The handle http://hdl.handle.net/1887/22714 holds various files of this Leiden University dissertation.

**Author:** Skinner, Simon Peter  
**Title:** Classical and paramagnetic NMR spectroscopy techniques applied to different protein systems  
**Issue Date:** 2013-12-05
Chapter 1

Introduction
The importance of NMR in protein structural biology

Proteins undergo wide range of motions in terms of distances and time scales, ranging from bond vibrations on the ps timescale to conformational rearrangements on the µs to ms timescales. The distances covered by these motions can range from $10^{-2}$ to > 10 Å. NMR spectroscopy is one of the only biophysical methods that can provide high-resolution structural and dynamic information about proteins, the other methods being X-ray crystallography, and, more recently, electron microscopy (EM). X-ray crystallography and EM can be used to obtain static structures of proteins and X-ray crystallography can yield dynamic information via the interpretation of crystallographic B-factors, which are sensitive to the mean square displacement of atoms due to their thermal motions ($T$). However, the timescales of these motions cannot easily be obtained from these data. Investigating protein dynamics is of particular interest for understanding how they impact on protein function, for example, conformational exchange processes leading to ligand binding, allosteric processes, enzyme catalysis and protein folding. A major challenge in NMR spectroscopy is obtaining data of sufficient resolution such that individual resonances can be unequivocally identified in the spectra. This is especially difficult with large macromolecules. This is not only attributable to the high number of resonances present in the spectra of these molecules. The greater the size of a molecule, the more slowly it tumbles in solution, and, therefore, the transverse relaxation rate of the protein is high, resulting in broad resonances. Advances in NMR spectroscopy over the last three decades have provided means to overcome these issues. Multidimensional heteronuclear NMR spectroscopy has enabled circumvention of the problems of spectral crowding and the introduction of TROSY and CRINEPT experiments made the acquisition of spectra with sharper lines possible. These developments in NMR spectroscopy have facilitated the study of very large macromolecular systems (>100 kDa). Moreover, the development of paramagnetic NMR techniques over the past decade has offered new avenues into the study of protein dynamics.
NMR assignment of proteins

A prerequisite for any detailed NMR study is that NMR assignments must be obtained for the nuclei of the protein in question. Depending on the size of the protein, different assignment strategies are used and, as a consequence, different isotope labelling strategies are employed to facilitate the assignment procedures. For very small proteins (<10 kDa), isotopic labelling of the amide nitrogens using $^{15}$N can be sufficient to obtain a backbone assignment for proteins using so-called “Wüthrich Walks (2)”. This assignment strategy involves acquiring data from $J$-correlation spectroscopy experiments, such as $^1$H-$^1$H COntent Spectroscopy (COSY) and $^1$H-$^1$H Total COntent Spectroscopy (TOCSY) as well as dipolar coupling based correlation in $^1$H-$^1$H Nuclear Overhauser Effect Spectroscopy (NOESY). Greater spectral resolution is afforded by linking TOCSY and NOESY spectra to a $^{15}$N-$^1$H HSQC spectrum (3-6). These experiments enable the classification of spin systems based on spins that can be correlated in these spectra. A spin system is a set of proton nuclei, which are linked by $^3J$ coupling and each amino acid is separated from the next by a carbonyl group, such that each amino acid has one or more spin systems. Examples of amino acids with more than one spin system include: Phe and Tyr. Phe has two spin systems wherein one spin system contains the protons attached to the aromatic ring and another spin system contains the $H^a$, $H^\beta$ and $H^N$ protons, and Tyr has three spin systems, where the backbone and aliphatic protons form one spin system, the aromatic protons form a second spin system and the phenolic proton forms the third spin system (Figure 1.1). The reason for that these are separate spin systems is that there is no $^3J$ coupling between the $H^\beta$ and the aromatic protons. Each spin system has its own TOCSY/COSY pattern, which enables identification the amino acid type. Phe, His and Tyr aromatic $^1$H resonances cannot be observed in HSQC-TOCSY or HSQC-NOESY spectra, only in $^1$H-$^1$H TOCSY and NOESY spectra or $^{13}$C-separated spectra.
Once the amino acids types been classified NOESY experiments (7, 8) can then be used to determine which spin systems are sequentially linked. NOEs can be observed between nuclei up to ~5 Å from each other, and, therefore, amide H\textsuperscript{N} sequential links can be established. Intra-residue NOEs are observable for H\textsuperscript{N}, H\textsuperscript{α} and H\textsuperscript{β} protons and inter-residue NOEs are observable between adjacent H\textsuperscript{N} protons, adjacent H\textsuperscript{N} and H\textsuperscript{α} protons and between adjacent H\textsuperscript{N} and side chain protons (Figure 1.2).

The combination of TOCSY and NOESY spectra permits the identification of a residue, \(i\), and the residue preceding it in the sequence, residue \(i-1\) or the residue succeeding it, residue \(i+1\). In some cases, non-sequential short-range (\(i\) to \(i+2\), \(i\) to \(i+3\), \(i\) to \(i+4\)) and long-range NOEs can be detected and used in NMR assignment and structure determination.
Figure 1.2 Sequential $^1$H NMR assignment of a valine–alanine (V–A) peptide segment in a polypeptide chain. The dotted lines connect groups of hydrogen atoms that are $^3J$ coupled. The solid arrows link pairs of hydrogen atoms in neighbouring amino acid residues that are separated by short through-space distances, $d_{\alpha N}$, $d_{\beta N}$, $d_{\gamma \alpha}$ and $d_{NN}$, and, therefore, neighbouring spin systems can be connected by these “sequential NOEs”.

For larger proteins, the overlap of resonances in TOCSY and NOESY spectra make it impossible to achieve complete assignment. In order to assign large proteins, a triple resonance approach is normally employed, which involves $^{13}$C-$^{15}$N labelling of proteins and the use of heteronuclear triple resonance experiments, which correlate the backbone amide, $C\alpha$ nuclei, side chain and backbone carbonyl nuclei. These types of experiments exploit the $^1J$ and $^2J$ couplings that occur between adjacent NMR active nuclei (Figure 1.3) and usually transfer coherence from the amide protons to a $^{13}$C carbon in the backbone or side chain of an amino acid. One of the main advantages of triple resonance experiments is the greater peak dispersion afforded by the third dimension in the spectrum compared to two-dimensional spectra. In addition, the $C\alpha$ and $C\beta$ nuclei have characteristic chemical shifts and, therefore, individual residue types can be easily identified.
Figure 1.3 $^1J$ (solid lines) and $^2J$ (dashed lines) coupling constants between protein nuclei used for magnetisation transfer in $^{13}C$, $^{15}N$ labelled proteins. (Figure adapted from (11))

These types of experiments are often named after the correlations that are involved in acquiring the spectrum, for example, an HNCACB (9) detects correlations between the amide proton and its attached nitrogen, along with correlations between the amide nitrogen and the Cα and between the backbone Cα and the side chain Cβ of both the residue in question via $^1J$ coupling and the residue which precedes it sequentially via the $^2J$ $^{15}N$-$C'$$-^{13}Cα$ coupling (Figure 1.4a). An HNcaCO (10) detects correlations between the amide proton and its attached nitrogen and the intraresidue carbonyl carbon via the $^1J$ $^{15}N$$-^{13}Cα$ coupling and the $^1J$ $^{13}Cα$$-^{13}C'$ coupling, along with correlations between the amide nitrogen and the carbonyl carbon of the previous residue via the $^2J$ $^{15}N$$-C'$$-^{13}Cα$ coupling and the $^1J$ $^{13}Cα$$-^{13}C'$ coupling (Figure 1.4b) (lower case letters in experiment names indicate nuclei that are used for coherence transfer, but not detected).
Figure 1.4 Magnetisation transfer pathways in the HNCACB (a) and HNcaCO (b) experiments.

In HNCACB spectra, the peaks of the inter-residue resonances can be of considerably lower intensities than the peaks of intra-residue resonances, by optimising the coherence transfer times for the $^1J_{^{15}N-^{13}C\alpha}$ coupling (Figure 1.3) and the delay for N-Cα scalar coupling evolution is set to $1/4J_{N\alpha}$. The same is true for the HNcaCO experiment, the peaks of inter-residue C’ resonances usually have much lower intensities than those of the intra-residue C’ resonances. This is because the intra-residue $J_{N\alpha}$ and $J_{C\alpha CO}$ couplings are exploited to obtain intra-residue correlations, whereas the much weaker $^2J_{N\alpha}$ is exploited to obtain the inter-residue correlations and the N-Cα scalar coupling evolution period is set to $1/4^1J_{N\alpha}$.

The general approach used in triple-resonance assignment (for review see (11)) is the use of a pair of experiments that complement each other, i.e. an HNCACB and an HNcoCACB (12) along with an HNcaCO and an HNCO (13). In the HNcoCACB experiment, only inter residue $(i-1)$ Cα and Cβ coherences are limited to the amide of residue, $i$, and the peak intensities are greater than those of the inter-residue resonances observed in an HNCACB because the weak $^2J$ coupling is avoided. In the HNCO experiment, only the inter-residue $(i-1)$ C’ coherences are observed for amide $i$. 

17
In spite of the increased spectral resolution allowed by using triple resonance experiments, proteins larger than 20 kDa present more of a challenge due to spectral crowding caused increased line broadening resulting from slower tumbling of large proteins. Furthermore, the triple resonance experiments have long preparation periods, leading to large coherence losses for large proteins. There are two strategies, often employed together, to reduce both spectral crowding and line broadening effects, namely Transverse Relaxation Optimised SpectroscopY (TROSY) (14) coupled with selective deuteration or perdeuteration of all non-exchangeable protons. Deuteration of side chain protons is achieved using deuterated amino acid precursors during protein expression and/or growing in media containing up to 100% $^2$H$_2$O. Deuteration reduces the dipole-dipole relaxation contributions that cause the large $^1$H line broadening observed in spectra of large proteins. This leads to longer-lived coherences and thus more sensitivity in triple resonance experiments as well as sharper resonances, reducing spectral crowding.

In order to discuss TROSY experiments, an understanding of the relaxation mechanisms that occur within proteins is necessary. TROSY experiments are based on constructive use of interference between dipole-dipole (DD) and chemical shift anisotropy (CSA) contributions to relaxation. Each spin $\frac{1}{2}$ nucleus has a magnetic moment associated with it, which in turn gives rise to a local magnetic field and this local field can interact with other spins. Therefore, two spins are required for DD interactions, one, which “creates” the field, and another, which experiences it and the roles of the two spins are reversible. The size of a DD interaction depends on the inverse cube of the distance between the two interacting nuclei and the direction of the vector joining the two nuclei, measured relative to the direction of the static magnetic field.

As a molecule tumbles in solution the directions of the internuclear vectors change, which causes to changes in the local magnetic fields experienced by each spin. These changes in orientation caused by tumbling of the molecules
are responsible for relaxation resulting from DD interactions. In essence, DD interactions turn molecular motion into a fluctuating magnetic field, which causes transitions of spins from one energy level to another. Transverse relaxation due to DD interactions can be described by Equation 1.1, here $J(\omega)$ is the spectral density function (15):

$$R_{2}^{\text{dp}} = d^2 \left( \frac{1}{10} J(0) + \frac{3}{20} J(\omega_{0,1}) + \frac{3}{20} J(\omega_{0,1} - \omega_{0,2}) + \frac{3}{40} J(\omega_{0,1}) + \frac{3}{20} J(\omega_{0,1} + \omega_{0,2}) \right)$$

(1.1)

where

$$d = \frac{\mu_0 \hbar y_1 y_2}{4\pi r_{NN}^3}$$

The chemical shift of a nucleus is a function of the orientation of the molecule in the magnetic field, provided the nucleus is not at the centre of tetrahedral or octahedral symmetry. As the molecule tumbles in solution, the chemical shift, and hence the magnetic field produced by the electrons surrounding the nucleus, is constantly changing; and this can cause relaxation of the nucleus. The relaxation rate due to CSA is proportional to the square of the gyromagnetic ratio and of the magnetic field strength, as well to the strength of the chemical shift anisotropy. Transverse relaxation due to CSA can be described by Equation 1.2 for an axially symmetric shielding tensor (15):

$$R_{2}^{\text{CSA}} = c^2 \left( \frac{2}{45} J(0) + \frac{1}{30} J(\omega_0) \right)$$

(1.2)

where

$$c = \gamma B_0 (\sigma_1 - \sigma_2)$$

In systems containing multiple spins, different relaxation mechanisms can interfere with each other, which is known as cross-correlated relaxation.
Cross-correlated relaxation effects cause the two signals in a doublet to have different linewidths, due to the fact that NMR linewidths are proportional to the transverse relaxation rate of the nucleus in question (Figure 1.5b) (Equations 3 and 4) (15).

\[
R_{xy}^{(1)} = \frac{1}{10} d^2j(0) + \frac{2}{45} c^2 j(0) + \frac{2}{15} cd (3 \cos^2 \theta - 1) j(0) \tag{1.3}
\]

\[
R_{xy}^{(2)} = \frac{1}{10} d^2j(0) + \frac{2}{45} c^2 j(0) - \frac{2}{15} cd (3 \cos^2 \theta - 1) j(0) \tag{1.4}
\]

TROSY experiments make use of cross-correlation interference between CSA and DD interactions to select the magnetisation component with the narrowest linewidth and, thereby, enable the acquisition of spectra with sharper signals, especially for large proteins, compared to conventional non-TROSY experiments. For a coupled spin-1/2 system, I-S, in a protein, the $T_2$ relaxation rates are dominated by the DD interaction between I and S, as well as the CSA of each individual spin. As shown in equations 1.3 and 1.4, the cross-term can be positive or negative. If the cross-term is negative, the rate of relaxation is reduced and therefore, a sharp line is observed for the doublets of spins I and S. Due to the fact that CSA is proportional to the static field strength, the CSA and DD interactions are almost cancelled for an amide spin pair, when the magnetic field strength is 23.5 T ($^{1}$H larmor frequency of 1.0 GHz) (14, 16). The $^{15}$N-$^{1}$H TROSY-HSQC selects only the $^{12}$S and $^{13}$I transitions through a combination of phase cycling and echo-antiecho quadrature detection.

An issue commonly encountered with large triple labelled proteins, is their instability in solution. A full set of triple resonance experiments can take weeks to record and the protein may not be sufficiently stable to permit this. The use of Band Selective Excitation Short Transient (BEST) (17, 18) experiments has sought to overcome this difficulty. In these experiments, shaped pulses are centred on the amide proton spectral region and selective $^{1}$H manipulation using these pulses ensures minimal perturbation of the
proton spins resonating outside the excited spectral window (19-22).

![Energy level diagram]

\[
\begin{align*}
S^{34} &= \omega_5 + \pi J_{15} \\
S^{12} &= \omega_5 - \pi J_{15}
\end{align*}
\]

Figure 1.5 (a) Energy level diagram of a two spin $\frac{1}{2}$ system and showing the identity of the components of the 2D multiplet expressed as single transition basis operators, I$^{13}$, I$^{24}$, S$^{12}$, S$^{34}$. (b) Schematic diagram of an HSQC spectrum showing the four peaks observed in an HSQC spectrum acquired without decoupling and the neutral peak observed in a decoupled HSQC.

The band selectivity of the $^1$H shaped pulses used in BEST experiments ensures that unobserved protons remain in their equilibrium state and therefore provide a thermal bath of proton spin polarisation, which thereby enables faster $T_1$ relaxation of the excited $^1$H spins via dipolar interactions between the detected amide and the unperturbed aliphatic protons between consecutive transients; this yields an increased sensitivity at high repetition rates (21). Such experiments enable the acquisition of multidimensional NMR experiments in as little as 20% of the time required for a traditional experiment. NMR studies of large proteins often require that the protein samples are deuterated. However, 100% deuteration of samples would eliminate the efficacy of BEST experiments. A significant sensitivity gain has been reported even for a 75% deuterated sample of an 18 kDa protein (18). The combination of BEST and TROSY techniques allows the fast acquisition of high-resolution NMR spectra for the assignment of large proteins with partial deuteration.
Obtaining a backbone assignment for a protein is often sufficient to characterise backbone dynamics, protein-protein and protein-ligand complexes, if the structure of the protein is known beforehand. However, if a structure is not available, NMR spectroscopy can be used to solve the structures of proteins up to 82 kDa in size (23). In order to perform structure calculations and obtain solution structures of proteins by NMR, an assignment of the side chains is required. Side chain assignment involves the use of TOCSY and COSY based experiments to correlate the backbone and Cβ assignments from triple resonance spectra or HSQC-TOCSY experiments, to spectra of side chain atoms. Herefrom, structure calculation software packages, such as CYANA (24) can be used to automatically assign NOESY spectra. The intensity of peaks in NOESY spectra are proportional to the sixth root of the distance between the two nuclei from which the NOE derives. As a result, NOE crosspeaks can be used to derive ambiguous distance restraints for protein structure calculation. Complete assignment of NOESY spectra is a time-consuming process and nowadays, software programs such as CYANA are used to perform NOESY assignment, on the basis of chemical shift data. CYANA performs initial assignment of the NOESY spectra and uses the principal of network anchoring (25) to rank assignments from which initial ensemble of structures can be generated. Network anchoring utilises the fact that correctly assigned constraints form a self-consistent subset in any network of distance constraints that is sufficiently dense for the determination of a 3D protein structure. The initial assignment and network anchoring scores are used generate an initial ensemble for structure calculation. The structure of the protein is obtained by iterative cycles of NOE assignment and refinement of the initial ensemble to produce a final ensemble with a low RMSD value corresponding to the final structure. The process of protein structure solution is discussed in more detail in Chapter 4.
Paramagnetic effects observable by NMR

A paramagnetic centre is an abstract concept, which assumes that one or more unpaired electrons are effectively located at a single point; this is the so-called “point-dipole approximation.” The presence of paramagnetic centres in proteins can cause changes in the positions and/or intensities of the observed resonances in NMR spectra, compared to a spectrum of the same protein devoid of a paramagnetic centre. In paramagnetic systems, there are many different experimentally observable NMR effects, however only three will be discussed here, namely, pseudocontact shifts (PCSs), residual dipolar couplings (RDCs) and paramagnetic relaxation enhancements (PREs). PREs result from an isotropic or an anisotropic electron $g$-factor and, therefore, can be observed in any paramagnetic system. PCSs and RDCs can only be observed in systems with an anisotropic electron $g$-factor and depend on a second rank magnetic susceptibility tensor, often referred to as the $\chi$ tensor.

If is assumed that a molecule tumbles isotropically, the magnitude of the PCS, $\delta_{\text{PCS}}$, is given by (26):

$$\delta_{\text{PCS}} = \frac{1}{12\pi \rho t^2} \left[ \Delta \chi_{ax} \left( 3(r_x^2 r_i^2) - r_i^4 \right) + \frac{3}{2} \Delta \chi_{rh} \left( r_x^2 r_i^2 - (r_y^2 r_i^2) \right) \right] \tag{1.5}$$

where $\Delta \chi_{ax}$ and $\Delta \chi_{rh}$ represent the axial and rhombic components of the second rank magnetic susceptibility tensor, $r_i = \sqrt{x_i^2 + y_i^2 + z_i^2}$ (the coordinates of nucleus, $i$, in the tensor frame and the paramagnetic centre at the origin) and $r_x^2$, $r_y^2$ and $r_z^2$, represent the unit vectors that determine the orientation of the magnetic susceptibility tensor. Equation 5 can be fitted using a five-parameter fit: $f(\Delta \chi_{ax}, \Delta \chi_{rh}, \alpha, \beta, \gamma)$, if the position of the paramagnetic centre is known, or an eight parameter fit: $f(\Delta \chi_{ax}, \Delta \chi_{rh}, \alpha, \beta, \gamma, x, y, z)$, where $\alpha, \beta$ and $\gamma$ are three Euler angles that determine the orientation of the magnetic susceptibility tensor in the protein frame and $x, y$ and $z$ are the Cartesian coordinates of the paramagnetic centre in the molecular frame.
PCS display an $r^{-3}$ distance dependence, which means that the distance range for an experimental detection of PCS is relatively long, for example, distances observable using Tm$^{3+}$ can extend up to ~60 Å (27). If the paramagnetic centre causing the PCSs is static, large PCSs can be observed; however, motion of a paramagnetic centre, relative to the protein, leads to a reduction in the observable PCSs.

A protein containing a paramagnetic centre with an anisotropic $g$-tensor will undergo partial alignment in the magnetic field, and this will give rise to RDCs. If the anisotropic magnetic susceptibility tensor and the $^1$H-heteroatom bond (A-B) are assumed to be static, RDCs are given by (26):

$$RDC = \frac{8}{15\mu_B^3} \frac{\gamma A \gamma R H}{16\pi^3\mu_B^3} \left[ \Delta \chi_{ax} (3(\vec{r}_A \cdot \vec{r}_H)^2 - \gamma_i^2) + \frac{3}{2} \Delta \chi_{rh} (\vec{r}_A \cdot \vec{r}_H)^2 - (\vec{r}_A \cdot \vec{r}_H)^2 \right] \enspace (1.6)$$

As with PCS, a downscaling of the RDCs is observed if the paramagnet giving rise to the partial alignment is not fixed, relative to the protein. This can be accounted for by inclusion of the order parameter $S^2$ in the calculation of the RDC pre-factor (28).

Two mechanisms contribute to PREs, namely the Solomon mechanism and the Curie spin mechanism. The Solomon mechanism is a dipolar mechanism, which takes into account relaxation driven by the lifetime of the electronic spin states. Solomon relaxation is predominant for slowly tumbling molecules with long lifetimes of the electronic spin state, as in the case of Gd$^{3+}$, Mn$^{2+}$ or nitroxide radicals. The transverse ($\Gamma_{\perp}$) PRE rate due to Solomon relaxation is described by the Solomon-Bloembergen (SB) equation (29, 30):

$$\Gamma_{\perp,SB} = \frac{i}{15} \left( \frac{\mu_0}{4\pi} \right)^2 \frac{\gamma_A^2 \gamma_R^2 \mu_B^2 (5+1)}{r^6} \left( 4\tau_c + \frac{3\tau_c}{1+(\omega_0^{1/2}\tau_c)^2} \right) \enspace (1.7)$$

Where $\tau_c$ is correlation time defined as $(\tau^{-1}_r + \tau^{-1}_s)^{-1}$, $\tau_r$ is the rotational correlation time of the macromolecule and $\tau_s$ is the effective electron relaxation time of the paramagnetic centre. The Solomon-Bloembergen theory makes the assumption that electron relaxation is not coupled to
molecular tumbling, which is reasonable since the electron relaxation time is comparable or shorter than the rotational correlation time of a macromolecule. Moreover, the dipole-dipole interaction vectors are assumed to be rigid in the molecular frame. If the paramagnetic centre is mobile, there are two ways in which this can be handled, either using multiple conformations of the paramagnetic centre when calculating PREs (31, 32) or by inclusion of the model free formalism in the Solomon-Bloembergen equation. For details see (33):

\[
\Gamma_{1,SBMF} = \frac{1}{15} \left( \frac{\mu_0}{4\pi} \right)^2 \frac{\gamma_e^2 \gamma_p^2 \mu_B^2 S(S+1)}{r^6} \left( 4(S^2 \tau_e + (1-S^2) \tau_e) + \frac{S^2 \tau_e}{1+(\omega/\tau_p)^2} + \frac{(1-S^2) \tau_e}{1+(\omega/\tau_p)^2} \right) \tag{1.8}
\]

Where \(\tau_e\) is the total correlation time, defined as \(\tau_e^{-1} + \tau_s^{-1} + \tau_i^{-1}\) and \(\tau_i\) is the correlation time for internal motion of nucleus, \(i\). Curie spin relaxation is the dominant component of the \(^1H\) transverse \((\Gamma_2)\) PRE rates for metal ions with anisotropic \(g\)-tensors and very short electron relaxation times \(\tau_e \ll \tau_s\), such as Fe\(^{3+}\), Tm\(^{3+}\), Yb\(^{3+}\) and Dy\(^{3+}\). The \(\Gamma_2\) rate due Curie-spin relaxation is given by (34):

\[
\Gamma_{2,\text{Curie-spin}} = \frac{1}{5} \left( \frac{\mu_0}{4\pi} \right)^2 \frac{\gamma_e^2 \gamma_p^2 \mu_B^2 S(S+1)^2}{(3k_B T)^2 r^6} \left( 4\tau_e + \frac{3\tau_p}{1+(\omega/\tau_p)^2} \right) \tag{1.9}
\]

Where \(\tau_p\) is the rotational correlation time of the macromolecule.

**Paramagnetic NMR to the study proteins and protein complexes**

The utility of paramagnetism in the field of protein NMR was first recognized in the 1960s. For example, \(^1H\) NMR studies of cytochrome \(c\) revealed different sets of resonances in the reduced and oxidised forms (35). Subsequently, this observation was used to obtain structural information pertaining to the origin of these effects in the same protein system (36). \(^1H\) NMR studies (37-39) and
$^{13}$C NMR studies (39) of ferredoxins and $^1$H NMR studies of rubredoxins (40) provided further evidence for the utility of paramagnetism in NMR studies of proteins. Over the last two decades, the use of paramagnetic NMR for the structural, functional and dynamic studies of proteins and protein complexes has risen exponentially. The first application of this technique to solve the structure of a protein-protein complex was the solution of the cytochrome $f$ and plastocyanin (41), which used the paramagnetic effects derived from a native low-spin haem iron in cytochrome $f$. Subsequently, this technique has been developed and not only applied to the solution of the structures of other protein-protein complexes, but also to investigate the transient nature of these complexes. Some metalloproteins contain a paramagnetic centre that can be used for NMR studies, such as Fe$^{3+}$ in ferredoxins (42, 43). In many cases, the paramagnetic centre is not present in the wild type protein and therefore, it is incorporated into the protein by one of two principal methods, either into a natural metal binding site or via an attached tag containing a paramagnetic centre. For metalloproteins the substitution of the metal ion with a paramagnetic metal is a classical approach, where the sidechains of Asp, Glu, His, Gln, Ser, Thr, Asn, His, Met and the backbone carbonyl groups typically coordinate the metal ions, for example the incorporation lanthanide ions (Dy$^{3+}$, Tb$^{3+}$ and Er$^{3+}$) into the metal binding site of the N-terminal domain of the ε subunit bound to the θ subunit of the E. coli DNA polymerase I (44). For non-metalloproteins, paramagnetic centres can be incorporated using small peptides or paramagnetic tags that bind paramagnetic metal ions. Metal-binding peptides can be attached to either the N or C-terminus of proteins (45-50), often inducing only small PCSs due to flexibility, or via a thiol reactive cysteine (48, 49). Recently, lanthanide binding peptides have also been incorporated into loop regions of proteins (51). Many paramagnetic tags attach via a single-cysteine residue engineered into the protein structure (52-60). Attachment via a single cysteine can lead to weak paramagnetic effects due to flexibility at the attachment site and the presence of stereoisomers of the probes attached to the protein leading to different tensors being observed for the same protein species, which can lead to double peaks. Attachment via two cysteine residues engineered to the surface (27, 61) can remove the
problem of probe flexibility and thus lead to the observation of stronger paramagnetic effects and therefore effects at longer distances. Such probes also have well-defined, predictable locations on the protein surface, which is required to obtain reliable paramagnetic effects.

Application of paramagnetic NMR for the assignment of proteins

There are many software packages available for the determination of the $\chi$ tensor from PCS data given the 3D structure of the protein: Fantasia (62), Fantasian (63), the PARArestraints module for Xplor-NIH (64, 65) and Numbat (28). These programs either perform a five-parameter $\chi$ tensor fit, assuming a known position of the paramagnetic centre, or do a complete eight-parameter fit (see above). In all cases, the diamagnetic and paramagnetic NMR assignments are required to perform the fitting. Software packages to enable assignment of proteins through the use of PCS data have been developed: Platypus (66), Echidna (67), and Possum (68). Platypus provides backbone amide assignment from chemical shift data of the diamagnetic and paramagnetic spectra using a known probe position, while simultaneously fitting the five tensor parameters. Echidna provides backbone amide assignment of a paramagnetic spectrum based on the diamagnetic assignment of the chemical shifts by fitting the magnitudes and Euler angles of the $\chi$ tensor, where the metal position is known. Possum is a method developed to automatically assign methyl groups of a protein using diamagnetic and paramagnetic chemical shifts and a known paramagnetic tensor. In this thesis, a new method to obtain the assignment of protein nuclei on the basis of PCS from multiple paramagnetic centres is presented. Each paramagnetic data set consists of a set of PCSs for each peak in the diamagnetic spectra of the protein in question and, therefore, these values, when combined, provide sufficient data for a novel and efficient methodology for simultaneous $\chi$ tensor refinement and protein assignment (See chapter 6).
Using NMR to study dynamic processes in proteins

Prior to the introduction of isotope labelling strategies in the late 1980s, one-dimensional $^{13}$C experiments were developed to study dynamics (69-74). However, the low natural abundance of non-proton NMR active nuclei, $^{13}$C and $^{15}$N, required labelling strategies which enabled uniform incorporation of $^{15}$N and/or $^{13}$C, coupled with multipulse, multidimensional NMR experiments. A variety of heteronuclear experiments for the measurement of different relaxation parameters on different timescales have been developed over the past three decades. The majority of these experiments have used $^{15}$N, $^{13}$C, $^1$H and $^2$H nuclei as probes for protein dynamics, although a few studies have utilised $^{19}$F (75) and $^{31}$P (76). Motions on the ps-ns timescale can be investigated with classical NMR relaxation experiments to measure spin-spin ($T_2$) relaxation, spin-lattice ($T_1$) relaxation and heteronuclear NOEs for each residue in the protein (77). The combination of these three data values can then be used for Model-Free analysis (78, 79) to determine the degree and rate of internal motion. The majority of these studies have focused on the N-H bond vectors of the backbone or selected side chain residues, in spite of the fact that a similar methodology can be used with $^{13}$C and $^2$H to obtain a more complete understanding of these fast timescale motions. For example, analysing both the $^{15}$N-1H and $^{13}$C$^\text{\alpha}$-$^{13}$C$^\text{\text{\text{\text{\text{O}}}}}$ bond vectors (80, 81) can yield a more comprehensive dynamic profile of backbone dynamics on the ps-ns timescale (82-84). Since these vectors point in different directions, they can sense different reorientational motions of the peptide plane. In some cases, an estimate of the exchange rate on the μs to ms timescale, $R_{ex}$, can be extracted from these analyses; however, the determination of the actual values of $R_{ex}$ requires different classes of experiments, namely, relaxation dispersion and EXSY experiments.

Relaxation dispersion and EXSY experiments are used to determine the rate constant for exchange, $k_{ex}$, between two or more states, the major and minor states. Depending on whether resonances can be observed for the major and minor states and the values of $k_{ex}$, different experiments are required.
If the peaks of the minor state are not visible in the acquired spectra, relaxation dispersion experiments are required to study the exchange process(es). There are two main classes of relaxation dispersion experiments, which can be employed depending on the values of $k_{\text{ex}}$, namely $R_{1\rho}$ and Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion experiments.

$R_{1\rho}$ experiments can be used when $k_{\text{ex}}$ is between 5000 and 50,000 s\(^{-1}\), such processes are said to occur in the intermediate to fast-exchange regime, wherein signals are only slightly broadened due to exchange ($k_{\text{ex}} \geq \Delta \omega$). In this experiment the magnetisation is spin-locked in the rotating frame using an RF field (85). The relaxation rate constant for the component of the magnetization along the direction of the effective field in the rotating frame is called $R_{1\rho}$ and depends on the amplitude of the applied RF field, $\omega_1$, the population average chemical shift, $\Omega = p_A \Omega_A + p_B \Omega_B$ and the effective field in the rotating frame, $\omega_{\text{eff}}$ is given by:

$$\omega_{\text{eff}} = \sqrt{(\Omega^2 + \omega_1^2)} (1.10)$$

Conventionally, $\omega_{\text{eff}}$ is assumed to be the same for all sites, which requires that the resonance offset of the kth resonance, $\Omega_k = \omega_k - \omega_{\text{rf}} > \Delta \omega$ for all sites, or that $\omega_1 > \Delta \omega$ (86). In the rotating frame, the tilt angle of the effective field is given by $\tan \theta = \omega_1 / \Omega$. In the fast exchange limit, ($k_{\text{ex}} / \Delta \omega \to \infty$), $R_{1\rho}$ is given by (87, 88):

$$R_{1\rho} = R_{1\rho}(\omega_{\text{eff}} \to \infty) + \sin^2 \theta \frac{p_A p_B \Delta \omega^2 k_{\text{ex}}}{k_{\text{ex}}^2 + \omega_{\text{eff}}^2} (1.11)$$

An approximate expression exists, which is valid for all timescales, provided that an on-resonance RF field is applied, $\omega_{\text{eff}} = \omega_1$ and $p_A >> p_B$:

$$R_{1\rho} = R_{1\rho}(\omega_{\text{eff}} \to \infty) + \sin^2 \theta \frac{p_A p_B \Delta \omega^2 k_{\text{ex}}}{k_{\text{ex}}^2 + p_A^2 \Delta \omega^2 + \omega_1^2} (1.12)$$

In the slow exchange regime, $R_{1\rho}(\omega_{\text{eff}} \to \infty)$ is given by:
\[ R_{1p}(\omega_{\text{eff}} \to \infty) = \cos^2 \theta R_{1A}^0 + \sin^2 \theta R_{2A}^0 \quad (1.13) \]

And in the fast exchange regime by:

\[ R_{1p}(\omega_{\text{eff}} \to \infty) = \cos^2 \theta (p_A R_{1A}^0 + p_B R_{1B}^0) + \sin^2 \theta (p_A R_{2A}^0 + p_B R_{2B}^0) \quad (1.14) \]

When a strong RF field is applied on-resonance, increasing \( \omega_{\text{eff}} \) is the same as increasing \( \omega_1 \), which quenches \( R_{\text{ex}} \). However, \( \omega_{\text{eff}} \) can also be increased by fixing \( \omega_1 \) and moving the carrier frequency off resonance to quench \( R_{\text{ex}} \); as \( \Omega \) increases, \( \theta \) goes to zero, hence the magnetization spends more time along the z axis, therefore, according to equation 1.10, as \( \omega_{\text{eff}} \) increases, exchange is moved into the fast exchange regime, which quenches \( R_{\text{ex}} \). From these conclusions, it follows that there are two main types of \( R_{1p} \) experiments, on-resonance experiments \((89)\) and off-resonance experiments \((90)\). In on-resonance experiments, the RF transmitter frequency is positioned close to the resonances of interest and the minimum value of \( \omega_1 \) is large enough such that \( \theta > 70^\circ \) over the spectral range of interest. In off-resonance experiments, the RF transmitter frequency is positioned far enough off-resonance in order that \( \theta < 70^\circ \). Relaxation dispersion curves are obtained by either varying \( \omega_1 \) or the carrier offset, or both, thereby, varying \( \omega_{\text{eff}} \). Off-resonance \( R_{1p} \) experiments have been used to study protein-ligand exchange in the Abp1p SH3 domain \((91)\), chemical exchange processes in nucleic acids \((92)\), the dynamic behaviour of the L99A mutant of T4 lysozyme \((93)\), the folding pathway of the G48M mutant of the \( \gamma \)SH3 domain \((94)\) and the dynamics of the E140Q mutant of calmodulin \((95)\). The majority of these studies used \( ^{15} \text{N} \) nuclei \((96)\) or \( ^1 \text{H}^\alpha \) nuclei \((91)\) as probes for dynamics. A combination of on- and off-resonance \( R_{1p} \) experiments has also been used to characterise backbone dynamics in ubiquitin \((97)\).
These types of experiment are particularly powerful for studying fast chemical exchange, however, cross-relaxation effects between the amide and aliphatic protons during the spin-lock can give rise to errors (92). The use of deuteration can circumvent this issue along with inverting only the aliphatic protons in the middle of the spin lock, which inverts the sign of the cross-relaxation term and leads to a cancellation of the cross-relaxation effects (92). All published $R_{1p}$ experiments are based on the assumption that the transverse rates of the major and minor states are the same. Recently, an analytical expression was published for $R_{1p}$ for cases where the transverse relaxation rates of both states are not equal (98).

CPMG relaxation dispersion experiments are used for systems where $k_{ss}$ does not exceed 3000 s$^{-1}$. This type of experiment was first described in the 1950s (99-101); however, it was not widely used for studying protein dynamics until the mid-1990s. This is likely due to the advances in experimental design (92, 102-104) and isotopic labelling strategies (105, 106), which accompanied the growth in the use of CPMG relaxation dispersion to study dynamic motions in proteins.

CPMG relaxation dispersion experiments have been used to study the μs-ms dynamics in a variety of different systems. Dynamics processes that can occur on this time scale include side-chain reorientation, secondary structure changes, loop motion and hinged domain movements (107-110). The motions can mediate ligand binding and release as shown for product release from the enzymes CLpP (111) and ribonuclease A (112), folding and unfolding events in the fyn SH3 domain (113), the kix domain (114), the phosphorylated kinase inducible activation domain (pKID) of the transcription factor CREB (115) and the PBX-homeodomain (116); allostERIC regulation, such as cyclic AMP (cAMP) binding to the catabolite activator protein (CAP) (117) and the rate of catalytic turnover, as demonstrated for the enzymes adenylate kinase (118) and cyclophilin A (119).
In CPMG RD experiments, the relaxation properties of transverse coherences are monitored during a series of $\pi$ pulses. Transverse coherence can be due to single quantum (SQ), multiple quantum (MQ), zero quantum (ZQ) and double quantum (DQ) transitions \(120, 121\). In the constant time (CT) CPMG experiment, the effective relaxation rate $R_{2}^{\text{eff}}$ is measured as a function of the frequency ($\nu_{\text{CPMG}}$) at which $\pi$ pulses are applied during the constant time period ($T_{\text{relax}}$). $\pi$ pulses are applied in blocks of $\tau_{\text{CPMG}} - \pi - \tau_{\text{CPMG}}$ such that 

$$2N(\frac{3}{2}\pi + \tau_{\text{CPMG}}) = T_{\text{relax}}.$$ 

The CPMG pulsing frequency, $\nu_{\text{CPMG}}$ and effective $R_{2}$ field, $R_{2}^{\text{eff}}$ are calculated as follows:

$$\nu_{\text{CPMG}} = \frac{1}{4\tau_{\text{CPMG}}} \quad (1.15)$$

$$R_{2}^{\text{eff}} = \frac{1}{T_{\text{relax}}} \ln \frac{I}{I_{0}} \quad (1.16)$$

where $I$ is the intensity of the peak in the spectrum recorded with the relaxation delay $T_{\text{relax}}$ and $I_{0}$ is the intensity in a reference spectrum recorded without the relaxation delay. This experiment works due to the fact that $\pi$ pulses function to refocus the chemical shift evolution and therefore, the application of a series of $\pi$ pulses reduces the amount of chemical shift evolution. If these pulses are applied at an appropriate rate, this can move exchange from the slow to the fast exchange regime and the effects of chemical exchange are only quenched when $\nu_{\text{CPMG}}$ is sufficiently high. The analysis of relaxation dispersion curves produced from these experiments, enable the determination of the exchange rate, $k_{\text{ex}}$, the population of the minor state, $p_{B}$ and the chemical shift difference between the two states, $\Delta \omega$ in the slow exchange regime ($k_{\text{ex}} << \Delta \omega$) by using Carver-Richards equation:

$$R_{2}^{\text{eff}} = R_{2}^{0} + \frac{k_{\text{ex}}}{2} - \nu_{\text{CPMG}} \cosh^{-1}[D_{+}\cosh(\eta_{+}) - D_{-}\cos(\eta_{-})] \quad (1.17)$$

with
or in the fast exchange regime ($k_{ex} \gg \Delta \omega$):

$$R_{z}^{eff} = R_{z}^{0} + \frac{p_A p_B \Delta \omega}{k_{ex}} \left[ 1 - \frac{4 \nu_{CPMG}}{k_{ex}} \tanh \left( \frac{k_{ex}}{4 \nu_{CPMG}} \right) \right]$$ (1.18)

In the fast exchange regime, $p_B$ and $\Delta \omega$ cannot be separated.

Accurate analysis of RD data can be impeded by the fact that there are often many sets of model parameters, which fit the dispersion curve well. One such issue is the determination of the sign of $\Delta \omega$, since in all expressions it is a square parameter. This issue is commonly overcome by either comparing peak positions in spectra recorded at different $B_0$ values or by comparing peak positions in HSQC and HMQC spectra (122). The exchange induced shift, $\delta^{ex}$ can be expressed as:

$$\delta^{ex} = k_{ex} \frac{\xi}{(1 + \rho)^2 + \xi^2}$$ (1.19)

where $\rho = \frac{\Delta R}{k_b}$ and $\xi = \frac{\Delta \omega}{k_b}$, where $\Delta R$ is the difference in transverse relaxation rates between the two states. To discuss the effect of changing $B_0$ in the case of $^{15}$N spins, it is useful to express the shifts in frequency in ppm, such that $\Delta \omega = \gamma_N B_0 \Delta \tilde{\omega}$ and $\delta^{ex}_N = \gamma_N B_0 \tilde{\delta}^{ex}_N$. All variables with a tilde are in ppm. Substituting these values into equation 1.19 gives:

$$\tilde{\delta}^{ex}_N = \frac{k_{ex} \gamma_N}{(1 + \rho)^2 + (\gamma_N B_0 \xi_N)^2}$$ (1.20)
It, therefore follows, that the positions of peaks recorded at magnetic field $B_0^A$ will be shifted with respect to their positions in spectra recorded at a field strength of $B_0^B$ and the shift difference between these two positions, $\tilde{\delta}_N$, is given by:

$$\tilde{\delta}_N = \frac{k_{15N}^{B,N}}{(1+\rho)^2 + (\gamma N^B \delta N_N^B)^2} - \frac{k_{2}^{15N}}{(1+\rho)^2 + (\gamma N^B \delta N_N^B)^2} \quad (1.21)$$

The method of comparing the frequencies of resonances recorded in single and multiple quantum experiments is based on the fact chemical exchange can affect both the nitrogen and proton spins and therefore, cross-peaks in an HMQC spectrum will be shifted in the $^{15}$N dimension relative to the equivalent peaks in an HSQC spectrum. This is due to the fact that exchange averaging in single and multiple quantum experiments is different.

Once the signs of the $\Delta\omega$ parameters have been determined, the chemical shifts of the excited state can be determined and these data can be used to solve the structure of the excited state of a protein. This has been demonstrated for the L99A mutant of T4 lysozyme (123).

EXSY spectroscopy, also known as, the ZZ exchange experiment has been in use since the late 1970s (124) to study systems where $k_{ex}$ falls in the range of 0.1 – 100 s$^{-1}$ and resonances for both states can be observed. Dynamics processes that occur on this timescale include slow conformation changes, such as domain movement, as demonstrated for the archaeal proteasome (125); the formation of protein-nuclei acid complexes (126), and topological interconversion of secondary structure elements, as shown for human lympholactin (127), PhoPQ-activated gene P (128) and the leucine zipper Gcn4 (129).
In the $^{15}$N-$^1$H ZZ-exchange experiment, magnetisation is usually transferred from an amide proton to the attached amide nitrogen, where it is frequency labelled during the nitrogen chemical shift evolution period. The nitrogen magnetisation is then converted to $N_z$ magnetisation, such that there is nonequilibrium magnetisation along the $z$-axis. During a mixing period, magnetisation is transferred between the two protein states and thereafter, transferred back to the amide proton for detection. The resultant spectra will contain four peaks for every amide in the protein, two autopeaks for the resonances in states A and B and two crosspeaks resulting from magnetisation transfer during the mixing period \(86\). Acquisition of a series of 2D spectra with different mixing times is used to generate build-up curves from the four measured intensities (two from the autopeaks and two from the crosspeaks). These data are then fit to an exchange model from which $k_{ex}$ values and the populations of the two states can be determined. For two-state exchange, three unique build-up curves can be used to extract these parameters, along with the $R_1$ rates of each state \(86\):

\[
I_{AA}(T) = p_A(p_A + p_B e^{-k_{ex}T})e^{-R_1T} \quad (1.22)
\]
\[
I_{BB}(T) = p_B(p_B + p_A e^{-k_{ex}T})e^{-R_1T} \quad (1.23)
\]
\[
I_{AB}(T) = I_{BA}(T) = p_A p_B(1 - e^{-k_{ex}T})e^{-R_1T} \quad (1.24)
\]

Spectral crowding and/or poor sensitivity can limit the utility of EXSY spectroscopy, since it introduces additional signals into the acquired spectra. This is generally not a significant limitation, since many EXSY studies only require a few structural probes, as evidenced in the study of the $\alpha_7$ annulus of the 20S proteasome core particle \(125\). A TROSY-selective version of the $^{15}$N-$^1$H ZZ exchange experiment has been introduced, enabling larger systems to be studied using this method \(130\).
Paramagnetic Relaxation Dispersion

Relaxation dispersion experiments are extremely useful for investigating minor, “invisible” states of proteins and provide information about the kinetics and thermodynamics of dynamic processes. However, the chemical shift differences between the major and minor states, obtained from these experiments, are difficult to interpret in terms of structure. It is possible to extract anisotropic chemical shift differences and dipolar couplings from relaxation dispersion experiments (104) and these can be of more utility in the determination of structures of lowly populated states (123). Moreover, pseudocontact shifts (PCS) resulting from a magnetic susceptibility of paramagnetic centres also contain structural information (131). Due to the fact that PCSs are determined by the position of the nucleus within a paramagnetic tensor frame of the metal, the PCS gradient across the protein can be employed as a reference frame for the study of protein dynamics on the appropriate time scale. If a region of a protein is undergoing a chemical exchange process in a paramagnetic field, such as internal motions relative to the paramagnetic centre, an average PCS will be observed for the nuclei in the region and, therefore, line broadening will be observed, if this motion is occurring on the μs – ms timescale (132). Fluctuating PCSs can be analysed using relaxation dispersion techniques and thereby, provide a route to determining the structures of minor states in proteins. Motion detected in this manner may not, however, result from internal motions of the protein itself, but rather from movement of a paramagnetic centre artificially incorporated into the protein’s structure, for example, using a paramagnetic tag as was demonstrated for CLaNP-5 attached to pseudoazurin and cytochrome c (133).

The proteins

P450cam

Cytochromes P450, so-called due to their characteristic absorption at 450 nm when CO is bound (134), are type b haemoproteins found throughout the three kingdoms of life. These proteins form a superfamily, which is subdivided
into families and subfamilies based on primary sequence comparisons (135). Their roles in mammals include steroid biosynthesis as well as the metabolism of xenobiotics by hydroxylation to make them more polar, and therefore easier for the organism to excrete. As a consequence of their ubiquity and role in xenobiotic metabolism, it is not surprising that over 10,000 cytochrome P450 genes have been identified and, from these, many P450 isozymes have been isolated and some of their associated functions have been identified.

In order for cytochromes P450 to hydroxylate various exogenous and endogenous compounds they require electrons to be donated and, therefore, are often found as part of multi-electron transfer chains. The nature of these electron transfer chains can vary greatly between different organisms (136).

Cytochromes P450 provide a major route by which pharmaceuticals and other foreign substances are metabolised and consequently, many drugs are substrates of one or more CYPs. It is important in drug development to ensure that compounds have a sufficiently adequate half-life, such that they are not metabolised prior to eliciting the desired effect. It is also vital that any potential drug is not a strong inhibitor of a CYP as this could lead to toxic levels of the compound itself and perhaps other compounds metabolized by the same CYP building up. This is a fine balance that needs to be struck in drug development and can amount to a significant challenge. Many mammalian cytochromes P450 have active sites whose volumes are far greater than the substrates and, in addition, cytochromes P450 in higher organisms are often promiscuous with regards to the substrates that they will bind. Prior to 2003, when the first crystal structure of a mammalian cytochrome P450 isozyme was solved (137), homology models based on the structures of bacterial isozymes were used (138), despite the fact that there is only ~20% sequence homology between bacterial and mammalian cytochromes P450 (139). The most studied of these bacterial enzymes is cytochrome P450101A1, also known as P450cam, isolated from Pseudomonas putida and investigations
using this enzyme are still used to drive studies of cytochromes P450 in higher organisms. The first structure of this P450 enzyme was solved in 1985 \((140)\) in both substrate-bound and substrate-free forms, both of which were in the so-called “closed conformation” and this raised the questions as to how substrate can enter this enzyme and products can leave. A more recent structure of the substrate-free form \((141)\) showed a more open conformation, thus the enzyme must open for the substrate to bind near to the haem by displacing a bound water molecule. Then the enzyme closes while the hydroxylation reaction occurs and re-opens to release the product. Figure 1.6 shows an overlay of the open and closed conformations of P450cam. To this end, the presence of channels in P450 enzymes has been investigated, principally using theoretical methods, such as random expulsion molecular dynamics (REMD) \((142-144)\), although EPR spectroscopy has also been used to study camphor binding to P450cam \((145)\). However, experimental evidence for the existence of these channels in solution and the mechanisms by which these proteins open and close is scarce.

Figure 1.6 Overlay of the open conformation of P450cam in red (PDB code 3L62) \((141)\) and the closed conformation of P450cam in cyan (PDB code 2CPP) \((146)\). The black arrow indicates the substrate entry channel.
T4 lysozyme

T4 lysozyme (Figure 1.7) is an 18.7 kDa endoacetyl-muramidase produced in the cells of *E. coli* after infection with bacteriophage T4. The three dimensional structure of this enzyme is organized into two domains joined by a long helix. The active site cavity, where the hydrolysis reaction occurs, is located between the two domains. In the resting state, this cavity is appears to be occluded, based on crystallographic data of the wild-type protein. However, crystal structures of point mutations of T4 lysozyme revealed a series of conformations in which the active site is more open. The transformation from the closed to open states has been referred to as hinge bending, which consists of a rotation of one domain relative to the other about an axis running through the interface of the two domains. Direct evidence for the hinge-bending motion in solution has been obtained using electron paramagnetic resonance (EPR) spectroscopy measurements of the distances between the two domains (147). It has been shown both by NMR (148) and molecular dynamics (MD) studies (149) that domain motion in T4 lysozyme occurs at equilibrium. The actual timescale for hinge-bending is as yet unknown, although an upper estimate of ~160 μs has been proposed based on NMR data (148) and classical MD studies showed that the transition from the open state to the closed state can occur within 1 ns (149). Förster resonance energy transfer (FRET) studies have also been used to investigate the conformational dynamics of T4 lysozyme. These studies indicated that the formation of the enzyme-substrate complex involves the population of six intermediate conformations and occurs on the ms timescale. Moreover, the L99A mutant of the protein has been showed to be highly dynamic on the ms timescale (150) and this mutant has been used to solve the structure of a minor transiently formed “invisible” state of this protein (123). A more recent study employing Fluorescence Correlation Spectroscopy has added further weight to the suggestion that T4 lysozyme populates multiple intermediate states (151).
Avr2

The fungus *Cladosporium fulvum* is an asexual extracellular fungal pathogen of tomato plants and during infection small, cysteine-rich proteins are secreted, including avirulence factor 2 (Avr2). The Avr2 gene encodes a pre-protein of 78 amino acids, which matures into a 58 amino acid protein containing eight cysteine residues. The mature protein induces hypersensitive response (HR) in tomato plants carrying the cognate Cf-2 resistance gene (153, 154). Avr2 inhibits at least four tomato cysteine proteases including Rcr3, Pip1, aleurain and TDI65 and, therefore, plays an offensive role in virulence by targeting and inhibiting host proteases that are important for host defence (155-157). The role of Avr2 has not only been demonstrated for *C. fulvum*, but also for other fungal tomato pathogens, including *Botrytis cinerea*, and *Cercosporidium dahliae*. Avr2 interacts with the plant cognate factor, Cf-2, in an indirect manner. Avr2 binds to Rcr3imp, which is a cysteine protease originating from *Lycopersicon pimplinellifolium* (155, 156), and Cf-2 recognises this complex and localised cell-death results, preventing further incursion of this pathogen (158). The exact mechanism of Avr2’s role is not yet known, however, structural modification of Rcr3 by Avr2, rather than Rcr3 inhibition is
the most likely cause of Cf-2-mediated defence signaling, since a natural variant of Rcr3 occurs in *Lycopersicon esculentum* (Rcr3esc), which causes spontaneous HR in the presence of Cf-2 in an Avr2-independent manner. Point mutations, deletions and transposon insertions have been shown to circumvent Avr2-triggered Cf-2-mediated HR (153), however, the structure of the Avr2 protein is necessary to understand the full impact of these changes.

**EiAPR**

*Enteromorpha intestinalis* adenosine-5’-phosphosulphate reductase (EiAPR) is considered to be an archetypal “plant-type” sulphonucleotide reductase. It consists of an N-terminal reductase domain, containing an iron sulphur cluster and a C-terminal domain. It shares 52% sequence identity with *Pseudomonas aeruginosa* adenosine-5’-phosphosulphate reductase (PaAPR), however, this identity is mainly restricted to the reductase domain of this enzyme. EiAPR differs mechanistically from PaAPR in that PaAPR uses thioredoxin as an electron donor, whereas the C-terminal domain provides this function in EiAPR. All plant and algal phosphosulphate reductases have this two domain structure (according to nucleotide sequencing), but neither the reason for this structure nor the mechanism of interaction of the two domains are fully understood. It is intriguing that even though thioredoxin plays an important role in chloroplast metabolism, these enzymes have evolved to contain a C-terminal domain that freed sulphate reduction from the thioredoxin-based electron transport system.

**Thesis outline**

This thesis explains the breadth of modern NMR methods for studying protein structure and dynamics. The first part focusses on classical NMR methods in several collaborative projects and the second part investigates new avenues for paramagnetic NMR.

Chapter 2 details the NMR assignment the C-terminal domain of EiAPR, using classical heteronuclear 3D NMR spectroscopy and the results can be used,
as part of an on-going collaborative project, to elucidate how this protein behaves in concert with its interaction partner at an atomic level.

In Chapter 3, the NMR assignments of a 47 kDa protein, P450cam, using TROSY and BEST-TROSY NMR experiments is presented. This protein exists in multiple forms depending on the presence of different ligands and, therefore, assignments were obtained for protein samples containing three different ligands. These assignments formed the basis for further studies of its complex with its redox partner (Chapter 5).

Chapter 4 describes the structures of a small protein, Avr2, solved at pH 7 using 2D NOESY spectra. This protein exists in two forms at pH 7 and the differences in structure between the two forms are evaluated.

In Chapter 5, paramagnetically-tagged $^{15}$N-Leu labelled P450cam is used to determine which form P450cam occupies when in complex with its binding partner, putidaredoxin. The spectra are assigned on the basis of data obtained in Chapter 3 and additional assignments were obtained on the basis of pseudocontact shifts.

Chapter 6 describes a piece of software, which uses paramagnetic NMR data, namely pseudocontact shifts, to perform de novo assignment of protein nuclei.

In Chapter 7, paramagnetic relaxation dispersion NMR is used to evaluate which factors influence two armed probe mobility. CLaNP-5 was attached to a mutant of T4 lysozyme and CPMG relaxation dispersion spectra were recorded for the dia- and paramagnetically tagged samples. All data were fitted to a two state relaxation dispersion model. No probe mobility was observed for this mutant of this protein.

Finally, conclusions are presented about the different protein systems studied and the advantages of the different methods employed. The prospects of the results obtained for each protein system are discussed and possible avenues of study are presented. The use of paramagnetic NMR for assignment of large protein complexes and studying protein dynamics is also discussed.