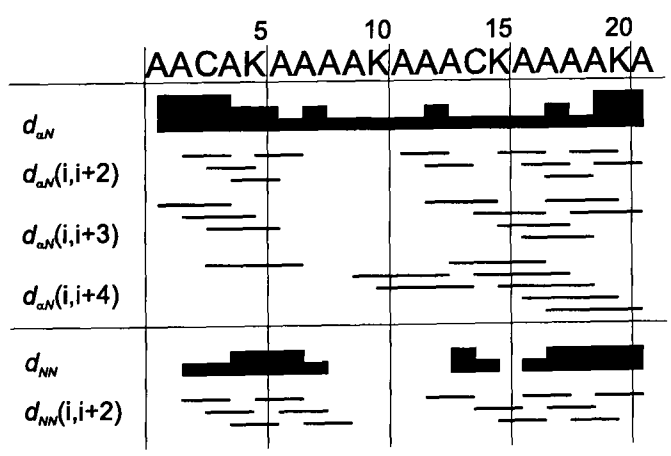
(a) 4K(3,7)

(b) 4K(3,11)

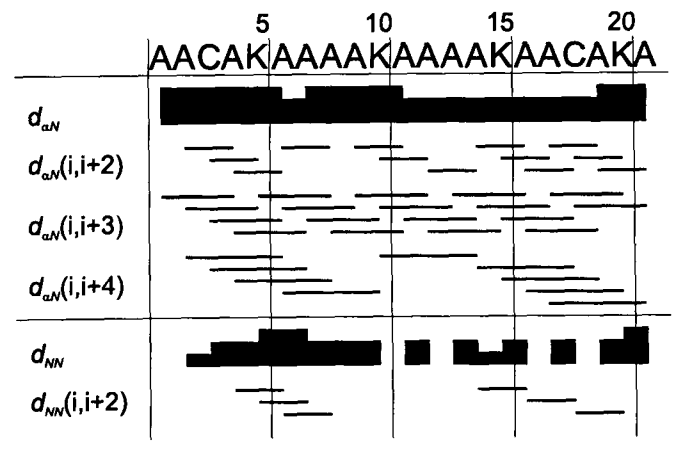
(c) 4K(3,12)



(d) 4K(3,13)



(e) 4K(3,14)



(f) 4K(3,18)

Figure 8.10: Short and medium range NOE analysis for $d_{\alpha N}(i,j)$ and $d_{NN}(i,j)$ for the series of 4K peptides. The thickness of the bar for the $d_{\alpha N}$ and d_{NN} NOEs indicates the relative magnitude of the observed resonance, categorised as strong, medium or weak. These assignments were performed manually, and in the case of any ambiguity the NOE was assigned was made to the weaker category.

approximately four identifiable backbone interactions per residue of the peptides.

The analysis of these NOEs is generally consistent with the formation of helical secondary structure. Characteristically, observation of a series of $d_{\alpha N}(i, i+3)$ NOEs along with strong sequential $d_{\alpha N}$ is indicative of helical secondary structure. Supporting evidence is provided by the observation of $d_{\alpha N}(i, i+2)$ and $d_{\alpha N}(i.i+4)$ interactions, although these are usually much weaker. In general, these NOEs can be observed within each of the peptides. Identification is sometimes hampered by overlapping signals. This is a particular problem for 4K(3,7) and 4K(3,14), where there are regions covering five amino acids that have few unambiguously assigned NOEs. Despite this, these results strongly indicate the formation of a series of helices.

Similar to the H^{α} to H^{N} interactions, the interactions between the amide protons of different amino acid residues can be used to obtain structural information. In the case of the helices, the important interactions are d_{NN} and $d_{NN}(i, i+2)$. The distance between amide protons separated by two residues is similar to the distance between the H^{α} and H^{N} protons of the same pair of residues. A number of these interactions have been found for each of the 4K peptides, further increasing the evidence for the formation of helical secondary structure.

Overall, the information gained from the detailed analysis of the NOESY experiments performed on the 4K peptides strongly indicates the formation of a helix for each of these peptides. This is true, even to the ends of the peptides. Sequential, and medium range interactions, indicative of a helix can be clearly identified right to the ends of the peptides.

8.5 Last thoughts / Conclusions

The data presented in this chapter are all consistent with the formation of helical secondary structure by this series of peptides. Strong evidence for global helicity

is provided by the circular dichroism spectra (Figure 8.2) and by the interpretation of the H^{α} chemical shifts in terms of the chemical shift index (Figure 8.9). These observations are further supported by the observed patterns of backbone NOEs obtained from the NOESY spectra (Figure 8.10). The models shown in Figure 8.11 show the six peptides studied in this chapter in idealised α-helix structures.

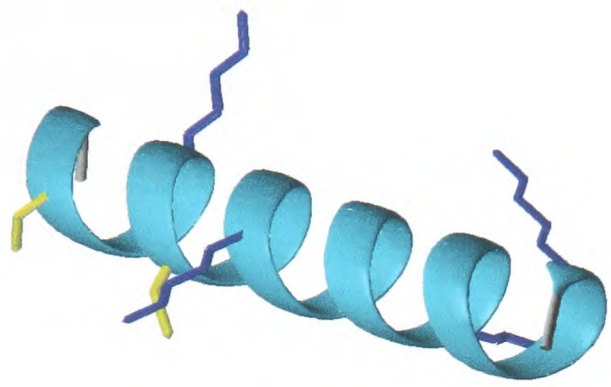
Unfortunately, neither the CSI analysis nor the interpretation of the NOE information can accurately reveal which type of helix, *i.e.* whether π-helix, α-helix, or 3_{10}-helix, is adopted by the peptides. Recently, Freedberg *et al.* showed that, by selective isotopic labelling of the backbone N^{H} and CO at particular residues, a slightly modified form of the triple resonance HNCO experiment can be used to determine which type of helix is present [280]. This approach utilises coherence transfer via the hydrogen bonds present in the helix. Observation of an HNCO cross peak for a given pattern of isotope labelling gives directly the type of helix formed.

The data presented in this chapter can be used as a starting point for structure calculations using information obtained from the NMR experiments as structural restraints. The NOE information given in Figure 8.10 can be input as constraints in restricted molecular dynamics (MD) calculations. The peptides are then allowed to fold in the MD force field subject to these constraints.

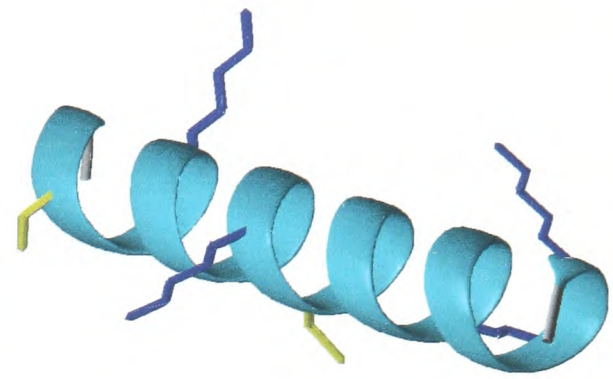
Additional constraints on the ϕ backbone torsion angle, *i.e.* the angle of rotation about the C^{α}—N^{H} bond, can be added to the MD calculations. This angle can be obtained from NMR experiments by measuring the three-bond scalar coupling between the H^{N} and H^{α} protons of a particular residue. The ϕ torsion angle is related to this coupling constant by the Karplus equation [74, 281]:

$$^{3}J_{H^{N}H^{\alpha}} = A\cos^{2}(|\phi - 60°|) + B\cos(|\phi - 60°|) + C \qquad (8.7)$$

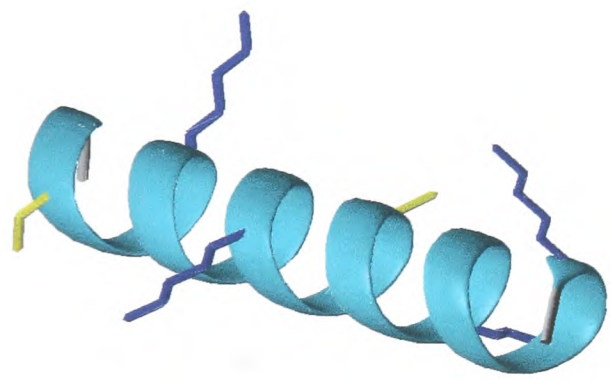
where $A = 6.4$, $B = -1.4$ and $C = 1.9$ are empirical constants. Commonly, the



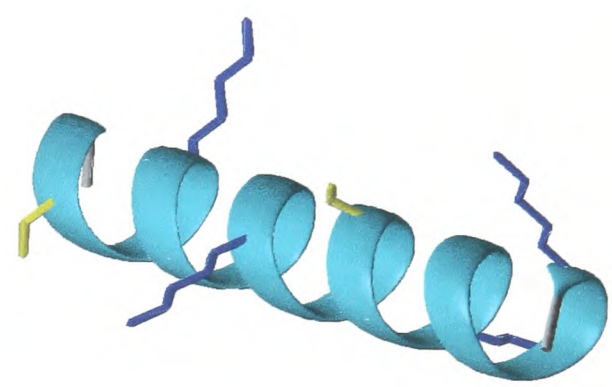
(a) 4K(3,7)



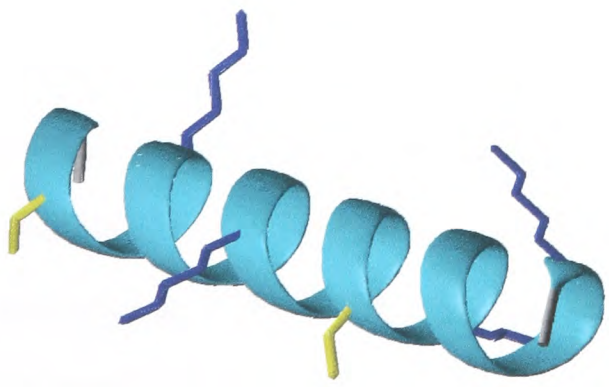
(b) 4K(3,11)



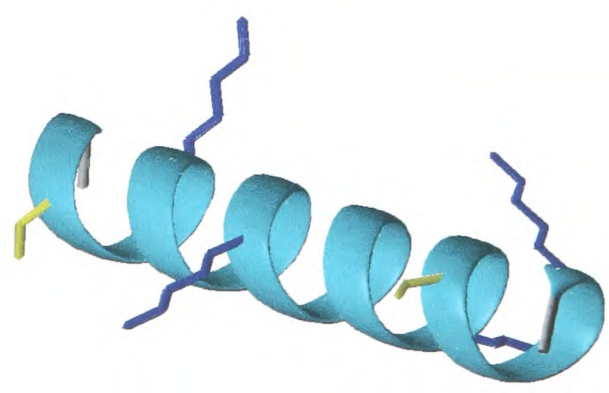
(c) 4K(3,12)



(d) 4K(3,13)



(e) 4K(3,14)



(f) 4K(3,18)

Figure 8.11: Idealised α-helix conformations for the 4K peptides. The side chains are shown in a fully extended conformation, with lysine shown in blue and cysteine shown in yellow. The figures were constructed and drawn using `MOLMOL` [89].

scalar coupling is measured using the COSY experiment or a double quantum-filtered variant. Unfortunately due to the limited amino acid type in these peptides, the chemical shift dispersion in the $H^{\alpha} \leftrightarrow H^{N}$ region of the spectrum leads to partial cancellation of the resonances due to the antiphase doublet nature of the cross peaks in the COSY experiment. This limits the accuracy of the measured scalar couplings, and hence their validity as structural constraints in the MD calculations.

It is also possible, using NMR methods, to determine how flexible the helix is along its length. This is generally performed using NMR relaxation measurements of the backbone carbon and nitrogen nuclei. As an initial experiment, a ^{13}C-^{1}H HSQC experiment was performed on 4K(3,18), using the ^{13}C at natural abundance, to test the possibility of measuring T_1, T_2 and the $^{13}C\{^{1}H\}$ NOE of the carbon nuclei at the C^{α} position. Each of these parameters is sensitive to backbone motions of the peptide. Figure 8.12 shows the ^{13}C-^{1}H HSQC spectrum of 4K(3,18) recorded at natural abundance. Using the H^{α} chemical shift assignments obtained for 4K(3,18), it is possible to transfer some of these to the C^{α} chemical shifts. From this, seven resonances can be clearly identified, three at the N-terminus, and four at the C-terminus of the peptide. Unfortunately, due to some resonances having coincident H^{α} chemical shifts and the lack of significant chemical shift dispersion in parts of the ^{13}C dimension, it is not possible to unambiguously assign the C^{α} of each residue. This therefore limits the applicability of measuring ^{13}C relaxation parameters as a probe of the peptide structure and dynamics, except at the termini of the peptide. It is possible that isotopic double labelling would introduce sufficient resolution in three dimensions ($^{1}H/^{13}C/^{15}N$) to enable these experiments to be performed. Cost effective synthesis of these peptides, however, precludes this approach.

The results presented in this chapter will be extremely useful in aiding the interpretation of EPR experiments performed on spin-labelled variants of these peptides. Once molecular dynamics simulations have been performed to determine the struc-

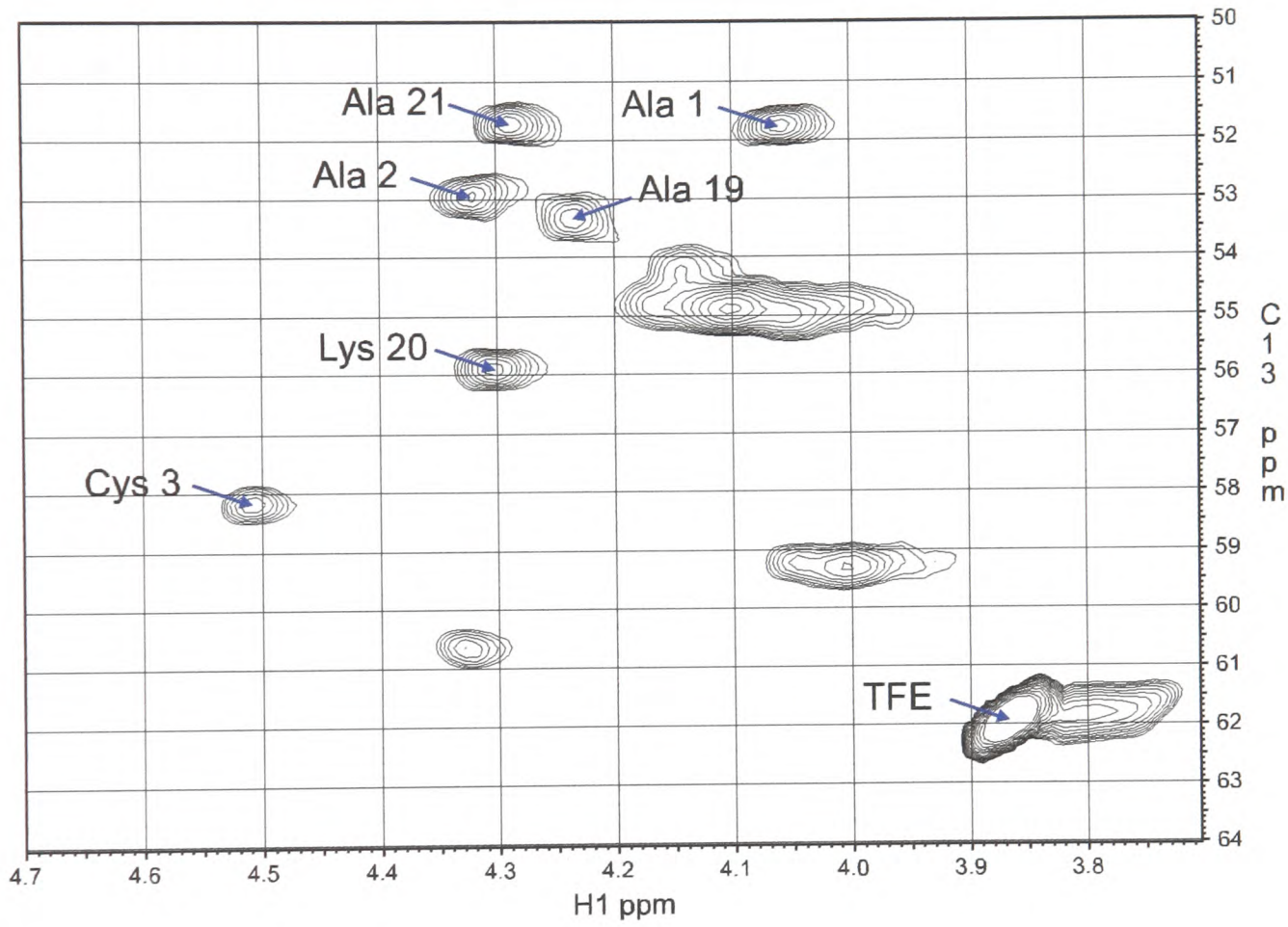


Figure 8.12: 600 MHz ^{13}C-^{1}H HSQC spectrum of 4K(3,18) showing the C^{α}/H^{α} region. The spectrum was recorded using a spectral width of 12 kHz and 1020 complex points in the directly detected dimension (^{1}H) and a spectral width of 9 kHz and 64 complex points in the indirectly detected dimension (^{13}C). The spectrum was recorded at natural ^{13}C abundance, with 64 transients per t_1 increment. Shifted sine-squared window functions were used in both dimensions.

tures, distances obtained from the EPR experiments can be compared to those expected from the structures obtained from these NMR experiments.

Appendix A

Product Operator Description of the Radical Pair Mechanism

The radical pair mechanism described in Chapter 1 can be recast [55] in terms of the product operator formalism [54]:

The triplet $|T_0\rangle$ spin-state of the radical pair can be described by the following expression:

$$|T_0\rangle\langle T_0| = \frac{1}{4}E - I_z S_z + \{ZQ\}_x \tag{A.1}$$

where E is the identity operator, $I_z S_z$ is electron two-spin order and:

$$\{ZQ\}_x = I_x S_x + I_y S_y \tag{A.2}$$

$$\text{and} \quad \{ZQ\}_y = I_y S_x - I_x S_y \tag{A.3}$$

is electronic Zero-Quantum coherence. By analogy, the singlet state is described by:

$$|S\rangle\langle S| = \frac{1}{4}E - I_z S_z - \{ZQ\}_x \tag{A.4}$$

The initial density operator[1] of the triplet-born radical pair, ignoring the $|T_+\rangle$ and $|T_-\rangle$ states, can be written in the following form:

$$\begin{aligned}\hat{\rho}(0) &= \frac{1}{2}|T_0\alpha\rangle\langle T_0\alpha| + \frac{1}{2}|T_0\beta\rangle\langle T_0\beta| \\ &= \frac{1}{4}E - I_zS_z + \{ZQ\}_x \end{aligned} \tag{A.5}$$

where α and β correspond to the nuclear spin states of a spin-$\frac{1}{2}$ nucleus coupled to electron I with isotropic hyperfine coupling a_{IR}. This density operator is then allowed to evolve under the electronic Zeeman Hamiltonian ($\hat{H}_Z = (\omega_I I_z + \omega_S S_z)$) and the electron-nuclear hyperfine Hamiltonian ($\hat{H}_{hf} = \pi a_{IR} 2 I_z R_z$):

$$\begin{aligned}\hat{\rho}(0) \xrightarrow{(\omega_I I_z + \omega_S S_z)t} & \frac{1}{4}E - I_zS_z \\ & + \{ZQ\}_x \cos(\omega_I - \omega_S)t \\ & + \{ZQ\}_y \sin(\omega_I - \omega_S)t \end{aligned} \tag{A.6}$$

$$\begin{aligned}\xrightarrow{\pi a_{IR} 2 I_z R_z t} & \frac{1}{4}E - I_zS_z \\ & + \{ZQ\}_x [\cos(\omega_I - \omega_S)t \cos(\pi a_{IR} t)] \\ & -2R_z \sin(\omega_I - \omega_S)t \sin(\pi a_{IR} t)] \\ & + \{ZQ\}_y [2R_z \cos(\omega_I - \omega_S)t \sin(\pi a_{IR} t)] \\ & + \sin(\omega_I - \omega_S)t \cos(\pi a_{IR} t)] \end{aligned} \tag{A.7}$$

$$= \hat{\rho}(t) \tag{A.8}$$

The nuclear polarisation in the recombination products, formed through the

[1]The density operator is defined as $\hat{\rho}(t) = \overline{|\psi(t)\rangle\langle\psi(t)|}$, where the overbar indictes averaging over the ensemble.

singlet channel, is then given by the trace of $\hat{\rho}(t)$ with:

$$|S\rangle\langle S| R_z = \left(\frac{1}{2}|S\alpha\rangle\langle S\alpha| + \frac{1}{2}|S\beta\rangle\langle S\beta|\right) R_z \quad (A.9)$$

$$= \left(\frac{1}{4}E - I_z S_z - \{ZQ\}_x\right) R_z \quad (A.10)$$

therefore:

$$\mathcal{P}_R^{\mathrm{R}} = \mathrm{Tr}\left[\hat{\rho}(t)|S\rangle\langle S| R_z\right] \quad (A.11)$$

$$\propto \frac{1}{4}\sin(\omega_I - \omega_S)t\sin(\pi a_{IR}t) \quad (A.12)$$

The polarisation in the escape products, which is equal and opposite in phase to the recombination products, is given by the trace of $\hat{\rho}(t)$ with $|T_0\rangle\langle T_0|R_z$.

$$\mathcal{P}_R^{\mathrm{E}} \propto -\frac{1}{4}\sin(\omega_I - \omega_S)t\sin(\pi a_{IR}t) \quad (A.13)$$

Clearly, from Equations A.12 and A.13 no polarisation is produced if the g-values of the two radicals are the same, or if there is no hyperfine coupling. The phase of the polarisation, as predicted by Kaptein's sign rules [56], is also evident from the signs of Δg and the hyperfine coupling.

Appendix B

Source Code Listings

This Appendix contains Fortran 77 source code listings for the programs used in Chapter 7.

B.1 Listing of `cidnpint.f`

```
      program cidnpint

* A program to calculate the CIDNP intensity field dependance.
*
* Options of different kinetic models.
*   a = exponential model
*   b = t^(-3/2) truncated diffusion model
*   c = full t^(-3/2) diffusion model (including exponential term)
*
* Can model 10 spin-1/2 nuclei on each radical (set using maxnuc parameter)
* Can model 5 spin-1 nuclei on each radical (set using maxnuc1 parameter)
* Calculates polarisation of a selected spin-1/2 nucleus.
*
* NB: Program requires input paameters in Tesla, however, comverts to
* angular frequencies for main calculations.
*
* Compile using:
*       g77 cidnpint.f -o cidnpint
*
* IJD
* 20/7/04

* Declaration of variables

      implicit none

      integer i, j
      integer maxnuc, nucint, nnuc, dnuc, anuc
      integer maxnuc1, nnuc1, dnuc1, anuc1, nnuct

      parameter (maxnuc = 10, maxnuc1=5)

      double precision mub, hbar, deltag, pi, onuc, g
      double precision b0start, b0end, numb0, b0, b0inc
      double precision dhfc(maxnuc), ahfc(maxnuc), hfc(2*maxnuc), p
```

```
      double precision dhfc1(maxnuc1), ahfc1(maxnuc1), hfc1(2*maxnuc1)
      double precision pattern((2*maxnuc-1)+(2*maxnuc1))
      double precision hfcmod((2*maxnuc-1)+(2*maxnuc1))
      double precision ks, alpha
      double precision omegaplus, omegaminus

      character*1 model
      character*6 fname

      integer base((2*maxnuc-1)+(2*maxnuc1))


* Set loop counters to zero

      i = 0
      j = 0


* Zero all arrays

      do i=1, maxnuc
            dhfc(i) = 0.0d0
            ahfc(i) = 0.0d0
      end do

      do i=1, maxnuc1
            dhfc1(i) = 0.0d0
            ahfc1(i) = 0.0d0
      end do

      do i=1, 2*maxnuc
            hfc(i) = 0.0d0
      end do

      do i=1, 2*maxnuc1
            hfc1(i) = 0.0d0
      end do

      do i=1, (2*maxnuc-1)+(2*maxnuc1)
            pattern(i) = 0.0d0
            base(i) = 0
      end do


* Zero the polarisation

      p = 0.0d0


* Define some constants

      pi     = 4.0d0 * datan(1.0d0)
      mub    = 9.427401d-24
      hbar   = 1.05457d-34
      g      = 2.0024


* Get parameters from the user

      write(6,*)'Enter the output filename'
      read *, fname
      write(6,*)'Enter the initial field strength'
      read *, b0start
      write(6,*)'Enter the final field strength'
      read *, b0end
      write(6,*)'Enter the number of field points'
      read *, numb0
      write(6,*)'Enter the difference in g-value (d-a)'
```

```
read *, deltag

write(6,*)'Which kinetic model?'
read *, model
if (model .eq. 'a') then
        write(6,*)'Enter the reaction rate constant'
        read *, ks
else if (model .eq. 'b') then
        write(6,*)'No model specific parameters to enter'
else if (model .eq. 'c') then
        write(6,*)'Enter diffusion model alpha parameter'
        read *, alpha
end if

write(6,*)'Enter the number of spin-1/2 nuclei on donor'
read *, dnuc
write(6,*)'Enter the number of spin-1/2 nuclei on acceptor'
read *, anuc
write(6,*)'Enter the donor spin-1/2 nucleus of interest'
read *, nucint
write(6,*)

if (dnuc .ne. 0) then
        write(6,*)'Enter the hfc for each donor nucleus (mT)'

        do i=1, dnuc
                read *, dhfc(i)
        end do
end if

if (anuc .ne. 0) then
        write(6,*)'Enter the hfc for each acceptor nucleus (mT)'

        do i=1, anuc
                read *, ahfc(i)
        end do
end if

write(6,*)'Enter the number of spin-1 nuclei on donor'
read *, dnuc1
write(6,*)'Enter the number of spin-1 nuclei on acceptor'
read *, anuc1
write(6,*)

if (dnuc1 .ne. 0) then
        write(6,*)'Enter the hfc for each donor nucleus (mT)'

        do i=1, dnuc1
                read *, dhfc1(i)
        end do
end if

if (anuc1 .ne. 0) then
        write(6,*)'Enter the hfc for each acceptor nucleus (mT)'

        do i=1, anuc1
                read *, ahfc1(i)
        end do
end if

write(6,*)


* Calculate some other variables

        nnuc = dnuc + anuc
        nnuc1 = dnuc1 + anuc1
```

```
        nnuct = nnuc + nnuc1


* Write out a parameters file

        open(2, file=fname//'.par')

        write(2,100) b0start, b0end, numb0, deltag, dnuc,
     &                  anuc, nucint, dnuc1, anuc1

        if (model .eq. 'a') then
                write(2,*)'Exponential Model used'
                write(2,101) ks
        else if (model .eq. 'b') then
                write(2,*)'Truncated Diffusion Model used'
                write(2,*)'  No model specific parameters needed'
        else if (model .eq. 'c') then
                write(2,*)'Full Diffusion Model used'
                write(2,102) alpha
        end if

100     format('B0 Field Strength start point (T)      = ', d12.5,/,
     &         'B0 Field Strength end point (T)        = ', d12.5,/,
     &         'Number of B0 Field points              = ', d12.5,/,
     &         'Difference in g-values (d-a)           = ', d12.5,/,
     &         'Number of spin-1/2 nuclei on donor     = ', i12,/,
     &         'Number of spin-1/2 nuclei on aceptor   = ', i12,/,
     &         'Spin-1/2 Nucleus of interest           = ', i12,/,
     &         'Number of spin-1 nuclei on donor       = ', i12,/,
     &         'Number of spin-1 nuclei on acceptor    = ', i12)

101     format('Reaction Rate constant                 = ', d12.5)

102     format('Diffusion model alpha (2m/p)           = ', d12.5)


* Stitch the arrays dhfc and ahfc together into hfc and scale to Tesla units

        do i=1, dnuc
                hfc(i) = dhfc(i) / 1.0d3
        end do

        do i=1, anuc
                hfc(i+dnuc) = -ahfc(i) / 1.0d3
        end do

        do i=1, dnuc1
                hfc1(i) = dhfc1(i) / 1.0d3
        end do

        do i=1, anuc1
                hfc1(i+dnuc1) = -ahfc1(i) / 1.0d3
        end do


* Convert the hyperfine couplings to angular frequency units for use in the
* rest of the calculations.

        do i=1, nnuc
                hfc(i) = hfc(i) * 2.8d10 * 2.0d0 * pi
        end do

        do i=1, nnuc1
                hfc1(i) = hfc1(i) * 2.8d10 * 2.0d0 * pi
        end do


* Cut out the spin-1/2 nucleus of interest from the list of hyperfine constants
```

```
        do i=1, nnuc - 1
                if (i .lt. nucint) then
                        hfcmod(i) = hfc(i)
                else
                        hfcmod(i) = hfc(i+1)
                end if
        end do


* Add the spin-1 nuclei to the vector of hyperfine couplings

        do i=1, nnuc1
                hfcmod(i+(nnuc-1)) = hfc1(i)
        end do


* Set the right base to use when calculating the array of spin states

        do i=1, nnuct-1
                pattern(i) = 0.0d0
                if (i .lt. nnuc) then
                        base(i) = 2
                else
                        base(i) = 3
                end if
        end do


* Open an output file

        open(3, file=fname//'.dat')

* Start the field loop

        do 800 b0inc=1, numb0

        if (numb0 .eq. 1) then
                b0 = (b0start + b0end) / 2.0d0
        else
                b0 = (b0inc - 1.0d0)
     &                  * (b0end - b0start) / (numb0 - 1.0d0)
     &                  + b0start
        end if


* Zero the polarisation before calculating the main result

        p = 0.0d0


* Perform the main calculation.
* Use the spinstates subroutine to generate all the spin-states of the nuclei

        do i=0, (((2 ** (nnuc - 1)) * (3 ** nnuc1)) - 1)

                call spinstates(i, base, nnuct - 1, pattern)


* Dot the spin-state pattern vector with the chopped list of hyperfine coupling
* constants.

                call rcdot(pattern, hfcmod, onuc, nnuct - 1)


* Calculate the ST0 oscillation frequencies
```

```
                omegaplus = (deltag * mub * b0) / hbar
     &                          + 0.5d0 * hfc(nucint) + onuc

                omegaminus = (deltag * mub * b0) / hbar
     &                          - 0.5d0 * hfc(nucint) + onuc


* Calculate the polarisation depending on what model was selected
* Exponential model version

                if (model .eq. 'a') then

                p = p + omegaplus**2.0d0 /
     &                  (ks**2.0d0 + 4.0d0 * omegaplus**2.0d0)
     &                - omegaminus**2.0d0 /
     &                  (ks**2.0d0 + 4.0d0 * omegaminus**2.0d0)


* Truncated diffusion model version

                else if (model .eq. 'b') then

                p = p + dsqrt(dabs(omegaplus)) - dsqrt(dabs(omegaminus))


* Full diffusion model version
* Need to convert oscillation frequencies to Tesla to avoid machine precision
* problems with taking the exponential of a negative big number

                else if (model .eq. 'c') then

                omegaplus = omegaplus / (2.8d10 * 2.0d0 * pi)
                omegaminus = omegaminus / (2.8d10 * 2.0d0 * pi)
                p = p + (1.0d0 - dexp(- alpha * dsqrt(dabs(omegaplus))) *
     &                  dcos(alpha * dsqrt(dabs(omegaplus))))
     &                - (1.0d0 - dexp(- alpha * dsqrt(dabs(omegaminus))) *
     &                  dcos(alpha * dsqrt(dabs(omegaminus))))

                end if
           end do


* Scale the polarisation by 2^(N-1)+3^M
* N is number of spin-1/2 nuclei and M is the number of spin-1

           p = p / (2.0d0 ** (nnuc - 1.0d0) * 3.0d0 ** nnuc1)


* Write out the data

           write(3,*) b0, p


* End the field loop

800        continue


* Close all open files

           close(2)
           close(3)


* End of the main program

           stop
```

```
        end


*************************
*
* Subroutines and functions
*
*************************

* Subroutine rcdot
* Calculates the dot product of a row vector with a column vector of the same
* dimension

        subroutine rcdot(in1, in2, out, size)


* Declare the subroutine varibles

        integer i, size

        double precision in1(size), in2(size), out


* Zero the output to make sure we don't do something stupid

        out = 0.0d0


* Do the dot product

        do i=1, size
                out = out + (in1(i) * in2(i))
        end do


* End the subroutine

        return
        end


*************************

* Subroutine spinstates
* Generate the spin states from an integer

        subroutine spinstates(decnum, base, size, out)


* Declare the subroutine variables

        integer decnum, size, i

        integer temp(size), base(size)

        double precision out(size)


* Reset the output base-number (base = 2I+1)

        do i=1, size
                out(i) = (real(base(i))-1.0d0)/2.0d0
        end do


* Set a counter one

        i=1
```

```
* Assign first element of out to the decimal number

        temp(1) = decnum


* Convert the decimal number to base-b
* Convert into a spinstate (2m+1 values)

        do while (temp(i) .ge. base(i))
                temp(i+1) = temp(i) / base(i)
                temp(i)   = temp(i) - (base(i) * temp(i+1))
                out(i) = (real(base(i))-1)/2.0d0 - real(temp(i))
                i = i + 1
        end do

        out(i) = (real(base(i))-1)/2.0d0 - real(temp(i))

* End the subroutine

        return
        end
```

B.2 Listing of spec-sim.f

```
        program specsim

* A program to simulate an NMR spectrum
* Program calculates the FID, then performs an FT to generate the NMR spectrum
* Compile against the fftw library using the following statement:
*       g77 spec-sim.f -o spec-sim -lfftw -lm
* Program uses generalised FID
* Based on nmr_sym from Peter
* Zero frequency is at the left hand end of the spectrum
* Maximum length of FID defined by maxfid
* Maximum number of multiplets defined by maxmult
*
* Converts x-axis to ppm
*
* IJD
* 31/5/04

* Declaration of Variables

        implicit none

        integer maxfid, maxmult

        parameter (maxfid = 131072, maxmult=20)

        double precision pi, sw, dw, at, digres, lw, t2
        double precision timept(maxfid), rfid(maxfid), ifid(maxfid)
        double precision nnuc(maxmult), fmult(maxmult), mult(maxmult)
        double precision j(maxmult), freqpt(maxfid)
        double precision t2relax, jmodu, ccsmodu, scsmodu, scaler
        double precision transfreq, carrppm, reffreq

        double complex fid(maxfid), spec(maxfid)

        integer i, m, n, np, npf, nmult, plan, strlen

        character*1 printfid
        character*200 fname, filepar, filefid, fileft
```

```
* Declaration of variables for the using with the FFTW library

        integer FFTW_FORWARD,FFTW_BACKWARD
        parameter (FFTW_FORWARD=-1,FFTW_BACKWARD=1)

        integer FFTW_REAL_TO_COMPLEX,FFTW_COMPLEX_TO_REAL
        parameter (FFTW_REAL_TO_COMPLEX=-1,FFTW_COMPLEX_TO_REAL=1)

        integer FFTW_ESTIMATE,FFTW_MEASURE
        parameter (FFTW_ESTIMATE=0,FFTW_MEASURE=1)

        integer FFTW_OUT_OF_PLACE,FFTW_IN_PLACE,FFTW_USE_WISDOM
        parameter (FFTW_OUT_OF_PLACE=0)
        parameter (FFTW_IN_PLACE=8,FFTW_USE_WISDOM=16)

        integer FFTW_THREADSAFE
        parameter (FFTW_THREADSAFE=128)


* Declare the used FFTW libraries as external

        external fftw_f77_create_plan
        external fftw_f77_one
        external fftw_f77_destroy_plan


* Set loop counters to zero initially

        i = 0
        m = 0

* Define some constants

        pi = 4.0d0 * datan(1.0d0)


* Zero arrays

        do i=1, maxfid
                rfid(i)   = 0.0d0
                ifid(i)   = 0.0d0
                timept(i) = 0.0d0
        end do

        do i=1, maxmult
                nnuc(i)  = 0.0d0
                fmult(i) = 0.0d0
                mult(i)  = 0.0d0
                j(i)     = 0.0d0
        end do

* Get parameters from the user

        write(6,*)'Enter the output filename'
        read *, fname
        write(6,*)'Write out FID?'
        read *, printfid
        write(6,*)'Scale spectrum'
        read *, scaler
        write(6,*)
        write(6,*)'Enter the spectral width (Hz)'
        read *, sw
        write(6,*)'Enter the total number of points'
        read *, np
        write(6,*)'Enter the Fourier number'
        read *, npf
```

```
      write(6,*)'Enter the transmitter frequency (MHz)'
      read *, transfreq
      write(6,*)'Enter the carrier position (ppm)'
      read *, carrppm
      write(6,*)'Enter the resonance linewidth (Hz)'
      read *, lw
      write(6,*)
      write(6,*)'Enter the number of multiplets'
      read *, nmult
      write(6,*)

      if (nmult .ne. 0) then
            write(6,*)'For each multiplet enter the following:'
            write(6,*)'No of Nuclei, Frequency (Hz), Multiplicity, J (Hz)'

            do i=1, nmult
                  read *, nnuc(i), fmult(i), mult(i), j(i)
            end do
      end if


* Calculate some other variables

      dw      = 1.0d0 / sw
      at      = np * dw
      digres  = sw / npf
      t2      = 1.0d0 / (pi * lw)
      reffreq = transfreq / (1.0d0 + carrppm * 1.0d-6)


* Calculate the array of time points

      do i=1, np
            timept(i) = ((i - 1.0d0) * at) / (np - 1.0d0)
      end do


* Write out a parameters file

      filepar = fname(1:strlen(fname))//'.par'

      open(2, file=filepar)
      write(2,100) sw, scaler, np, npf, transfreq, carrppm, dw, at,
     &                  digres, lw, t2, nmult

100   format('Spectral Width (Hz)                  = ', d12.5,/,
     &       'Spectral scaling factor              = ', d12.5,/,
     &       'Number of Points                     = ', i12,/,
     &       'Fourier Number                       = ', i12,/,
     &       'Dwell Time                           = ', d12.5,/,
     &       'Transmitter Frequency (MHz)          = ', d12.5,/,
     &       'Carrier Position                     = ', d12.5,/,
     &       'Acquisition Time                     = ', d12.5,/,
     &       'Digital Resolution                   = ', d12.5,/,
     &       'Resonance Linewidth                  = ', d12.5,/,
     &       'T2 relaxation time                   = ', d12.5,/,
     &       'Number of multiplets                 = ', i12)

      if (nmult .ne. 0) then
            write(2,*)
            write(2,*)'Spin systems'
            write(2,*)
            write(2,*)'No of Nuclei, Frequency (Hz), Multiplicity, J (Hz)'
            do i=1, nmult
                  write(2,101) nnuc(i), fmult(i), int(mult(i)), j(i)
101               format(d12.5, d12.5, i12, d12.5)
            end do
      end if
```

```
* Calculate the FID, real and imaginary parts separately

        do i=1, nmult
                do m=1, np
                        rfid(m) = rfid(m) + (nnuc(i) *
     &                          ccsmodu(pi, fmult(i), timept(m)) *
     &                          t2relax(timept(m), t2) *
     &                          (jmodu(pi, j(i), timept(m))**(mult(i) - 1.0d0)))

                        ifid(m) = ifid(m) + (nnuc(i) *
     &                          scsmodu(pi, fmult(i), timept(m)) *
     &                          t2relax(timept(m), t2) *
     &                          (jmodu(pi, j(i), timept(m))**(mult(i) - 1.0d0)))
                end do
        end do


* Scale the first point of the FID

        rfid(1) = 0.5d0 * rfid(1)
        ifid(1) = 0.5d0 * ifid(1)


* Write out the FID, using alternating real and imaginary points

        filefid = fname(1:strlen(fname))//'.fid'

        if (printfid .eq. 'y') then
                open(3, file=filefid)
                do i=1, npf
                        write(3,*) rfid(i)
                        write(3,*) ifid(i)
                end do
        end if


* Combine the real and imaginary parts of the FID into one double complex array
* Zerofilling is done if npf > np

        do i=1, npf
                fid(i) = cmplx(rfid(i), ifid(i))
        end do


* Perform the FFT using the fftw library

        call fftw_f77_create_plan(plan,npf,FFTW_FORWARD,FFTW_ESTIMATE)
        call fftw_f77_one(plan,fid,spec)
        call fftw_f77_destroy_plan(plan)


* Create a frequency array to use for spectrum x-axis scale

        do i=1, npf
                freqpt(i) = ((transfreq * 1.0d6) + (sw / 2.0d0)
     &                      - ((i - 1.0d0) * sw) / (npf - 1.0d0)) / 1.0d6
        end do


* Convert the x-axis into a ppm scale

        do i=1, npf
                freqpt(i) = ((freqpt(i) - reffreq) / reffreq) * 1.0d6
        end do
```

```
* Write out the real part of the spectrum and the frequency axis

        fileft = fname(1:strlen(fname))//'.ft'

        open(4, file=fileft)
        do i=1, npf
                write(4,*) freqpt(i), dble(spec(i)) * scaler
        end do


* Close all open files

        close(2)
        close(3)
        close(4)


* End of main program

        stop
        end


**************************
*
* Subroutines and functions
*
**************************

* Function ccsmodu
* Calculates the Chemical Shift modulation - real part, cosine modulation

        double precision function ccsmodu(pi, freq, time)

        double precision pi, freq, time

        ccsmodu = dcos(2.0d0 * pi * freq * time)

        return
        end


* Function scsmodu
* Calculates the Chemical Shift modulation - imaginary part, sine modulation

        double precision function scsmodu(pi, freq, time)

        double precision pi, freq, time

        scsmodu = dsin(2.0d0 * pi * freq * time)

        return
        end


* Function Jmodu
* Calculates the J-modulation, cosine modulation

        double precision function jmodu(pi, j, time)

        double precision pi, j, time

        jmodu = dcos(pi * j * time)

        return
        end
```

```
* Function T2relax
* Calculates the T2 relaxation decay

        double precision function t2relax(time, t2)

        double precision time, t2

        t2relax = dexp(-time/t2)

        return
        end


* Function strlen
* Finds the logical length of a string

        integer function strlen(st)

        integer i

        character st*(*)

        i = len(st)

        do while (st(i:i) .eq. ' ')
                i = i - 1
        end do

        strlen = i

        return
        end
```

Appendix C

Tables of Hyperfine Coupling Constants Used in Chapter 7

This Appendix contains tables of g-values and isotropic hyperfine coupling constants used in Chapter 7. Hyperfine coupling constants are quoted in mT, and all nitrogen atoms are ^{14}N unless otherwise noted.

C.1 Flavin Mononucleotide

C.1.1 Neutral Radical

Source	g-value	H1′r	H5	H6	8-CH_3	N5	N10
[254, 255]	2.0030	–	–	−0.17	0.24	0.78	0.37
[282–284]	2.0036	0.390	−0.769	−0.158	0.2675	0.393	0.212

Table C.1: Isotropic hyperfine coupling constants and the isotropic g-value for the neutral flavin radical. [254, 255] are from X-band EPR spectra of flavin neutral radicals. [282, 284] are from EPR, ENDOR and TRIPLE resonance spectroscopy of DNA photolyase. [283, 284] are from DFT calculations at the B3LYP/EPR-II level of theory.

C.1.2 Radical Anion

Source	g-value	H1′r	H5	H6	8-CH_3	N5	N10
[258, 259]	2.0034	–	–	−0.35	0.40	0.73	0.31

Table C.2: Isotropic hyperfine coupling constants and the isotropic g-value for the anionic flavin radical. [258, 259] are from X-band EPR spectra of flavin anion radicals.

C.2 Tyrosine

C.2.1 Neutral Radical

Source	g-value	H2	H3	H5	H6	H_2^β
[256]	2.0041	0.15	−0.65	−0.65	0.15	0.77

Table C.3: Isotropic hyperfine coupling constants and the isotropic g-value for the neutral tyrosine radical. [256] are from the EPR spectra taken during the photooxidation of tyrosine.

C.3 Tryptophan

C.3.1 Neutral Radical

Source	g-value	H2	H4	H5	H6	H7	H$^{\beta}$1	H$^{\beta}$2	N1
[248]	–	−0.05	−0.40	0.03	−0.32	−0.08	0.29	1.08	0.25
[247]	–	0.11	−0.42	–	−0.34	−0.06	2.48	2.52	0.41
[244, 284]	2.0028	–	−0.40	–	−0.36	–	1.36	2.83	0.40

Table C.4: Isotropic hyperfine coupling constants and the isotropic g-value for the neutral tryptophan radical. [248] are from DFT calculations at the PWP86/ILGO-III level of theory on 3-ethyl indole. [247] are from DFT calculations at the B3LYP/EPR-III level of theory on 3-methyl indole. [244, 284] are from EPR and ENDOR spectroscopy of ribonucleotide reductase.

C.3.2 Radical Cation

Source	NH (H1)	H2	H4	H5	H6	H7	H$^{\beta}$1	H$^{\beta}$2	N1
[248]	−0.34	−0.47	−0.51	0.06	−0.35	−0.21	0.33	0.84	0.25
[247]	−0.52	−0.49	−0.58	–	−0.42	−0.17	2.65	1.22	0.22

Table C.5: Isotropic hyperfine coupling constants for the cationic tryptophan radical. [248] are from DFT calculations at the PWP86/ILGO-III level of theory on 3-ethyl indole. [247] are from DFT calculations at the B3LYP/EPR-III level of theory on 3-methyl indole.

Appendix D

^{1}H Chemical Shift Assignments for the 4K Peptides

This appendix contains the chemical shift assignments for the 4K peptides characterised in Chapter 8. All assignments were performed using a combination of TOCSY spectroscopy with a mixing time of 80 ms and NOESY spectroscopy with a mixing time of 150 ms, as described in Chapter 8. Complete assignment of the lysine side chains and stereospecific assignment of the cysteine H^{β} protons was not performed. All assignments are referenced to the CH_2 protons of trifluoroethanol at 3.88 ppm.

D.1 4K(3,7)

Number	Residue	H^N	H^α	H^β	Other
1	Ala	—	4.05		
2	Ala	8.60	4.34	1.39	
3	Cys	8.24	4.54	3.09, 2.96	SH: 2.16
4	Ala	8.56	4.12	1.47	
5	Lys	8.15	4.04		
6	Ala	7.70	4.17	1.50	
7	Cys	7.96	4.12	3.07, 2.91	SH: 2.30
8	Ala	8.07	4.09	1.48	
9	Ala	8.26	4.06	1.50	
10	Lys	8.13	4.00		
11	Ala	8.09	4.10	1.49	
12	Ala	8.22	4.12	1.51	
13	Ala	8.26	4.05	1.50	
14	Ala	8.22	4.10	1.51	
15	Lys	8.01	4.02		
16	Ala	8.13	4.10	1.49	
17	Ala	8.15	4.12	1.50	
18	Ala	7.90	4.15	1.49	
19	Ala	7.76	4.21	1.47	
20	Lys	7.65	4.29		
21	Ala	7.83	4.31	1.44	

Table D.1: ^{1}H chemical shift assignments for 4K(3,7)

D.2 4K(3,11)

Number	Residue	H^N	H^α	H^β	Other
1	Ala	—	4.06		
2	Ala	8.65	4.33	1.40	
3	Cys	8.18	4.52	3.06, 2.95	SH: 2.13
4	Ala	8.46	4.17	1.47	
5	Lys	8.05	4.05		
6	Ala	7.70	4.14	1.47	
7	Ala	7.90	4.07	1.48	
8	Ala	8.15	4.11	1.53	
9	Ala	8.16	4.05	1.49	
10	Lys	8.18	3.99		
11	Cys	8.21	4.11	3.08, 2.88	SH: 2.39
12	Ala	8.31	4.06	1.50	
13	Ala	8.22	4.14	1.51	
14	Ala	8.30	4.04	1.51	
15	Lys	8.12	4.03		
16	Ala	8.19	4.10	1.51	
17	Ala	8.06	4.09	1.48	
18	Ala	7.90	4.15	1.48	
19	Ala	7.75	4.21	1.44	
20	Lys	7.66	4.28		
21	Ala	7.82	4.30	1.51	

Table D.2: ^{1}H chemical shift assignments for 4K(3,11)

D.3 4K(3,12)

Number	Residue	H^N	H^α	H^β	Other
1	Ala	—	4.06		
2	Ala	8.65	4.33	1.41	
3	Cys	8.19	4.53	3.06, 2.96	SH: 2.14
4	Ala	8.47	4.18	1.48	
5	Lys	8.04	4.05		
6	Ala	7.71	4.14	1.48	
7	Ala	7.89	4.06	1.49	
8	Ala	8.12	4.13	1.49	
9	Ala	8.16	4.06	1.49	
10	Lys	8.08	4.00		
11	Ala	8.16	4.05	1.49	
12	Cys	8.38	4.09	3.11, 2.87	SH: 2.57
13	Ala	8.28	4.10	1.51	
14	Ala	8.16	4.11	1.49	
15	Lys	8.06	4.01		
16	Ala	8.29	4.12	1.51	
17	Ala	8.09	4.10	1.49	
18	Ala	7.90	4.16	1.49	
19	Ala	7.77	4.21	1.49	
20	Lys	7.67	4.29		
21	Ala	7.85	4.32	1.45	

Table D.3: ^{1}H chemical shift assignments for 4K(3,12)

D.4 4K(3,13)

Number	Residue	H^N	H^α	H^β	Other
1	Ala	—	4.07		
2	Ala	8.66	4.33	1.41	
3	Cys	8.20	4.52	3.06, 2.96	SH: 2.13
4	Ala	8.47	4.18	1.48	
5	Lys	8.06	4.06		
6	Ala	7.71	4.15	1.48	
7	Ala	7.91	4.06	1.49	
8	Ala	8.10	4.11	1.50	
9	Ala	8.25	4.10	1.51	
10	Lys	8.18	3.96		
11	Ala	8.09	4.14	1.50	
12	Ala	8.23	4.13	1.56	
13	Cys	8.22	4.15	3.12, 2.88	SH: 2.56
14	Ala	8.25	4.08	1.56	
15	Lys	8.19	4.00		
16	Ala	8.10	4.02	1.50	
17	Ala	8.00	4.11	1.49	
18	Ala	7.90	4.16	1.49	
19	Ala	7.75	4.22	1.48	
20	Lys	7.69	4.30		
21	Ala	7.84	4.30	1.44	

Table D.4: ^{1}H chemical shift assignments for 4K(3,13)

D.5 4K(3,14)

Number	Residue	H^N	H^α	H^β	Other
1	Ala	—	4.06		
2	Ala	8.65	4.33	1.40	
3	Cys	8.19	4.52	3.06, 2.96	SH: 2.14
4	Ala	8.47	4.18	1.48	
5	Lys	8.06	4.05		
6	Ala	7.71	4.14	1.47	
7	Ala	7.92	4.07	1.48	
8	Ala	8.07	4.08	1.50	
9	Ala	8.11	4.11	1.49	
10	Lys	8.12	4.05		
11	Ala	8.14	4.14	1.50	
12	Ala	8.13	4.13	1.50	
13	Ala	8.21	4.06	1.51	
14	Cys	8.34	4.15	3.12, 2.87	SH: 2.56
15	Lys	8.21	4.05		
16	Ala	8.18	4.00	1.51	
17	Ala	8.00	4.10	1.50	
18	Ala	7.92	4.15	1.48	
19	Ala	7.76	4.22	1.48	
20	Lys	7.67	4.29		
21	Ala	7.84	4.30	1.44	

Table D.5: ^{1}H chemical shift assignments for 4K(3,14)

D.6 4K(3,18)

Number	Residue	H^N	H^α	H^β	Other
1	Ala	—	4.06		
2	Ala	8.64	4.32	1.41	
3	Cys	8.18	4.51	3.07, 2.96	SH: 2.14
4	Ala	8.45	4.17	1.47	
5	Lys	8.04	4.04		
6	Ala	7.71	4.13	1.48	
7	Ala	7.90	4.05	1.43	
8	Ala	8.10	4.10	1.51	
9	Ala	8.22	4.05	1.51	
10	Lys	8.12	4.00		
11	Ala	8.08	4.11	1.51	
12	Ala	8.19	4.10	1.50	
13	Ala	8.19	4.12	1.50	
14	Ala	8.23	4.07	1.52	
15	Lys	8.08	4.00		
16	Ala	8.02	4.10	1.53	
17	Ala	8.00	4.15	1.53	
18	Cys	7.87	4.33	3.05, 2.94	SH: 2.54
19	Ala	7.85	4.25	1.47	
20	Lys	7.80	4.31		
21	Ala	7.91	4.29	1.43	

Table D.6: ^{1}H chemical shift assignments for 4K(3,18)

Bibliography

[1] C. M. Dobson. Protein folding and misfolding. *Nature*, 426:884–890, 2003.

[2] R. Sitia and I. Braakman. Quality control in the endoplasmic reticulum protein factory. *Nature*, 426:891–894, 2003.

[3] A. L. Goldberg. Protein degradation and protection against misfolded or damaged proteins. *Nature*, 426:895–899, 2003.

[4] D. J. Selkoe. Folding proteins in fatal ways. *Nature*, 426:900–904, 2003.

[5] F. E. Cohen and J. W. Kelly. Therapeutic approaches to protein-misfolding diseases. *Nature*, 426:905–909, 2003.

[6] C. Levinthal. Are there pathways for protein folding? *J. Chim. Phys.*, 65:44–45, 1968.

[7] C. Levinthal. How to fold graciously. In P. Debrunner, J. C. M. Tsibris, and E. Münck, editors, *Mössbauer spectroscopy in biological systems*, pages 22–23. University of Illinois at Urbana-Champagne, 1969.

[8] L. Qui, S. A. Pabit, A. E. Roitberg, and S. J. Hagen. Smaller and faster: the 20-residue Trp-cage protein folds in 4 μs. *J. Am. Chem. Soc.*, 124(44):12952–12953, 2002.

[9] J. K. Myers and T. G. Oas. Preorganised secondary structure as an important determinant of fast protein folding. *Nat. Struct. Biol.*, 8(6):552–558, 2001.

[10] D. M. Vu, J. K. Myers, T. G. Oas, and R. B. Dyer. Probing the folding and unfolding dynamics of secondary and tertiary structures in a three-helix bundle protein. *Biochemistry*, 43(12):3582–3589, 2004.

[11] D. Estapé and U. Rinas. Folding kinetics of the all-β-sheet protein human basic fibroblast growth factor, a structural homolog of interleukin-1β. *J. Biol. Chem.*, 274(48):34083–34088, 1999.

[12] D. Bashford, F. E. Cohen, M. Karplus, I. D. Kuntz, and D. L. Weaver. Diffusion-collision model for the folding kinetics of myoglobin. *Proteins*, 4:211–227, 1988.

[13] P. S. Kim and R. L. Baldwin. Intermediates in the folding reactions of small proteins. *Annu. Rev. Biochem.*, 59:631–660, 1990.

[14] M. Karplus and D. L. Weaver. Protein folding dynamics: the diffusion-collision model and experimental data. *Protein Sci.*, 3:650–668, 1994.

[15] D. B. Wetlaufer. Nucleation, rapid folding and globular intrachain regions in proteins. *Proc. Natl. Acad. Sci. USA*, 70:697–701, 1973.

[16] D. B. Wetlaufer. Nucleation in protein folding — confusion of structure and process. *Trends Biochem. Sci.*, 15:414–415, 1990.

[17] K. Kuwajima. The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins*, 6:87–103, 1989.

[18] O. B. Ptitsyn. How the molten globule became. *Trends Biochem. Sci.*, 20:376–379, 1995.

[19] K. A. Dill and H. S. Chan. From Levinthal to pathways to funnels. *Nat. Struct. Biol.*, 4(1):4–6, 1997.

[20] C. Frieden, S. D. Hoeltzli, and I. J. Ropson. NMR and protein folding: equilibrium and stopped-flow studies. *Protein Sci.*, 2:2007–2014, 1993.

[21] A. K. Bhuyan and J. B. Udgaonkar. Real-time NMR measurements of protein folding and hydrogen exchange dynamics. *Curr. Sci.*, 77(7):942–952, 1997.

[22] N. A. J. van Nuland, V. Forge, J. Balbach, and C. M. Dobson. Real-time NMR studies of protein folding. *Acc. Chem. Res.*, 37:773–780, 1998.

[23] C. M. Dobson and P. J. Hore. Kinetic studies of protein folding using NMR spectroscopy. *Nat. Struct. Biol.*, 5:504–507, 1998.

[24] T. R. Killick, S. M. V. Freund, and A. R. Fersht. Real-time NMR studies on folding of mutants of barnase and chymotrypsin inhibitor 2. *FEBS Lett.*, 423:110–112, 1998.

[25] S. D. Hoeltzli and C. Frieden. Refolding of [6-^{19}F] tryptophan-labelled *Escherichia coli* dihydrofolate reductase in the presence of ligand: a stopped-flow NMR spectroscopy study. *Biochemistry*, 37:387–398, 1998.

[26] P. J. Hore, S. L. Winder, C. H. Roberts, and C. M. Dobson. Stopped-flow photo-CIDNP observation of protein folding. *J. Am. Chem. Soc.*, 119(21):5049–5050, 1997.

[27] K. Hun Mok, T. Nagashima, I. J. Day, J. A. Jones, C. J. V. Jones, C. M. Dobson, and P. J. Hore. Rapid sample-mixing technique for transient NMR and Photo-CIDNP spectroscopy: Applications to real-time protein folding. *J. Am. Chem. Soc.*, 125:12484–12492, 2003.

[28] C. E. Lyon, J. A. Jones, C. Redfield, C. M. Dobson, and P. J. Hore. Two-dimensional ^{15}N-^{1}H photo-CIDNP as a surface probe of native and partially structured proteins. *J. Am. Chem. Soc.*, 121(27):6505–6506, 1999.

[29] T. Kühn and H. Schwalbe. Monitoring the kinetics of ion-dependent protein folding by time-resolved NMR spectroscopy at atomic resolution. *J. Am. Chem. Soc.*, 122:6169–6174, 2000.

[30] J. Wirmer, T. Kühn, and H. Schwalbe. Millisecond time resolved photo-CIDNP NMR reveals a non-native folding intermediate on the ion-induced refolding pathway of bovine α-lactalbumin. *Angew. Chem. Int. Ed.*, 40(22):4248–4251, 2001.

[31] C. E. Lyon. *Photo-CIDNP and Protein Folding.* DPhil thesis, University of Oxford, 1999.

[32] C. E. Lyon, E-S. Suh, C. M. Dobson, and P. J. Hore. Probing the exposure of tyrosine and tryptophan residues in partially folded proteins and folding intermediates by CIDNP pulse labelling. *J. Am. Chem. Soc.*, 124(44):13018–13024, 2002.

[33] D. Canet, C. E. Lyon, R. M. Scheek, G. T. Robillard, C. M. Dobson, P. J. Hore, and N. A. J. van Nuland. Rapid formation of non-native contacts during the folding of HPr revealed by real-time photo-CIDNP NMR and stopped-flow fluorescence experiments. *J. Mol. Biol.*, 350:397–407, 2003.

[34] C. Steegborn, H. Schneider-Hassloff, M. Zeeb, and J. Balbach. Cooperativity of a protein folding reaction probed at multiple chain positions by real-time 2D NMR spectroscopy. *Biochemistry*, 39:7910–7919, 2000.

[35] M. Mizuguchi, G. J. Kroon, P. E. Wright, and H. J. Dyson. Folding of a β-sheet protein monitored by real-time NMR spectroscopy. *J. Mol. Biol.*, 328:1161–1171, 2003.

[36] S. M. Harper, L. C. Neil, and K. H. Gardner. Structural basis of a phototropin light switch. *Science*, 301:1541–1544, 2003.

[37] J. T. M. Kennis, I. H. M. van Stokkum, S. Crosson, M. Gauden, K. Moffat, and R. van Grondelle. The LOV2 domain of phototropin: a reversible photochromic switch. *J. Am. Chem. Soc*, 126(14):4512–4513, 2004.

[38] S. M. Harper, L. C. Neil, I. J. Day, P. J. Hore, and K. H. Gardner. Conformational changes in a phototropin LOV domain monitored by time-resolved NMR spectroscopy. *J. Am. Chem. Soc.*, 126:3390–3391, 2004.

[39] F. Bloch. Nuclear induction. *Phys. Rev.*, 70(7,8):460–474, 1946.

[40] F. Bloch, W. W. Hansen, and M. Packard. The nuclear induction experiment. *Phys. Rev.*, 70(7,8):474–485, 1946.

[41] E. M. Purcell, H.C. Torrey, and R. V. Pound. Resonance absorption by nuclear magnetic moments in a solid. *Phys. Rev.*, 69:37–38, 1946.

[42] R. R. Ernst, G. Bodenhausen, and A. Wokaun. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions.* Oxford Science Publications, first edition, 1990.

[43] J. Bargon, H. Fischer, and U. Johnsen. Kernresonanz-emissionslinien während rascher radikalreaktionen I: aufnahmeverfahren und beispiele. *Z. Naturforsch.*, 22a:1551–1555, 1967.

[44] H. R. Ward and R. G. Lawler. Nuclear magnetic resonance emission and enhanced absorption in rapid organometallic reactions. *J. Am. Chem. Soc.*, 89:5518–5519, 1967.

[45] J. Bargon and H. Fischer. Kernresonanz-emissionslinien während rascher radikalreaktionen II: chemisch induzierte dynamische kernpolarisation. *Z. Naturforsch.*, 22a:1556–1562, 1967.

[46] R. G. Lawler. Chemically induced dynamic nuclear polarisation. *J. Am. Chem. Soc.*, 89:5519–5521, 1969.

[47] H. Fischer and J. Bargon. Chemically induced nuclear polarisation during thermal decomposition of peroxides and azo compounds. *Acc. Chem. Res.*, 2:110–114, 1969.

[48] G. L. Closs. A mechanism explaining nuclear spin polarisation in radical combination reactions. *J. Am. Chem. Soc.*, 91:4552–4554, 1969.

[49] R. Kaptein and L. J. Oosterhoff. Chemically induced dynamic nuclear polarisation III: (anomalous multiplets of radical coupling and disproportionation products). *Chem. Phys. Lett.*, 4:214–216, 1969.

[50] R. Kaptein and L. J. Oosterhoff. Chemically induced dynamic nuclear polarisation II: (relation with anomalous ESR spectra). *Chem. Phys. Lett.*, 4:195–197, 1969.

[51] U. E. Steiner and T. Ulrich. Magnetic field effects in chemical kinetics and related phenomena. *Chem. Rev.*, 89:51–147, 1989.

[52] R. Kaptein. Chemically induced dynamic nuclear polarisation: theory and applications in mechanistic chemistry. *Adv. Free Radical Chem.*, 5:319–380, 1975.

[53] R. Kaptein, K. Dijkstra, and K. Nicolay. Laser photo-CIDNP as a surface probe for proteins in solution. *Nature*, 274:293–294, 1978.

[54] O. W. Sørensen, G. W. Eich, M. H. Levitt, G. Bodenhausen, and R. R. Ernst. Product operator formalism for the description of NMR pulse experiments. *Prog. Nucl. Mag. Res. Sp.*, 16:163–192, 1983.

[55] P. J. Hore and R. W. Broadhurst. Photo-CIDNP of biopolymers. *Prog. NMR. Spec.*, 25:345–402, 1993.

[56] R. Kaptein. Simple rules for chemically induced dynamic nuclear polarisation. *J. Chem. Soc. D Chem. Comm.*, 14:732–733, 1971.

[57] P. J. Hore, S. Stob, J. Kemmink, and R. Kaptein. An exception to the CIDNP sign rules. *Chem. Phys. Lett.*, 98(5):409–413, 1983.

[58] S. Stob and R. Kaptein. Photo-CIDNP of the amino acids. *Photochem. Photobiol.*, 49(5):565–577, 1989.

[59] Yu. P. Tsentalovich, J. J. Lopez, P. J. Hore, and R. Z. Sagdeev. Mechanisms of reactions of flavin mononucleotide triplet with aromatic amino acids. *Spectrochim. Acta. A*, 58:2043–2050, 2002.

[60] R. Kaptein. Photo-CIDNP studies of proteins. In L. G. Berliner and J. Reubens, editors, *Biological Magnetic Resonance*, volume 4, pages 145–191. Plenum Press, New York, 1982.

[61] S. L. Winder, R. W. Broadhurst, and P. J. Hore. Photo-CIDNP of amino acids and proteins: effects of competition for flavin triplets. *Spectrochim. Acta. A*, 51:1753–1761, 1995.

[62] A. T. Alexandrescu, R. W. Broadhurst, C. Wormald, C-L. Chyan, J. Baum, and C. M. Dobson. ^{1}H-NMR assignments and local environments of aromatic residues in bovine, human and guinea pig variants of α-lactalbumin. *Eur. J. Biochem.*, 210:699–709, 1992.

[63] F. J. J. de Kanter and R. Kaptein. CIDNP transfer via dipolar relaxation and spin-spin coupling. *Chem. Phys. Lett.*, 62(3):421–426, 1979.

[64] P. J. Hore, M. R. Egmond, H. T. Edzes, and R. Kaptein. Cross-relaxation effects in the Photo-CIDNP spectra of amino acids and proteins. *J. Magn. Reson.*, 49:122–150, 1982.

[65] I. Kuprov and P. J. Hore. Chemically amplified ^{19}F–^{1}H nuclear Overhauser effects. *J. Magn. Reson.*, 168:1–7, 2004.

[66] A. G. Cochran, N. J. Skelton, and M. A. Starovasnik. Tryptophan zippers: stable, monomeric β-hairpins. *Proc. Natl. Acad. Sci. USA*, 98(10):5578–5583, 2001.

[67] N. L. Gregory, T. D. W. Claridge, and M. Leonard. Thermoelectric cooling for NMR sample temperature control. *J. Magn. Reson.*, 124:228–231, 1997.

[68] J. E. Scheffler, C. E. Cottrell, and L. J. Berliner. An inexpensive, versatile sample illuminator for Photo-CIDNP on any NMR spectrometer. *J. Magn. Reson.*, 63:199–201, 1985.

[69] I. Kuprov and P. J. Hore. Uniform illumination of optically dense NMR samples. *J. Magn. Reson.*, 171:171–175, 2004.

[70] P. F. Heelis. The photochemistry of flavins. In F. Müller, editor, *The Chemistry and Biochemistry of Flavoenzymes*, volume 1, pages 171–193. CRC Press, Boca Raton, 1991.

[71] F. Müller. Free flavins: synthesis, chemical and physical properties. In F. Müller, editor, *The Chemistry and Biochemistry of Flavoenzymes*, volume 1, pages 1–71. CRC Press, Boca Raton, 1991.

[72] P. J. Hore. Solvent suppression. In N. J. Oppenheimer and T. L. James, editors, *Methods in Enzymology*, volume 176, pages 64–77. Academic Press, San Diego, 1989.

[73] Tsang-Lin Hwang and A. J. Shaka. Water suppression that works. Excitation sculpting using arbitrary waveforms and pulsed field gradients. *J. Magn. Reson. A.*, 112:275–279, 1995.

[74] J. Cavanagh, W. J. Fairbrother, A. G. Palmer III, and N. J. Skelton. *Protein NMR spectroscopy: principles and practice.* Academic Press, first edition, 1996.

[75] M. H. Levitt. *Spin Dynamics.* John Wiley and Sons, Ltd, first edition, 2001.

[76] A. Bax and D. G. Davis. MLEV-17-based two-dimensional homonuclear magnetisation transfer spectroscopy. *J. Magn. Reson.*, 65:355–360, 1985.

[77] L. E. Kay, P. Keifer, and T. Saarinen. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.*, 114:10663–10665, 1992.

[78] A. G. Palmer III, J. Cavanagh, P. E. Wright, and M. Rance. Sensitivity improvement in proton-detected two-dimensional heternuclear correlation NMR spectroscopy. *J. Magn. Reson.*, 93:151–170, 1991.

[79] J. Cavanagh, A. G. Palmer III, P. E. Wright, and M. Rance. Sensitivity improvement in proton-detected two-dimensional heteronuclear relay spectroscopy. *J. Magn. Reson.*, 91:429–436, 1991.

[80] A. L. Davis, J. Keeler, E. D. Laue, and D. Moskau. Experiments for recording pure-absorption heteronuclear correlation spectra using pulsed field gradients. *J. Magn. Reson.*, 98:207–216, 1992.

[81] J. Boyd, N. Soffe, B. John, D. Plant, and R. Hurd. The generation of phase-sensitive 2D ^{15}N-^{1}H spectra using gradient pulses for coherence-transfer-pathway selection. *J. Magn. Reson.*, 98:660–664, 1992.

[82] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR.*, 6:277–293, 1995.

[83] B. A. Johnson and R. A. Blevins. NMRView: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR*, 4:603–614, 1994.

[84] S. M. Kelly and N. C. Price. The use of circular dichroism in the investigation of protein structure and function. *Curr. Protein Peptide Sci.*, 1:349–384, 2000.

[85] F. R. Salemme. Structure and function of cytochromes *c*. *Annu. Rev. Biochem.*, 46:299–329, 1977.

[86] A. Hori, F. Hayashi, Y. Kyogoku, and H. Akutsu. A photo-chemically induced dynamic nucelar polarisation NMR study on rabbit and bovine cytochrome b_5. *Eur. J. Biochem.*, 174:503–508, 1988.

[87] E. J. Tomlinson and S. J. Ferguson. Conservation of a *c* type cytochrome to a *b* type that spontaneously forms *in vitro* from apo protein and heme: Implications for *c* type cytochrome biogenesis and folding. *Proc. Natl. Acad. Sci. USA.*, 97(10):5156–5160, 2000.

[88] J. Hasegawa, T. Yoshida, T. Yamazaki, Y. Sambongi, Y. Yu, Y. Igaeashi, T. Kodama, K. Yamazaki, Y. Kyogoku, and Yu. Kobayashi. Solution structure of thermostable cytochrome *c*-552 from *Hydrogenobacter thermophilus* determined by ^{1}H-NMR spectroscopy. *Biochemistry*, 37:9641–9649, 1998.

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

[90] R. Wain, T. A. Pertinhez, E. J. Tomlinson, L. Hong, C. M. Dobson, S. J. Ferguson, and L. J. Smith. The cytochrome *c* fold can be attained from a compact apo state by occupancy of a nascent heme binding site. *J. Biol. Chem.*, 276(49):45813–45817, 2001.

[91] E. J. Tomlinson and S. J. Ferguson. Loss of either of the two heme-binding cysteines from a class I *c*-type cytochrome has a suprisingly small effect on physicochemical properties. *J. Biol. Chem.*, 275(42):32530–32534, 2000.

[92] T. A. Pertinhez, M. Mouchard, E. J. Tomlinson, R. Wain, S. J. Ferguson, C. M. Dobson, and L. J. Smith. Amyloid fibril formation by a helical cytochrome. *FEBS Lett.*, 495:184–186, 2001.

[93] D. Eliezer and P. E. Wright. Is apomyoglobin a molten globule? Structural characterisation by NMR. *J. Mol. Biol.*, 263:531–538, 1996.

[94] E. F. Karan, B. S. Russell, and K. L. Bren. Characterization of *Hydrogenobactor thermophilus* cytochromes $c_5$52 expressed in the cytoplasma and periplasm of *Escherichia coli*. *J. Biol. Inorg. Chem.*, 7:260–272, 2002.

[95] I. Bertini and C. Luchinat. *NMR of paramagnetic molecules in biological systems.* The Benjamin/Cummins Publishing Company Inc., first edition, 1986.

[96] R. Wain. *Studies of Protein Folding and Aggregation.* DPhil thesis, University of Oxford, 2003.

[97] R. Wain, C. Redfield, S. J. Ferguson, and L. J. Smith. NMR analysis shows that a *b*-type variant of *Hydrogenobactor thermophilus* Cytochrome c_{552} retains its native structure. *J. Biol. Chem.*, 279:15177–15182, 2004.

[98] G. R. Moore and R. J. P. Williams. Nuclear-magnetic-resonance studies of eukaryotic cytochrome *c*. Assignment of resonances of aromatic amino acids. *Eur. J. Biochem.*, 103:493–502, 1980.

[99] G. R. Moore and R. J. P. Williams. Nuclear-magnetic-resonance studies of eukaryotic cytochrome *c*. Assignment of resonances of aliphatic amino acids. *Eur. J. Biochem.*, 103:503–512, 1980.

[100] G. R. Moore and R. J. P. Williams. Nuclear-magnetic-resonance studies of ferrocytochrome *c*. pH and temperature dependence. *Eur. J. Biochem.*, 103:513–521, 1980.

[101] G. R. Moore and R. J. P. Williams. The stability of ferricytochrome *c*. Temperature dependence of its NMR spectrum. *Eur. J. Biochem.*, 103:523–532, 1980.

[102] A. P. Boswell, G. R. Moore, R. J. P. Williams, J. C. W. Chien, and L. C. Dickinson. Nuclear magnetic resonance studies of the phenylalanine residues of eukaryotic cytochrome *c*. *J. Inorg. Biochem.*, 13:347–352, 1980.

[103] B. Lee and F. M. Richards. The interpretation of protein structures: estimation of static accessibilities. *J. Mol. Biol.*, 55:379–400, 1971.

[104] S. J. Hubbard, S. F. Campbell, and J. M. Thornton. Molecular recognition: conformational analysis of limited proteolytic sites and serine proteinase protein inhibitors. *J. Mol. Biol.*, 220:507–530, 1991.

[105] O. B. Morozova, A. V. Yurkovskaya, R. Z. Sagdeev, K. H. Mok, and P. J. Hore. Time-resolved CIDNP study of native-state bovine and human α-lactalbumin. *J. Phys. Chem. B*, 108:15355–15363, 2004.

[106] M. T. Stankovich. Redox properties of flavins and flavoproteins. In F. Müller, editor, *The Chemistry and Biochemistry of Flavoenzymes*, volume 1, pages 401–425. CRC Press, Boca Raton, 1991.

[107] M. S. Chao. The sulfite/dithionite couple: its standard potential and Pourbaix diagram. *J. Electrochem. Soc.*, 133(5):954–955, 1986.

[108] M. M. E. Snel, R. Kaptein, and B. de Kruijff. Interaction of apocytochrome-*c* and derived polypeptide fragments with sodium dodecyl sulfate micelles monitored by photochemically induced dynamic nuclear polarisation ^{1}H NMR and fluorescence spectroscopy. *Biochemistry*, 30:3387–3395, 1991.

[109] J. Grimaldi, J. Baldo, C. McMurray, and B. D. Sykes. Stopped-flow nuclear magnetic resonance spectroscopy. *J. Am. Chem. Soc.*, 94(22):7641–7645, 1972.

[110] R. O. Kühne, T. Schaffhauser, A. Wokaun, and R. R. Ernst. Study of transient chemical reactions by NMR. Fast stopped-flow fourier transform experiments. *J. Magn. Reson.*, 35(1):39–67, 1979.

[111] J. F. McGarrity, J. Prodolliet, and T. Smyth. Rapid injection NMR: a simple technique for the observation of reactive intermediates. *Org. Magn. Resonance*, 17(1):59–65, 1981.

[112] J. F. McGarrity and J. Prodolliet. High-field rapid injection NMR: observation of unstable primary ozonide intermdiates. *J. Am. Chem. Soc.*, 49:4465–4470, 1984.

[113] D. B. Green, J. Lane, and R. M. Wing. A standard session stopped-flow NMR tube. *Appl. Spectrosc.*, 41(5):847–851, 1987.

[114] W. A. McGee and L. J. Parkhust. A combined nuclear magnetic resonance and absorbance stopped-flow apparatus for biochemical studies. *Anal. Biochem.*, 189:267–273, 1990.

[115] P. A. Keifer, S. H. Smallcombe, E. H. Williams, K. E. Salomon, G. Mendez, J. L. Belletire, and C. D. Moore. Direct-injection NMR: A flow NMR techniques for the analysis of combinatorial chemistry libraries. *J. Comb. Chem.*, 2:151–171, 2000.

[116] M. Hamang, A. Sanson, L. Liagre, V. Forge, and P. Berthault. Fast mixing device inside a nuclear magnetic resonance magnet: a tool for observing early steps in protein folding. *Rev. Sci. Inst.*, 71:2180–2183, 2000.

[117] K. Maeda, C. E. Lyon, J. J. Lopez, M. Čemažar, C. M. Dobson, and P. J. Hore. Improved photo-CIDNP methods for studying protein structure and folding. *J. Biomol. NMR.*, 16:235–244, 2000.

[118] M. Piotto, V. Saudek, and V. Sklenár. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR*, 2(6):661–665, 1992.

[119] V. Sklenár, M. Piotto, R. Leppik, and V. Saudek. Gradient-taliored water suppression for ^{1}H-^{15}N HSQC experiments optimized to retain full sensitivity. *J. Magn. Reson. A*, 102:241–245, 1993.

[120] M. Liu, X-a. Mao, C. Ye, H. Huang, J. K. Nicholson, and J. C. Lindon. Improved WATERGATE pulse sequences for solvent suppression in NMR spectroscopy. *J. Magn. Reson.*, 132:125–129, 1998.

[121] T. D. W. Claridge. *High-resolution NMR techniques in organic chemistry.* Pergamon, first edition, 1999.

[122] R. J. Ogg, P. B. Kingsley, and J. S. Taylor. WET, a T_1- and B_1-insensitive water-suppression method for *in vivo* localised ^{1}H NMR spectrocopy. *J. Magn. Reson. B.*, 104:1–10, 1994.

[123] R. H. Griffey and D. P. Flamig. VAPOR for solvent-suppressed, short-echo, volume-locaised proton spectroscopy. *J. Magn. Reson.*, 88(1):161–166, 1990.

[124] C. T. Moonen and P. C. M. van Zijl. Highly effective water suppresssion for *in vivo* proton NMR spectroscopy (DRYSTEAM). *J. Magn. Reson.*, 88:28–41, 1990.

[125] A. Haase, J. Frahm, W. Hänicke, and D. Matthaei. ^{1}H NMR chemical shift (CHESS) selective imaging. *Phys. Med. Biol.*, 30(4):341–344, 1985.

[126] Y-y. Lin, N. Lisitza, S. Ahn, and W. S. Warren. Resurrection of crushed magnetization and chaotic dynamics in solution NMR spectroscopy. *Science*, 290:118–121, 2000.

[127] S. Y. Huang, Y-y. Lin, N. Lisitza, and W. S. Warren. Signal interferences from turbulent spin dynamics in solution nuclear magnetic resonance spectroscopy. *J. Chem. Phys.*, 116(23):10325–10337, 2002.

[128] W. S. Warren. Personnal communication, 2002.

[129] T. Ernst and J. Hennig. Improved water suppression for localised *in vivo* ^{1}H spectroscopy. *J. Magn. Reson. B*, 106(2):181–186, 1995.

[130] D. G. Gadian. *NMR and its applications to living systems.* Oxford Science Publications, second edition, 1995.

[131] P. Mansfield and P. G. Morris. NMR imaging in biomedicine. In J. S. Waugh, editor, *Advances in Magnetic Resonance*, volume Supplement 2. Academic Press, 1982.

[132] F. C. Champion and N. Davy. *Properties of matter.* Blackie and son, third edition, 1965.

[133] D. R. Lide. *CRC Handbook of chemistry and physics.* Chemical rubber publishing company, seventy first edition, 1990-1991.

[134] M. Arai and K. Kuwajima. Role of the molten globule state in protein folding. In C. R. Matthews, editor, *Adv. Protein Chem.*, volume 53, pages 209–282. Academic Press, 2000.

[135] J. Baum, C. M. Dobson, P. A. Evans, and C. Hanley. Characterization of a partly folded protein by NMR methods: studies on the molten globule state of guinea pig α-lactalbumin. *Biochemistry*, 28:7–13, 1989.

[136] K. Kuwajima. The molten globule state of α-lactalbumin. *FASEB Journal*, 10:102–109, 1996.

[137] V. E. Bychkova, R. Berni, G. L. Rossi, V. P. Kutyshenko, and O. B. Ptitsyn. Retinol-binding protein is in the molten globule state at low pH. *Biochemistry*, 31(33):7566–7571, 1992.

[138] J. Balbach, V. Forge, Wai Shun Lau, J. A. Jones, N. A. van Nuland, and C. M. Dobson. Detection of residue contacts in a protein folding intermediate. *Proc. Natl. Acad. Sci. USA.*, 94:7182–7185, 1997.

[139] C. Blake, L. Johnson, G. Mair, A. North, D. Phillips, and V. Sarma. Crystallographic studies of the activity of hen egg-white lysozyme. *Proc. Roy. Soc.*, B167:378, 1967.

[140] L. J. Smith, M. J. Sutcliffe, C. Redfield, and C. M. Dobson. Structure of hen lysozyme in solution. *J. Mol. Biol.*, 229:930–944, 1993.

[141] C. Redfield, F. M. Poulsen, and C. M. Dobson. Complete assignment of the ^{1}H NMR spectrum of the aromatic residues of Lysozyme. *Eur. J. Biochem.*, 128:527–531, 1982.

[142] C. Redfield and C. M. Dobson. Sequential ^{1}H NMR assignments and secondary structure of hen egg white lysozyme in solution. *Biochemistry*, 27:122–136, 1988.

[143] Y. Wang, T. C. Bjorndahl, and D. S. Wishart. Complete ^{1}H and non-carbonylic ^{13}C assignments of native hen egg-white lysozyme. *J. Biol. NMR*, 17:83–83, 2000.

[144] H. Schwalbe, S. B. Grimshaw, A. Spencer, M. Buck, J. Boyd, C. M. Dobson, C. Redfield, and L. J. Smith. A refined solution structure of hen lysozyme determined using residual dipolar coupling data. *Protein Sci.*, 10:677–688, 2001.

[145] B. Lai, A. Cao, and L. Lai. Organic cosolvents and hen egg-white lysozyme folding. *Biochim. Biophys. Acta*, 1543:115–122, 2000.

[146] M. Kotik, S. E. Radford, and C. M. Dobson. Comparison of the refolding of hen lysozyme from dimethyl sulfoxide and guanidinium chloride. *Biochemistry*, 34:1714–1724, 1995.

[147] S. Bhattacharjya and P. Balaram. Effects of organic solvents on protein structures: observation of a structured helical core in hen egg-white lysozyme in aqueous dimethylsulfoxide. *Proteins*, 29:492–507, 1997.

[148] S. C. Mande and M. E. Sobhia. Structural characterisation of protein-denaturant interactions: crystal structures of hen egg-white lysozyme in complex with DMSO and guanidinium chloride. *Protein Eng.*, 13(2):133–141, 2000.

[149] M. Buck, S. E. Radford, and C. M. Dobson. A partially folded state of Hen Egg White Lysozyme in trifluoroethanol: structural characterisation and implications for protein folding. *Biochemistry*, 32:669–678, 1993.

[150] M. Buck, H. Schwalbe, and C. M. Dobson. Characterisation of conformational preferences in a partly folded protein by heteronuclear NMR spectroscopy: assigment and secondary structure analysis of Hen Egg-white Lysozyme in trifluoroethanol. *Biochemistry*, 34:13219–13232, 1995.

[151] H. Lu, M. Buck, S. E. Radford, and C. M. Dobson. Acceleration of the folding of Hen Lysozyme by trifluoroethanol. *J. Mol. Biol.*, 265:112–117, 1997.

[152] F. Chiti, N. Taddei, P. Webster, D. Hamada, T. Fiaschi, G. Ramponi, and C. M. Dobson. Acceleration of the folding of acylphosphotase by stabilisation of local seconadary structure. *Nat. Struct. Biol.*, 6:380–387, 1999.

[153] A. R. Fersht. Nucleation mechanism in protein folding. *Curr. Opin. Struct. Biol.*, 7:3–9, 1997.

[154] C. M. Dobson, A. Šali, and M Karplus. Protein folding: a perspective from theory and experiment. *Angew. Chem. Int. Ed.*, 37:868–893, 1998.

[155] A. C. W. Pike, K. Brew, and K. R. Acharya. Crystal structures of guinea-pig, goat and bovine α-lactalbumin highlight the enhanced conformational flexibility of regions that are significant for its action in lactose synthase. *Structure*, 4(6):691–703, 1996.

[156] R. Wijesinha-Bettoni, C. M. Dobson, and C. Redfield. Comparison of the structural and dynamical properties of holo and apo bovine α-lactalbumin by NMR spectroscopy. *J. Mol. Biol.*, 307:885–898, 2001.

[157] E. D. Chrysina, K. Brew, and K. R. Acharya. Crystal structures of Apo- and Holo-bovine α-Lactalbumin at 2.2-Å resolution reveal on effct of calcium on inter-lobe interactions. *J. Biol. Chem.*, 275(47):37201–37209, 2000.

[158] J. Ren, D. I. Stuart, and K. R. Acharya. α-Lactalbumin possesses a distinct zinc binding site. *J. Biol. Chem.*, 268(26):19292–19298, 1993.

[159] K. Koga and L. J. Berliner. Structural elucidation of a hydrophobic box in bovine α-lactalbumin by NMR: nuclear overhauser effects. *Biochemistry*, 24:7257–7264, 1985.

[160] J. A. Carver, R. A. Linder, C. Lyon, D. Canet, H. Hernandez, C. M. Dobson, and C. Redfield. The interaction of the molecular chaperone α-crystallin with unfolding α-lactalbumin: a structural and kinetic spectroscopic study. *J. Mol. Biol.*, 318:815–827, 2002.

[161] L. J. Berliner and R. Kaptein. Nuclear magnetic resonance characterisation of aromatic residues of α-lactalbumins. Laser photo chemically induced dynanic nuclear polarisation nuclear magnetic resonance studies of surface exposure. *Biochemistry*, 20:799–807, 1981.

[162] S. Improta, H. Molinari, A. Pastore, R. Consonni, and L. Zetta. Probing protein structure by solvent perturbation of NMR spectra: a comparison with photochemically induced dynamic nuclear polarisation techniques applied to native α-lactalbumin. *Eur. J. Biochem.*, 227:78–86, 1995.

[163] S. Improta, H. Molinari, A. Pastore, R. Consonni, and L. Zetta. Probing protein structure by solvent perturbation of NMR spectra: photochemically induced dynamic nuclear polarisation and paramagnetic perturbation techniques applied to the study of the molten globule state of α-lactalbumin. *Eur. J. Biochem.*, 227:87–96, 1995.

[164] K. Kuwajima, M. Mitani, and S. Sugai. Characterisation of the critical state in protein folding: effects of guandine hydrochloride and specific Ca^{2+} binding on the folding of α-lactalbumin. *J. Mol. Biol.*, 206:547–561, 1989.

[165] M. Arai, K. Ito, T. Inobe, M Nakao, K. Maki, K. Kamagata, H. Kihara, Y. Amemiya, and K. Kuwajima. Fast compaction of the α-lactalbumin during folding studied by stopped-flow X-ray scattering. *J. Mol. Biol.*, 321:121–132, 2002.

[166] C. M. Dobson and P. A. Evans. Protein folding kinetics from magnetisation transfer nuclear magnetic resonance. *Biochemistry*, 23:4267–4270, 1984.

[167] K. Kuwajima, Y. Hiraoka, M. Ikeguchi, and S. Sugai. Comparision of the transient folding intermediates in lysozyme and α-lactalbumin. *Biochemistry*, 24:874–881, 1985.

[168] B. A. Schulman, P. S. Kim, C. M. Dobson, and C. Redfield. A residue-specific NMR view of the non-cooperative unfolding of a molten globule. *Nat. Struct. Biol.*, 4(8):630–634, 1997.

[169] M. W. Lassalle, H. Li, H. Yamada, K. Akasaka, and C. Redfield. Pressure-induced unfolding of the molten globule of all-Ala α-lactalbumin. *Protein Sci.*, 12:66–72, 2003.

[170] B. A. Schulman, C. Redfield, Z-y. Peng, C. M. Dobson, and P. S. Kim. Different subdomains are most protected from hydrogen exchange in the molten globule and native states of human α-lactalbumin. *J. Mol. Biol.*, 253:651–657, 1995.

[171] J. Klein-Seetharaman, M. Oikawa, S. B. Grimshaw, J. Wirmer, E. Duchardt, T. Ueda, T. Imoto, L. J. Smith, C. M. Dobson, and H. Schwalbe. Long-range interactions within a nonnative protein. *Science*, 295:1719–1722, 2002.

[172] A. Matagne and C. M. Dobson. The folding process of hen lysozyme: a perspectuve from the 'new view'. *Cell. Mol. Life Sci.*, 54:363–371, 1998.

[173] T. Kiefhaber. Kinetic traps in lysozyme folding. *Proc. Natl. Acad. Sci. USA*, 92:9029–9033, 1995.

[174] P. J. Connolly and J. C. Hoch. Photochemical degredation of tryptophan residues during CIDNP experiments. *J. Magn. Reson.*, 95(1):165–173, 1991.

[175] F. Y-H. Wu and D. B. McCormick. Flavin-sensitised photooxidation of tryptophan and tyrosine. *Biochim. Biophys. Acta*, 236:479–486, 1971.

[176] P. G. Johnson, A. P. Bell, and D. B. McCormick. Flavin-sensitised photooxidation of histidine. *Photochem. Photobiol.*, 21:205–208, 1975.

[177] T. R. Hopkins and J. D. Spikes. Conformational changes of lysozyme during photodynamic inactivation. *Photochem. Photobiol.*, 12:175–184, 1970.

[178] T. R. Hopkins and J. D. Spikes. Conformational changes in ribonclease during photodynamic inactivation. *Radiat. Res.*, 51:293–301, 1972.

[179] C. B. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181(4069):223–230, 1973.

[180] M. Saunders, A. Wishnia, and J. G. Kirkwood. The nuclear magnetic resonance spectrum of ribonuclease. *J. Am. Chem. Soc.*, 79:3289–3290, 1957.

[181] J. Santoro, C. González, M. Bruix, J. L. Neira, J. L. Neito, J. Herranz, and M. Rico. High-resolution three-dimensional structure of Ribonuclease A in solution by nuclear magnetic resonance spectroscopy. *J. Mol. Biol.*, 229:722–734, 1993.

[182] L-n. Lin and J. F. Brandts. Determination of cis-trans proline isomerisation by trypsin proteolysis: applications to a model pentapeptide and to oxidised ribonuclease A. *Biochemistry*, 22:553–559, 1983.

[183] L-n. Lin and J. F. Brandts. Isomerisation of proline-93 during the unfolding and refolding of ribonuclease A. *Biochemistry*, 22:559–563, 1983.

[184] A. R. Fersht. *Structure and mechanism in protein science.* W. H. Freeman and Company, first edition, 1999.

[185] J. L. Neira and M. Rico. Folding studies on ribonuclease A, a model protein. *Folding and Design*, 2(1):R1–R11, 1997.

[186] J. B. Udgaonkar and R. L. Baldwin. NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. *Nature*, 335:694–699, 1988.

[187] T. Kieifhaber, A. M. Labhardt, and R. L. Baldwin. Direct NMR evidence for an intermediate preceeding the rate-limiting step in the unfolding of ribonuclease A. *Nature*, 375:513–515, 1995.

[188] K. Akasaka, A. Naito, and H. Nakatani. Temperature-jump NMR study of protein folding: ribonuclease A at low pH. *J. Biomol. NMR.*, 1:65–70, 1991.

[189] R. G. Biringer and A. L. Fink. Observation of intermediates in the folding of ribonuclease A at low temperature using proton nuclear magnetic resonance. *Biochemistry*, 21:4748–4755, 1982.

[190] R. G. Biringer and A. L. Fink. Intermediates in the refolding of ribonuclease at subzero tempterature 1: monitoring by nitrotyrosine absorbance. *Biochemistry*, 27:301–311, 1988.

[191] L-n. Lin and J. F. Brandts. Mechanism for the unfolding and refolding of ribonuclease A: kinetic studies utilising spectroscopic methods. *Biochemistry*, 22:564–573, 1983.

[192] J. Juneja and J. B. Udgaonkar. Characterisation of the unfolding of ribonuclease A by pulsed hydrogen exchange study: evidence for competing pathways for unfolding. *Biochemistry*, 41:2641–2654, 2002.

[193] J. Backmann, H. Fabian, and D. Naumann. Temperature-jump-induced refolding of ribonuclease A: a time-resolved FTIR spectroscopic study. *FEBS Lett.*, 364:175–178, 1995.

[194] G. Panick and R. Winter. Pressure-induced unfolding/refolding of ribonuclease A: static and kinetic fourier transform infrared spectroscopy study. *Biochemistry*, 39:1862–1869, 2000.

[195] R. G. Biringer, C. M. Austin, and A. L. Fink. Intermediates in the refolding of ribonuclease at subzero tempterature 2: monitoring by inhibitor binding and catalytic activity. *Biochemistry*, 27:311–315, 1988.

[196] F. X. Schmid and R. L. Baldwin. The rate of interconversion between the two unfolded forms of ribonuclease A does not depend on GdnCl concentration. *J. Mol. Biol.*, 135:199–215, 1979.

[197] P. S. Kim and R. L. Baldwin. Structural intermediates trapped during the folding of ribonuclease A by aide proton exchange. *Biochemistry*, 19:6124–6129, 1980.

[198] W. A. Houry, D. M. Rothwarf, and H. A. Scheraga. A very fast phase in the refolding of disulphide-intact ribonuclease A: implications for the refolding and unfolding pathways. *Biochemistry*, 33:2516–2530, 1994.

[199] A. L. Fink, W. D. Anderson, J. E. Hattersley, and B. S. Lustig. The effect of methanol and temperature on the kinetics of refolding of ribonuclease A. *FEBS Lett.*, 236(1):190–194, 1988.

[200] M. Buck. Trifluoroethanol and colleagues: cosolvents come of age. Recent studies with peptides and proteins. *Quart. Rev. Biophys.*, 31(3):297–353, 1998.

[201] E. R. G. Main and S. E. Jackson. Does trifluoroethanol affect folding pathways and can it be used as a probe of structure in transition states? *Nat. Struct. Biol.*, 6(9):831–835, 1999.

[202] F. Ahmad and C. C. Bigelow. Estimation of the free energy of stabilisation of ribonuclease A, lysozyme, α-lactalbumin and myoglobin. *J. Biol. Chem.*, 257(21):12935–12938, 1982.

[203] P. W. Mui, Y. Konishi, and H. A. Scheraga. Kinetics and mechanisms of the refolding of denatured ribonuclease A. *Biochemistry*, 24:4481–4489, 1985.

[204] S. Ramboarina and C. Redfield. Structural characterisation of the human α-lactalbumin molten globule at high temperature. *J. Mol. Biol.*, 330(5):1177–1188, 2003.

[205] M. Rico, M. Bruix, J. Santoro, C. Gonzalez, J. L. Neira, J. L. Nieto, and J. Herranz. Sequential ^{1}H-NMR assignments and solution structure of bovine pancreatic ribonuclease A. *Eur. J. Biochem.*, 183:623–638, 1989.

[206] D. J. Segel, A. L. Fink, K. O. Hodgson, and S. Doniach. Protein denaturation: a small-angle X-ray scattering study of the ensemble of unfolded states of cytochrome *c*. *Biochemistry*, 37:12443–12451, 1998.

[207] D. Shortle. Composites of local structure propensities: evidence for local encoding of long-range structure. *Protein Sci.*, 11:18–26, 2002.

[208] W. A. Houry and H. A. Scheraga. Nature of the unfolded state of ribonuclease A: effects of cis-trans X–Pro peptide bond isomerization. *Biochemistry*, 35:11719–11733, 1996.

[209] R. W. Dodge and H. A. Scheraga. Folding and unfolding kinetics of the proline-to-alanine mutants of bovine pancreatic ribonuclease A. *Biochemistry*, 35:1548–1559, 1996.

[210] A. Yoshimura and S. Kato. The mechanism of lumiflavin-sensitised photooxygenation of tryptophan accelerated by adenine. *Bull. Chem. Soc. Jpn.*, 46:1141–1144, 1973.

[211] A. Yoshimura and T. Ohno. Lumiflavin-sensitised photooxygenation of indole. *Photochem. Photobiol.*, 48(5):561–565, 1988.

[212] J. M. Christie, T. E. Swartz, R. A. Bogomolni, and W. R. Briggs. Phototropin LOV domains exhibit distinct roles in regulating photoreceptor function. *Plant J.*, 32:205–219, 2002.

[213] C. W. M. Kay, E. Schleicher, A. Kuppig, H. Hofner, W. Rüdiger, M. Schleicher, M. Fischer, A. Bacher, S. Weber, and G. Richter. Blue light perception in plants. *J. Biol. Chem.*, 278(13):10973–10982, 2003.

[214] I. G. Zhulin, B. L. Talyor, and R. Dixon. PAS domain S-boxes in archaea, bacteria and sensors for oxygen and redox. *Trends Biochem. Sci.*, 22:331–333, 1997.

[215] E. Huala, P. W. Oeller, E. Liscum, I-S. Han, E. Larsen, and W. R. Briggs. *Arabidopsis* NPH1: a protein kinase with putative redox-sensing domain. *Science*, 278:2120–2123, 1997.

[216] Y-z. Gu, J. B. Hogenesch, and C. A. Bradfield. The PAS superfamily: sensors of environmental and developmental signals. *Annu. Rev. Pharmacol. Toxicol.*, 40:519–561, 2000.

[217] J. Rutter, C. H. Michnoff, S. M. Harper, K. H. Gardner, and S. L. McKnight. PAS Kinase: an evolutionarily conserved PAS domained-regulated serine/threonine kinase. *Proc. Natl. Acad. Sci. USA*, 98(16):8991–8996, 2001.

[218] J. Vreede, M. A. van der Horst, K. J. Hellingwerf, W. Crielaard, and D. M. F. van Aalten. PAS domains. *J. Biol. Chem.*, 278(20):18434–18439, 2003.

[219] J. M. Christie, P. Reymond, G. K. Powell, P. Bernasconi, A. A. Raibekas, E. Liscum, and W. R. Briggs. *Arabidopsis* NPH1: a flavoprotein with the properties of a photoreceptor for phototropism. *Science*, 282:1698–1701, 1998.

[220] M. Saloman, J. M. Christie, E. Knieb, U. Lempert, and W. R. Briggs. Photochemical and mutational analysis of the FMN-binding domains of plant blue light receptor, phototropin. *Biochemistry*, 39:9401–9410, 2000.

[221] M. Saloman, W. Eisenreich, H. Dürr, E. Schleicher, E. Kneib, V. Massey, W. Rüdiger, F. Müller, A. Bacher, and G. Richer. An optomechanical transducer in the blue light receptor phototropin from *Avena sative*. *Proc. Natl. Acad. Sci. USA*, 98(22):12357–12361, 2001.

[222] M. Kasahara, T. E. Swartz, M. A. Olney, A. Onodera, N. Mochizuki, H. Fukuzawa, E. Asamizu, S. Tabata, H. Kanegae, M. Takano, J. M. Christie, A. Nagatani, and W. R. Briggs. Photochemical properties of the flavin mononucleotide-binding domains of the phototropins from Arabidopsis, Rice and *Chlamydomonas reinhardtii*. *Plant Physiol.*, 129:762–773, 2002.

[223] S. Crosson and K. Moffat. Structure of a flavin-binding plant photoreceptor domain: insights into light-mediated signal transduction. *Proc. Nat. Acad. Sci. USA*, 98(6):2995–3000, 2001.

[224] J. M. Christie, M. Saloman, K. Nozue, M. Wada, and W. R. Briggs. LOV (light, oxygen, or voltage) domains of the blue-light photoreceptor phottropin (nph1): binding sites for the chromophore flavin mononucleotide. *Proc. Natl. Acad. Sci. USA*, 96:8779–8783, 1999.

[225] P. Düx, G. Rubinstenn, G. W. Vuister, R. Boelens, F. A. A. Mulder, K. Hård, W. D. Hoff, A. R. Kroon, W. Crielaard, K. J. Hellingwerf, and R. Kaptein. Solution structure and backbone dynamics of the photoactive yellow protein. *Biochemistry*, 37:12689–12699, 1998.

[226] G. Rubinstenn, G. W. Vuister, F. A. A. Mulder, P. E. Düx, R. Boelens, K. J. Hellingwerf, and R. Kaptein. Structural and dynamic changes of photoactive yellow protein during its photocycle in solution. *Nat. Struct. Biol.*, 5(7):568–570, 1998.

[227] S. Crosson and K. Moffat. Photoexcited state of a plant photoreceptor domain reveals a light-driven molecular switch. *Plant Cell*, 14:1067–1075, 2002.

[228] T. Iwata, S. Tokutomi, and H. Kandori. Photoreaction of the cysteine S-H group in the LOV2 domain of *adiantum* phytochrome3. *J. Am. Chem. Soc.*, 124(40):11840–11841, 2002.

[229] J. M. Walker and R. Rapley. *Molecular biology and biotechnology*. The Royal Society of Chemistry, forth edition, 2000.

[230] C. Walsh, J. Fisher, R. Spencer, D. W. Graham, W. T. Ashton, J. E. Brown, R. D. Brown, and E. F. Rogers. Chemical and enzymatic properties of riboflavin analogues. *Biochemistry*, 17(10):1942–1951, 1978.

[231] S. Kasai, R. Miura, and K. Matsui. Chemical structure and some properties of roseoflavin. *Bull. Chem. Soc. Jpn.*, 48(10):2877–2880, 1975.

[232] K. Matsui and S. Kasai. Roseoflavin, nekoflavin and schizoflavin. In F. Müller, editor, *The Chemistry and Biochemistry of Flavoenzymes*, volume 1, pages 105–120. CRC Press, Boca Raton, 1991.

[233] W. T. Ashton, D. W. Graham, R. D. Brown, and E. F. Rogers. Deaza isosteres of riboflavin and lumichrome. *Tetrahedron Lett.*, 18(30):2551–2554, 1977.

[234] B. J. Fritz, S. Kasai, and K. Matsui. Photochemical properties of flavin derivatives. *Photochem. Photobiol.*, 45(1):113–117, 1987.

[235] J. J. Lopez. *Photo-CIDNP Studies of Amino-acids and Proteins*. DPhil thesis, University of Oxford, 2001.

[236] M. A. G. Carter. *Photo-CIDNP of Amino Acids and Proteins*. Part II thesis, University of Oxford, 2000.

[237] M. Nakagawa, Y. Yokoyama, S. Kato, and T. Hino. Dye-sensitised photooxidation of tryptophan. *Tetrahedron*, 41(11):2125–2132, 1985.

[238] R. Nilsson, P. B. Merkel, and D. R. Kearns. Kinetic properties of the triplet states of methylene blue and other photosensitising dyes. *Photochem. Photobiol.*, 16:109–116, 1972.

[239] S. M. Harper and K. H. Gardner. Personnal communication, 2004.

[240] J. Balbach, N. A. Forge, V. van Nuland, S. L. Winder, P. J. Hore, and C. M. Dobson. Following protein folding in real time using NMR spectroscopy. *Nat. Struct. Biol.*, 2(10):865–870, 1995.

[241] J. Balbach. Compaction during protein folding studies by real-time NMR diffusion experiments. *J. Am. Chem. Soc.*, 122:5887–5888, 2000.

[242] Y. F. Li, P. F. Heelis, and A. Sancar. Active site of DNA photolyase: tryptophan-306 is the intrinsic hydrogen atom donor essential for flavin radical photoreduction and DNA repair in vitro. *Biochemistry*, 30:6322–6329, 1991.

[243] J. E. Huyett, P. E. Doan, R. Gurbiel, A. L. P. Houseman, M. Sivaraja, D. B. Goodin, and B. M. Hoffman. Compound ES of cytochrome *c* peroxidase contrains a Trp π-cation radical: characterisation by CW and pulsed Q-band ENDOR spectroscopy. *J. Am. Chem. Soc.*, 117:9033–9041, 1995.

[244] F. Lendzian, M. Sahlin, F. MacMillan, R. Bittl, R. Fiege, S. Pötsch, B-M. Stöberg, A. Gräslund, W. Lubitz, and G. Lassmann. Electronic structure of the neutral tryptophan radicals in ribonucleotide reductase studied by EPR and ENDOR spectroscopy. *J. Am. Chem. Soc.*, 118:8111–8120, 1996.

[245] S. E. Walden and R. A. Wheeler. Distinguishing features of indolyl radical and radical cation: implications for tryptophan radical studies. *J. Phys. Chem.*, 100:1530–1535, 1996.

[246] S. E. Walden and R. A. Wheeler. Structural and vibrational analysis of indolyl radical and indolyl radical cation from density functional methods. *J. Chem. Soc. Perkin Trans.*, 2:2663–2672, 1996.

[247] P. J. O'Malley and D. A. Ellson. The calculation of ^{1}H, ^{13}C, ^{14}N isotropic and anisotropic hyperfine interactions for the 3-methyl indole cation and neutral radicals using hybrid density functional methods: models for in vivo tryptophan-based radicals. *Chem. Phys. Lett.*, 260:491–498, 1996.

[248] F. Himo and L. A. Eriksson. Theoretical study of model tryptophan radicals and radical cations: comparision with experimental data of DNA photlyase, cytochrome *c* peroxidase and ribonucleotide reductase. *J. Phys. Chem. B.*, 101:9811–9818, 1997.

[249] M. M. Hertel. *A CIDNP study of the tryptophan radical.* Diploma thesis, Free University of Berlin, 2001.

[250] O. B. Morozova, A. V. Yurkovskaya, Yu. P. Tsentalovich, M. D. E. Forbes, and R. Z. Sagdeev. Time-resolved CIDNP study of intramolecular charge transfer in the dipeptide tryptophan-tryrosine. *J. Phys. Chem. B*, 106:1455–1460, 2002.

[251] O. B. Morozova, A. V. Yurkovskaya, Yu. P. Tsentalovich, M. D. E. Forbes, P. J. Hore, and R. Z. Sagdeev. Time-resolved CIDNP study of electron transfer reactions in proteins and model compounds. *Mol. Phys.*, 100(8):1187–1195, 2002.

[252] F. J. Adrian. Radical pair mechanism of chemically induced magnetic polarisation. In L. T. Muus, editor, *Chemically Induced Magnetic Polarisation: proceedings of the NATO Advanced Study Institute held at Sogesta, Urbino, Italy*, volume 5, pages 77–105. D. Reidel Publishing Company, 1977.

[253] K. A. McLauchlan and U. E. Steiner. The spin correlated radical pair as a reaction intermediate. *Mol. Phys.*, 73(2):241–263, 1991.

[254] A. Ehrenberg and L. E. G. Eriksson. Electron spin resonance study on flavin free radicals in non-alkaline media. *Arch. Biochem. Biophys.*, 105:453–454, 1964.

[255] F. Müller, P. Hemmerich, A. Ehrenberg, G. Palmer, and V. Massey. The chemical and electronic structure of the neutral flavin radical as revealed by electron spin resonance spectroscopy of chemically and isotopically substituted derivates. *Eur. J. Biochem.*, 14:185–196, 1970.

[256] M. Tomkiewicz, R. D. McAlpine, and M. Cocivera. Photooxidation and decarboxylation of tyrosine studied by EPR and CIDNP techniques. *Can. J. Chem.*, 50:3849–3856, 1972.

[257] C. E. Lyon, J. J. Lopez, B-M. Cho, and P. J. Hore. Low field CIDNP of amino acids and proteins: characterisation of transient radicals and NMR sensitivity enhancement. *Mol. Phys.*, 100(8):1261–1269, 2002.

[258] A. Ehrenberg, L. E. G. Eriksson, and F. Müller. Electron-spin resonance studies of flavins. In E. C. Slater, editor, *Flavins and Flavoproteins*, volume 8, pages 37–48. Elsevier, Amsterdam, 1966.

[259] A. Ehrenberg, F. Müller, and P. Hemmerich. Basicity, visible spectra and electron spin resonance of flavosemiquinone anions. *Eur. J. Biochem.*, 2:286–293, 1966.

[260] Yu. P. Tsentalovich, O. B. Morozova, A. V. Yurkovskaya, and P. J. Hore. Kinetics and mechanism of the photochemical reaction of 2,2′-dipyridyl with tryptophan in water: time-resolved CIDNP and laser flash photolysis study. *J. Phys. Chem. A*, 103:5362–5368, 1999.

[261] Yu. P. Tsentalovich and O. B. Morozova. Laser flash photolysis and time resolved CIDNP study of the photoreaction of 2,2′-dipyridyl with *N*-acetyl-tyrosine in aqueous solution. *J. Photochem. Photobiol. A*, 131:33–40, 2000.

[262] Yu. P. Tsentalovich, O. B. Morozova, A. V. Yurkovskaya, P. J. Hore, and R. Z. Sagdeev. Time-resolved CIDNP and laser flash photolysis study of the photoreactions of *N*-acetyl-histidine with 2,2′-dipyridyl in aqueous solution. *J. Phys. Chem. A*, 104:6912–6916, 2000.

[263] 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.

[264] W. L. Hubbell, A. Gross, R. Langen, and M. A. Lietzow. Recent advances in site-directed spin labeling of proteins. *Curr. Opin. Struct. Biol.*, 8:649–656, 1998.

[265] W. Jahnke, S. Rüdisser, and M. Zurini. Spin label enhanced NMR screening. *J. Am. Chem. Soc.*, 123:3149–3150, 2001.

[266] Z. Qin, S. L. Wertz, J. Jacob, Y. Savino, and D. S. Cafiso. Defining protein–protein interactions using site-directed spin-labelling: the binding of protein kinase C substrates to calmodulin. *Biochemistry*, 35:13272–13276, 1996.

[267] T. G. Kutateladze, D. G. S. Capelluto, C. G. Ferguson, M. L. Cheever, A. G. Kutateladze, G. D. Prestwich, and M. Overduin. Multivalent mechanism of membrane insertion by the FYVE domain. *J. Biol. Chem.*, 279(4):3050–3057, 2004.

[268] G. R. Eaton, S. S. Eaton, and L. J. Berliner, editors. *Biological Magnetic Resonance*, volume 19. Kluwer Publishing, first edition, 2000.

[269] A. Schweiger and G. Jeschke. *Principles of pulse electon paramagnetic resonance.* Oxford University Press, first edition, 2001.

[270] O. Schiemann, N. Piton, Y. Mu, G. Stock, J. W. Engles, and T. F. Prisner. A PELDOR-based nanometer distance ruler for oligonucleotides. *J. Am. Chem. Soc.*, 126:5722–5729, 2004.

[271] J. E. Banham. Personnal communication, 2003.

[272] L. Serrano. The relationship between sequence and structure in elementary folding units. *Adv. Protein. Chem.*, 53:49–85, 2000.

[273] J. H. Jones. *Amino acid and peptide synthesis.* Oxford Science Publications, second edition, 2002.

[274] K. Wüthrich, M. Billeter, and W. Braun. Polypeptide secondary structure determination by nuclear magnetic resonance observation of short proton-proton distances. *J. Mol. Biol.*, 180:715–740, 1984.

[275] K. Wüthrich. *NMR of proteins and nucleic acids.* Wiley-Interscience, first edition, 1986.

[276] D. S. Wishart, B. D. Sykes, and F. M. Richards. The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry*, 31:1647–1651, 1992.

[277] D. S. Wishart, B. D. Sykes, and F. M. Richards. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J. Mol. Biol.*, 222:311–333, 1991.

[278] D. S. Wishart and B. D. Sykes. Chemical shifts as a tool for structure determination. In T. L. James and N. J. Oppenheimer, editors, *Methods in Enzymology*, volume 239, pages 363–392. Academic Press, San Diego, 1994.

[279] G. Merutka, J. H. Dyson, and P. E. Wright. Random coil 1H chemical shifts obtained as a function of temperature and trifluoroethanol concentration for the peptide series GGXGG. *J. Biomol. NMR*, 5:14–24, 1995.

[280] D. I. Freedberg, R. M. Venable, A. Rossi, T. E. Bull, and R. W. Pastor. Discriminating the helical forms of peptides by NMR and molecular dynamics simulation. *J. Am. Chem. Soc.*, 126(33):10478–10484, 2004.

[281] M. Karplus. Contact electron-spin coupling of nuclear magnetic moments. *J. Chem. Phys*, 30:11–15, 1959.

[282] C. W. M. Kay, R. Feicht, K. Schulz, P. Sadewater, A. Sancar, A. Bacher, K. Möbius, G. Richter, and S. Weber. EPR, ENDOR and TRIPLE resonance spectroscopy on the neutral flavin radical in *Escherichia coli* DNA photolyase. *Biochemistry*, 38:16740–16748, 1999.

[283] S. Weber, K. Möbius, G. Richter, and C. W. M. Kay. The electronic structure of the flavin cofactor in DNA photolyase. *J. Am. Chem. Soc.*, 123:3790–3798, 2001.

[284] F. Cintolesi, T. Ritz, C. W. M. Kay, C. R. Timmel, and P. J. Hore. Anisotropic recombination of an immobilized photoinduced radical pair in a 50 μT magnetic field: a model avian photomagnetoreceptor. *Chem. Phys.*, 294:385–399, 2003.