



Universiteit
Leiden
The Netherlands

Classical and paramagnetic NMR spectroscopy techniques applied to different protein systems

Skinner, S.P.

Citation

Skinner, S. P. (2013, December 5). *Classical and paramagnetic NMR spectroscopy techniques applied to different protein systems*. Retrieved from <https://hdl.handle.net/1887/22714>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/22714>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



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

**Classical and paramagnetic NMR spectroscopy
techniques applied to different protein systems.**

Simon P. Skinner

Classical and Paramagnetic NMR spectroscopy
techniques applied to different protein systems

Simon P. Skinner

Doctoral Thesis, Leiden University, 2013

ISBN number: 978-94-6203-477-8

© 2013, Simon Peter Skinner

Cover designed by Myra Albersen: www.albersendesign.exto.nl

Printed by CPI-Wöhrmann Print Service - Zutphen

Classical and paramagnetic NMR spectroscopy techniques applied to different protein systems.

Simon P. Skinner

Proefschrift

ter verkrijging van

de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus Prof. mr. C. J. J. M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op donderdag 5 december 2013
klokke 16.15 uur

door

Simon Peter Skinner

Geboren te London, Verenigd Koninkrijk
in 1985

Promotiecommissie

Promotor: Prof. Dr. M. Ubbink

Overige Leden: Dr. R. A. L. van der Hoorn (MPI Köln)
Prof. Dr. J. Brouwer
Prof. Dr. G. W. Vuister (University of Leicester)
Dr. J. A. R. Worrall (University of Essex)

The investigations described in this thesis were performed at the Protein Chemistry department of the Leiden Institute of Chemistry, Universiteit Leiden, Leiden, Nederland.

Financial support for the research was provided by the Netherland Organisation for Scientific research (NWO) VICI grant 700.58.441

***To my family who have
always supported me***

Contents

Chapter 1	General Introduction	11
Chapter 2	Backbone Assignment of the C-terminal domain <i>Enteromorpha intestinalis</i> adenosine 5'-phosphosulfate reductase (EiAPR)	43
Chapter 3	Backbone and side-chain methyl assignments of cytochrome P450cam with different ligands bound	57
Chapter 4	Solution structures of two forms of Avirulence factor 2 (Avr2) at pH 7	79
Chapter 5	The influence of Pdx binding on the opening and closing of P450cam in solution	95
Chapter 6	PARAssign – paramagnetic NMR assignment on the basis of pseudocontact shifts	111
Chapter 7	Paramagnetic relaxation dispersion of the K147C/T151C mutant of T4 lysozyme	137
Chapter 8	General discussion, conclusions and perspectives	149
	Reference List	155
	Nederlandse Samenvatting	169
	English Summary	173
	List of Publications	177
	Curriculum vitae	178

List of abbreviations

1-PI	1-phenylimidazole
Avr2	Avirulence Factor 2
BEST	Band Selective Short Transient
CLaNP	Caged Lanthanide NMR Probe
CNS	Crystallography and NMR System
COSY	Correlation Spectroscopy
CPMG	Carr-Purcell-Meiboom-Gill
CRINEPT	Cross-correlated Relaxation Enhanced Polarisation Transfer
CSA	Chemical Shift Anisotropy
CSP	Chemical Shift Perturbation
CT	Constant Time
CYP	Cytochrome P450
DD	Dipole-dipole
DEER	Double Electron Electron Resonance
DQ	Double Quantum
EiAPR	<i>Enteromorpha intestinalis</i> adenosine-5'-phosphosulphate reductase
EPR	Electron Paramagnetic resonance
EXSY	Exchange Spectroscopy
HMQC	Heteronuclear Multiple Quantum Coherence
HSQC	Heteronuclear Single Quantum Coherence
Ln	Lanthanide
MALDI-TOF	Matrix-assisted laser desorption/ionisation time of flight
MQ	Multiple Quantum
MS	Mass Spectrometry
NADH	reduced Nicotinamide Dinucleotide
NMR	Nuclear Magnetic Spectroscopy
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
PAZ	Pseudoazurin
PCS	Pseudocontact Shift
PDB	Protein Databank
Pdx	Putidaredoxin
PRE	Paramagnetic Relaxation Enhancement
RD	Relaxation Dispersion
RDC	Residual Dipolar Coupling
RMSD	Root Mean Square Deviation
RF	Radio Frequency
SB	Solomon Bloembergen
SBMF	Solomon Bloembergen Model Free
SQ	Single Quantum
TOCSY	Total Correlation Spectroscopy
TROSY	Transverse Relaxation Optimised Spectroscopy

List of symbols

B_0	static magnetic field
g	electron g -factor
h	planck's constant
$J(\omega)$	Spectral density function
k_a	exchange rate of state A
k_B	Boltzmann constant
k_{ex}	exchange rate constant
p_a	population of state A
r_{AB}	distance between nucleus A and nucleus B
R_2^{eff}	effective transverse relaxation rate
R_{ex}	rate of chemical exchange broadening
S	spin quantum number
S^2	order parameter
T	Temperature
T_{relax}	constant time period in RD experiments
$\Gamma_{2,SB}$	transverse PRE rate
$\Delta\chi_{ax}$	axial component of the magnetic susceptibility tensor
$\Delta\chi_{rh}$	rhombic component of the magnetic susceptibility tensor
$\Delta\omega$	chemical shift difference between two sites
Ω	population-weighted average chemical shift
Ω_k	resonance offset of the k th resonance
γ_A	gyromagnetic ratio of nucleus, A
μ_0	permeability of a vacuum
μ_B	Bohr magneton
ν_{CPMG}	CPMG pulsing frequency
τ_c	total correlation time
τ_{CPMG}	CPMG delay
τ_r	rotational correlation time
τ_s	effective electron relaxation time
ω_1	amplitude of applied RF field
ω_I	larmor frequency of nucleus I
ω_{eff}	effective field in the rotating frame

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 μ s 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 (1). 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 ^1H - ^1H COrrrelation SpectroscopY (COSY) and ^1H - ^1H Total COrrrelation SpectroscopY (TOCSY) as well as dipolar coupling based correlation in ^1H - ^1H Nuclear Overhauser Effect SpectroscopY (NOESY). Greater spectral resolution is afforded by linking TOCSY and NOESY spectra to a ^{15}N - ^1H 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^α , H^β 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^β 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 ^1H resonances cannot be observed in HSQC-TOCSY or HSQC-NOESY spectra, only in ^1H - ^1H TOCSY and NOESY spectra or ^{13}C -separated spectra.

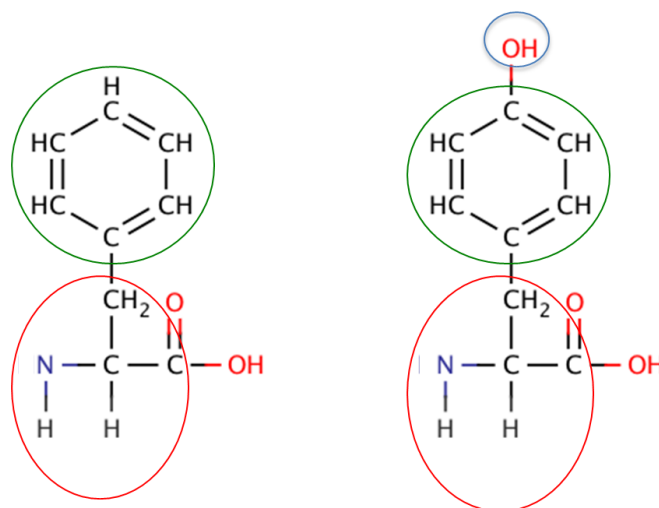


Figure 1.1 Multiple spins systems in Phe and Tyr residues. The atoms encircled in red, green and blue represent separate spin systems.

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 $\sim 5 \text{ \AA}$ from each other, and, therefore, amide H^N sequential links can be established. Intra-residue NOEs are observable for H^N , H^α and H^β protons and inter-residue NOEs are observable between adjacent H^N protons, adjacent H^N and H^α protons and between adjacent H^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.

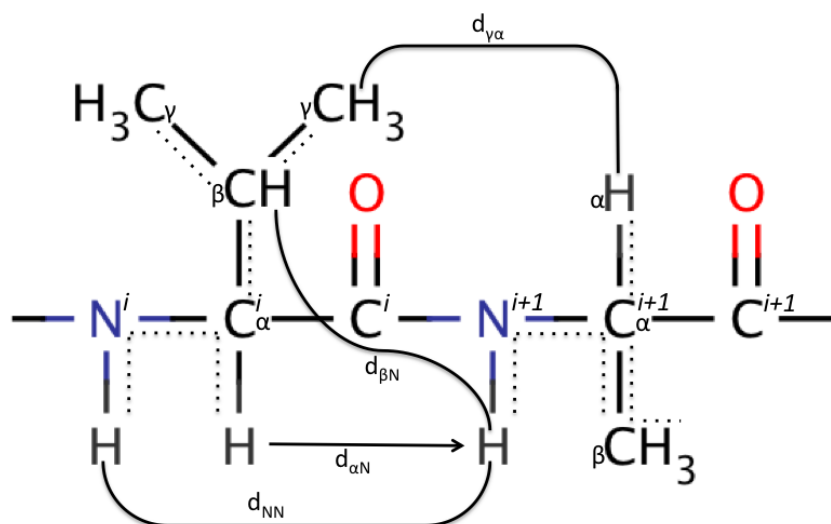


Figure 1.2 Sequential ^1H NMR assignment of a valine–alanine (V–A) peptide segment in a polypeptide chain. The dotted lines connect groups of hydrogen atoms that are 3J coupled. The solid arrows link pairs of hydrogen atoms in neighbouring amino acid residues that are separated by short through-space distances, $d_{\alpha\text{N}}$, $d_{\beta\text{N}}$, $d_{\gamma\alpha}$ 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, $\text{C}\alpha$ 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 $\text{C}\alpha$ and $\text{C}\beta$ nuclei have characteristic chemical shifts and, therefore, individual residue types can be easily identified.

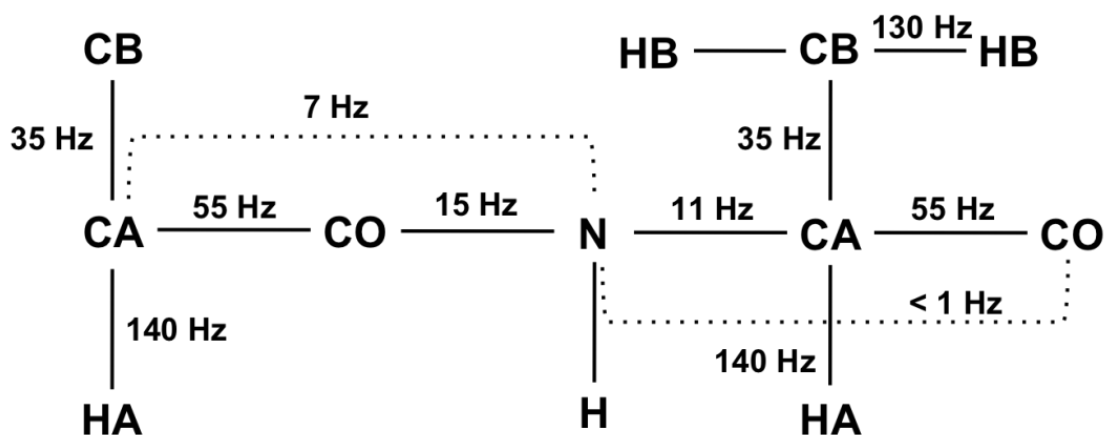


Figure 1.3 1J (solid lines) and 2J (dashed lines) coupling constants between protein nuclei used for magnetisation transfer in ^{13}C , ^{15}N labelled proteins. (Figure adapted from (11))

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 $\text{C}\alpha$ and between the backbone $\text{C}\alpha$ and the side chain $\text{C}\beta$ of both the residue in question via 1J coupling and the residue which precedes it sequentially via the 2J $^{15}\text{N}-\text{C}'-\text{C}\alpha$ 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 $^{15}\text{N}-\text{C}\alpha$ coupling and the 1J $^{13}\text{C}\alpha-\text{C}'$ coupling, along with correlations between the amide nitrogen and the carbonyl carbon of the previous residue via the 2J $^{15}\text{N}-\text{C}'-\text{C}\alpha$ coupling and the 1J $^{13}\text{C}\alpha-\text{C}'$ coupling (Figure 1.4b) (lower case letters in experiment names indicate nuclei that are used for coherence transfer, but not detected).

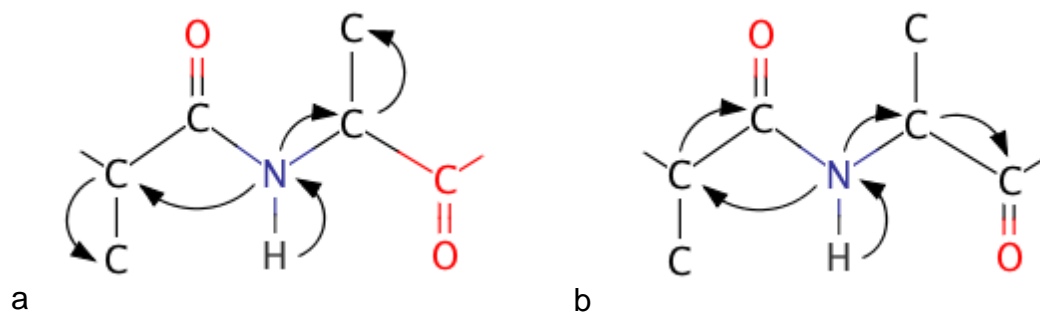


Figure 1.4 Magnetisation transfer pathways in the HNCACB (a) and HNcaCO (b) experiments.

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 $^1J^{15\text{N}-^{13}\text{C}\alpha}$ coupling (Figure 1.3) and the delay for N-C α scalar coupling evolution is set to $1/4J_{\text{NC}\alpha}$. 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 $^1J_{\text{NC}\alpha}$ and $^1J_{\text{C}\alpha\text{CO}}$ couplings are exploited to obtain intra-residue correlations, whereas the much weaker $^2J_{\text{NC}\alpha}$ is exploited to obtain the inter-residue correlations and the N-C α scalar coupling evolution period is set to $1/4^1J_{\text{NC}\alpha}$.

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 ($i-1$) C α and C β 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 .

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% $^2\text{H}_2\text{O}$. Deuteration reduces the dipole-dipole relaxation contributions that cause the large ^1H 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):

$$R_2^{dip} = d^2 \left(\frac{1}{10} J(0) + \frac{3}{20} J(\omega_{0,1}) + \frac{3}{20} J(\omega_{0,1} - \omega_{0,2}) + \frac{3}{40} J(\omega_{0,1}) + \frac{3}{20} J(\omega_{0,1} + \omega_{0,2}) \right) \quad (1.1)$$

where

$$d = \frac{\mu_0 h \gamma_1 \gamma_2}{4\pi r_{NH}^3}$$

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):

$$R_2^{CSA} = c^2 \left(\frac{2}{45} J(0) + \frac{1}{30} J(\omega_0) \right) \quad (1.2)$$

where

$$c = \gamma B_0 (\sigma_{\parallel} - \sigma_{\perp})$$

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).

$$R_{xy}^{(1)} = \frac{1}{10} d^2 J(0) + \frac{2}{45} c^2 J(0) + \frac{2}{15} cd(3 \cos^2 \theta - 1) J(0) \quad (1.3)$$

$$R_{xy}^{(2)} = \frac{1}{10} d^2 J(0) + \frac{2}{45} c^2 J(0) - \frac{2}{15} cd(3 \cos^2 \theta - 1) J(0) \quad (1.4)$$

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- $\frac{1}{2}$ 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 (^1H larmor frequency of 1.0 GHz) (14, 16). The ^{15}N - ^1H TROSY-HSQC selects only the S^{12} and I^{13} 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 ^1H manipulation using these pulses ensures minimal perturbation of the

proton spins resonating outside the excited spectral window (19-22).

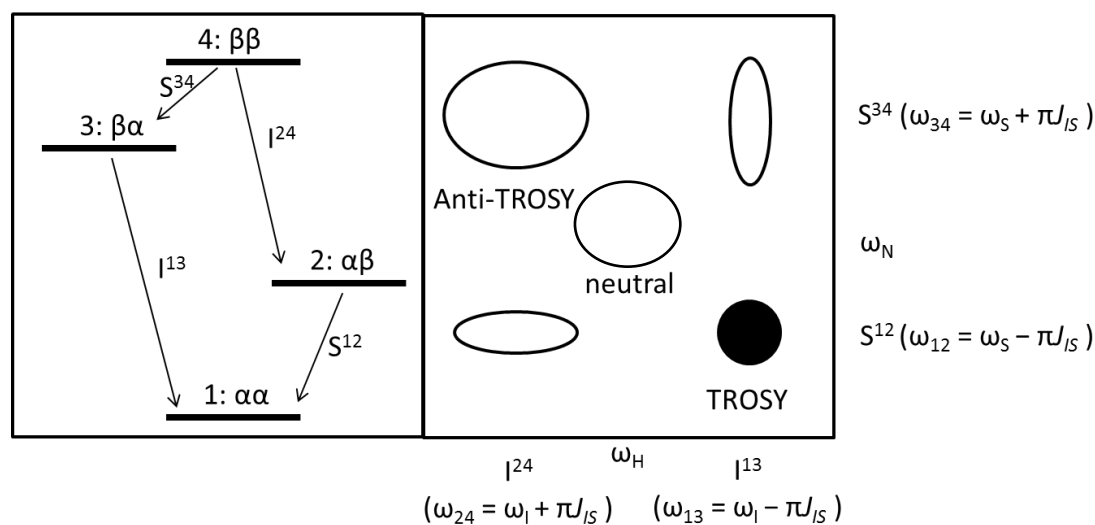


Figure 1.5 (a) Energy level diagram of a two spin $\frac{1}{2}$ system and showing the identity of the components of the 2D multiplet expressed as single transition basis operators, I^{13} , I^{24} , S^{12} , S^{34} . (b) Schematic diagram of an HSQC spectrum showing the four peaks observed in an HSQC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

The band selectivity of the ^1H 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 ^1H 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 C β 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 χ tensor. If it is assumed that a molecule tumbles isotropically, the magnitude of the PCS, δ_{PCS} , is given by (26):

$$\delta_{PCS} = \frac{1}{12\pi r_i^3} \left[\Delta\chi_{ax} (3(\vec{r}_z \cdot \vec{r}_i)^2 - r_i^2) + \frac{3}{2} \Delta\chi_{rh} (\vec{r}_x \cdot \vec{r}_i)^2 - (\vec{r}_y \cdot \vec{r}_i)^2 \right] \quad (1.5)$$

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 \vec{r}_x , \vec{r}_y and \vec{r}_z , 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 α, β and γ 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 $\sim 60 \text{ \AA}$ (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 ^1H -heteroatom bond (A-B) are assumed to be static, RDCs are given by (26):

$$RDC = \frac{B_0^2}{15k_B T} \frac{\gamma_A \gamma_B \hbar}{16\pi^3 r_{AB}^3} \left[\Delta\chi_{ax} (3(\vec{r}_x \cdot \vec{r}_i)^2 - r_i^2) + \frac{3}{2} \Delta\chi_{rh} (\vec{r}_x \cdot \vec{r}_i)^2 - (\vec{r}_y \cdot \vec{r}_i)^2 \right] \quad (1.6)$$

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 (Γ_2) PRE rate due to Solomon relaxation is described by the Solomon-Bloembergen (SB) equation (29, 30):

$$\Gamma_{2SB} = \frac{1}{15} \left(\frac{\mu_0}{4\pi} \right)^2 \frac{\gamma_I^2 g^2 \mu_B^2 S(S+1)}{r^6} \left(4\tau_c + \frac{3\tau_c}{1+(\omega_I \tau_c)^2} \right) \quad (1.7)$$

Where τ_c is correlation time defined as $(\tau_r^{-1} + \tau_s^{-1})^{-1}$, τ_r is the rotational correlation time of the macromolecule and τ_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):*

$$\Gamma_{2,SBMF} = \frac{1}{15} \left(\frac{\mu_0}{4\pi} \right)^2 \frac{\gamma_I^2 g^2 \mu_B^2 S(S+1)}{r^6} \left(4(S^2 \tau_c + (1 - S^2) \tau_t) + 3 \left[\frac{S^2 \tau_c}{1 + (\omega_I \tau_c)^2} + \frac{(1 - S^2) \tau_t}{1 + (\omega_I \tau_t)^2} \right] \right) \quad (1.8)$$

Where τ_c is the total correlation time, defined as $\tau_r^{-1} + \tau_s^{-1} + \tau_i^{-1}$ and τ_i is the correlation time for internal motion of nucleus, i . Curie spin relaxation is the dominant component of the ^1H transverse (Γ_2) PRE rates for metal ions with anisotropic g -tensors and very short electron relaxation times $\tau_r \ll \tau_s$, such as Fe^{3+} , Tm^{3+} , Yb^{3+} and Dy^{3+} . The Γ_2 rate due Curie-spin relaxation is given by (34):

$$\Gamma_{2,curie-spin} = \frac{1}{5} \left(\frac{\mu_0}{4\pi} \right)^2 \frac{\omega_I^2 g^4 \mu_B^4 S^2 (S+1)^2}{(3k_B T)^2 r^6} \left(4\tau_r + \frac{3\tau_r}{1 + (\omega_I \tau_r)^2} \right) \quad (1.9)$$

Where τ_r 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 ^1H 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 contains 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 χ 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 χ 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 dia- 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 χ 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 χ 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 , ^1H and ^2H 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 ^2H to obtain a more complete understanding of these fast timescale motions. For example, analysing both the ^{15}N - ^1H and $^{13}\text{C}^\alpha$ - ^{13}CO 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_{ex} , namely $R_{1\rho}$ and Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion experiments.

$R_{1\rho}$ experiments can be used when k_{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 ($k_{ex} \geq \Delta\omega$). 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, ω_1 , the population average chemical shift, $\Omega = p_A\Omega_A + p_B\Omega_B$ and the effective field in the rotating frame, ω_{eff} is given by:

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

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

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

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

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

In the slow exchange regime, $R_{1\rho}(\omega_{eff} \rightarrow \infty)$ is given by:

$$R_{1\rho}(\omega_{eff} \rightarrow \infty) = \cos^2\theta R_{1A}^0 + \sin^2\theta R_{2A}^0 \quad (1.13)$$

And in the fast exchange regime by:

$$R_{1\rho}(\omega_{eff} \rightarrow \infty) = \cos^2\theta(p_A R_{1A}^0 + p_B R_{1B}^0) + \sin^2\theta(p_A R_{2A}^0 + p_B R_{2B}^0) \quad (1.14)$$

When a strong RF field is applied on-resonance, increasing ω_{eff} is the same as increasing ω_1 , which quenches R_{ex} . However, ω_{eff} can also be increased by fixing ω_1 and moving the carrier frequency off resonance to quench R_{ex} ; as Ω increases, θ goes to zero, hence the magnetization spends more time along the z axis, therefore, according to equation 1.10, as ω_{eff} increases, exchange is moved into the fast exchange regime, which quenches R_{ex} . From these conclusions, it follows that there are two main types of $R_{1\rho}$ 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 ω_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 ω_1 or the carrier offset, or both, thereby, varying ω_{eff} . Off-resonance $R_{1\rho}$ 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 *fyn* SH3 domain (94) and the dynamics of the E140Q mutant of calmodulin (95). The majority of these studies used ^{15}N nuclei (96) or $^1\text{H}^\alpha$ nuclei (91) as probes for dynamics. A combination of on- and off-resonance $R_{1\rho}$ 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 $R_{1\rho}$ 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 $R_{1\rho}$ for cases where the transverse relaxation rates of both states are not equal (98).

CPMG relaxation dispersion experiments are used for systems where k_{ex} 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 π 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^{eff} is measured as a function of the frequency (ν_{CPMG}) at which π pulses are applied during the constant time period (T_{relax}). π pulses are applied in blocks of $\tau_{CPMG} - \pi - \tau_{CPMG}$ such that $2N(\frac{1}{2}\pi + \tau_{CPMG}) = T_{relax}$. The CPMG pulsing frequency, ν_{CPMG} and effective R_2 field, R_2^{eff} are calculated as follows:

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

$$R_2^{eff} = \frac{1}{T_{relax}} \ln \frac{I_0}{I} \quad (1.16)$$

where I is the intensity of the peak in the spectrum recorded with the relaxation delay T_{relax} and I_0 is the intensity in a reference spectrum recorded without the relaxation delay. This experiment works due to the fact that π pulses function to refocus the chemical shift evolution and therefore, the application of a series of π 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 ν_{CPMG} is sufficiently high. The analysis of relaxation dispersion curves produced from these experiments, enable the determination of the exchange rate, k_{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 ($k_{ex} \ll \Delta\omega$) by using Carver-Richards equation:

$$R_2^{eff} = R_2^0 + \frac{k_{ex}}{2} - \nu_{CPMG} \cosh^{-1}[D_+ \cosh(\eta_+) - D_- \cos(\eta_-)] \quad (1.17)$$

with

$$D_{\pm} = \frac{1}{2} \left[\pm 1 + \frac{\Psi + 2\Delta\omega^2}{\sqrt{(\Psi^2 + \zeta^2)}} \right]$$

$$\eta_{\pm} = \frac{[\pm\Psi + \sqrt{(\Psi^2 + \zeta^2)}]^{\frac{1}{2}}}{2\sqrt{2}\nu_{CPMG}}$$

$$\Psi = k_{ex}^2 - \Delta\omega^2$$

$$\zeta = -2\Delta\omega k_{ex} (p_A - p_B)$$

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

$$R_2^{eff} = R_2^0 + \frac{p_A p_B \Delta\omega^2}{k_{ex}} \left[1 - \frac{4\nu_{CPMG}}{k_{ex}} \tanh\left(\frac{k_{ex}}{4\nu_{CPMG}}\right) \right] \quad (1.18)$$

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, δ^{ex} can be expressed as:

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

where $\rho = \frac{\Delta R}{k_b}$ and $\xi = \frac{\Delta\omega}{k_b}$, where ΔR 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 $\delta_N^{ex} = \gamma_N B_0 \tilde{\delta}_N^{ex}$. All variables with a tilde are in ppm.

Substituting these values into equation 1.19 gives:

$$\tilde{\delta}_N^{ex} = \frac{k_a \tilde{\xi}_N}{(1+\rho)^2 + (\gamma_N B_0 \tilde{\xi}_N)^2} \quad (1.20)$$

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, $\tilde{\sigma}_N$, is given by :

$$\tilde{\sigma}_N = \frac{k_a \xi_N}{(1+\rho)^2 + (\gamma_N B_0^A \xi_N)^2} - \frac{k_a \xi_N}{(1+\rho)^2 + (\gamma_N B_0^B \xi_N)^2} \quad (1.21)$$

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 $0.1 - 100 \text{ s}^{-1}$ 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 - ^1H 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 α_7 annulus of the 20S proteasome core particle (125). A TROSY-selective version of the ^{15}N - ^1H 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.

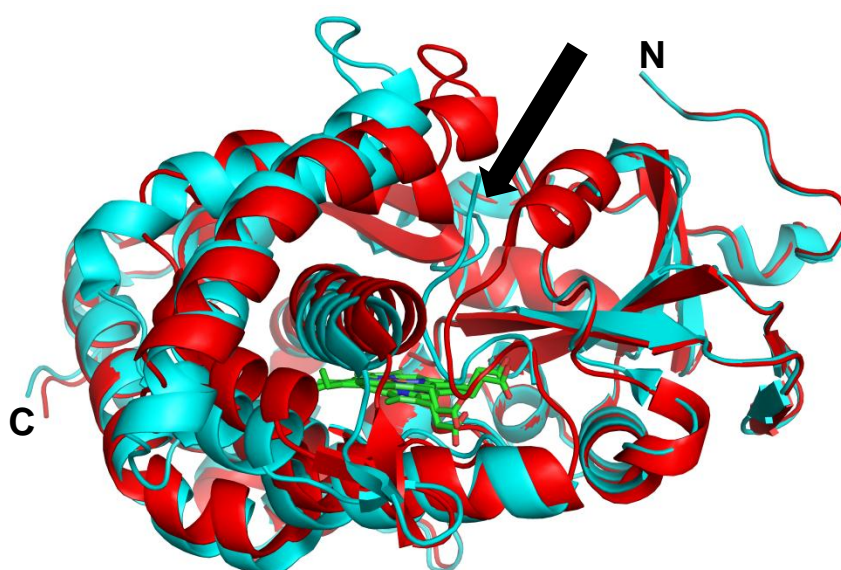


Figure 1.6 Overlay of the open conformation of P450cam in red (PDB code 3L62) (141) and the closed conformation of P450cam in cyan (PDB code 2CPP) (146). The black arrow indicates the substrate entry channel.

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 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 μ s 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).

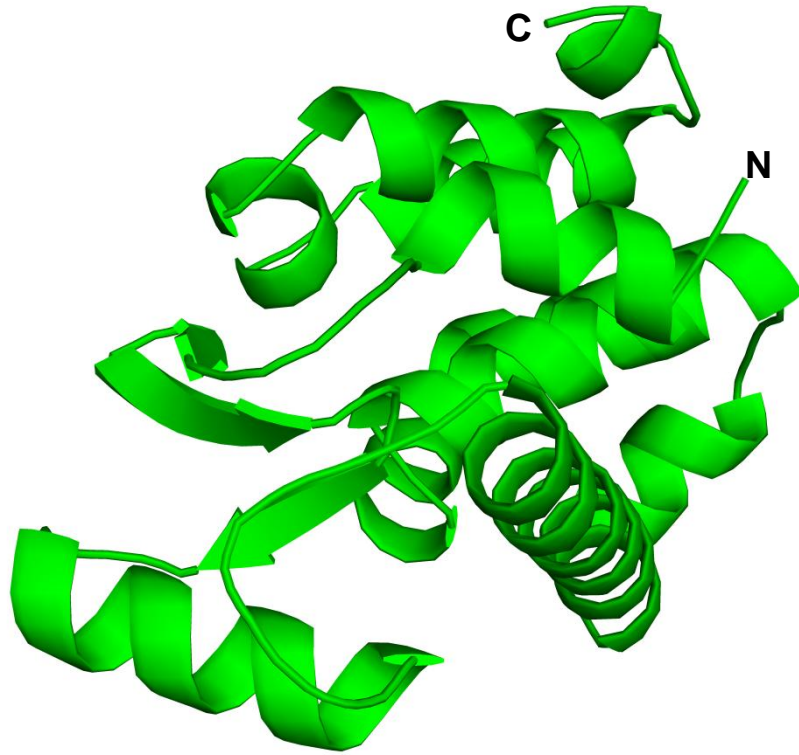


Figure 1.7 Structure of T4 lysozyme (PDB code 215L) (152).

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 *Cercitium dahliae*. Avr2 interacts with the plant cognate factor, Cf-2, in an indirect manner. Avr2 binds to Rcr3^{pimp}, which is a cysteine protease originating from *Lycopersicon pimpinellifolium* (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* (Rcr3^{esc}), 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 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.

Chapter 2

***Backbone Assignment of the C-terminal domain
Enteromorpha intestinalis adenosine 5'-
phosphosulfate reductase (EiAPR)***

Abstract

5'-Adenylylsulfate (APS) reductase from *Enteromorpha intestinalis* (EiAPR) is made up of two domains, which function to reduce APS to sulfite. The carboxyl-terminal domain functions as a thioredoxin, mediating the transfer of electrons to the APS reduction site on the N-terminal domain. In order to study the behaviour of the C-terminal domain of EiAPR, an NMR assignment of this domain was obtained. The sequence of the C-terminal domain is homologous to a thioredoxin for which the structure has been solved and the data showed that the secondary structure predicted from TALOS+ is consistent with that of this thioredoxin.

Introduction

Sulphate reduction occurs via two sequential electron transfer reactions. Firstly, a two-electron reduction occurs and this is catalysed by one of many diverse, yet related sulphonucleotide reductases, which reduce sulphate to sulphite. Secondly, a six-electron reduction of sulphite to sulphide is catalysed by sulphite reductase. In spite of this common two step procedure, all known assimilatory sulphonucleotide reductases are structurally and mechanistically diverse, but they all share a common ancestry. The first characterized sulphonucleotide reductase was CysH from *Escherichia coli* (159). This enzyme reduces 3'-phospho-5'-adenylylsulphate (PAPS) using thioredoxin or glutaredoxin as an electron donor. This enzyme is characteristic of the PAPS-dependent assimilatory sulphonucleotide reductases found in different eubacteria, fungi and archaea (160, 161).

Another member of the sulphonucleotide reductase family of enzymes is that isolated from *Pseudomonas aeruginosa* (PaAPR), which uses thioredoxin as an electron donor, but instead of reducing PAPS, it reduces 5'-adenylylsulphate (APS) (162). There is a high level of sequence homology between CysH and PaAPR, including a conserved cysteine residue in the active site. However, PaAPR contains a 4[Fe-S]²⁺ iron-sulphur cluster, whereas CysH does not, and PaAPR contains four additional conserved cysteine residues that are absent in

CysH. PaAPR has a thioredoxin binding site located near the C-terminus and a catalytic domain near the N-terminus (Figure 2.1).

Two models have been proposed for the roles of the cysteine residues in PaAPR, one of which proposes that all four cysteine residues are ligands for the iron-sulphur cluster (163) and a second which proposes that only three of the four cysteins ligate the iron-sulphur cluster (164). PaAPR is a typical example of APS-utilising sulphonucleotide reductases which occur in diverse eubacteria and archaea (165). The sulphonucleotide reductase found in *Bacillus subtilis* shows sequence and structural features which place it in the same group as PaAPR, with the exception that it can reduce both PAPS and APS (166).

A third class of sulphonucleotide reductases is characterised by the protein isolated from *Enteromorpha intestinalis*, the plant-type enzyme (167). This enzyme contains a $4[\text{Fe-S}]^{2+}$ iron-sulphur cluster and has a high level sequence homology with the amino terminal domain of PaAPR, the so-called “reductase domain.” It also contains a C-terminal domain, which is homologous to thioredoxin, but is not found in either CysH or PaAPR (168). In a related APS-reductase found in *Arabidopsis thaliana*, the C-terminal domain has been shown to confer the ability to use glutathione as an electron donor, suggesting that it functions as glutaredoxin, not a thioredoxin (169). EiAPR and PaAPR are 52% identical in their amino acid sequences and also have very similar calculated isoelectric points (EiAPR: 6.91, PaAPR: 6.26) (170). The precise reaction mechanism of sulphonucleotide reductases is not fully understood and it may vary among different types, as exemplified by the diversity in existing structures.

The kinetic properties of CysH indicate that it reduces PAPS via a Ping-Pong mechanism (159), wherein thioredoxin reduces a disulphide bridge formed between the two active site cysteine residues, one on each subunit of the CysH dimer and uses the stored electrons to reduce PAPS and release sulphite. Two mechanisms for PaAPR activity have been proposed. One proposes that the

active site cysteine (analogous to that in CysH) forms part of an intramolecular disulphide when the enzyme is oxidized and this is the target of reduction by thioredoxin (168). An alternative mechanism proposes that there is no intramolecular disulphide bridge and thioredoxin reduction occurs in a later step. However, both models agree on the formation of a transient disulphide linked intermediate between thioredoxin and PaAPR (163).

The function of the two domain structure of plant-type APS reductases is not fully understood and there are many examples of enzymes that have a fused domain structure with thioredoxin and glutaredoxin (171-173). All plant and algal APS reductases appear to have this overall structure (according to nucleotide sequencing comparisons). This suggests that this structure is present in all chloroplast containing host species and this may be an adaptation to the environment of the chloroplast stroma. Studies with the *A. thaliana* enzyme showed that the C-terminal domain functions as a glutaredoxin and that the efficiency of APS reduction is greatest when the C-terminal domain is located on the same polypeptide chain as the reductase domain and that catalytic efficiency of this reaction is greatly reduced when the C-terminal and reductase domains are expressed separately and then combined in an APS reductase assay (169).

Due to the fact that thioredoxin is a co-substrate in bacterial APS and PAPS reductases and thioredoxin plays a central role in chloroplast metabolism, it is unclear why plant-type APS reductases evolved to include a C-terminal domain and thereby, freed sulphate reduction from the thioredoxin based electron transport system. Moreover, the mechanism of interaction between the reductase and C-terminal domains of plant-type APS reductases is not understood. Using EiAPR as a model protein for this type of enzyme, NMR spectroscopy can be used to study these intramolecular interactions at the atomic level. Prior to any detailed NMR studies, an assignment of the protein nuclei is required and, therefore, the HN, N and C α nuclei of the backbone and the side chain C β nuclei of the C-terminal domain of EiAPR have been assigned using heteronuclear 3D NMR spectroscopy.

Methods and Experiments

A ^{15}N ^{13}C labelled sample of the C-terminal domain (residues 293-423) of EiAPR (Uniprot ID: O81350) with an N-terminal 6-His tag was kindly supplied by Prof David Knaff (Texas Technical University, USA). The protein contained six mutations: A347G, F348L, P354A, P359A, F364L and G365R. The sample was buffered exchanged into 30 mM Tris HCl pH 8.01 using a PD10 column (GE Healthcare). The NMR experiments for backbone assignment were performed with a sample of concentration 400 μM in 8%/92% $\text{D}_2\text{O}/\text{H}_2\text{O}$ at 300 K on a Bruker DMX 600 spectrometer equipped with a TCI-Z-GRAD cryoprobe and a ^1H Larmor frequency of 600 MHz.

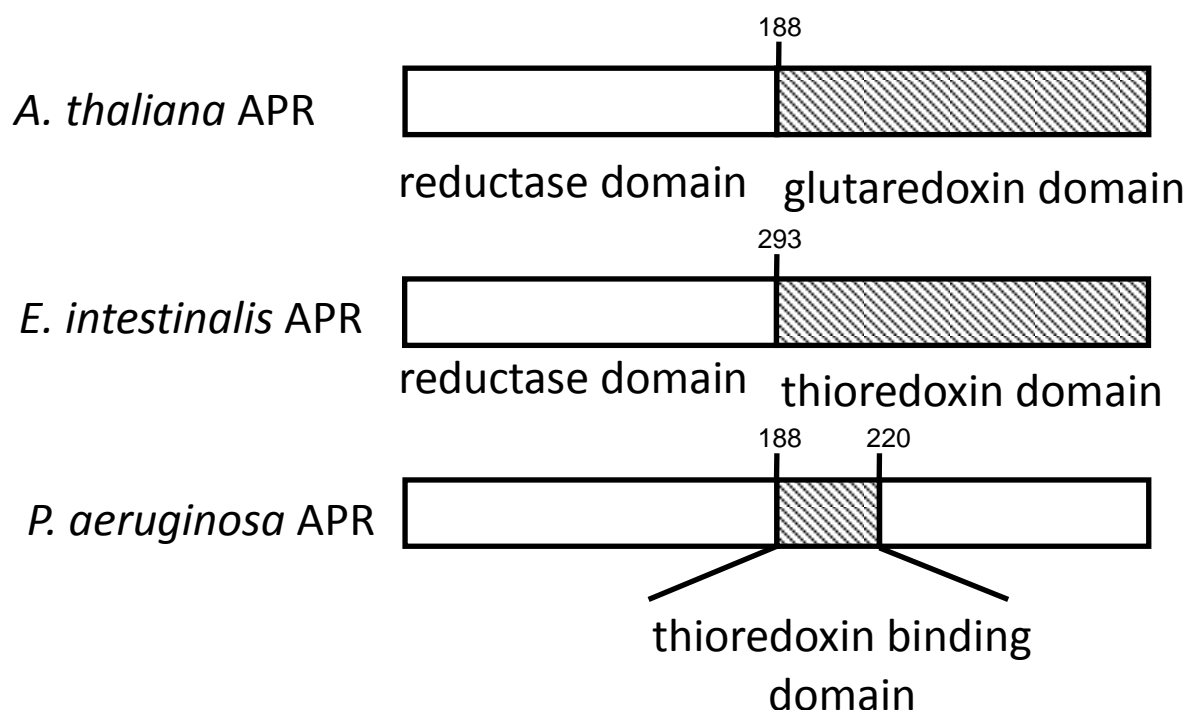


Figure 2.1 Schematic diagrams of APR from *A. thaliana*, *P. aeruginosa* and *E. intestinalis* showing the two domain structure of plant APRs compared to the one domain structure of a bacterial APR.

^1H , ^{13}C and ^{15}N sequential resonance assignments were obtained using a 3D HNCACB experiment and a 3D HN(CO)CACB experiment, both with 40 and 100 complex points and spectral widths of 2.25 kHz and 9.81 kHz in the ^{15}N and ^{13}C

dimensions, respectively. Data were processed in nmrPipe (174) and evaluated in CCPN Analysis 2.1.5 (175).

Results and Discussion

The chemical shifts were obtained for 99% of the backbone HN, $^{13}\text{C}\alpha$, ^{15}N , (without considering the proline residues) and the sidechain $^{13}\text{C}\beta$ nuclei (Table 2.1). The high quality of the NMR data is seen in the assigned ^1H - ^{15}N HSQC spectrum (Figure 2.2). The construct used for C-terminal domain of this protein begins at residue 293, hence V1 in the presented assignment corresponds to V293 in the full-length protein. Secondary structure prediction was performed based on the $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ chemical shifts using TALOS+ (176) and compared to the structure of a homologous protein, the third thioredoxin domain of Protein disulphide isomerase A4 from *Mus musculus* (PDB code 2DJ3) obtained from searching the PDB with the sequence of the C-terminal domain of EiAPR, the alignment is shown in Figure 2.3. The predicted secondary structure of EiAPR is consistent with that of 2DJ3, indicating reliability of the assignment and suggesting that the structures of the two proteins are very similar (Figure 2.4). Thus, it is concluded that the structure of the C-terminal domain of EiAPR is thioredoxin-like.

The APS reductase isolated from *A. thaliana* uses glutathione as an electron donor, which suggests that this enzyme functions as a glutaredoxin, rather than a thioredoxin. The fact that EiAPR has a C-terminal thioredoxin-like domain, indicates a secondary evolutionary divergence within sulphonucleotide reductases. The bacterial enzymes CysH and PaAPR both bind free thioredoxin and use the electrons generated from the reduction of a disulphide bridge to reduce PAPS and release sulfite. APR from *A. thaliana* and *E. intestinalis* have a two-domain structure consisting of an reductase domain containing an Fe-S cluster and a C-terminal reductase domain. The results presented in this study provide further evidence that the C-terminal domain of EiAPR is thioredoxin-like and this may be an adaptation to the environment of the chloroplast stroma, even

though thioredoxin is a co-substrate in chloroplast metabolism. The evolutionary advantage conveyed by this two domain structure has yet to be elucidated.

The way in which the two domains of EiAPR interact is not known and, therefore, further studies are required to determine the mechanistic details of sulphonucleotide reduction by this enzyme. The assignments obtained in this study can be used to determine which residues on the C-terminus are affected by interaction with the reductase domain and the dynamic behavior of the C-terminal domain. Moreover, a homology model of EiAPR could be produced using 2DJ3 as a template and used for chemical shift mapping. The data presented in this study provide an excellent starting point for mechanistic studies of this enzyme, from which the significance of the two domain structure could be determined.

Table 2.1 Backbone HN, N and C α assignments and side-chain C β assignments of EiAPR

Residue	HN	N	C α	C β	Residue	HN	N	C α	C β
1 Val		122.23	62.03	32.86	25 Asp	7.87	119.84	51.17	41.74
2 Asp	8.42	124.15	54.37	41.25	26 Lys	8.70	121.78	60.39	32.78
3 Ser	8.17	116.62	58.47	63.66	27 Ala	8.26	118.35	54.85	18.00
4 Ala	8.32	125.82	52.89	18.93	28 Thr	8.19	115.90	66.84	67.38
5 Glu	8.18	119.60	56.75	29.93	29 Leu	8.38	123.17	58.59	40.44
6 Ala	8.08	124.11	52.85	19.03	30 Glu	8.77	117.62	59.54	29.57
7 Glu	8.17	119.62	56.68	30.04	31 Lys	7.60	121.47	59.38	32.42
8 Ala	8.09	124.85	52.50	18.86	32 Leu	8.25	118.54	57.03	43.37
9 Lys	8.10	120.81	56.12	33.00	33 Ala	9.97	126.39	56.40	17.73
10 Ala	8.16	125.31	52.33	19.13	34 Gly	7.79	101.74	45.70	
11 Glu	8.21	120.24	55.93	30.36	35 Gly	7.39	103.17	44.13	
12 Ala	8.22	127.02	50.12	17.97	36 Asp	7.56	123.26	52.87	39.20
13 Pro			62.55	31.98	37 Arg	7.47	120.90	55.33	32.03
14 Asp	8.20	118.62	52.54	41.88	38 Asp	9.08	123.51	54.91	38.68
15 Leu	8.01	120.33	52.50	44.37	39 Lys	7.64	115.72	54.22	34.63
16 Trp	9.58	116.44	60.45	24.75	40 Asp	8.68	124.21	55.75	41.26
17 Thr	8.42	113.38	61.70	69.33	41 Thr	7.09	116.25	61.54	72.82
18 Gly	8.71	111.69	45.98		42 Met	9.27	129.79	54.03	37.08
19 Gly	8.05	109.65	44.78		43 Val	9.31	124.52	60.21	35.26
20 Ala	10.11	127.88	53.04	19.42	44 Val	7.78	124.38	58.24	33.19
21 Val	8.09	120.21	62.35	32.37	45 Leu	9.48	130.55	53.78	41.29
22 Glu	8.69	130.11	56.35	31.32	46 Tyr	8.37	120.40	54.83	43.77
23 Ala	8.93	128.36	51.79	19.30	47 Ala	6.98	119.19	48.43	19.29
24 Leu	8.15	122.23	53.88	43.71	48 Pro			65.06	32.11

Table 2.1 continued

49 Trp	5.93	110.57	53.10	29.61	76 Lys	8.84	118.78	54.10	35.64
50 Cys	6.35	127.90	56.91	29.23	77 Tyr	8.98	118.90	58.23	41.96
51 Pro			64.19	31.86	78 Asn	8.08	125.90	52.10	35.35
52 Phe	9.74	127.48	60.50	38.70	79 Ala	8.32	131.00	51.75	21.53
53 Cys	9.76	130.28	64.57	29.21	80 Asp	7.96	119.14	57.18	42.59
54 Gln	8.38	117.39	58.79	27.10	81 Ala	7.56	117.46	51.17	18.46
55 Gly	7.91	106.13	46.38		82 Asp			54.60	40.48
56 Leu	7.36	120.16	54.94	42.90	83 Arg	8.08	120.55	54.76	38.36
57 Glu	7.20	118.31	60.46	27.05	84 Glu	8.68	120.46	60.54	28.75
58 Pro			65.85	30.63	85 Tyr	8.96	121.10	61.09	37.80
59 Asp	7.65	120.12	56.93	40.70	86 Ser	7.84	115.90	63.47	62.43
60 Tyr	8.44	120.73	62.23	38.69	87 Glu	8.70	123.62	59.43	28.84
61 Glu	8.09	118.41	58.73	29.41	88 Ser	7.74	119.45	61.29	62.34
62 Ala	7.59	120.80	55.18	17.84	89 Leu	6.90	120.51	54.56	42.27
63 Val	7.60	119.11	66.40	30.72	90 Gly	7.27	102.08	43.99	
64 Ala	7.68	121.82	55.36	18.56	91 Leu	7.49	120.85	52.85	41.43
65 Gln	8.30	116.56	58.54	28.39	92 Lys	8.80	124.92	60.22	30.01
66 Glu	8.06	117.14	58.01	30.55	93 Thr			58.90	70.71
67 Ala	8.70	119.86	52.03	20.07	94 Phe	8.29	119.95	55.03	40.85
68 Gly	7.56	105.88	45.17		95 Pro			62.33	34.63
69 Gly	8.57	109.86	45.66		96 Thr	8.29	115.42	62.47	72.32
70 Ala			54.58	18.31	97 Ile	9.34	129.96	60.43	39.64
71 Gly	8.34	102.71	44.99		98 Ile	9.07	118.97	58.73	42.16
72 Leu	7.24	121.71	54.72	43.93	99 Phe	9.99	123.83	56.77	44.11
73 Arg	8.82	128.75	54.56	30.06	100 Leu	8.80	127.00	51.09	41.03
74 Val	8.91	129.92	63.13	31.17	101 Pro			61.79	32.21
75 Ala	9.30	127.97	50.44	24.49	102 Lys	9.69	125.73	58.10	32.48

Table 2.1 continued

103 Gly	8.82	112.30	45.42		119 Ser	8.33	116.24	61.14	63.03
104 Ser	7.56	113.31	56.56	65.24	120 Met	8.57	121.71	60.55	35.76
105 Asp	7.88	115.12	52.26	40.08	121 Thr	8.36	112.56	66.64	68.43
106 Lys	8.13	119.65	57.02	33.76	122 Met	8.40	121.72	58.90	32.44
107 Val	8.62	123.42	62.18	33.77	123 Trp	7.73	122.53	60.87	28.64
108 Val	9.06	128.5	56.02	34.56	124 Ala	7.95	118.95	54.39	17.67
109 Lys	8.77	128.16	57.93	38.16	125 Ala	8.12	119.86	54.51	17.86
110 Phe			52.31	40.90	126 Ala	7.54	120.27	54.13	17.91
111 Glu	8.28	120.45	56.03	30.37	127 Leu	7.03	117.60	55.36	41.47
112 Ser	6.13	112.93	57.08	63.43	128 Gly	7.47	106.15	45.26	
115 Arg			54.96	29.97	129 Ala	8.22	124.41	52.40	18.77
116 Thr	7.79	112.53	58.54	72.04	130 Arg	8.14	120.26	55.55	30.54
117 Val	9.04	122.82	67.62	31.59	131 Glu	7.92	126.03	57.84	30.93
118 Glu	8.77	120.31	59.91	29.54					

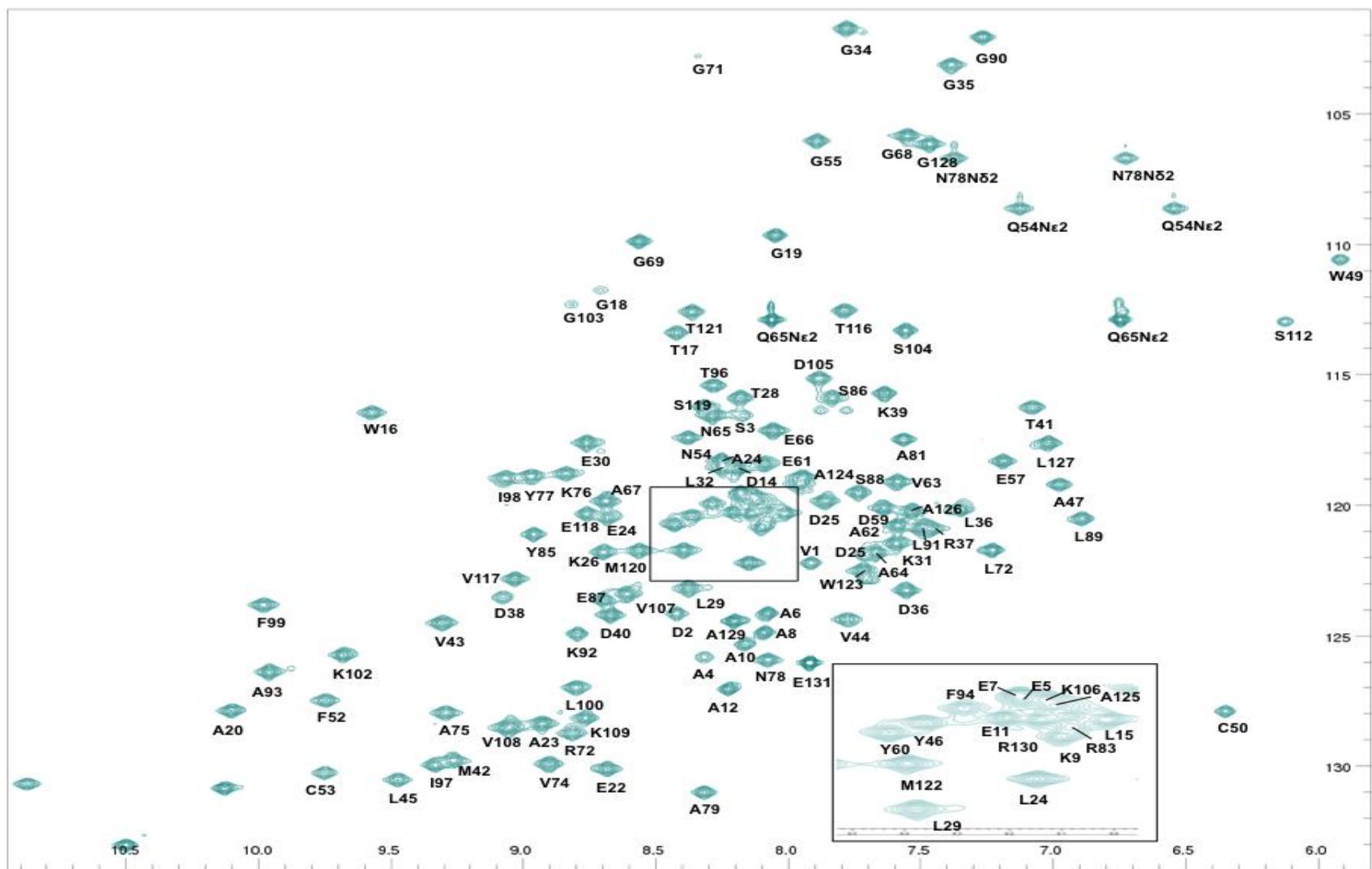


Figure 2.2 Assigned HSQC spectrum of EiAPR. Labels indicate the assignment at 300 K, 30 mM Tris-HCl pH 8.0

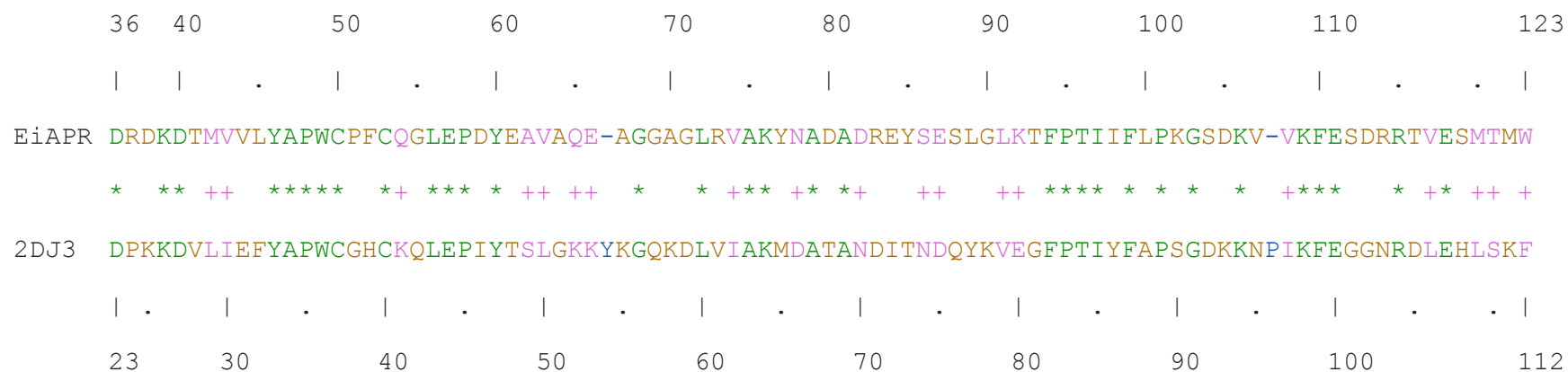


Figure 2.3 Alignment of EiAPR and the third thioredoxin domain of Protein disulphide isomerase A4 from *Mus musculus* showing identities (green) are marked with a '*', similarities (pink) are marked with a '+', mismatches (orange) and deletions (blue). The alignment contains 32/90 identities and 51/97 similarities.

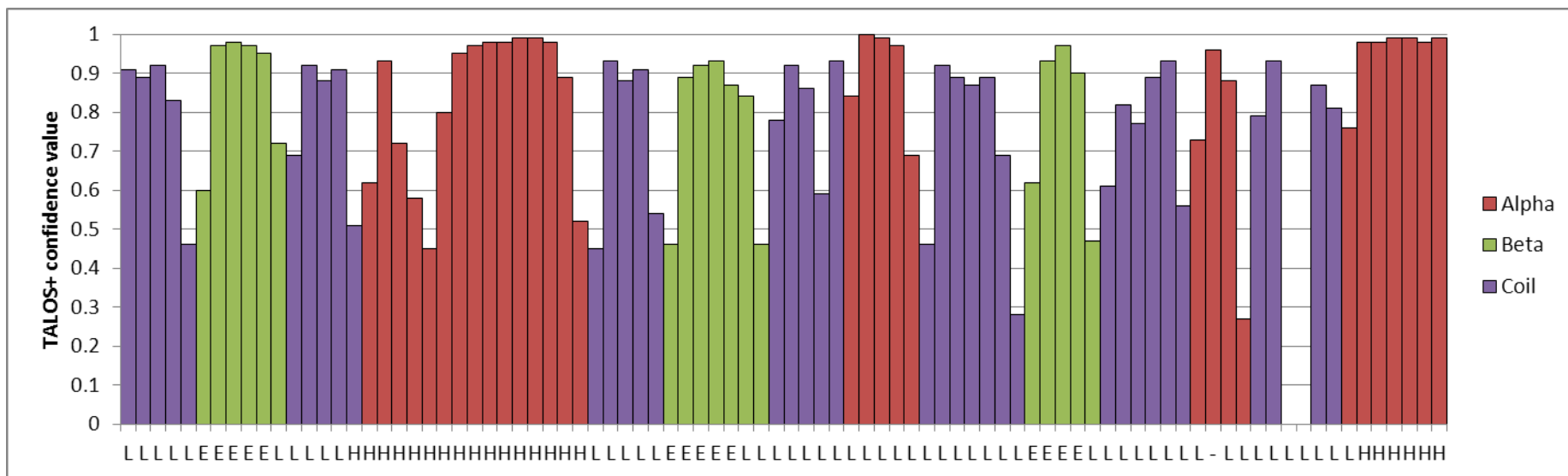


Figure 2.4 Comparison of predicted secondary structure and prediction confidence values from TALOS+ and the structure of 2DJ3. Residues predicted as alpha helix are shown in red, those predicted as beta sheet are shown in green and those predicted as random coil are shown in purple. The secondary structure of 2DJ3 is plotted on the x axis, H indicates helix, E indicates beta sheet and L indicates random coil.

Chapter 3

***Backbone and side-chain methyl assignments
of cytochrome P450cam with different ligands
bound***

Abstract

^{15}N - ^{13}C - ^2H samples of the C334A mutant of P450cam from *Pseudomonas putida* were produced by heterologous expression in *Escherichia coli* grown on minimal medium enriched with $^{15}\text{NH}_4\text{Cl}$ and ^{13}C - ^2H -glucose in D_2O . CHD₂-methyl labelled samples of P450cam were produced using the same techniques by substituting ^{13}C - ^2H -glucose with ^{13}C - ^1H -glucose. These samples were produced with various ligands bound and backbone HN, N, C α and CO and side-chain C β NMR assignments were obtained for substrate-free and ligand bound forms of this enzyme. Moreover, side-chain methyl assignments were obtained for the CHD₂ methyl labelled samples of P450cam.

Introduction

In order to characterise a protein by NMR it is essential to obtain resonance assignments for the backbone of the entire protein or the residues of interest. Advances in protein expression techniques in the 1990s (177) have made it relatively facile to produce isotopically enriched protein samples, which can be used for NMR studies. Moreover, a number of 3D heteronuclear experiments, developed in the last 20 years (for review see (11)), have enabled efficient resonance assignment of the backbone of proteins and thereby, identification of ^{15}N - ^1H resonance pairs in HSQC spectra. These spectra are an excellent fingerprint of protein structure. This fingerprint can be exploited to study protein-ligand binding and protein dynamics, depending on the timescale of the motions. Larger proteins have high relaxation rates, which result in broad lines in the NMR spectra. The use of perdeuteration of protein samples can reduce the relaxation rates of hydrogen nuclei (178) and when used in combination with TROSY-based experiments, spectra with sharp lines and thus better peak dispersion can be acquired (14). BEST experiments (17) are extremely useful when protein samples are unstable, since these types of experiments make it possible to obtain high resolution NMR spectra in a

fraction of the time required for traditional data acquisition. Side chain methyl groups can also be useful probes for the dynamic behaviour of proteins, and these can be assigned using CHD₂-selective-¹³C-TOCSY experiment (179), using C α and C β assignments for these nuclei. CHD₂ methyl groups can be introduced into protein samples by over-expression of proteins in minimal media containing ¹³C-¹H glucose and 100% D₂O.

Cytochrome P450cam has a molecular mass of 46.6 kDa and exists in at least three distinct states, open, closed and intermediate (141, 180). In order to probe these states on the atomic level, it is necessary to obtain NMR data for the protein in these states. In addition, the haem group of P450cam can exist in different spin and coordination states (181). In the substrate free form, P450cam exists in a six-coordinate 92% low spin state, with a water molecule coordinated to the haem. This state is rather unstable in solution and therefore, difficult to probe using NMR experiments. When camphor binds to the protein, the water molecule is displaced, creating a five-coordinate high-spin haem and the protein exists principally in the closed form (140, 141, 146, 180). Many nuclei are unobservable in NMR spectra of the camphor bound form of P450cam, due to the fact that the high-spin haem causes paramagnetic relaxation of these nuclei. This can be circumvented by binding cyanide to the haem, since camphor and cyanide can co-exist in the active site of the enzyme (182) and cyanide acts as a weak sixth ligand, creating a low spin iron state. P450cam has also been crystallised in the closed form with 1-phenylimidazole bound, and the imidazole ring coordinates to the haem, also keeping it in a low-spin state (183).

Backbone HN, N, C α and CO assignments along with side-chain C β have been obtained in the cyanide/camphor bound state and the 1-phenylimidazole bound state using TROSY and BEST-TROSY triple resonance experiments, respectively. Side-chain methyl assignments for the 1-phenylimidazole form were also obtained using a CHD₂-selective TOCSY experiment. These NMR

data can be used to study the dynamic behaviour of P450cam in solution and complex formation between putidaredoxin and P450cam.

Methods and Materials

^{15}N ^{13}C ^2H enriched samples of P450cam were kindly supplied by Klaas de Lint (Leiden University).

NMR experiments on the camphor bound form

The NMR experiments for backbone assignment were performed with a sample of concentration 530 μM in 20 mM HEPES pH 7.5, 100 mM NaCN, 1 mM camphor and 8%/92% $\text{D}_2\text{O}/\text{H}_2\text{O}$ at 298 K. ^1H , ^{13}C and ^{15}N sequential resonance assignments were obtained using 3D TROSY-HNCACB, 3D TROSY-HNcaCO and 3D TROSY-HNCO experiments acquired on a Bruker Avance III spectrometer equipped with a TCI-Z-GRAD cryoprobe operating a Larmor frequency of 950 MHz, along with a 3D TROSY-HNcoCACB acquired on a Bruker Avance I spectrometer equipped with a TCI-Z-GRAD cryoprobe operating a Larmor frequency of 800 MHz. All experiments were recorded with the assistance of Dr Frank Löhr at the University of Frankfurt, Germany.

NMR experiments on the 1-phenylimidazole bound form: The NMR experiments for backbone assignment were performed with a sample of concentration 530 μM in 20 mM HEPES pH 7.5, 100 mM 1-phenylimidazole and 8%/92% $\text{D}_2\text{O}/\text{H}_2\text{O}$ at 298 K. ^1H , ^{13}C and ^{15}N sequential resonance assignments were obtained using 3D BEST-TROSY-HNCACB, 3D BEST-TROSY-HNcaCO, 3D BEST-TROSY-HNcoCACB and 3D BEST-TROSY-HNCO experiments acquired on a Bruker Avance III spectrometer equipped with a TCI-Z-GRAD cryoprobe operating a Larmor frequency of 950 MHz. All experiments were recorded with the assistance of Dr Frank Löhr at the University of Frankfurt, Germany. Side-chain methyl resonance assignments were obtained using CHD_2 -selective- ^{13}C -TOCSY and CHD_2 -CT- ^{13}C HSQC experiments acquired on a Bruker Avance I spectrometer equipped with a TCI-Z-GRAD cryoprobe operating a ^1H Larmor frequency of 800 MHz.

Assignment of camphor-free P450cam: A nine-point titration was performed on a sample of P450cam with CN bound in order to obtain assignments for the camphor-free form. The NMR sample contained 340 μM protein in 20 mM HEPES pH 7.5, 100 mM KCN and 8%/92% $\text{D}_2\text{O}/\text{H}_2\text{O}$ at 298 K. ^{15}N -TROSY-HSQC experiments were performed on a Bruker DMX 600 spectrometer equipped with a TCI-Z-GRAD cryoprobe. Nine aliquots of camphor were added to the protein sample, to obtain camphor concentrations of 68, 119, 170, 221, 272, 340, 408, 510 and 680 μM in the protein sample, respectively. The final titration point containing 680 μM camphor was used to transfer assignments from the CN/camphor bound form to the camphor-free/CN bound form. The spectral widths and numbers of complex points used for each experiment are shown in Table 3.1. All experiments were acquired with 1024 complex points in the direct dimension. All data were processed in Topspin 2.1 (Bruker Biospin) or nmrPipe (174) and spectral assignment and analysis were carried out using CCPN analysis 2.1.5 (175).

Table 3.1 Numbers of complex points and spectral width in the indirect dimensions used to acquire the 3D spectra used for the assignments of P450cam.

Experiment	Complex points	Spectral width (kHz) ^{15}N	Complex points	Spectral width (kHz) ^{13}C
TROSY				
HSQC	256	4.05		
HNcoCACB	128	3.00	76	12.82
HNCACB	160	3.57	128	15.38
HNcaCO	160	3.57	64	3.33
HNCO	160	3.57	64	3.33
BEST-TROSY				
HSQC	256	4.05		
HNcoCACB	128	3.00	64	12.82
HNCACB	128	3.57	100	15.38
HNcaCO	128	3.57	40	3.33
HNCO	128	3.57	40	3.33
CHD₂				
HSQC	256	7.35		
CC-TOCSY	114	4.03	148	12.97

Results and Discussion

Spin systems were identified from intra-residual C α and C β nuclei in the HNCACB and sequence specific connectivities were obtained from inter-residual connectivities in the HNCACB and HNcoCACB experiments along with the HNcaCO and HNCO experiments. The camphor/CN bound form of P450cam was assigned to a level of 74% for the backbone amides (Figure 3.1) and 81% for the backbone C α and CO nuclei and side-chain C β nuclei (Table 3.2). A complete assignment could not be obtained due to the fact that some resonances could not be found in the spectra, possibly due to paramagnetic broadening: 43 missing assignments were within 20 Å of the haem iron and 26 of these were within 16 Å of the haem iron. Moreover, these missing assignments created difficulty in obtaining sequential connectivities of sufficient length, such that the resonances could not be unequivocally assigned to a single part of the protein sequence. Finally, 14 prolines could not be assigned in the HNCACB/HNcoCACB spectra and some short stretches of residues are flanked by prolines, making their unequivocal assignment difficult. The CN/camphor bound assignments were used to obtain assignments for 55% of the backbone amides from the titration experiments (Figure 3.2, Table 3.3). The lower level assignment could be due, in part, to reduced peak dispersion owing to the lower field at which the spectra were recorded. The spectrum of CN bound P450cam is shifted when compared to that of the camphor/CN bound form. The effects may be pseudocontact shifts caused by a change in the spin state of the haem or chemical shift perturbations due to a change in protein structure, such as open-to-closed. The $\Delta\delta_{avg}$ values (184) of the spectral changes observed on camphor binding were plotted onto the structures of the open and closed forms of P450cam (Figure 3.5). Many significant CSPs were observed on the B', F, G and I helices and the E/F loop. These CSPs may be indicative of the protein closing when camphor binds. Additional CSPs were observed in the vicinity of the haem, which may be due to a change in the spin state of the haem. Moreover, in accordance with previously reported data (185), camphor

binding occurred in slow exchange on the NMR timescale, since two distinct states were observed for many peaks, rather than a population averaged chemical shift, which would be observed if camphor bound were in the fast exchange regime. The 1-phenylimidazole form was assigned to a level of 66% for the backbone amides (Figure 3.3), 75% for the backbone C α and side-chain C β nuclei and 73% of the backbone carbonyl groups (Table 3.4). The same assignment difficulties encountered in assigning this form as in assigning CN/camphor bound form, preventing complete assignment of the protein. These assignments were used to assign 44% of the side chain methyl groups of Ala, Thr, Ile, Val and Leu (Figure 3.4, Table 3.5). The spectra of the 1-phenylimidazole form are significantly different from those of the camphor/CN bound form. The structures of camphor/CN bound P450cam and 1-phenylimidazole bound P450cam are almost identical (RMSD =0.3 Å), therefore, these changes can only be attributed to a change in size and orientation of the magnetic susceptibility tensor of the haem. Due to the fact that CN and 1-phenylimidazole cause the haem to be low spin, it is evident that these two ligands induce two different low spin forms. Since CN binds weakly and 1-phenylimidazole binds strongly, the existence of two different low spin haem tensors is highly plausible.

The assignments obtained can be used as probes for the dynamic behaviour of P450cam since they cover the B', F, G and I helices and the F/G loop. These structural motifs have been heavily implicated in the process of P450cam opening and closing and these assignments in combination with selective labelling of backbone amides can be used to reveal the timescale on which these motions occur and provide experimental insight into how these motions occur at the atomic level (Chapter 5).

Table 3.2 Backbone HN, N, C α , CO assignments and side-chain C β assignments for the CN/camphor bound form of P450cam.

Residue	HN	N	CO	C α	C β	Residue	HN	N	CO	C α	C β		
4	Glu	8.41	124.03	176.62	56.34	29.49	62	Gly	7.46	112.03	173.01	43.74	
5	Thr	8.17	116.59	174.54	61.67	69.27	63	His	7.09	121.65	171.77	55.04	28.5
6	Ile	8.12	124.32	176.28	60.76	37.69	64	Trp	8.25	118.32	180.03	56.04	32.13
7	Gln	8.38	125.16	176.03	55.35	28.67	67	Thr			173.51	61.63	68.12
8	Ser	8.28	117.92	174.39	58	63.03	68	Arg	6.23	114.54	177.28	51.45	32.78
9	Asn	8.34	121.32	175.05	52.78	38.29	69	Gly	10.14	111.88	174.76	46.71	
12	Leu	10.25	125.46		62.09	41.01	70	Gln	8.75	119.56	177.8	58.67	27.63
13	Ala	7.62	128.32	176.8	50.25	20.53	71	Leu	6.24	116.69	177.62	58.64	42.63
17	Pro			177.06	64.12	31.2	72	Ile	7.44	120.47	176.96	65.49	37.93
18	His	7.64	111.53	175.45	55.85	30.16	73	Arg	7.92	115.46	178.39	59.66	29.59
19	Val	6.8	127.38	172.76	59.86	31.18	74	Glu	7.42	117.09	180.38	59.13	29.71
20	Pro			178.45	61.26	31.65	75	Ala	8.04	121.04	179.78	54.16	17.26
21	Glu	8.6	121.18	178.74	59.2	28.68	76	Tyr	7.62	115.09	178.29	60.93	36.88
22	His	7.85	113.86	176.14	57.81	28.64	77	Glu	7.75	116.57	177.52	58.13	29.42
23	Leu	7.17	119.39	175.73	53.51	40.31	78	Asp	7.53	119.98	175.6	51.53	38.57
24	Val	7.08	120.57	177.09	62.88	30.78	79	Tyr	7.31	121.64	175.99	57.23	36.45
25	Phe	9.17	133.38	174.67	57.76	39.79	80	Arg	7.21	121.14	178.54	57.9	27.83
26	Asp	8.14	126.19	174.97	55.25	39.52	81	His	6.78	112.99	172.65	58.61	28.44
27	Phe	8.25	124.99	175.24	57.51	40.86	82	Phe	7.85	119.06	174.53	55.3	36.5
28	Asp	8.14	133.18	174.89	51.4	40.21	83	Ser	8.85	117.39	176.92	56.43	63.35
30	Tyr	7.93	118.22	175.71	59.21	38.16	84	Ser	9.68	126.59	172.41	59.22	61.11
31	Asn	6.83	116.5	169.41	51.63	38.4	85	Glu	8.05	119.35	175.13	58.5	29.02
32	Pro			178.58	62.5	30.84	90	Pro			176.42	61.56	34.96
33	Ser	7.86	118.15	175.23	59.86	62.39	91	Arg	9.42	124.2	177.13	60.16	29.85
34	Asn	8.75	118.06	175.34	53.23	37.2	92	Glu	9.41	115.99	179.65	59.61	27.81
35	Leu	7.27	120.04	178.4	57.14	41.29	93	Ala	7.36	123.36	178.09	53.85	17.08
36	Ser	8.55	114.47	174.82	60.29	61.97	94	Gly	7.47	104.45	176.11	48.79	
37	Ala	7.77	124.46	176.96	52.47	18.35	95	Glu	9.05	120.83	178.75	58.79	28.54
38	Gly	7.4	109.58	175.91	42.77		96	Ala	7.01	120.46	177.06	52.41	18.35
39	Val			176.66	66.71	31.33	97	Tyr	8.33	117.56	174.45	52.96	37.84
40	Gln	7.61	122.31	177.68	61.77	24.38	98	Asp	7.73	122.08	176.07	53	40.76
41	Glu	9.63	120.48	179.44	60.82	27.25	99	Phe	8.68	129.86	176.33	50.04	
42	Ala	8.51	122.97	181.09	54.46	17.06	101	Pro			178.62	62.61	34.29
43	Trp	8.35	120.29	177.71	60.56	27.37	102	Thr	7.24	115.83	173.29	66.23	67.77
44	Ala	8.21	120.3	179.71	53.67	16.61	103	Ser	8.07	111.05	172.44	58.74	63.48
45	Val	7.03	119.57	177.48	64.23	30.11	104	Met	7.57	119.52	174.15	54.24	35.22
46	Leu	7.47	119.25	174.63	55.41	40.63	105	Asp	8.69	119.8	175.43	51.84	41.46
47	Gln	7.4	112.05	176.58	53.41	27.7	107	Pro			178.86	64.91	31.18
48	Glu	6.76	121.11	178.06	56.9	29.58	108	Glu	8.98	117.22	179.65	59.59	28.02
49	Ser	8.59	115.67	174.68	60.93	62.63	109	Gln	7.49	119.61	177.99	61.7	30.68
50	Asn	8.27	116.52	174.64	53.1	37.25	110	Arg	7.54	117.29	178.06	60.54	29.61
51	Val	7.68	125.9	175.17	60.53	31.61	111	Gln	8.3	116.76	178.05	58.08	27.11
56	Trp	8.86	122.93	175.24	55.05	31.1	112	Phe	7.03	117.98	177.82	61.36	39.21
57	Thr	7.86	120.05	171.9	55.16	69.48	113	Arg	8.65	121.24	177.2	59.99	29.68
58	Arg	7.92	122.62	176.82	55.52	29.83	114	Ala	7.79	120.09	180.12	54.79	17.12
59	Cys	8.02	119.18	174.66	58.69	28.66	115	Leu	6.99	119.98	178.13	57.19	40.63
60	Asn	9.3	114.98	174.54	55.39	37.67	116	Ala	7.91	120.03	179.86	54.57	16.72
61	Gly	8.84	103.98	174.48	44.91		117	Asn	8.72	116.47	177.38	55.81	38.69

Table 3.2 continued

Residue	HN	N	CO	Ca	C β	Residue	HN	N	CO	Ca	C β		
118	Gln	7.58	116.46	176.84	57.85	28.02	182	Thr	6.91	115.9	177.47	62.62	61.15
119	Val	7.16	108.9	175.63	61.64	31.59	183	Asp	8.21	122.78	178.11	56.54	39.56
120	Val			178.51	63.42	30.64	184	Gln	7.32	114.27	177.39	56.01	26.89
121	Gly	8.66	113.05	174.17	45.00		185	Met	7.68	115.8	176.77	58.78	33.52
122	Met	8.19	118.44	175.26	54.05	31.03	186	Thr	7.82	107.12	176.1	62.77	69.69
123	Pro			179.43	65.61	30.24	187	Arg	8.37	122.89	172.76	53.98	31.24
124	Val	7.15	116.73	177.63	64.21	30.88	189	Asp			178.05	52.73	40.47
125	Val	7.28	121.55	178.12	66.59	29.8	190	Gly	8.74	110.49	175.48	44.62	
126	Asp	8.1	117.64	179.28	56.94	39.53	191	Ser	8.81	119.66	174.60	60.57	62.73
127	Lys	7.11	118.75	178.34	57.91	31.42	192	Met	8.35	121.1	176.38	55.11	36.84
128	Leu	7.65	119.42	177.49	54.3	42.58	193	Thr	8.91	116.83	174.64	60.9	70.32
129	Glu	7.38	121.21	177.31	60.52	28.69	194	Phe	8.64	119.33	179.89	55.04	37.77
130	Asn	8.51	116.38	177.43	56.32	36.98	196	Glu	7.64	117.92	179.58	58.54	29.43
131	Arg	7.76	121.42	178.84	57.43	28.4	197	Ala	8.00	125.44	178.38	54.91	17.2
132	Ile	8.24	120.57	177.1	65.68	37.17	198	Lys	8.39	119.19	177.51	58.92	29.9
133	Gln	8.01	118.39	177.99	59.21	27.43	199	Glu	7.91	117.26	178.87	58.87	28.17
134	Glu	8.27	118.99	179.74	59.04	29.33	202	Tyr	7.62	117.12	176.43	58.28	30.04
135	Leu	8.24	121.24	178.47	57.45	40.52	203	Asp	7.54	117.55	179.17	57.1	41.43
137	Cys			176.77	64.15	26.31	204	Tyr	9.02	120.57	177.24	57.13	31.16
138	Ser	8.15	115.27	178.23	61.1	62.18	205	Leu	7.5	119.13	177.63	61.31	37.22
139	Leu	8.43	122.67	180.27	57.58	41.81	206	Ile	7.36	116.72	178.88	57.42	40.97
140	Ile			178.05	65.94	37.09	207	Pro			179.52	64.93	30.2
141	Glu	8.48	119.68	178.51	57.83	27.76	208	Ile	6.6	120.01	177.99	64.01	37.5
142	Ser	7.42	112.57	176.27	60.95	62.38	209	Ile	8.28	120.92	177.39	64.93	36.47
143	Leu	7.35	119.84	178.96	56.31	42.05	210	Glu	7.68	117.16	179.82	58.84	28.16
145	Pro			177.3	64.48	30.67	211	Gln	7.71	118.89	179.2	58.59	27.66
146	Gln	8.34	113.95	178.57	57.36	28.89	212	Arg	8.07	118.53	177.16	57.25	28.01
147	Gly	6.48	101.71	170.6	44.21		213	Arg	7.59	118.82	177.76	57.63	29.93
148	Gln	5.87	111.23	174.54	53.71	29.17	214	Gln	7.39	118.03	177.01	57.19	28.19
149	Cys	8.27	110.54	171.86	56.45	30.14	215	Lys	7.87	119.17	171.77	52.93	32.18
150	Asn	9.13	119.79	176.18	51.63	36.75	216	Pro			178.81	63.43	31.93
151	Phe	9.38	124.03	176.78	62.85	39.59	217	Gly	8.46	111.1	173.62	43.29	
152	Thr	7.27	109.55	174.06	64.59	66.78	218	Thr	8.52	110.04	173.85	60.6	68.17
153	Glu	6.22	118.49	177.78	57.31	30.67	219	Asp	8.07	119.13	174.76	52.31	42.25
163	Ile			176.51	62.55	34.72	220	Ala	8.48	118.04	179.09	54.89	18.83
164	Phe	7.81	121.3	176.87	62.39	37.85	221	Ile	7.48	114.58	177.76	64.39	35.13
165	Met	8.9	116.48	178.86	57.53	29.2	225	Ala					18.07
166	Leu	7.3	120.61	180.13	57.25	40.38	225	Ala	8.89	116.3	176.54	53.73	
167	Leu	7.69	123.17	178.31	57.19	40.73	226	Asn	7.14	112.73	174.79	52.48	39.74
168	Ala	8.25	118.83	174.99	52.16	17.54	227	Gly	7.59	109.86	171.77	44.87	
169	Gly	7.62	108.46	175.33	46.88		228	Gln	8.08	118.12	175.2	54	30.96
170	Leu	8.29	122.36	173.78	50.88	41.48	229	Val	8.88	117.47	174.9	59.34	33.53
171	Pro			178.31	62.14	31.54	230	Asn	8.89	121.52	175.78	53.64	37.4
172	Glu	8.7	123.08	179.65	58.99	28.51	231	Gly	8.57	103.37	173.18	45.36	
173	Glu	9.38	119.81	177.1	58.59	27.82	232	Arg	7.95	119.11	172.9	52.16	29.39
174	Asp	7.55	119.03	176.09	54.77	39.75	234	Ile	8.22	127.02	175.84	61.61	38.69
175	Ile	7.56	119.2	173.64	67.53		235	Thr	9.03	119.88	176.14	60.62	70.61
180	Tyr			173.96	62.47	31.9	236	Ser	9.03	118.32		62.61	61.1
181	Leu	7.98	117.63	174.55	55.6	40.23							

Table 3.2 continued

Residue	HN	N	CO	Ca	C β	Residue	HN	N	CO	Ca	C β		
238	Glu		179.04	58.58	29.2	300	Arg	8.39	117.48	173.98	51.99	36.17	
239	Ala	8.58	119.43	179.81	54.67	17.1	301	Ile	9.4	121.39	172.82	58.59	41.3
240	Lys	8.09	121.8	176.9	59.66	30.29	302	Leu	8.7	131.18	181.34	51.65	39.27
241	Arg	7.46	119.09	179.8	58.35	40.6	303	Thr	8.58	116.39	174.1	61.86	69.03
242	Met	8.13	119.94	178.8	57.02	33.29	304	Ser	7.54	112.8	171.46	55.98	64.91
243	Cys			176.75	64.8	26.66	305	Asp	8.46	121.11	177.22	54.56	38.75
244	Gly	8.77	105.68	171.93	47.47		306	Tyr			172.04	56.05	40.77
245	Leu	6.17	121.39	178.56	55.86	39.57	307	Glu	7.96	128.48	172.75	54.44	28.91
250	Gly			176.13	44.82		308	Phe	8.43	129.65	173.62	54.08	41.73
251	Leu	7.65	119.1	176.88	56.23	41.78	309	His	9.03	123.55	175.1	56.58	25.65
254	Val			177.14	67.64	31.14	310	Gly	8.34	102.64	173.5	44.71	
255	Val	7.03	118.59	178.46	64.29	32.54	311	Val	7.43	123.8	174.7	60.85	32.25
256	Asn	7.9	115.37	176.62	56.8	40.49	312	Gln	8.37	125.83	173.58	55.57	27.27
257	Phe	9.09	121.61	178.56	63.86	39.91	313	Leu	8.57	128.31	177.42	53.2	40.53
258	Leu	9.02	118.72	179.12	58.02	41.45	314	Lys	8.52	124.75	175.54	53.08	33.44
259	Ser	7.3	115.22		63.04	62.03	315	Lys	8.69	121.77	177.55	58.33	31.66
260	Phe			177.85	57.07	39.74	316	Gly	8.54	117.26	174.25	44.65	
261	Ser	7	108.41	174.04	64.65	69.15	317	Asp	8.93	122.76	176.08	55.13	41.08
262	Met	6.46	118.53	175.21	56.98	32.58	318	Gln	9.42	118.94	174.95	55.34	32.62
266	Ala			178.53	53.53	18.29	319	Ile	8.79	119.16	172.01	57.33	41.75
267	Lys	7.26	116.12	176.65	56.53	33.14	322	Pro			175.55	63.12	30.67
268	Ser	7.16	114.61		53.76	60.84	323	Gln	9.64	134.89	178.4	61.95	23.45
269	Pro			178.99	63.87	33.93	324	Met	7.66	113.64	176.82	55.96	34.24
270	Glu	8.73	127.82	177.87	60	28.76	325	Leu	7.21	113.86	180.25	55.86	40.29
271	His	7.64	113.42	178.26	57.49	26.82	327	Gly			175.22	45.77	
272	Arg	7.41	117.82	179.48	60.48	29.16	328	Leu	6.72	119.59	173.56	53.22	39.92
274	Glu			178.46	58.06	31.63	329	Asp	6.65	118.69	178.68	53.24	42.23
275	Leu	7.88	119.6	178.84	54.52	43.04	330	Glu	9.43	131.88	177.09	56.85	27.68
276	Ile	7.61	121.57	180.92	63.73	36.1	331	Arg	8.67	118.82	177.86	57.34	29.03
277	Glu	7.63	118.46	177.76	66.38	30.82	332	Glu			176.42	56.54	30.9
278	Arg	8.89	116.29	178.87	57.17	28.98	333	Asn	7.79	116.49	171.59	51.84	42.47
279	Pro			179.77	64.2	30.5	334	Ala	8.5	124.18	178.34	51.97	17.85
280	Glu	10.02	122.02	177.66	57.82	26.44	335	Ala	9.19	122.86	176.63		16.28
281	Arg	8.55	120.92	176.38	56.05	29.54	336	Pro			179.87	63.69	32.42
282	Ile	7.83	119.98	174.78	67.69	34.95	337	Met	8.7	115.77	178.07	53.77	29.4
283	Pro			179.07	66.86	29.78	338	His	8.08	123.47	174.19	54.99	30.02
284	Ala	7.78	122.48	181.06	54.54	17.41	339	Val	8.1	126.88	174.9	61.38	29.45
285	Ala	8.55	122.81	179.05		17	340	Asp	10.15	131.4	176.7	51.35	41.54
288	Glu			174.42	63.84	31.16	341	Phe	9.57	124.42	177.21	55.49	35.18
289	Leu	5.71	117.08	172.46	53.52	39.49	342	Ser	8.71	116.16	174.01	57.54	62.73
290	Leu	5.9	121.57	174.58	57.07	40.72	343	Arg	7.21	124.27	177.15		30.46
291	Arg	7.45	120.61	177.44	58.47	28.38	344	Gln			176.3	57.39	28.44
292	Arg	8.18	119.35	175.12	56.44	28.44	345	Lys	8.35	120.71	174.7	54.96	31.98
293	Phe	8.36	115.14	171.44	55.86	35.95	346	Val	8.64	126.92	175.24	62.43	30.89
294	Ser	5.54	109.8	175.22	60.09	63.1	347	Ser	8.05	123.46	174.23	56.14	62.88
296	Val			172.46	61.21	30.9	348	His	8.19	121.68	173.64	55.59	31.68
297	Ala	8.68	127.45	172.53	51.73	21.48	349	Thr			177.81	59.39	68.49
298	Asp	7.44	119.81	173.16	50	37.18	350	Thr	11.75	129.05	173.46	66.38	67.12
299	Gly	7.51	102.46	172.7	44.96		351	Phe	8.07	114.71	172.28	57.13	40.96

Table 3.2 continued

Residue	HN	N	CO	Ca	C β	Residue	HN	N	CO	Ca	C β		
352	Gly	6.88	104.63	172.28	42.94	389	Gln	8.46	122.44	175.06	53.88	29.03	
353	His	7.80	118.18	172.81	56.66	33.26	390	Ile	8.60	126.57	176.07	58.96	34.53
354	Gly	8.03	114.57	175.85	43.87		391	Gln	9.59	129.68	174.85	54.15	30.45
355	Ser	9.39	120.09	176.75	61.48	62.60	392	His	8.91	125.03	174.94	52.59	34.44
356	His	8.52	119.45	171.43	55.4	27.12	393	Lys	8.95	119.54	174.56	54.82	35.48
357	Leu	6.92	119.24		61.69	42.76	394	Ser	8.94	117.74	173.76	55.24	65.35
359	Leu				59.23	32.14	395	Gly	7.06	110.76	171.82	45.76	
360	Gly	7.05	105.96	173.68	43.67		396	Ile	8.35	125.01	176.53	61.71	36.49
361	Gln	8.66	118.27		59.39	29.16	397	Val	6.34	113.55	177.49	60.87	34.36
361	Gln	7.50	118.72	178.26			398	Ser	8.65	126.08	172.49	61.27	63.82
362	His	7.01	115.51	177.63	55.54	40.07	399	Gly	8.93	102.46	171.65	45.02	
363	Leu	7.06	118.89	175.23	54.09	44.25	400	Val	7.07	118.17	176.51	60.71	33.06
370	Val	8.80	118.94	174.34	54.61	34.94	401	Gln	8.92	125.76	175.54	59.09	28.76
371	Thr	8.80	117.36	173.55	54.94	65.41	402	Ala	7.35	117.57	175.52	51.83	20.86
375	Trp			178.31	63.05	31.07	403	Leu	8.7	119.94	170.58		43.06
381	Asp	8.14	116.52	174.41	52.89	42.64	409	Pro			178.33	64.76	31.91
382	Phe	6.63	119.03	172.08	54.98	39.29	410	Ala	8.66	121.56	178.8	54.05	17.65
383	Ser			173.56	56.57	66.56	411	Thr	8.14	108.31	174.75	61.53	69.83
384	Ile	8.53	122.50	177.09	60.35	35.96	412	Thr	7.29	113.00	174.08	60.46	70.91
385	Ala	8.20	131.60	175.19	50.77	17.04	413	Lys	8.96	120.99	173.99	54.86	33.91
387	Gly			174.16		44.93	414	Ala	8.04	125.37	177.29	51.02	18.34
388	Ala	7.64	123.80	177.14	52.21	19.00	415	Val	7.97	127.04	181.17	63.66	32.67

Table 3.3 Backbone HN and N assignments of the KCN bound form of P450cam.

Residue	HN	N	Residue	HN	N	Residue	HN	N
4Glu	8.43	124.24	41Glu	9.72	120.81	84Ser	9.67	126.73
5Thr	8.20	117.09	45Val	7.02	119.89	85Glu	8.04	119.34
6Ile	8.16	124.71	47Gln	7.38	112.15	91Arg	9.44	124.38
8Ser	8.32	118.21	48Glu	6.69	121.09	92Glu	9.44	115.90
12Leu	10.25	125.64	49Ser	8.55	115.80	93Ala	7.37	123.74
13Ala	7.61	128.44	51Val	7.67	126.12	94Gly	7.45	104.56
18His	7.60	111.61	56Trp	8.87	122.99	95Glu	9.05	120.81
19Val	6.81	127.61	57Thr	7.86	120.16	96Ala	7.00	120.54
21Glu	8.62	121.27	58Arg	7.93	122.70	97Tyr	8.35	117.65
22His	7.84	114.01	60Asn	9.34	115.14	98Asp	7.76	122.22
23Leu	7.15	119.56	61Gly	8.88	104.11	99Phe	8.73	130.25
24Val	7.07	120.66	62Gly	7.47	112.22	102Thr	7.22	115.77
25Phe	9.17	133.60	63His	7.10	121.93	103Ser	8.03	111.28
26Asp	8.12	126.32	64Trp	8.28	118.47	108Glu	8.98	117.28
27Phe	8.26	125.12	68Arg	6.21	114.66	112Phe	6.98	118.24
28Asp	8.10	133.15	69Gly	10.13	111.97	115Leu	6.93	120.15
31Asn	6.79	116.51	71Leu	6.20	116.86	116Ala	7.91	120.17
33Ser	7.88	118.38	75Ala	8.06	121.04	118Gln	7.55	116.47
35Leu	7.22	120.06	77Glu	7.73	116.64	119Val	7.14	108.95
36Ser	8.60	114.76	78Asp	7.49	119.92	121Gly	8.67	113.30
37Ala	7.76	124.61	79Tyr	7.30	121.80	124Val	7.14	116.91
38Gly	7.38	109.90	80Arg	7.18	121.40	126Asp	8.10	117.74
40Gln	7.52	122.57	81His	6.75	113.04	127Lys	7.10	118.90

Table 3.3 continued

Residue	HN	N	Residue	HN	N	Residue	HN	N
129Glu	7.35	121.25	229Val	8.87	117.63	324Met	7.64	113.62
130Asn	8.53	116.52	230Asn	8.89	121.62	325Leu	7.19	113.88
131Arg	7.79	121.52	231Gly	8.58	103.49	328Leu	6.71	119.66
134Glu	8.29	119.15	232Arg	7.94	119.26	330Glu	9.44	132.03
138Ser	8.14	115.34	234Ile	8.22	127.18	331Arg	8.65	118.95
139Leu	8.41	122.82	235Thr	9.04	120.02	333Asn	7.78	116.67
142Ser	7.41	112.64	240Lys	8.05	121.80	334Ala	8.55	124.41
143Leu	7.32	119.91	241Arg	7.50	118.99	335Ala	9.22	123.13
146Gln	8.33	114.05	244Gly	8.78	105.96	337Met	8.72	115.89
147Gly	6.47	101.78	245Leu	6.14	121.60	339Val	8.07	127.03
148Gln	5.85	111.30	256Asn	7.91	115.52	340Asp	10.14	131.41
149Cys	8.23	110.54	257Phe	9.12	121.78	341Phe	9.63	124.63
150Asn	9.13	120.01	259Ser	7.26	115.14	343Arg	7.25	124.69
151Phe	9.30	123.89	261Ser	7.00	108.48	346Val	8.69	127.28
152Thr	7.27	109.51	262Met	6.44	118.71	352Gly	6.86	104.67
153Glu	6.23	118.62	267Lys	7.31	115.76	353His	7.79	118.38
165Met	8.80	116.11	270Glu	8.71	128.28	354Gly	8.05	114.70
169Gly	7.51	108.76	271His	7.61	113.13	357Leu	6.93	119.42
172Glu	8.70	123.18	276Ile	7.61	121.70	360Gly	7.02	105.96
173Glu	9.38	120.07	277Glu	7.62	118.68	361Gln	8.68	118.29
174Asp	7.56	119.15	280Glu	10.07	122.27	362His	7.00	115.55
181Leu	8.01	117.75	284Ala	7.81	122.77	363Leu	7.04	118.99
182Thr	6.82	115.97	289Leu	5.72	117.24	381Asp	8.15	116.59
183Asp	8.24	122.81	290Leu	5.86	121.64	382Phe	6.60	119.16
184Gln	7.30	114.43	292Arg	8.23	119.56	385Ala	8.20	131.74
185Met	7.63	116.29	293Phe	8.35	115.26	388Ala	7.63	123.98
187Arg	8.76	118.35	294Ser	5.52	109.82	389Gln	8.50	122.63
190Gly	8.76	110.75	298Asp	7.43	119.90	390Ile	8.63	126.82
193Thr	8.97	117.06	299Gly	7.49	102.52	391Gln	9.61	129.90
194Phe	8.81	123.87	301Ile	9.37	121.42	392His	8.96	125.38
196Glu	7.72	118.28	302Leu	8.75	131.44	393Lys	8.96	119.68
197Ala	8.00	125.59	303Thr	8.58	116.61	394Ser	8.97	117.82
198Lys	8.39	119.29	304Ser	7.54	112.97	395Gly	7.02	110.93
199Glu	7.76	117.25	305Asp	8.48	121.22	396Ile	8.42	125.51
202Tyr	7.57	117.34	307Glu	7.71	130.18	397Val	6.34	113.55
208Ile	6.53	120.04	308Phe	8.42	129.83	398Ser	8.66	126.29
211Gln	7.72	119.15	309His	8.95	123.76	399Gly	8.69	102.42
212Arg	8.03	118.57	310Gly	8.35	102.76	400Val	7.08	118.29
214Gln	7.39	118.04	311Val	7.43	124.04	401Gln	8.97	126.00
215Lys	7.87	119.29	312Gln	8.39	125.86	402Ala	7.31	117.31
217Gly	8.44	111.33	313Leu	8.57	128.37	410Ala	8.64	121.53
218Thr	8.55	110.02	314Lys	8.53	125.19	411Thr	8.13	108.37
220Ala	8.43	118.10	315Lys	8.70	121.95	412Thr	7.28	113.14
226Asn	7.13	112.88	316Gly	8.55	117.46	413Lys	8.97	121.11
227Gly	7.57	110.04	317Asp	8.93	122.83	414Ala	8.07	125.38
228Gln	8.08	118.40	318Gln	9.48	118.96	415Val	8.03	127.41

Table 3.4 Backbone HN, N, C α , CO assignments and side-chain C β assignments for the 1-phenylimidazole bound form of P450cam.

Residue	HN	N	CO	C α	C β	Residue	HN	N	CO	C α	C β		
2	Thr		174.55	61.72	69.11	61	Gly	7.56	111.33	172.95	43.72		
3	Glu	8.50	123.46	176.53	56.23	29.47	62	His	7.21	121.00	171.73	55.06	28.61
4	Thr	8.27	116.26	174.46	61.65	69.27	63	Trp	8.02	117.61	175.49	59.57	38.25
5	Ile	8.23	123.82	176.22	60.74	37.64	64	Ile	6.83	115.67		51.49	38.36
6	Gln	8.48	124.61	176.02	55.27	28.55	65	Ala	8.35	124.13	177.19	52.24	18.36
7	Ser	8.38	117.39		57.99	63.09	66	Thr	7.72	118.48	179.28	62.64	70.22
9	Ala	8.51	124.85	177.21	52.06	18.33	74	Ala			178.78	54.10	17.25
10	Asn	8.42	116.82	174.39	52.94	37.87	75	Tyr	7.70	114.60	178.02	60.81	36.89
11	Leu	7.84	121.41	176.00	52.92	40.63	76	Glu	7.77	116.13	177.21	58.01	29.24
12	Ala	8.81	129.48	176.24	50.01	16.76	77	Asp	7.56	119.17	175.47	51.40	38.55
16	Pro			176.98	64.03	31.10	78	Tyr			175.91	57.16	36.39
17	His	7.71	110.84	175.39	55.89	30.13	79	Arg	7.31	120.72	178.48	57.90	27.75
18	Val	6.89	126.80	172.65	59.83	31.14	80	His	6.87	112.39	172.69	58.66	28.45
19	Pro				61.30	31.59	81	Phe	7.89	118.19	174.43	55.28	36.58
20	Glu	8.70	120.58	178.67	59.17	28.35	82	Ser	8.98	116.60	177.13	56.71	63.21
21	His	7.92	113.23	176.08	57.83	28.63	83	Ser	10.00	126.59	172.45	59.89	61.47
22	Leu	7.24	118.80	175.65	53.53	40.32	84	Glu	8.26	118.23	175.01	58.49	28.93
23	Val	7.16	119.96	177.02	62.84	30.82	85	Cys	7.16	119.40		52.51	18.36
24	Phe	9.27	132.81	174.62	57.71	39.98	91	Glu	8.19	114.37	179.42	59.47	27.97
25	Asp	8.21	125.64	174.78	55.15	39.50	92	Ala	7.42	122.75	178.93	53.74	16.91
26	Phe	8.36	124.27	175.28	57.44	40.80	93	Gly	7.58	103.98	176.07	48.83	
27	Asp	8.25	132.40	175.19	51.34	40.13	101	Thr	7.53	109.54	177.74	62.36	68.14
33	Asn	7.09	112.79		53.11	37.21	103	Met	7.52	118.52	175.00	54.16	35.89
34	Leu	7.30	119.20	178.23	57.11	41.18	104	Asp	8.92	118.59	172.17	51.65	41.47
35	Ser	8.68	114.11	174.73	60.33	61.93	106	Pro			178.78	64.85	31.13
36	Ala	7.85	123.84	176.82	52.46	18.33	107	Glu	9.07	116.59	179.61	59.56	27.97
37	Gly	7.47	109.02	175.89	42.59	42.67	108	Gln	7.58	119.10	177.91	61.73	30.64
38	Val	7.55	119.75	176.89	65.48	37.84	109	Arg	7.61	116.71	177.98	65.99	29.59
39	Gln	8.04	114.98	178.28	59.59	29.46	110	Gln	8.40	116.11	177.95	57.96	27.00
40	Glu	7.50	116.50	180.35	59.05	29.61	111	Phe	7.10	117.09	177.53	61.18	39.27
41	Ala	8.13	120.36	179.82	54.14	17.25	112	Arg	8.56	121.09	176.94	59.90	29.45
42	Trp	8.47	119.64	177.55	60.51	27.21	113	Ala	7.74	119.29	180.13	54.71	17.04
43	Ala	8.34	119.81	177.60	53.66	16.67	114	Leu	7.07	119.21	178.13	57.14	40.52
44	Val	7.13	119.11	179.69	64.28	30.15	115	Ala	8.02	119.19	179.65	54.57	16.31
45	Leu	7.54	118.50	173.91	55.35	40.61	116	Asn	8.64	115.55	177.43	55.69	38.45
46	Gln	7.49	111.45	176.41	53.34	27.61	117	Gln	7.70	116.47	177.09	57.88	27.88
47	Glu	6.79	120.35	177.91	56.77	29.50	122	Pro			179.59	65.62	
48	Ser	8.65	115.09	174.63	60.95	62.65	123	Val	7.26	116.75	177.78	64.55	30.92
49	Asn	8.85	117.61		53.07	29.13	124	Val	7.48	120.93	178.15	66.51	29.92
50	Val	7.76	125.29	175.04	60.43	31.66	125	Asp	8.36	117.57	179.37	57.07	39.50
51	Pro			173.91	62.41	31.85	126	Lys	7.25	118.30	178.28	58.00	31.47
52	Asp	8.10	117.09	175.98	57.20	40.08	127	Leu	7.79	118.92	177.46	54.34	42.77
53	Leu	7.16	118.28	175.21	54.00	44.19	128	Glu	7.55	120.77	177.28	60.54	28.68
54	Val	8.51	112.97	174.15	57.85	35.20	129	Asn	8.68	115.91		56.34	36.89
55	Trp	8.97	122.37	175.12	55.06	31.21	139	Ile			177.92	65.86	37.12
57	Arg			174.66	55.58	29.91	140	Glu	8.59	119.11	178.36	57.81	27.67
58	Cys	8.12	118.52	174.65	58.72	28.67	141	Ser	7.50	111.91	176.21	60.95	62.44
59	Asn	9.44	114.44	174.46	55.44	37.67	142	Leu	7.39	119.34	179.92	56.27	40.45
60	Gly	8.96	103.40	174.46	44.83		143	Arg	8.41	118.70		61.53	26.95

Table 3.4 continued

Residue	HN	N	CO	Ca	C β	Residue	HN	N	CO	Ca	C β		
147	Gln		174.46	53.67	29.17	217	Thr	8.65	109.27	173.85	60.55	68.15	
148	Cys	8.32	109.82	171.61	56.44	30.17	218	Asp	8.16	118.49	174.68	52.34	42.22
149	Asn	9.23	119.24	176.05	51.68	36.65	219	Ala	8.49	117.34	179.15	54.90	18.64
150	Phe	9.38	123.12	176.80	63.01	39.69	220	Ile	7.62	113.81	177.65	64.57	35.04
151	Thr	7.29	107.77	176.82	64.04	66.22	221	Ser	8.77	118.71		62.21	62.96
160	Ile		177.47	56.39	36.93		222	Ile	7.66	120.75	178.78	63.65	35.97
161	Arg	7.91	120.65	178.90	57.44	28.39	223	Val	7.68	117.80	177.64	66.38	30.96
162	Ile	8.41	120.10	177.03	65.75	37.10	224	Ala	8.95	115.55	176.48	53.72	18.16
163	Phe	8.21	117.95	177.96	59.21	27.43	225	Asn	7.24	112.16	174.67	52.46	39.65
164	Met	8.42	118.55	179.65	59.03	29.30	226	Gly	7.68	109.24	171.65	44.87	
165	Leu	8.33	120.47	178.39	57.37	40.45	227	Gln	8.19	117.69	175.09	53.93	30.98
166	Leu	7.76	122.68	178.81	57.12	40.87	228	Val	8.97	116.97	174.80	59.25	33.53
167	Ala	8.29	117.58	174.67	51.86	17.88	229	Asn	9.00	121.22	175.71	53.57	37.41
168	Gly	7.65	108.36	175.19	47.01		230	Gly	8.64	102.70	173.07	45.27	
169	Leu	8.51	121.66	175.19	50.73	41.34	231	Arg	8.02	118.49	180.50	52.05	29.40
172	Glu	8.66	119.43	177.63	57.12	34.76	232	Pro				62.95	31.09
173	Asp	8.96	120.58	178.85	56.98	40.01	232	Pro			178.26	63.05	37.82
174	Ile	7.44	118.24	176.75	60.38	37.17	233	Ile	8.31	126.30	175.73	61.56	38.66
175	Pro				66.04	29.61	234	Thr	9.12	119.16	176.08	60.49	70.48
176	His	7.65	116.77	176.44	58.36	30.48	235	Ser	9.14	118.01	177.12	62.58	61.15
177	Leu	7.52	116.12	178.78	56.79	41.84	236	Asp	8.33	122.11	177.89	57.39	41.60
178	Lys	9.02	119.89	176.72	56.79	31.29	237	Glu	8.02	119.44	179.04	58.50	29.03
179	Tyr	7.39	117.86	177.59	61.21	37.22	238	Ala	8.64	119.02	179.93	54.65	17.19
180	Leu	7.25	115.59	179.17	57.29	41.32	239	Lys	8.23	121.49	176.77	59.64	30.27
181	Thr	8.22	110.24	178.06	66.36	67.40	256	Phe			177.58	62.66	40.06
182	Asp	8.07	123.91	177.71	56.13	39.10	257	Leu	9.09	115.61	178.36	57.55	41.64
183	Gln	7.22	113.90	177.17	55.74	26.61	258	Ser	6.87	114.39		62.77	61.73
184	Met			176.70	58.41	33.44	265	Ala			178.53	53.60	18.26
187	Pro			178.94	63.28	31.92	266	Lys	7.32	115.40	176.51	56.45	33.05
188	Asp	8.76	120.56	177.97	52.76	40.19	267	Ser	7.17	113.93		53.72	60.74
189	Gly	8.80	109.88	175.30	44.57		268	Pro			179.00	63.76	33.91
190	Ser	8.92	119.04	174.47	60.38	62.92	269	Glu	8.86	127.24	177.63	59.95	28.67
191	Met	8.48	120.86	174.46	55.05	36.89	270	His	7.72	113.58	178.13	57.68	26.76
192	Thr	9.02	116.34	177.13	60.81	70.20	271	Arg	7.45	115.92	179.24	60.34	28.82
193	Phe				62.06		274	Leu	6.89	118.83	173.52	53.17	39.98
194	Ala	8.80	118.89	179.75	55.10	17.39	275	Ile	6.77	118.13	178.57	53.18	42.14
195	Glu	7.74	117.36	179.60	58.50	29.37	276	Glu	9.53	131.28	176.94	56.77	27.64
196	Ala	8.02	124.53	178.38	54.88	17.03	277	Arg	8.75	118.25	177.78	57.27	28.99
197	Lys	8.35	117.88	177.51	59.00	30.07	278	Pro			179.70	64.17	30.47
198	Glu	8.19	116.47	178.75	58.80	28.18	279	Glu	10.13	121.46	177.59	57.85	26.46
199	Ala	7.73	120.83	180.97	54.36	17.97	280	Arg	8.65	120.45	176.24	56.03	29.54
206	Pro			177.84	64.02	37.59	281	Ile	7.94	119.33	174.79	67.70	34.91
207	Ile	8.30	120.34	177.26	65.06	36.54	282	Pro			179.06	66.91	29.70
208	Ile	7.82	116.58	179.79	58.82	28.01	283	Ala	7.95	122.33	180.98	54.55	17.44
209	Glu	7.82	118.42	179.11	58.51	27.57	284	Ala	8.69	122.21	179.01	54.61	16.95
210	Gln	8.12	117.84	176.99	57.11	28.08	285	Cys	8.78	115.77	175.12	62.31	25.55
211	Arg	7.97	118.55	171.66	52.86	32.16	286	Glu	7.45	119.36	177.84	59.23	28.96
215	Pro			178.76	63.39	31.90	287	Glu	7.74	118.91	179.40	56.90	28.03
216	Gly	8.52	110.52	171.29	43.26		291	Arg	7.64	112.45	176.55	55.88	34.15

Table 3.4 continued

Residue	HN	N	CO	Ca	C β	Residue	HN	N	CO	Ca	C β		
292	Phe	7.29	113.44	180.07	55.66	40.49	347	His	8.29	121.02	173.37	55.61	31.63
293	Ser	6.98	115.16	178.21	62.37	61.12	348	Thr	9.30	112.11		55.54	68.29
294	Leu	8.52	121.87	180.24	57.51	41.75	351	Gly			171.28	42.26	
295	Val	8.40	124.92	172.41	61.28	30.88	352	His	7.29	116.49	172.54	56.43	33.12
296	Ala	8.85	127.17	172.81	51.63	21.24	353	Gly	7.93	114.44	175.36	43.59	
297	Asp	7.85	119.60	173.05	50.33	38.01	354	Ser	9.34	119.22		61.09	62.08
298	Gly	8.06	102.64	172.66	45.13		358	Leu				59.18	32.31
299	Arg	8.68	117.77	174.01	52.10	36.29	359	Gly	7.25	105.77	173.56	43.79	
300	Ile	9.42	120.99	172.73	58.51	41.39	360	Gln	8.81	117.71		59.49	29.14
301	Leu			181.22	60.54	28.70	364	Arg	9.61	133.10	178.26	61.79	31.42
301	Leu	8.83	130.75		51.68	39.39	365	Arg	8.79	122.43	179.54	58.94	28.42
302	Thr	8.68	115.89	177.42	61.86	68.94	366	Glu	9.45	119.43	176.96	58.52	27.75
303	Ser	7.64	112.28	171.36	56.02	65.00	367	Ile	7.53	118.17	175.78	54.69	39.78
304	Asp	8.58	120.53	177.14	54.56	38.69	368	Ile	7.59	118.14	173.33	67.69	34.78
305	Tyr	8.45	125.50	171.93	56.11	40.86	379	Pro			176.75	63.99	31.87
306	Glu	7.78	129.43	172.89	54.36	28.94	380	Asp	8.23	115.75	174.34	52.82	42.56
307	Phe	8.52	129.03	173.43	54.01	41.80	381	Phe	6.67	118.36	172.00	54.92	39.29
308	His	9.03	123.10	175.13	56.68	25.86	382	Ser			173.45	56.61	66.66
309	Gly	8.43	102.02	173.44	44.64		383	Ile	8.59	121.80	176.97	60.21	35.87
310	Val	7.54	123.27	174.60	60.83	32.31	384	Ala	8.26	130.91	175.06	50.72	16.90
311	Gln	8.47	125.19	173.45	55.59	27.21	385	Pro			178.41	63.27	30.59
312	Leu	8.67	127.72	177.33	53.16	40.52	386	Gly	8.74	112.46	174.05	44.88	
313	Lys	8.61	124.13	175.47	53.17	33.47	387	Ala	7.71	123.18	177.01	52.16	18.93
314	Lys	8.80	121.22	177.47	58.33	31.65	388	Gln	8.54	121.97	174.89	53.81	28.91
315	Gly	8.65	116.66	174.21	44.66		389	Ile	8.71	126.05	175.90	58.55	34.25
316	Asp	9.04	122.10	176.02	55.10	41.15	390	Gln	9.68	129.05	174.80	53.89	30.64
317	Gln	9.59	118.12	174.91	55.20	32.83	391	His	9.16	124.89	178.58	52.57	34.32
318	Ile	8.92	118.31	172.19	57.35	41.89	392	Lys	9.03	119.02	174.40	54.74	34.80
319	Leu	9.08	130.40	174.30	54.19	42.55	393	Ser	8.96	117.21	174.40	55.04	65.29
320	Leu	8.60	130.60	173.00	50.09	40.96	394	Gly	7.10	109.21	171.99	45.17	
330	Arg			177.79	62.86	30.99	395	Ile	8.43	123.80		52.60	40.46
331	Glu	8.05	125.90	180.91		33.04	396	Val			176.91	61.15	33.89
331	Glu	7.82	115.87	176.33	56.70	30.72	397	Ser	8.11	125.04	171.93	60.85	64.17
332	Asn	7.86	115.87	171.49	51.80	42.34	398	Gly	8.76	102.10	170.66	44.52	
333	Ala	8.62	123.69	178.31	51.89	17.77	399	Val	8.82	119.77	175.88	60.80	33.46
334	Cys	9.29	122.29	176.50	51.97	16.16	400	Gln	9.04	124.47	175.48	59.35	28.47
337	His			174.20	55.07	30.24	401	Ala	7.39	116.36	170.28	51.78	20.76
338	Val	8.15	126.22	174.75	61.22	29.54	402	Leu	8.95	119.48	170.28	51.83	43.02
339	Asp	10.21	130.67	176.64	51.26	41.46	408	Pro			178.31	61.30	31.59
340	Phe	9.67	123.81	177.18	55.47	35.19	409	Ala	8.72	120.88	178.64	53.98	17.63
341	Ser	8.78	115.57	173.85	57.46	62.86	410	Thr	8.22	107.61	174.63	61.55	69.82
342	Arg	7.31	123.79	177.05	57.48	30.35	411	Thr	7.38	112.42	174.03	60.38	70.91
343	Gln	8.60	124.41	176.23	57.43	28.48	412	Lys	9.04	120.35	173.89	54.85	33.90
344	Lys	8.50	120.27	174.63	54.93	31.93	413	Ala	8.15	124.73	173.86	50.98	18.31
345	Val	8.75	126.47	175.18	62.41	30.74	414	Val	8.08	126.55	181.13	63.70	32.72
346	Ser	8.15	123.01	174.24	56.24	62.92							

Table 3.5 Side-chain assignments of the 1-phenylimidazole bound form of P450cam.

Residue	C β	H β	C γ 1	H γ 1	C γ 2	H γ 2	C δ 1	H δ 1	C δ 2	H δ 2
3 Thr					22.61	1.27				
5 Thr					20.98	1.19				
6 Ile					16.75	0.85	12.43	0.88		
10 Ala	18.06	1.35								
12 Leu							22.65	0.57	24.49	0.57
19 Val			18.74	-0.02	17.86	-1.36				
24 Val			21.15	0.92	24.18	0.94				
35 Val			22.72	0.69	24.49	0.57				
42 Ala	16.92	1.46								
45 Val			20.12	1.17	19.58	1.15				
46 Leu							25.58	0.41	20.48	-0.41
51 Val			20.43	0.68	21.47	0.67				
55 Val			22.21	0.77	17.03	0.77				
66 Ala	18.14	0.84								
67 Thr					21.35	1.16				
116 Ala	16.63	1.57								
124 Val			19.57	0.76	21.23	0.76				
125 Val			21.13	0.74	22.04	0.74				
128 Leu							24.62	0.61	22.20	0.61
132 Ile					17.69	0.26			12.60	0.61
135 Leu							24.16	0.76	23.26	0.79
139 Leu							22.30	1.19	25.94	1.19
140 Ile					17.31	1.08			12.26	0.63
143 Leu							22.13	1.00	24.61	1.04
152 Thr					21.56	1.12				
161 Ile					17.65	0.73			11.95	0.75
163 Ile					15.95	0.54			10.22	0.32
170 Leu							21.89	1.05	24.13	0.96
178 Leu							24.34	0.16	19.31	0.46
182 Thr					22.32	1.10				
193 Thr					21.33	1.27				
195 Ala	17.31	1.52								
200 Ala	17.93	1.53								
218 Thr					20.77	1.09				
220 Ala	18.93	1.03								
221 Ile					16.59	0.83			11.26	0.85
223 Ile					16.78	0.65			11.34	0.62
224 Val			22.78	0.93	21.34	1.18				
225 Ala	18.38	1.22								
229 Val			17.90	0.20	20.75	0.43				
234 Ile					16.35	0.99			13.15	0.84
235 Thr					21.19	1.32				
239 Ala	17.14	0.47								
282 Ile					16.27	0.80			13.52	0.88
284 Ala	17.42	1.43								
297 Ala	21.03	1.44								
301 Ile					15.40	0.93			13.82	0.95
302 Leu							20.53	0.62	22.07	0.90
303 Thr					20.82	1.27				
311 Val			21.13	0.79	22.14	0.79				
335 Ala	16.47	1.24								
339 Val			21.49	0.27	22.54	0.27				
346 Val			22.36	0.99	20.86	0.99				

Table 3.5 continued

Residue	C β	H β	C γ 1	H γ 1	C γ 2	H γ 2	C δ 1	H δ 1	C δ 2	H δ 2
349 Thr					18.69	0.79				
384 Ile					17.35	1.16			11.47	0.83
385 Ala	17.08	1.33								
388 Ala	18.61	1.40								
390 Ile					17.73	0.95			9.60	0.88
397 Val			21.95	0.74	20.29	0.74				
400 Val			20.09	0.67	21.09	-0.26				
403 Leu							22.91	0.74		
410 Ala	17.52	0.04								
411 Thr					21.30	1.32				
412 Thr					23.47	1.42				
414 Ala	18.58	1.29								
415 Val			16.36	0.72	20.73	0.78				

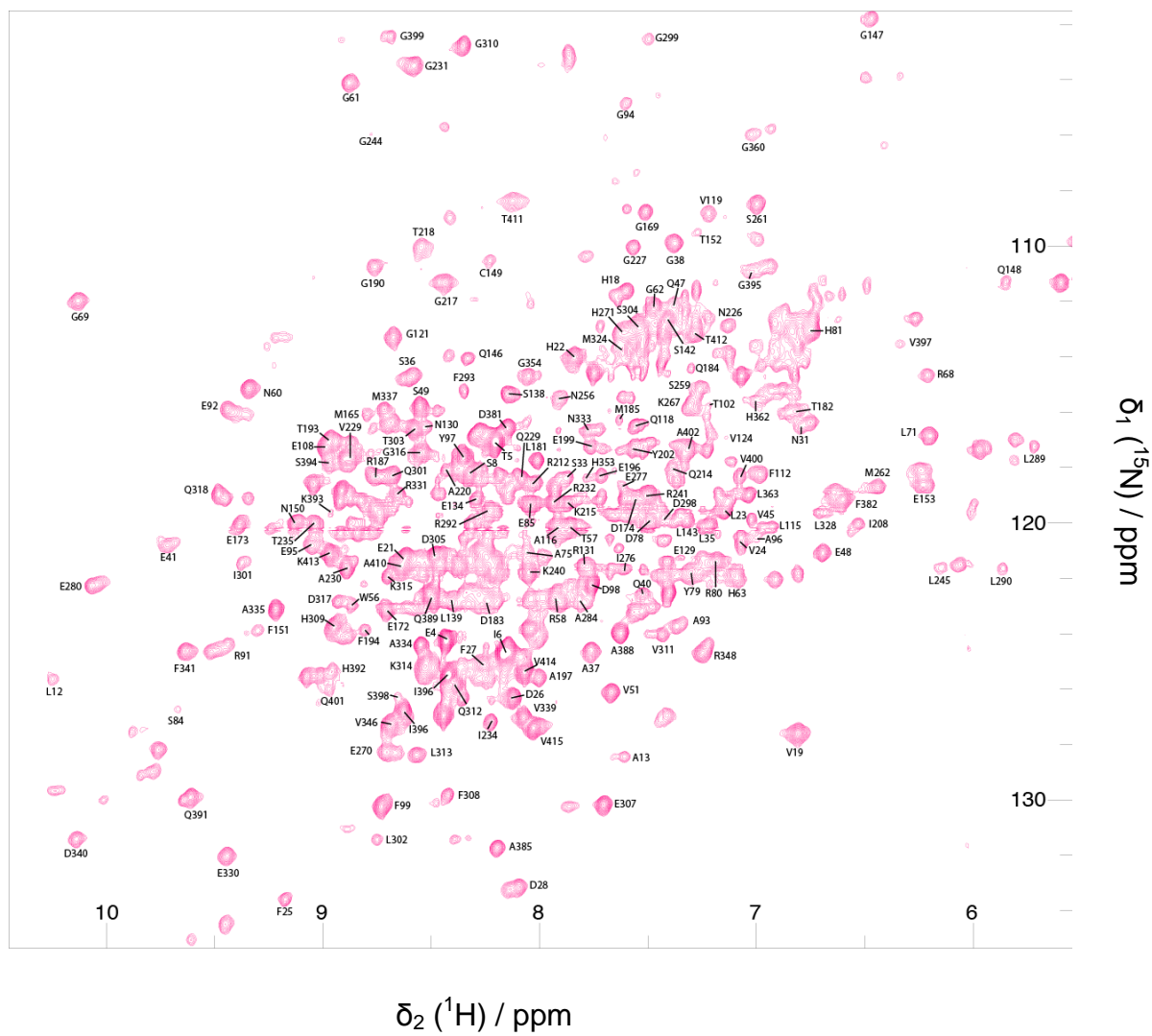


Figure 3.2 ^{15}N - ^1H TROSY HSQC of P450cam with KCN bound. Labels indicate assignment at 298 K, 20 mM HEPES pH 7.5, 100 mM KCN.

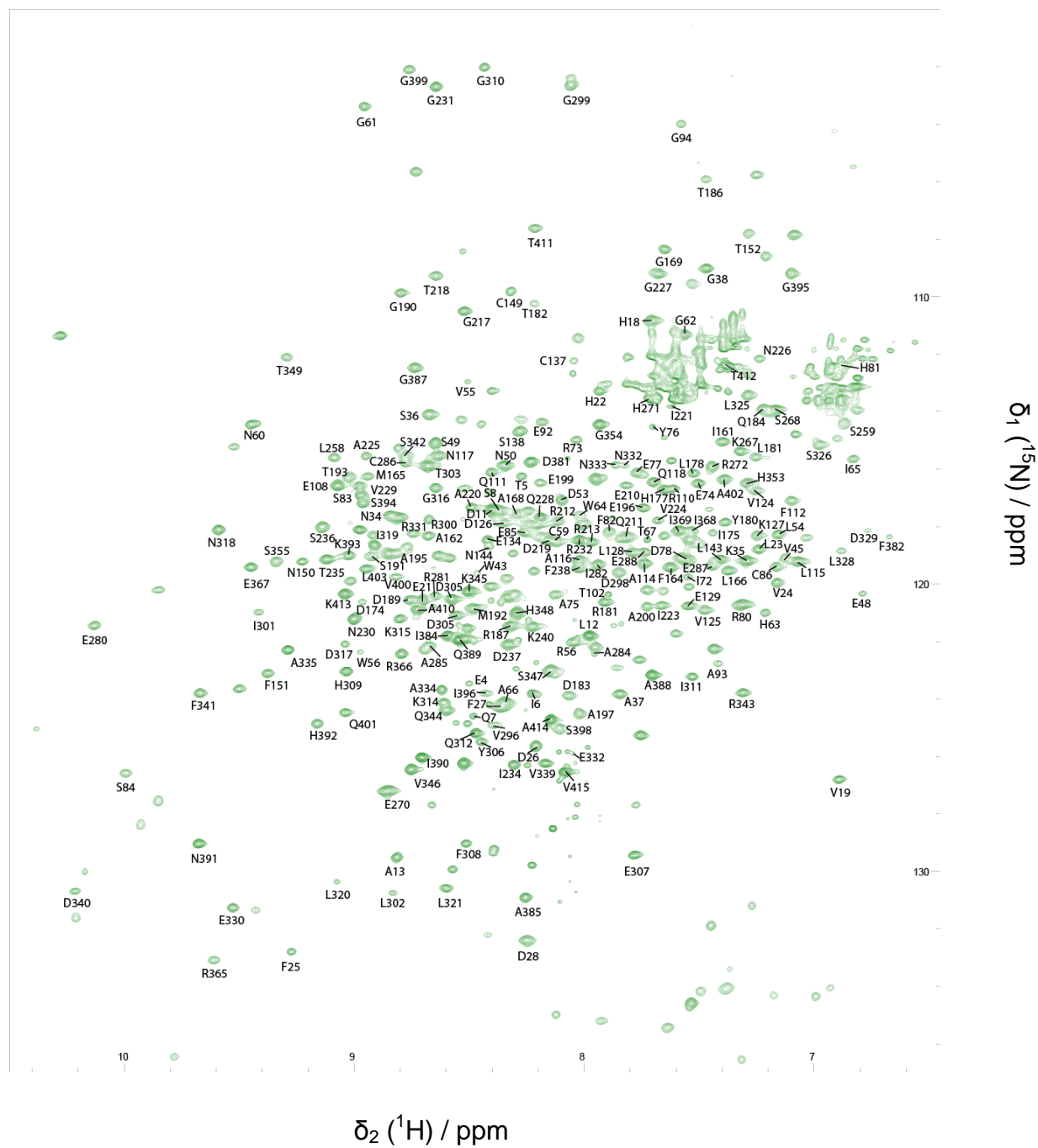


Figure 3.3 ^{15}N - ^1H TROSY HSQC of P450cam with phenylimidazole bound. Labels indicate assignment at 298 K, 20 mM HEPES pH 7.5, 100 mM phenylimidazole.

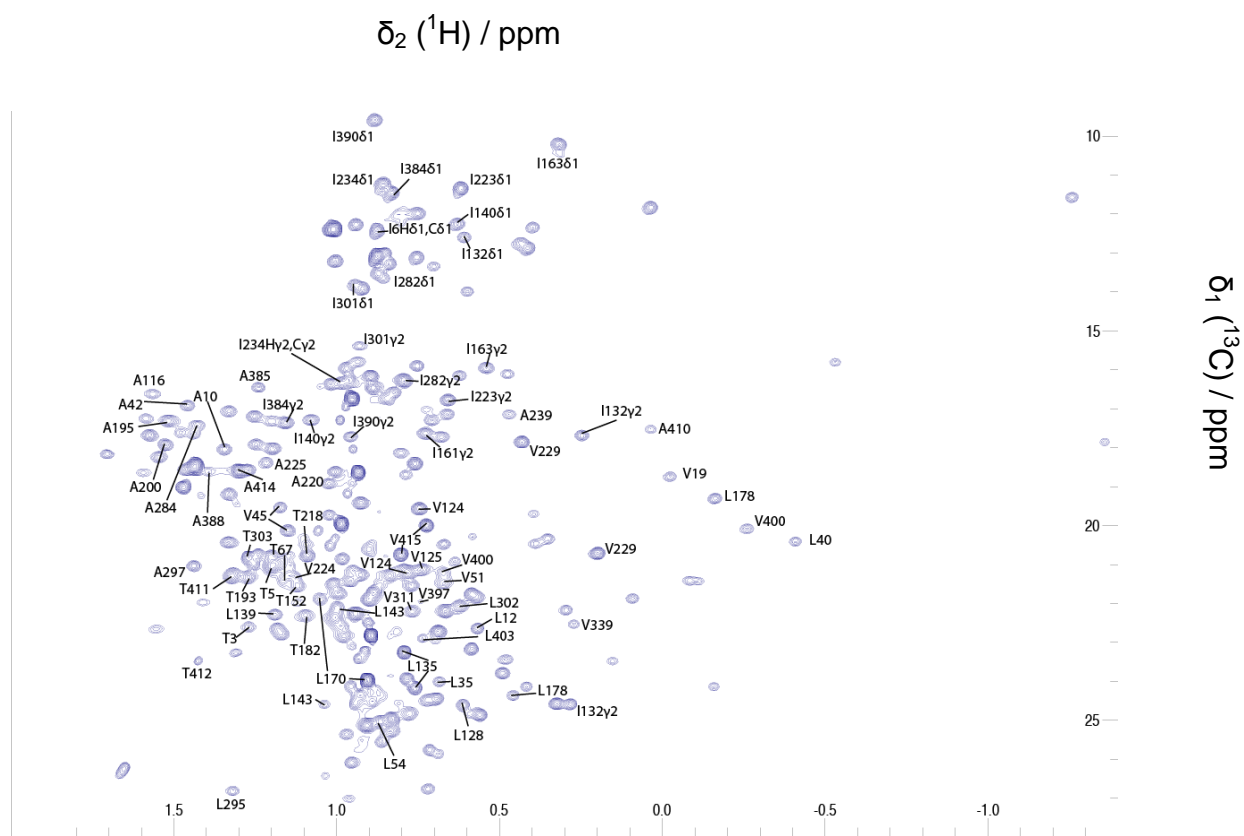


Figure 3.4 $^{13}\text{CHD}_2\text{-CT-HSQC}$ of P450cam. Labels indicate 298 K, 20 mM HEPES pH 7.5, 100 mM phenylimidazole.

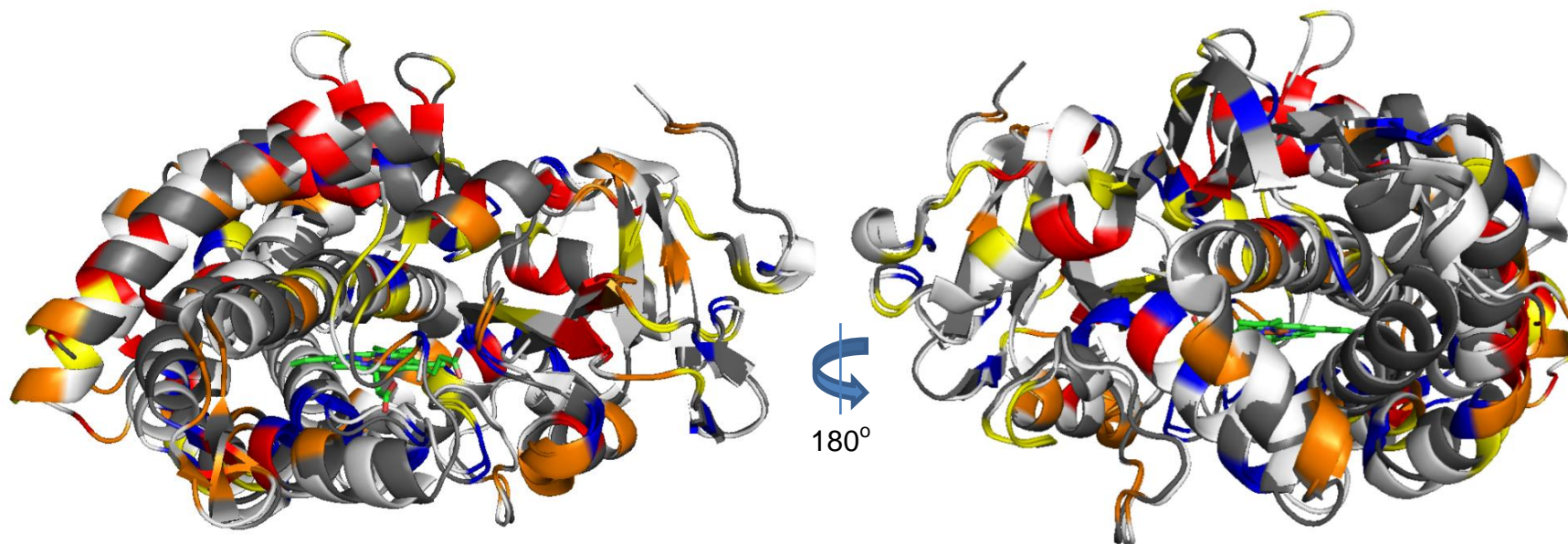


Figure 3.5 Overlay of the open and closed forms of P450cam with CSPs from camphor titration shown on both structures. The colour pattern represents the average size of the chemical shift perturbations. Red: $\Delta\delta_{avg} \geq 0.02$, Orange: $0.01 \leq \Delta\delta_{avg} < 0.02$, Yellow: $0.005 < \Delta\delta_{avg} < 0.01$, Blue $\Delta\delta_{avg} \leq 0.005$. Unassigned residues are colour white for the open form and grey for the closed form. The haem is shown as sticks.

Chapter 4

Solution structures of two forms of Avirulence factor 2 (Avr2) at pH 7

Abstract

Avirulence factor 2 (Avr2) is a protease inhibitor secreted by the pathogenic fungus *Cladosporium fulvum* during infection of the tomato plant, *Solanum lycopersicum*. The interaction between this protein and the two host papain-like cysteine proteases, Rcr3 and PIP1 induces a Cf-2 mediated hypersensitive response, however, the structural basis for this interaction is not known. The three dimensional structures of two forms of Avr2 have been solved at pH 7. The RMSDs to mean for the backbone atoms were 0.92 ± 0.12 and 0.87 ± 0.21 Å for the two forms, indicating that tight bundles were obtained for both structures. These structures provide significant insight into the conformation changes involved in the interconversion between the two forms and the form that binds to Rcr3 in the apoplast of *Solanum lycopersicum* is proposed.

Introduction

Cladosporium fulvum is an asexual fungus that causes leaf mould of the tomato plant, *Solanum lycopersicum* (186). During infection and colonisation of the leaf apoplast, a number of cysteine rich avirulence proteins are secreted, one example of which is Avr2. In resistant strains of *Solanum lycopersicum*, the protein Cf-2 mediates recognition of Avr2. Two papain-like cysteine proteases (PLCPs), Rcr3 and PIP1 bind to Avr2 and these binding events trigger a hypersensitive response in the leaf apoplast (155, 156). In this way, Rcr3 and PIP1 are essential components of a pathogen perception system, such that when the hypersensitive reaction occurs, further incursion of the fungus is avoided due to partial necrosis of the leaf. Models of PIP1 and Rcr3 have been generated and the substrate binding cleft has been identified in both proteins, reviewed in (187). However, the structure of Avr2, up to this point, had not been elucidated and therefore, it has not been possible to produce models for the binary complexes, Avr2-Rcr3 and Avr2-PIP1.

In spite of the absence of three dimensional structural data for Avr2, many groups have investigated the factors that affect the interaction of Avr2 with the two host PLCPs. Avr2 binds to Rcr3 at pH 5 and not at pH 7, and therefore, a pH dependent

conformational change of Avr2 must occur in order for binding to occur. Moreover, four point mutants of Avr2: Y21A, K17A, F54A and K20A and a deletion mutant $\Delta 54$; have been produced and the influence of these mutations on binding has been determined (188). The deletion mutant had a negative effect on binding to Rcr3; however, the point mutations did not significantly influence binding. The effect of Rcr3 mutations on Avr2 binding have been investigated. Three point mutants were produced: H148N, R151Q and N194D and these mutants can be inhibited by Avr2, although the N194D mutant showed less inhibition than the other mutant Rcr3 proteins.

In this study, the three dimensional structures of two forms of Avr2 have been solved using high resolution NMR spectroscopy. The backbone and side chain proton nuclei were assigned using 3D HSQC-TOCSY, 3D HSQC-NOESY, HNCACB, 2D NOESY and 2D TOCSY spectra. Two different forms were found to be present at pH 7. Ten structure calculations were performed for each form using CYANA (24). The resulting structures were refined using the RECOORD scripts (189) and the structures were validated using CING (190). A model for the pH dependent conformation changes involved in Avr2 binding is proposed.

Methods and Materials

NMR samples

A ^{15}N -labelled, ^{15}N ^{13}C labelled and unlabelled samples of Avr2 were kindly provided by Anneloes Blok (Leiden University), using an Avr2 plasmid kindly provided by Prof. Renier van der Hoorn (MPI, Köln).

NMR spectroscopy

^{15}N -HSQC, ^{15}N TOCSY-HSQC and ^{15}N NOESY-HSQC experiments were recorded by Prof. Dr. Marcellus Ubbink (Leiden University) using a 1.5 mM sample of ^{15}N -labelled Avr2 in 20 mM potassium phosphate buffer pH = 7, 6%/94% $\text{D}_2\text{O}/\text{H}_2\text{O}$. The TOCSY and NOESY mixing times were 80 ms and 120 ms, respectively. A ^{15}N ^{13}C HNCACB experiment was recorded by Dr Sagario Arias-Rivas (Leiden University)

using a 200 μM sample of Avr2 in 20 mM potassium phosphate pH = 7, 6%/94% $\text{D}_2\text{O}/\text{H}_2\text{O}$. The HNCACB and the HSQC-TOCSY/NOESY spectra were used to assign the backbone nuclei of the protein and some side-chain nuclei, where possible. The experiments used for backbone assignment were recorded on a Bruker Avance III spectrometer equipped with a TCI-Z-GRAD cryoprobe operating at a ^1H Larmor frequency of 600 MHz. ^1H - ^1H TOCSY experiments with mixing times of 60 and 20 ms and a ^1H - ^1H NOESY experiment with a mixing time of 200 ms were recorded by Dr Mathias A. S. Hass (Leiden University) and Dr Sagrario Arias-Rivas (Leiden University) using a 500 μM sample of unlabelled Avr2 in 20 mM potassium phosphate buffer pH = 7, 6%/94% $\text{D}_2\text{O}/\text{H}_2\text{O}$. These experiments were recorded on a Bruker Avance III spectrometer equipped with a TCI-Z-GRAD cryoprobe operating at a ^1H Larmor frequency of 900 MHz at the NWO-CW NMR facility (Utrecht University). The ^1H - ^1H TOCSY and NOESY experiments were used to assign the side chain protons of Avr2 and the aromatic side chain protons were assigned based on analysis of spin systems and intra-NOEs with 2D NOESY and 2D TOCSY spectra. If two sets of resonances were observed for the same residue, the intensities of the HSQC cross peaks were used to determine which resonances belonged to state A and state B. The chemical shift assignments and the NOESY peak list were used to perform structure calculations in CYANA. All experiments were recorded at 298 K. The spectral widths and numbers of complex points in the indirect dimensions in each 3D experiment are shown in Table 4.1 and for each 2D experiment in Table 4.2. All experiments were recorded with 1024 increments.

Table 4.1 Numbers of complex points and spectral widths in the indirect dimensions used to acquire the 3D spectra used for the assignment of Avr2.

Experiment	Complex points	Spectral width (kHz)	
		^{15}N	^1H
TOCSY-HSQC	40	1.46	7.81
NOESY-HSQC	40	1.46	7.81
HNCACB	32	1.64	9.36

Table 4.2 Numbers of complex points and spectral widths used to acquire the 2D spectra used for the assignment of Avr2.

Experiment	Complex points	Spectral width (kHz)
^1H - ^1H TOCSY (60 ms)	256	11.70
^1H - ^1H TOCSY (20 ms)	256	11.70
^1H - ^1H NOESY	380	11.70
^{15}N - ^1H HSQC	256	1.46

CYANA structure calculation

Using the assignments of state A and state B and their corresponding NOESY peak lists, structure calculation of the two forms was carried out using CYANA 3.0 (24). CYANA performs automated NOESY assignment over seven cycles using the CANDID algorithm, which creates one or more initial assignments for each NOESY cross peak based on chemical shift agreement within user defined tolerance ranges. For these structure calculations tolerances of 0.03 and 0.02 were used for amide protons and side-chain protons, respectively. The initial assignments were weighted with respect to following structure filtering criteria:

- Agreement between the values of the chemical shift list and the peak position.
- The compatibility of the assignments with structure from previous cycles (from cycle 2 onwards).
- Self-consistency of assignments within the entire NOE network using network anchoring.

In the first cycle, structure compatibility cannot be assessed because structural information is not available to the algorithm. Consequently, the structure calculations rely on a structure-independent NOE self-consistency check derived from network anchoring (see *chapter 1*). Once the assignments have been ranked, the NOESY peak volumes or intensities are used to derive upper distance bounds for the corresponding ambiguous or unambiguous distance restraints. Only cross peaks with at least one assignment with a satisfactory network anchoring score and compatible with the intermediate structures (cycle 2 onwards) are retained for subsequent calculations. The structure calculations themselves are performed using torsion angle dynamics. Distance constraints from NOEs with multiple assignments are included as ambiguous distance restraints. In the structure calculations of Avr2, additional

constraints were included in the form of unambiguous distance restraints to define the four disulphide bridges (188) and a set of dihedral angles predicted using TALOS+ (191). For both state A and state B, ten separate calculation runs were carried out using different random seeds. At each calculation step, 100 structures were calculated and the 20 with the lowest target functions were used in subsequent calculation cycles.

Water-refinement using RECOORD

The structure bundles produced by CYANA are calculated *in vacuo* and therefore, may not be representative of the actual solution structure of the protein. To overcome this potential difficulty, water-refinement of the structure bundles was carried out using the RECOORD scripts. These scripts generate an extended conformation of the protein structure using CNS (192, 193). The distance constraints generated by CYANA and dihedral angles from TALOS+ were used for simulated annealing and subsequent water-refinement in a shell of TIP3P water using a full Lennard Jones potential.

To combine the results of the ten CYANA runs per structure calculation, the final distance restraints were combined by selecting only those restraints present in at least five of the calculations and selecting the longest distance for each of these. Torsion angles predicted by TALOS+ were also included in the structure calculations. Moreover, hydrogen bond restraints were selected if they occurred in more than 10 structures of the bundle from each calculation and if they were predicted in at least five of the calculation runs. During the simulated annealing stage, 200 structures were calculated and the 100 structures with the lowest energies were carried through to the water refinement stage. 100 structures were calculated during the water refinement stage and PROCHECK-NMR (194) was used to evaluate the Ramachandran plot of each structure. The final structure bundles consisted of 20 structures that were selected from the 100 water refined structures on the basis of their overall energies and Ramachandran plot analysis. Structure validation of the final structure bundles was carried out using CING.

Results and Discussion

Resonance assignment

With the 3D ^{15}N ^1H NOESY-HSQC, ^{15}N ^1H TOCSY HSQC and ^{13}C ^{15}N HNCACB spectra, assignment of the amide groups for all residues except G1, A2, K3, D35 and L38 were assigned. It was possible to identify two distinct states for E16, K22, V46, Q47H ϵ and H50 using the 3D ^{15}N NOESY-HSQC, ^{15}N TOCSY-HSQC and the ^{15}N HSQC spectra. With the high resolution 2D homonuclear ^1H ^1H TOCSY spectra recorded with two different mixing times and a ^1H ^1H NOESY spectrum, in combination with the ^{15}N ^1H NOESY-HSQC and ^{15}N ^1H TOCSY HSQC, a near complete assignment (94%) of the side chain protons was obtained for those residues for which backbone assignments had been obtained. All aromatic ring protons were assigned. The ^1H - ^1H TOCSY spectra also revealed a second state for the Y22 backbone, side-chain and aromatic protons. The resonances of E16, K22, V46, Q47H ϵ and H50 were assigned to state A or state B based on the intensities of their amide resonances in the HSQC spectrum. The Y21 resonances were assigned to the two states on the basis of NOEs between its side chain protons and E16. Two distinct states were also observed for the aromatic ring protons of Y21 and these were separated between the two states using NOEs from the Y21 H β protons. The near complete ^1H assignments and separation of the resonances belonging to the two states enabled the determination of high quality solution structures for both state A and state B.

Solution structures of two states of Avr2

Pooling the restraints from the ten calculation runs for both states resulted in a total of 650 distance restraints and 10 hydrogen bond restraints for state A and a total of 665 distance restraints and 12 hydrogen bond restraints for state B. The CYANA target function values ranged from 0.47 to 1.74 \AA^2 , the RMSDs to mean for the backbone ranged from 0.48 to 0.85 \AA and the RMSDs to mean for heavy atoms ranged from 0.89 to 1.28 \AA (Tables 4.3 and 4.4). The numbers of distance restraints used for water refinement are shown in Table 4.5. Short range restraints are defined as $|i-j| \leq 1$, medium range restraints are defined as $1 < |i-j| < 5$ and long range restraints

are defined as $|i-j| \geq 5$. Talos was unable to predict reliable dihedral angles for the backbone of K3, K13, S18, G19, D36, Q47, L56 and C57 and, therefore, values for the dihedral angles for these residues were extracted from the ten pooled structure bundles and included in the water refinement calculations. In 5 of the 10 structure calculation runs for state B, dihedral angles in the disallowed region of the Ramachandran plot were present for S18 and D34. In both datasets there are only intra-residue NOEs for S18 and D34 is unassigned; this could explain the disallowed dihedral angles present in the five of the structure bundles.

One hundred structures were produced for both state A and state B of Avr2 after the water refinement process and each was analysed using PROCHECK-NMR. The final structure bundles consisted of 20 structures selected on the basis of energy values of the structures and their Ramachandran statistics. The final ensembles were validated using CING and the structure and validation statistics are shown in Table 4.5. The RMSDs to mean of the backbone of residues 5-56 of states A and B were 0.92 ± 0.12 and 0.87 ± 0.21 Å, and the RMSDs to mean of heavy atoms were 1.34 ± 0.12 Å and 1.37 ± 0.19 Å for state A and state B, respectively. Residue 3 and residues 56-58 were excluded from RMSD calculation due to the fact that they are unstructured in the ensembles.

CING validates structures using a variety of different software packages and combines the results into a ROG score. CING colours residues as red, orange or green, such that red residues are have problems, orange residues have potential probes and green residues have no problems. The CING ROG score of state A was 32 % / 41 % / 27 % and the ROG score of state B was 32 % / 32 % / 36 %. The red residues were considered problematic due to low PROCHECK_NMR scores caused by a deviation between the interresidual distance limits and the upper distance limits used for water refinement. S18 in state A was in the disallowed region of the Ramachandran plot, although there were no inter-residual NOEs in the data set for this residue. A well modelled structure should have a red ROG score below 50 % and a green ROG score greater than ~ 20 %, which is the case for both state A and B. These structures are therefore considered valid.

Table 4.3 Structural statistics for 10 ensembles of 20 structures of state B of Avr2

	Run 1	Run2	Run 3	Run 4	Run 5	Run 6	Run 7	Run 8	Run 9	Run 10
All restraints	651	652	633	631	638	604	629	644	648	642
Short range	347	345	348	346	343	325	338	340	358	347
Medium range	67	72	66	64	66	55	65	73	59	70
Long Range	237	235	219	221	229	224	226	231	231	225
CYANA target function (\AA^2)	0.65	0.59	0.87	1.59	1.74	0.80	1.60	0.47	1.50	0.50
Max violation for upper limits (\AA)	0.25 ± 0.01	0.35 ± 0.03	0.47 ± 0.06	0.38 ± 0.06	0.62 ± 0.08	0.39 ± 0.06	0.75 ± 0.15	0.26 ± 0.00	0.48 ± 0.11	0.25 ± 0.00
RMSD to mean (\AA)										
Backbone atoms	0.53 ± 0.10	0.54 ± 0.16	0.51 ± 0.15	0.48 ± 0.15	0.69 ± 0.19	0.85 ± 0.20	0.50 ± 0.17	0.53 ± 0.13	0.50 ± 0.06	0.59 ± 0.14
Heavy atoms	0.93 ± 0.09	0.92 ± 0.10	0.91 ± 0.09	0.89 ± 0.12	1.08 ± 0.16	1.28 ± 0.19	0.96 ± 0.16	0.91 ± 0.09	0.90 ± 0.09	1.02 ± 0.12
Ramachandran plot (%)										
Most favoured	56.1	65.0	58.2	63.4	60.9	67.2	61.5	68.8	66.0	66.2
Additionally allowed	42.2	32.9	38.3	36.1	37.6	32.0	36.6	29.3	33.8	31.7
Generously allowed	1.7	2.1	3.5	0.5	1.6	0.7	1.9	1.9	0.2	2.1
Disallowed	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 4.4 Structural statistics for 10 ensembles of 20 structures of state A of Avr2

	Run 1	Run2	Run 3	Run 4	Run 5	Run 6	Run 7	Run 8	Run 9	Run 10
All restraints	654	645	639	644	635	649	653	640	656	665
Short range	348	359	347	351	341	353	353	347	351	356
Medium range	73	67	69	76	78	76	76	81	75	71
Long Range	233	219	223	217	216	220	224	212	230	238
CYANA target function (\AA^2)	1.79	1.45	0.87	3.37	2.12	2.04	1.46	0.87	1.13	0.93
Max violation for upper limits (\AA)	0.27 ± 0.01	0.38 ± 0.04	0.40 ± 0.05	0.45 ± 0.06	0.59 ± 0.11	0.85 ± 0.13	0.37 ± 0.04	0.26 ± 0.01	0.54 ± 0.1	0.26 ± 0.00
RMSD to mean (\AA)										
Backbone atoms	0.50 ± 0.13	0.73 ± 0.16	0.61 ± 0.08	0.67 ± 0.13	0.62 ± 0.12	0.43 ± 0.13	0.54 ± 0.15	0.67 ± 0.17	0.76 ± 0.11	0.34 ± 0.09
Heavy atoms	0.90 ± 0.08	1.12 ± 0.15	1.01 ± 0.07	1.05 ± 0.11	1.02 ± 0.11	0.89 ± 0.15	0.91 ± 0.11	1.06 ± 0.14	1.16 ± 0.1	0.75 ± 0.11
Ramachandran plot (%)										
Most favoured	68.4	63.0	68.3	60.5	66.1	65.5	60.0	69.7	68.4	64.0
Additionally allowed	31.6	36.8	30.9	38.0	33.9	33.9	39.4	30.3	30.7	36.0
Generously allowed	0.0	0.1	0.1	0.0	0.0	0.7	0.0	0.0	0.8	0.0
Disallowed	0.0	0.1	0.7	1.6	0.0	0.0	0.6	0.0	0.1	0.0

Table 4.5 Structural statistics the water refined ensembles of 20 structures of state A and state B of Avr2

	State A	State B
All restraints	656	652
Short range	363	345
Medium range	78	72
Long Range	215	235
Max violation		
for upper	0.48	0.40
limits (Å)		
RMSD to mean		
(Å)		
Backbone		
atoms	0.92 ± 0.12	0.87 ± 0.21 Å
Heavy atoms	1.34 ± 0.12	1.37 ± 0.19
Ramachandran		
plot (%)		
Most favoured	77.7	72.6
Additionally		
allowed	19.5	24.7
Generously		
allowed	0.8	2.7
Disallowed	2.0	0.0
ROG score	32 % / 41 % / 27 %	32 % / 32 % / 36 %

An overlay of structures closest to the mean of state A and state B of Avr2 is shown in Figure 4.1. The two forms of Avr2 have a similar overall structure consisting of two antiparallel beta sheets, $\beta 1$ (residues 31-33 and 36-38), $\beta 2$ (residues 42-46 and 49-53) and a 3^{10} helix (residues 21-23) (Figure 4.2). In state B, $\beta 1$ is stabilised by four hydrogen bonds: A31-NH – T38-O, C33-NH – K36-O, K36-NH – C33-O and T38-NH – A31-O, $\beta 2$ is stabilised by three hydrogen bonds: E44-NH – L51-O, V46-NH – N49-O and N49-NH – V46-O and the 3^{10} helix is stabilised by a hydrogen bond between L23-NH – K20-O. Two additional hydrogen bonds are also present: K9-NH – T32-O

and G26-NH – C43-O. In state A, an additional hydrogen bond: K24-NH – Y21-O stabilises the 3¹⁰ helix and two hydrogen bonds: T42NH – C53-O and L51HN – E44-O stabilise β 2. The hydrogen bond C33-NH – K36-O is not present in state B. Hydrogen bonds observed at least 10 structures of states A and B are shown in Table 4.6.

Table 4.6 Hydrogen bonds present in at least 10 structures of both state A and state B with accompanying distances.

	State A		State B	
	No. models	Distance (Å)	No. models	Distance (Å)
K9 HN – T32 O	13	2.13 ± 0.022	18	2.10 ± 0.24
L23 HN – K20 O	20	2.19 ± 0.07	16	2.14 ± 0.17
K24 HN – Y21 O	20	2.16 ± 0.06	10	2.25 ± 0.09
G26 HN – C43 O	20	1.93 ± 0.28	19	1.89 ± 0.17
C29 HN – G41 O	11	2.11 ± 0.27	10	2.02 ± 0.21
A31 HN – T38 O	20	2.03 ± 0.25	19	2.10 ± 0.22
C33 HN – K36 O	17	1.93 ± 0.18	19	1.91 ± 0.21
K36 HN – C33 O	20	2.02 ± 0.16	19	2.02 ± 0.18
T38 HN – A 31 O	20	1.83 ± 0.15	19	1.87 ± 0.20
T38 Hy2 – K36 O	13	2.14 ± 0.22	11	2.16 ± 0.20
T42 HN – C53 O	20	2.08 ± 0.31	10	2.01 ± 0.22
C43 HN – A28 O	18	2.04 ± 0.21	19	1.98 ± 0.27
E44 HN – L52 O	20	1.92 ± 0.19	18	1.79 ± 0.13
V46 HN – N49 O	20	1.80 ± 0.10	19	1.74 ± 0.11
N49 HN – V46 O	19	1.96 ± 0.14	19	1.98 ± 0.19
L51 HN – E44 O	20	2.03 ± 0.14	19	1.87 ± 0.13
C53 HN – T42 O	18	1.82 ± 0.14	17	1.94 ± 0.23

In state A two additional hydrogen bonds were observed: K20 HN – K17 O (1.94 ± 0.28 Å) and G48 HN – N45 O δ 1 (2.11 ± 0.19 Å) in 17 and 14 models respectively. In state B, two hydrogen bonds to E16 O, from G19 HN (2.19 ± 0.18 Å) and Y22 HN (1.98 ± 0.31 Å) were observed in 13 and 10 models, respectively. An additional

hydrogen bond between A27 HN and K24 O was observed in 12 models with a distance of 2.12 ± 0.20 Å.

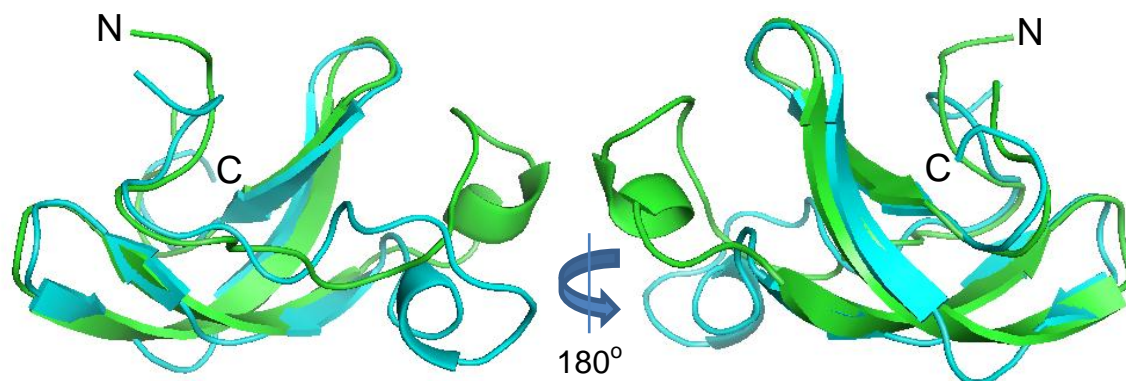


Figure 4.1 Overlay of the structures of state A (green) and state B (cyan) of Avr2. The overall structure consists of two antiparallel beta sheets and a 3^{10} helix. The N and C termini are indicated on the structures.

In state A, Y21 is solvent exposed and E16 is located inside the protein, possibly forming an ionic interaction with the imidazole of H50 (Figure 4.2). There are three lysines (K17, K15 and K20) whose $N\zeta$ groups could form cation- π interactions with the phenol ring of Y21. NOEs were observed between the $H\delta$ protons of K15 and the ring protons of Y21 and between the $H\gamma$ protons of K20 and the ring protons. The maximum distance for a cation- π interactions is 6 Å, and therefore, this type of interaction between K15/20 and Y21 is conceivable. Moreover, the hydrogen bond observed between K20 and K17 indicates that these residues are in close proximity and therefore, K17 could also participate in this interaction.

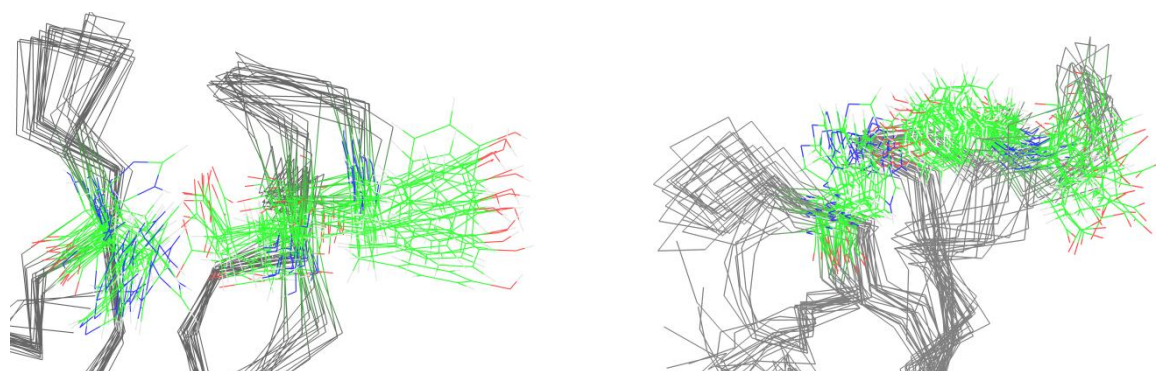


Figure 4.2 Positions of E16, Y21 and H50 in the ensembles of states A (left) and B (right) of Avr2. The backbone is shown in grey.

The Y21A, K17A and K20A mutants of Avr2 did not significantly affect binding to Avr2 and, therefore, although they may play a part in the stabilisation of this form of Avr2, these interactions are not the principal driving force. In 10 models, ionic interactions between the O δ s of E16 and the N δ /N ϵ of the H50 imidazole ring are present with a distance of 2.95 ± 0.27 Å. This strong interaction could hold the histidine in place, along with hydrogen bonds to C13 O (observed in 5 models). State A is most likely a low pH state, similar to that present at pH 5, since protonation of the three lysines and protonation of the H ϵ 2 of H50 would drive the aforementioned interactions to occur and the hydrogen bonds between K20 and K18 would further stabilise this conformation.

State B would therefore be a high pH form and the principal form at pH 7. In this form, Y21 is in the interior of the protein and it is likely that this conformation of Y21 is stabilised by the hydrogen bonds between its amide proton and the carbonyl of E16. Moreover, the second hydrogen bond to the carbonyl oxygen of E16, namely that with the amide proton of G20, could stabilise its position on the exterior of the protein. In this form, the N ζ s of K17, K15 and K20 and the N ϵ 2 of H50 would also be deprotonated. The deprotonation of H50 would weaken the interaction between H50 and E16 residues considerably and the deprotonation of the lysine N ζ s would remove any interactions with the phenol ring of Y21. The protonation/deprotonation events, ionic interaction cleavage and the resultant hydrogen bond formations could provide a reason as to why the two states are in slow exchange at pH 7 and also provide a basis for the pH dependent binding behaviour of Avr2 to Rcr3.

State A and state B both contained the four disulphide bridges, however, in state B, the S γ atoms of C12 and C29, and C29 and C43 are within the Van der Waals bonding distance in all 20 models. In state A, the S γ atoms of C43 and C52 are also within the Van der Waals bonding distance in all 20 models and in 8 models the S γ atoms of C12 and C52 are less than 2.1 Å from each other. It is possible that the disulphide bridges are over-restrained in these structures; however, the correct interactions are all formed in each model of both states A and B. It is possible that S-S bond exchange could occur in Avr2 at pH 7. The elucidation of the disulphide bridges was performed using MALDI-TOF MS at low pH and therefore, at pH 7 the

disulphide bridge pattern may be different. MS/MS analysis of Avr2 at pH 7 would be required to determine whether or not this is the case.

Previous mutation studies have indicated that mutation of Y21 does not significantly affect Avr2 binding to Rcr3, an observation which adds further weight to the proposition that in the low pH form of Avr2, Y21 is solvent exposed and not involved in any interactions with the rest of the protein. In order to determine, whether state A is indeed the low-pH form of Avr2, the structure of Avr2 at pH 5 needs to be solved and a pH titration would be required to determine the pK_a values of Y21, H50 and E16 and see how these fit to the model for conformational change. Moreover, if an ionic interaction between E16 and H51 is a key interaction involved in stabilising State A, the low pH form, an E16Q mutant should be inactive at pH 5. Further experimentation would be required to validate or invalidate these hypotheses.

Chapter 5

The influence of Pdx binding on the opening and closing of P450cam in solution

Abstract

Cytochrome P450cam catalyses the hydroxylation of camphor in a process involving two electron transfers from putidaredoxin and the structure of this complex has recently been solved using both NMR spectroscopy and X-ray crystallography. There is a controversy over whether P450cam is in an open, a closed or an intermediate conformation when Pdx binds in solution, since each of the X-ray crystallographic structures shows a different conformation of P450cam. To determine which state P450cam occupies on Pdx binding, a paramagnetic tag was attached to the G helix of P450cam and a Pdx titration was monitored using NMR spectroscopy. The data demonstrated that the spin state of the haem may change when Pdx binds to P450cam. Moreover, there was no residues that showed a significant change in PCS, which indicates that the enzyme may not open when Pdx binds, as recently publications have suggested and all observed PCS are consistent with the closed state.

Introduction

Cytochromes P450 (CYPs) are b-type haem containing monooxygenases found throughout the three kingdoms of life. These enzymes catalyse the regio- and stereospecific hydroxylation of various aliphatic and aromatic compounds, which are involved in a considerable number of metabolic processes, such as steroid biosynthesis and xenobiotic metabolism in mammals. The CYP superfamily has been extensively studied over the past five decades and an understanding of its mechanism is crucial for the development of pharmaceutical compounds that do not inhibit these enzymes. The archetypal member of this superfamily is CYP101A1 (more commonly known as P450cam) from *Pseudomonas putida*, which was the first enzyme for which the three dimensional structure was determined by X-ray crystallography (146). Prior to the publication of the first structure of a mammalian CYP, CYP2C9 (137), the structure of P450cam was used to produce homology models of these CYPs. P450cam catalyses the regio- and stereospecific hydroxylation of D-camphor to 5-exohydroxycamphor using two electrons from NADH

and molecular dioxygen. Putidaredoxin reductase oxidises NADH and transfers the electrons to putidaredoxin, which in turn shuttles electrons to P450cam.

P450cam has been reported to exist in at least three forms, open, closed and intermediate (180), however, which state P450cam occupies when in complex with Pdx is unknown. It is believed that during the catalytic cycle P450cam opens to bind camphor, closes to perform hydroxylation and then re-opens to release the product. Recently, a structure of the complex of P450cam tethered to Pdx obtained using X-ray crystallography has been published and it suggests that P450cam is in the open state when Pdx is bound (195). A second structure obtained using X-ray crystallography has been published, which suggests that P450cam occupies an intermediate state when Pdx binds (196). A Resonance Raman study published in 2001 proposed that a conformational population redistribution occurs when Pdx binds to P450cam and this was attributed to a perturbation of the structural equilibrium (197). A more recent DEER study advocates the opening of P450cam when Pdx binds in the first, but not the second electron transfer step and that a spin state change also occurs (198). It is somewhat counter-intuitive that P450cam would be open when Pdx binds, since this would mean that P450cam opens during the catalytic cycle and that reactive radical species could be released from this enzyme. The conformation of P450cam can be manipulated in solution using potassium ions and different ligands (199-201). In the absence of substrate, P450cam is in the open state and in the presence of substrate and a high concentration of potassium ions, P450cam is in the closed state (180). In this chapter, the state of P450cam, open or closed, was investigated by titrating Pdx into a solution of ¹⁵N leucine labelled P450cam with CLaNP-7 (61) attached to the G helix and PCS were measured. If the enzyme opened when Pdx binds, a change in PCS would be observed in the spectra.

Methods and Materials

A ¹⁵N-Leu labelled sample of the E195C/A199C/C334A mutant of P450cam was kindly provided by Dr. Monika Timmer (Leiden University) and this sample was divided in two and tagged with CLaNP-7 containing Lu and Yb by Yoshitaka Hiruma (Leiden University). The protein concentrations of the Lu-CLaNP-7 and Yb-CLaNP-7 tagged samples were 175 μ M and 300 μ M, respectively. To investigate complex

formation between Pdx and P450cam, microliter aliquots of a 4 mM Pdx stock solution were titrated into the NMR samples to approximately 2.5 molar equivalents and the effects on the diamagnetically and paramagnetically tagged proteins were observed. After each addition of protein ^{15}N - ^1H HSQC spectra were recorded.

All measurements were performed on a Bruker AV850 NMR spectrometer equipped with a TCI-Z-GRAD cryoprobe operating at 298 K. Each spectrum was recorded with 128 complex points and a spectral width of 2.9 kHz in the indirect dimension. Spectra of the paramagnetically tagged sample were acquired with 8 transients per increment and spectra of diamagnetically tagged sample were acquired with 32 transients per increment; a fourfold increase of transients was used for the spectra of the diamagnetically tagged sample to account for the lower P450cam concentration in this sample.

The spectra of the diamagnetically tagged samples were assigned using previously acquired NMR assignments and pseudocontact shifts were predicted using the previously published magnetic susceptibility values (196). These PCS predictions were used to assign the spectra of the paramagnetically tagged samples and additional assignments of the diamagnetically tagged spectrum also were obtained from the predicted PCSs. Using the same tensor, PCSs were predicted using the open (PDB code 4JWS (195)) and closed forms (PDB code (141)) of P450cam.

Results and Discussion

The spectra of a ^{15}N -uniformly labelled sample of P450cam are extremely crowded (Chapter 3) making detailed spectral analysis of very challenging. To overcome this difficulty, specific ^{15}N leucine labelling was used and spectra containing far fewer resonances were obtained. To determine if Pdx causes P450cam to open, Pdx was titrated into P450cam in the presence of substrate and a high concentration of potassium chloride. When the protein opens, the F and G helices move, therefore, by attaching a paramagnetic probe to the G helix and measuring pseudocontact shifts, it

can be determined if the protein indeed opens or remains closed when Pdx binds. The paramagnetic tensor of the E195C/A199C mutant with CLaNP-7 attached has previously been determined (196), thus this was used to predict pseudocontact shifts for different structures of P450cam alone, and for the complex.

In the spectrum of the diamagnetically tagged sample of the initial P450cam sample containing no Pdx, 27 resonances were assigned using previously obtained assignment data (Chapter 3). The resonance of L402 was not observed in the spectrum of the diamagnetically tagged sample in the presence of 2.5 molar equivalents of Pdx. In order to assign the paramagnetic spectra, the data were fitted using previously obtained tensor values for CLaNP-7 attached to P450cam using Numbat (28) and this tensor was refined using the experimentally obtained paramagnetic data. The fit of predicted to experimental pseudocontact shifts are shown in Figure 5.1. The low Q value (202) of 0.035, the line of best fit and R^2 indicate that the fit of the data is very good. Using pseudocontact shifts, 25 resonances were assigned in the spectrum of the paramagnetically tagged sample of P450cam; the resonances of L177 and L180 were broadened beyond detection, due to the fact that they are in close proximity of the paramagnetic tag. In the spectra of the final titration point of the paramagnetically tagged sample, 22 resonances were identified; L138, L200 and L204 could not be assigned in this spectrum. The assigned spectra are shown in Figures 5.2 – 5.5.

Chemical shift perturbations were observed in the spectra of the diamagnetically and paramagnetically tagged samples and $\Delta\delta_{avg}$ (184) values were plotted against residue number in Figure 5.6. The CSPs were also mapped onto overlays of the open and closed states as shown in Figure 5.7 to determine whether significant CSPs coincided with conformational changes between the open and closed states of P450cam. Significant CSPs are observed for both the para- and diamagnetically tagged samples for L70, L134, L142, L166, L169, L250, L257, L289, L301 and L324. The resonances of L138, L177, L180, L200 and L294 could not be assigned in the spectrum of the paramagnetically labelled sample. These CSPs indicate that Pdx binding causes either conformational changes or a spin-state change. Pochapsky

and co-workers have previously reported spectral changes on Pdx binding; however, they observed no spectral change for L142, whereas, in the current study, a significant CSP for this residue was obtained. Residues L166, L169, L177, L180 and L200 undergo conformational changes when P450cam goes from the closed to the open state. L250 and L257 are located on the I helix in the vicinity of the haem and also undergo conformational changes when P450cam opens. Pochapsky and co-workers also observed CSPs for these two residues when Pdx binds. L289, L301 and L324 are located near the Pdx binding site; however, these residues show no significant structural differences between the open and closed states of P450cam. Consequently, the CSPs observed in this study are consistent with the observations of Pochapsky and co-workers such that effects are observed from the Pdx binding site and elsewhere. These effects could be due to a conformational change in P450cam, a change in the spin-state of the haem, or a combination of the two effects.

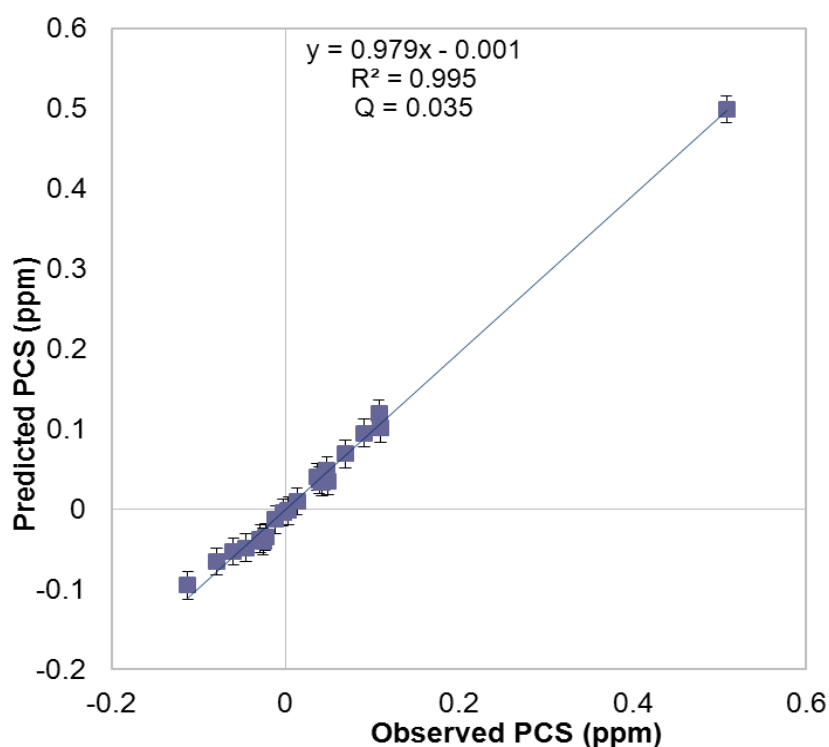


Figure 5.1 Experimentally determined PCS plotted against back-calculated PCS for the A195/E199C mutant of P450cam with CLaNP-7 attached. The solid line represents the least-squares linear regression line.

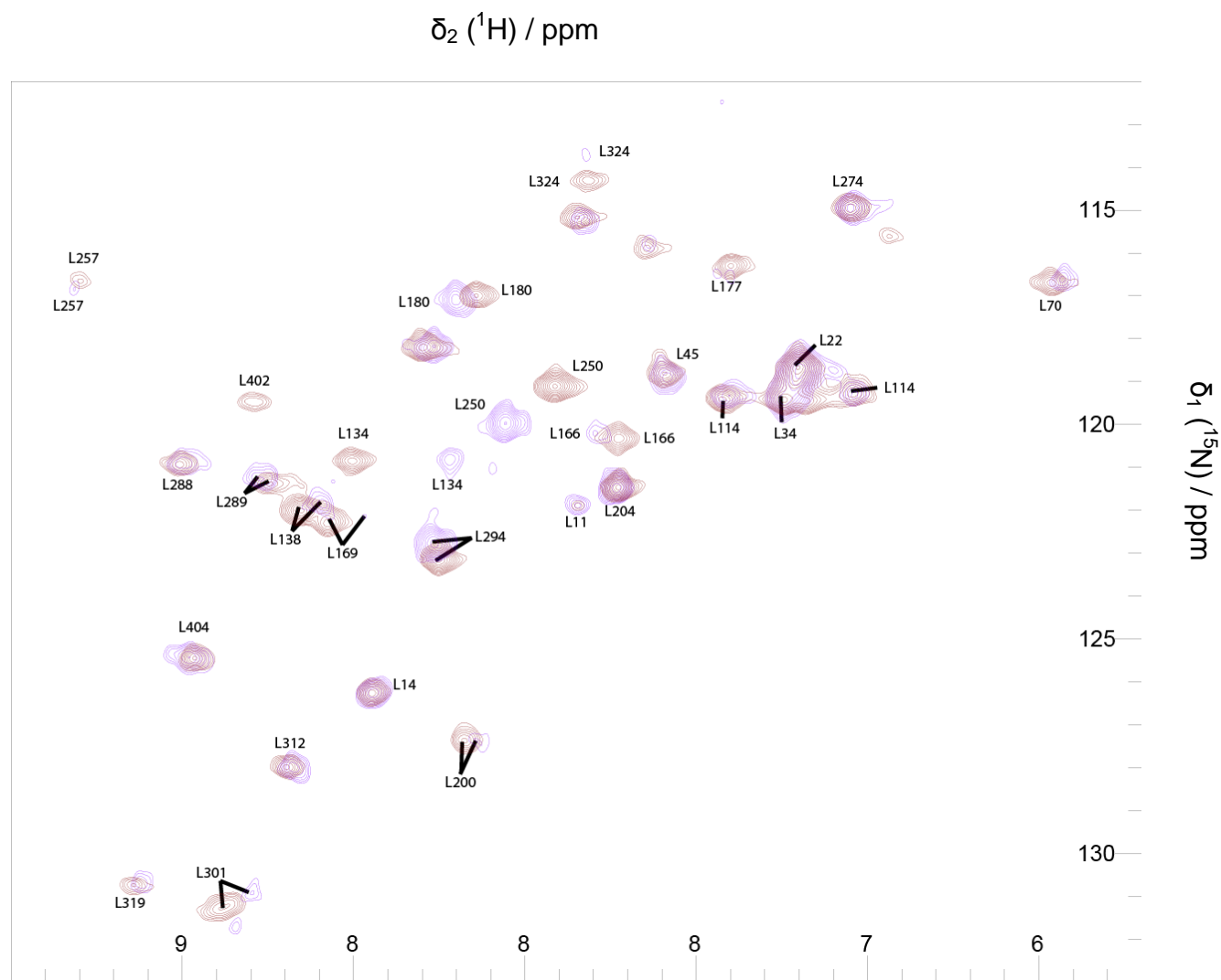


Figure 5.2 ^{15}N - ^1H HSQC spectra of the ^{15}N -Leu labelled A195/E199C mutant P450cam with Lu-CLaNP-7 attached in the presence (purple) and absence (brown) of 2.5 molar equivalents of Pdx. Labels indicate assignments at 298 K, 20 mM HEPES pH 7.5, 100 mM KCl.

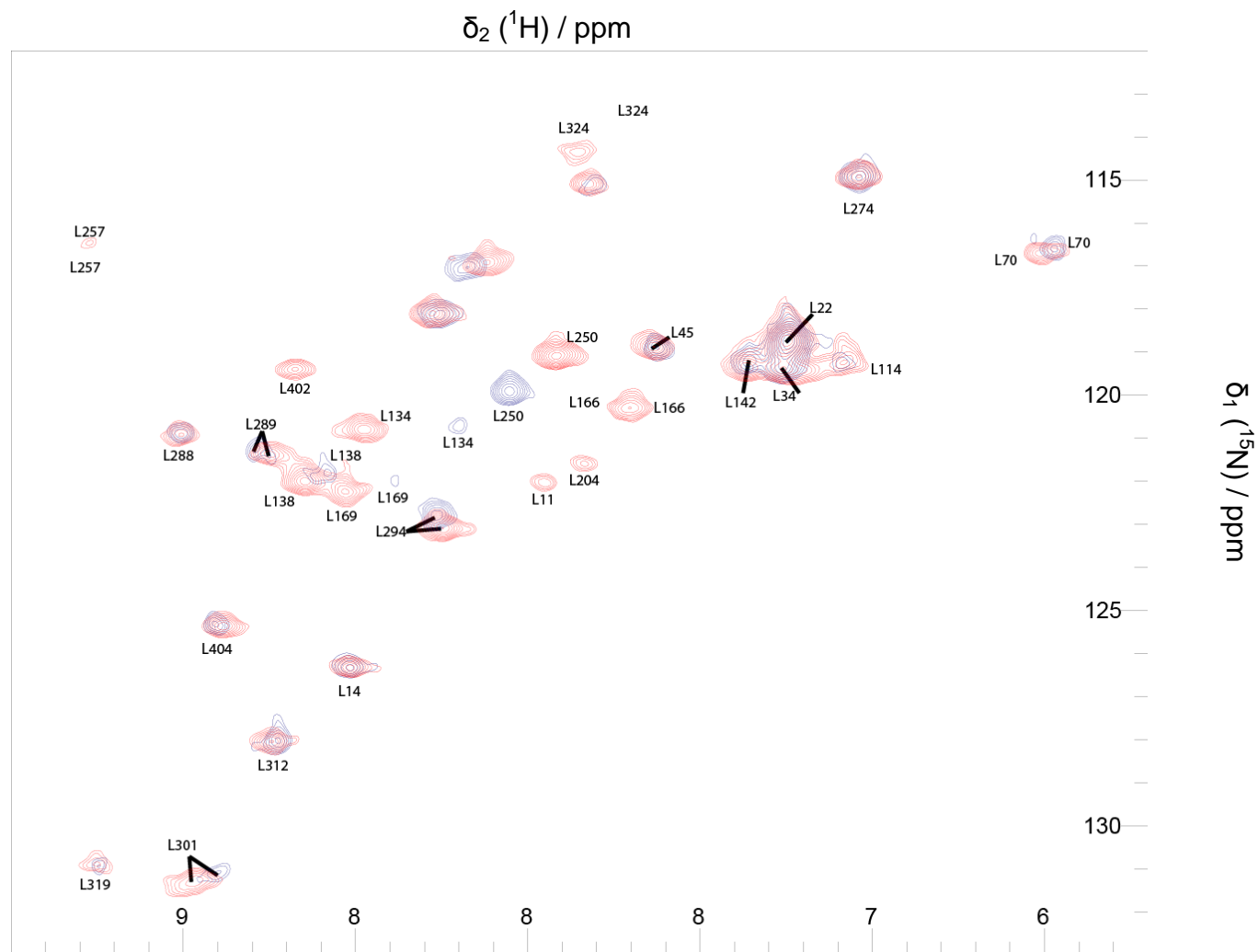


Figure 5.3 ^{15}N - ^1H HSQC spectra of the ^{15}N -Leu labelled A195/E199C mutant P450cam with Yb-CLaNP-7 attached in the presence (dark blue) and absence (red) of 2.5 molar equivalents of Pdx. Labels indicate assignments at 298 K, 20 mM HEPES pH 7.5, 100 mM KCl.

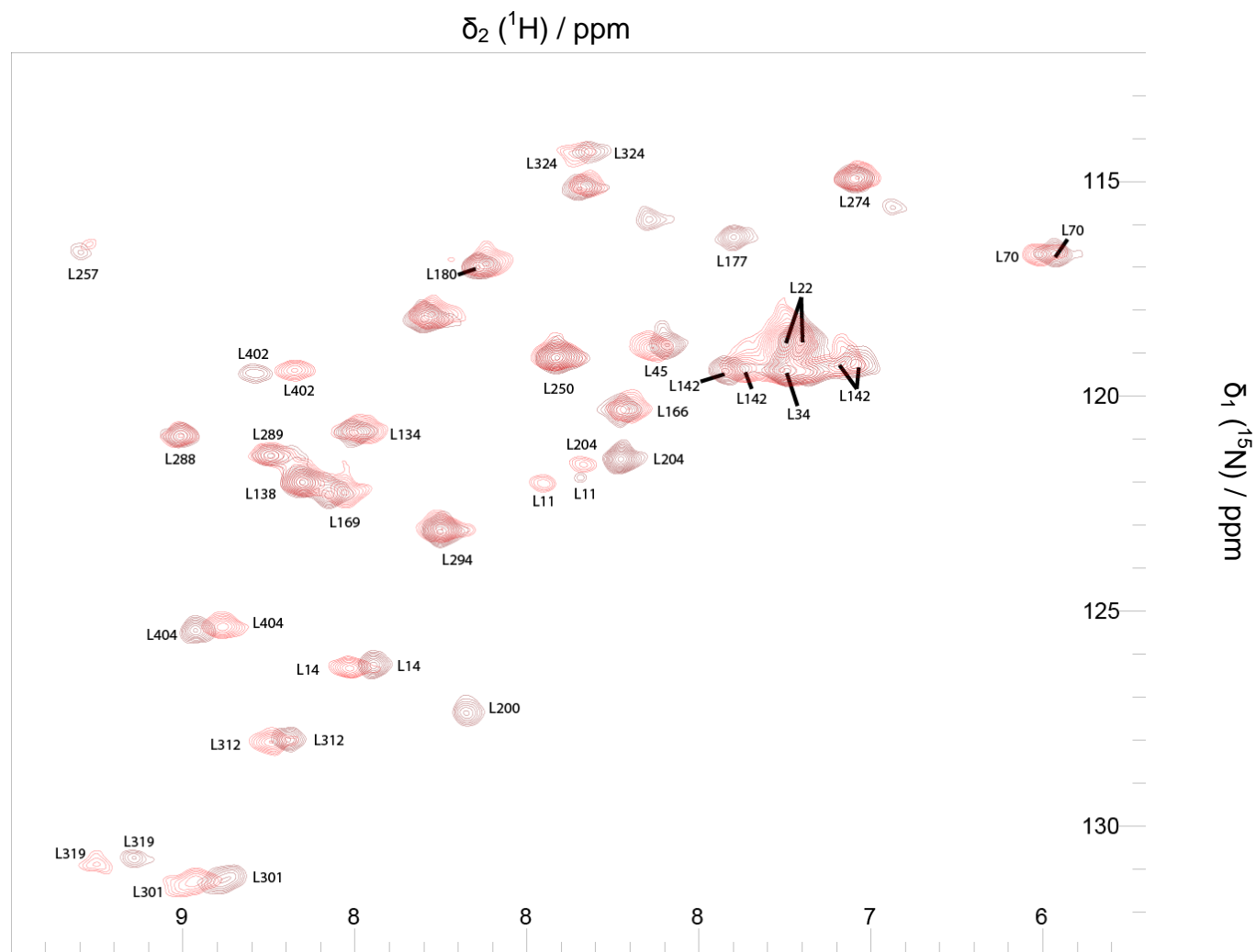


Figure 5.4 ^{15}N - ^1H HSQC spectra of the ^{15}N -Leu labelled A195/E199C mutant P450cam with CLaNP-7 attached. The spectrum with Lu-CLaNP-7 attached is shown in brown and the spectrum with Yb-CLaNP-7 attached shown in red. Labels indicate assignments at 298 K, 20 mM HEPES pH 7.5, 100 mM KCl.

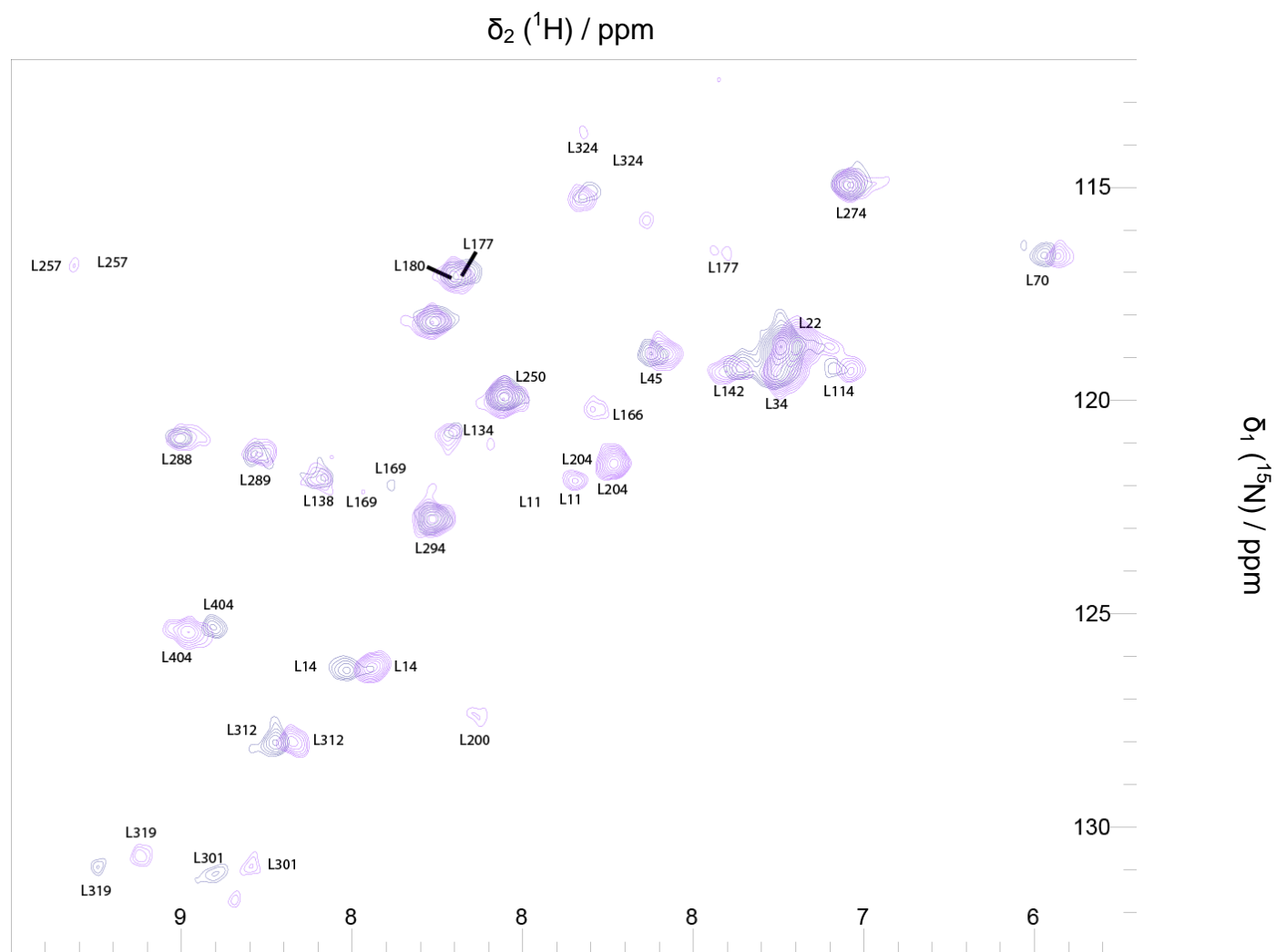


Figure 5.5 ^{15}N - ^1H HSQC spectra of the ^{15}N -Leu labelled A195/E199C mutant P450cam with CLaNP-7 attached in the presence of 2.5 molar equivalents of Pdx. The spectrum with Lu-CLaNP-7 attached is shown in purple and the spectrum with Yb-CLaNP-7 attached shown in dark blue. Labels indicate assignments at 298 K, 20 mM HEPES pH 7.5, 100 mM KCl.

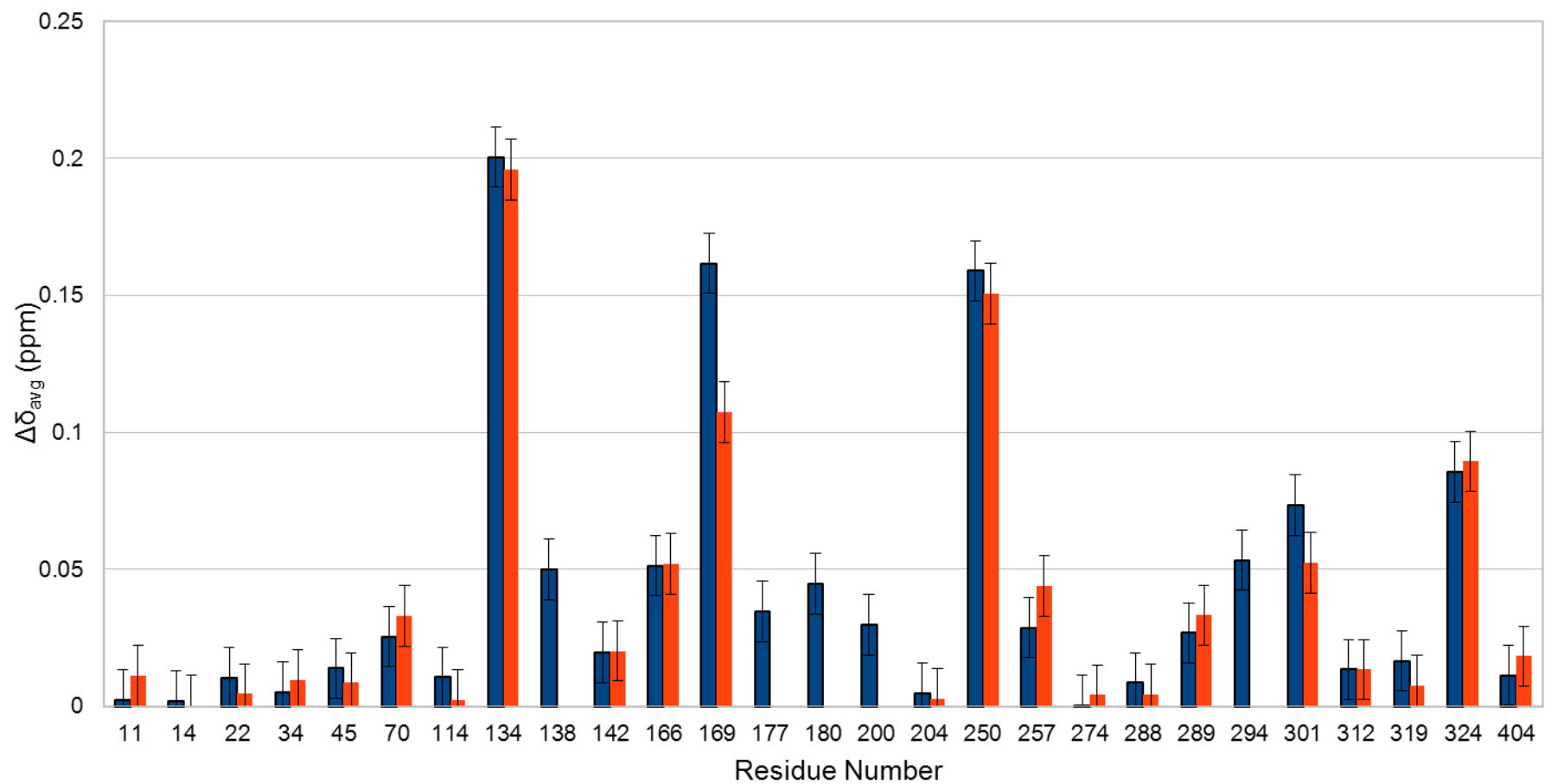


Figure 5.6 Chemical shift perturbations observed in the spectra of the diamagnetically (blue) and paramagnetically (red) tagged samples of P450cam after the Pdx titration.

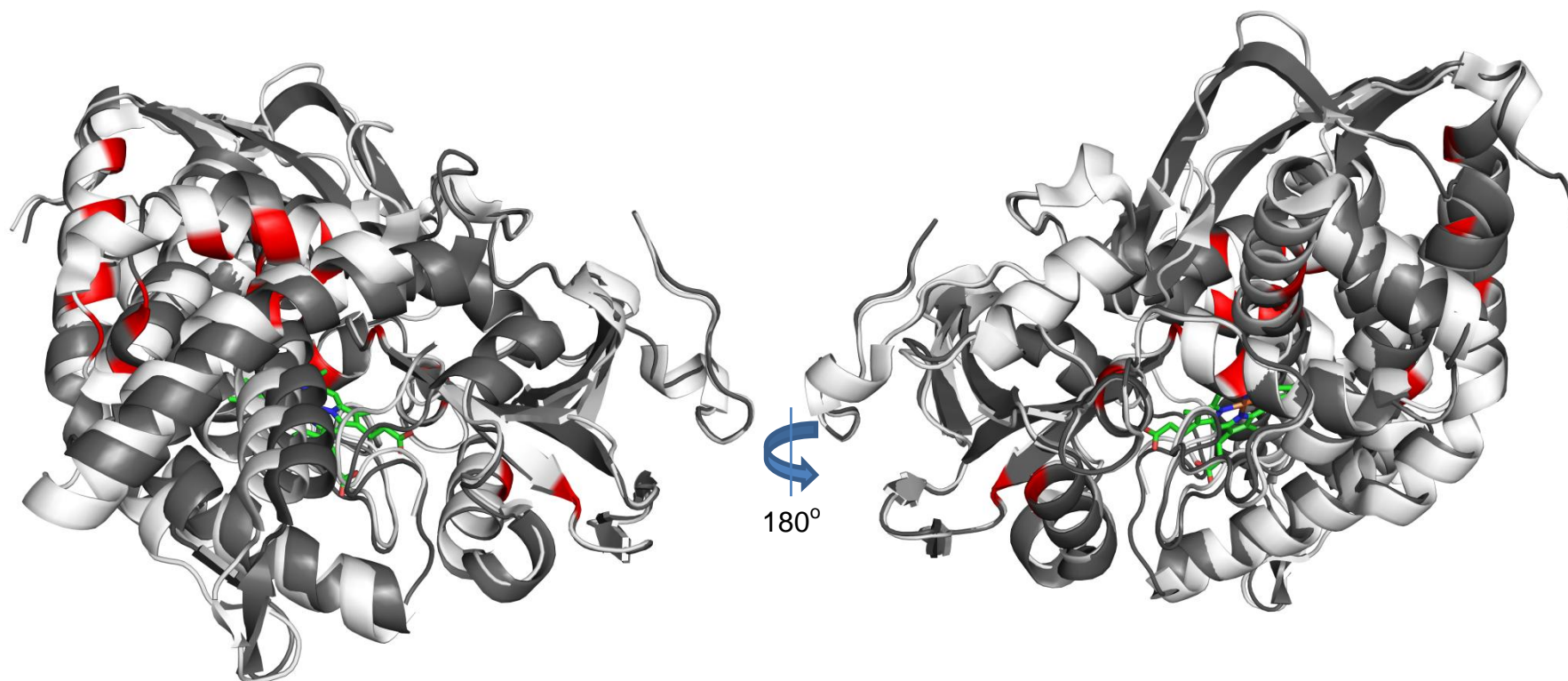


Figure 5.7 Chemical shift perturbations observed in the Pdx titration mapped onto overlays of the open (white) and closed (grey) states of P450cam. Residues for which a CSP was observed are shown in red and the haem, is shown as sticks.

To determine whether conformational changes were occurring when Pdx binds to P450cam, a paramagnetic tag was attached to the G helix of P450cam. This helix moves when P450cam opens and, therefore significant changes in PCS would be observed between the Pdx-free and Pdx bound forms of P450cam if P450cam did transition from the closed to the open form. The expected PCS for the open state of P450cam was predicted using previously published tensor values and the structure of the P450cam-Pdx complex (PDB code 4JWS (195)), which shows P450cam in the open state, when in complex with Pdx. PCS for the closed state were also predicted using a structure of P450cam in the closed state (PDB code 3L63 (141)). The experimentally observed PCS for the Pdx-free and Pdx-bound states of P450cam are shown in Figure 5.8, along with predicted PCS for the closed and open states of P450cam. These data clearly demonstrate that there are no significant changes in PCS between the Pdx bound and Pdx free forms of P450cam. Moreover, the predicted PCS for the closed state of P450cam fit better to the experimental data than the PCS predicted for the open state of P450cam in complex with Pdx as demonstrated by Q values of 0.159 and 0.321 for the predicted PCS of the closed and open states, respectively. The predicted PCS for L204 are considerably higher than the PCS observed for both Pdx-free and Pdx bound P450cam. This residue is close to the lanthanide and very sensitive to the precise tensor orientation. For objectivity the PCS in Figure 5.8 were calculated with the tensor derived for the entire P450cam dataset (196). The optimised tensor (Figure 5.1) only deviates by a few degrees, yet yields an excellent prediction for the PCS of L204. Moreover, PCS for L169 and L324 (marked with asterisks in Figure 5.8) could not be accurately determined. The peaks for these resonances in the spectrum of the paramagnetically tagged, Pdx bound P450cam were extremely weak and, therefore, the peak positions were deemed unreliable. Further experiments are required to determine the effects of Pdx binding on these two residues.

As with the previously published models, a considerable quantity of K^+ ions was present in the NMR samples and K^+ has been implicated in stabilising the closed state of this enzyme (199). This is the first study of this process that has been performed under ambient conditions in solution. The open state observed by both X-ray crystallography and proposed by DEER EPR spectroscopy may be an artefact of

the methodologies, such that the protein was trapped in the open state during crystallisation or freezing to 30 K for the X-ray and EPR experiments, respectively. Moreover, the Resonance Raman spectroscopy experiments that have been performed on this system were carried out at 213 K and therefore the same effect may have occurred.

The PCS results presented in this chapter provide strong evidence that P450cam does not adopt an open conformation when Pdx binds. This is in direct contradiction to the recently published models obtained from EPR and X-ray crystallography. However, significant chemical shift perturbations were observed for both diamagnetically and paramagnetically tagged samples, when Pdx was titrated into P450cam. It is therefore possible that the CSPs observed are due to a change in the spin state of the haem and not a conformational change: Further experimentation is required to validate this hypothesis.

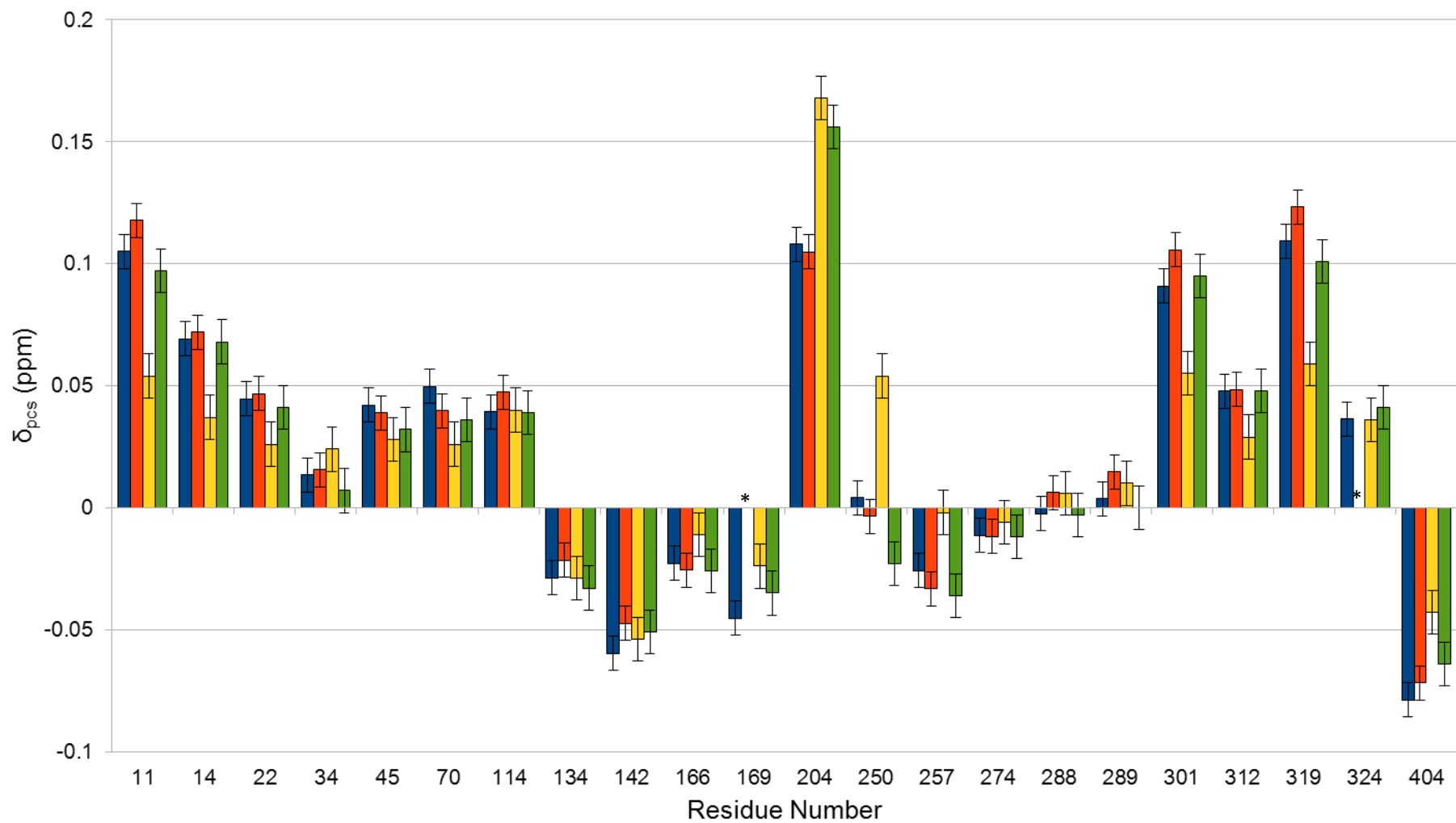


Figure 5.8 PCS for P450cam in the absence (blue) and presence (red) of Pdx and predicted PCS for the closed (yellow) and open (green) states of P450cam. Asterisks indicate PCS that could not be determined.

Chapter 6

PARAssign – paramagnetic NMR assignment on the basis of pseudocontact shifts

The results presented in this chapter have been published as:

Skinner, S. P., Moshev, M., Hass, M. A. S., Keizers, P.H. J. and Ubbink M.
(2013) *J Biomol NMR* **55**:379-89

Abstract

The use of paramagnetic NMR data for the refinement of structures of proteins and protein complexes is widespread. However, the power of paramagnetism for protein assignment has not yet been fully exploited. PARAssign is software that uses pseudocontact shift data derived from several paramagnetic centres attached to the protein to obtain amide and methyl assignments. The ability of PARAssign to perform assignment when the positions of the paramagnetic centres are known and unknown is demonstrated. PARAssign has been tested using synthetic data for methyl assignment of a 47 kDa protein, and using both synthetic and experimental data for amide assignment of a 14 kDa protein. The complex fitting space involved in such an assignment procedure necessitates that good starting conditions are found, both regarding placement and strength of paramagnetic centres. These starting conditions are obtained through automated tensor placement and user-defined tensor parameters. The results presented herein demonstrate that PARAssign is able to successfully perform resonance assignment in large systems with a high degree of reliability. This software provides a method for obtaining the assignments of large systems, which may previously have been unassignable, by using 2D NMR spectral data and a known protein structure.

Introduction

NMR spectroscopy is an invaluable technique to obtain structural and dynamic information about proteins and, thereby, enables a greater understanding of the functions that proteins carry out. A prerequisite for any detailed NMR study is that NMR assignments must be obtained for the nuclei of the protein in question. Traditionally, heteronuclear multidimensional (3D/4D) NMR spectra are used to obtain this information (16). For small proteins (<20 kDa), these experiments are sufficient to assign the nuclei. Larger proteins present more of a challenge due to their relaxation properties and spectral crowding. The line broadening problems can be overcome, for a large part, using TROSY and CRINEPT based experiments (14, 203-205).

Isotope labeling of methyl groups for larger proteins and protein complexes has also been used to circumvent the difficulties encountered when applying multidimensional NMR experiments on uniformly labeled proteins (105). Moreover, the internal mobility of methyl groups causes slow relaxation of methyl protons, and therefore sharp lines are observed in the resulting spectra. Selective methyl labeling does not provide sequential information, but assignment can be obtained, for example by selective mutation of a methyl containing amino acid for another, as was shown for the 20S proteasome (206). The method of selective mutation has been developed into a high-throughput technique, which combines automated site-directed mutagenesis, residue-type-specific-isotope labeling and fast NMR experiments (207). An additional approach is the so-called 'divide and conquer' approach (208), which involves assignment of nuclei of parts in a large complex and then transfer of the assignments to the spectra of the entire complex. However, many of these techniques require high concentrations of protein. Paramagnetic NMR can be complementary, or even an alternative to many of these methods. One of the major advantages of using paramagnetic NMR, in particular pseudocontact shifts (PCSs), for large proteins and protein complexes is that concentrations as low as 10 μ M can be used to obtain data of sufficient quality to observe these effects (209). Furthermore, pseudocontact shifts have already proven to be very useful in the assignment and structure solution of proteins using solid-state NMR techniques (210, 211).

Incorporation of a paramagnetic lanthanide ion either into a natural metal binding site or via an attached lanthanide binding tag (52-56, 58, 59, 212, 213) gives rise to observable paramagnetic effects such as PCSs, residual dipolar couplings (RDCs) and paramagnetic relaxation enhancements (PREs). PCSs are particularly useful due to the fact that they are easy to measure and provide long-range structural information. A PCS is a change in the observed Larmor frequency of a nuclear spin resulting from the time-averaged dipolar interaction between the observed nucleus and the anisotropy of the static magnetic moment of an unpaired electron spin in the paramagnetic centre.

PCSs depend on the distance between the centre and the nucleus in an anisotropic fashion and are described by a magnetic susceptibility tensor (χ tensor). In order to calculate PCSs theoretically for nuclei in a known 3D structure, it is necessary to know eight parameters, namely, the position of the paramagnetic centre (three Cartesian coordinates), the orientation of the χ tensor relative to the molecular frame (three Euler angles) and two anisotropy parameters that represent the axial and rhombic components of the χ tensor (214).

With many software packages the χ tensor can be determined 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 χ tensor fit, using a known position of the paramagnetic centre, or a complete eight-parameter fit. In all cases, the NMR assignments of the spectra of both the diamagnetic and paramagnetic protein samples are required to perform the fitting. Software packages to enable assignment of protein nuclei through the use of PCS data have been developed: Platypus (66), Echidna (67), and Possum (68). Platypus provides backbone amide assignment for resonances in the spectra of both the diamagnetic and paramagnetic samples using a known probe position, while simultaneously fitting the five tensor parameters. This has only been demonstrated, however, using residue-selectively ^{15}N -labeled samples and a small data set of 20 resonances. Echidna provides backbone amide assignment of a paramagnetic sample based on the spectral assignment of the equivalent diamagnetic sample, by fitting the magnitudes and Euler angles of the χ tensor: The metal position must be known for this procedure to be carried out. Possum is a method developed to automatically assign methyl groups of a protein using the resonance frequencies of the diamagnetic and paramagnetic samples, where the position of the paramagnetic centre, as well as the orientation and magnitudes of the tensor are known beforehand. PARAssign is a tool wherein the assignment of the nuclei of large proteins can be achieved using only 2D spectra and a three-dimensional structure of the

protein of interest with only the resonance frequencies observed for the diamagnetic and paramagnetic samples. In contrast to all previously published software, PARAssign uses the data obtained with several paramagnetic centres to determine the assignments. Thus, for each nucleus that needs to be assigned in the diamagnetic sample, a set of PCSs is available, providing sufficient data for a novel and efficient method for simultaneous χ tensor refinement and assignment of the nuclei. PARAssign requires neither prior knowledge of the positions of the paramagnetic centres the tensor orientations or any assignment information, only estimates of the magnitudes of the paramagnetic tensors, and in the case of methyl assignment, specification of the residue type.

Extensive testing of the program was undertaken using synthetic $^1\text{H}^{\text{N}}$ PCS data for pseudoazurin (PAZ) (125 residues) and P450cam (414 residues) to which selective methyl labeling schemes were applied. The robustness of the algorithm was further demonstrated using PAZ experimental data. All data were based on the use of Caged Lanthanide NMR Probe 5 (CLaNP-5) (27, 212). However, the use of the software is not restricted to data acquired with this probe and the use of other paramagnetic probes, reviewed elsewhere (215), is also possible. It should be noted though that many tags attached via a single arm to the protein induce much weaker paramagnetic effects, probably as a consequence of averaging effects due to considerable mobility of the tag relative to the protein. A large amplitude of tag movements could compromise the reliability of the assignment.

Theory

A PCS can be described by a second rank magnetic susceptibility tensor as described by Equation 1.5 (see *chapter 1*). Equation 1.5 can be fitted using a five parameter fit, such that $\delta_{\text{PCS}} = f(\Delta\chi_{\text{ax}}, \Delta\chi_{\text{rh}}, \alpha, \beta, \gamma)$ or an eight parameter fit, such that $\delta_{\text{PCS}} = f(\Delta\chi_{\text{ax}}, \Delta\chi_{\text{rh}}, \alpha, \beta, \gamma, x, y, z)$, where α , β and γ are three Euler angles used to rotate \vec{r}_x , \vec{r}_y and \vec{r}_z and x , y and z correspond to the Cartesian coordinates of the paramagnetic centre. Using the Z-Y-Z convention, a

rotation matrix, $R = R_z(\alpha)R_y(\beta)R_z(\gamma)$ can be constructed to use these angles in a fitting procedure. The fitting procedure employed in PARAssign is a sequential least squares programming procedure (216) as implemented in Scipy 0.8.0 (<http://www.scipy.org>). The target function used in this fitting procedure is:

$$\frac{1}{N} \sum_i^N (\delta_{\text{PCS},i}^{\text{pred}} - \delta_{\text{PCS},i}^{\text{exp}})^2 \quad (6.1)$$

where $\delta_{\text{PCS},i}^{\text{pred}}$ and $\delta_{\text{PCS},i}^{\text{exp}}$ represent the predicted and experimental PCSs, respectively, and N is the total number of PCSs used for the minimization.

Each paramagnetic centre is placed on the protein according to a set of vector equations based on the refined tensor positions of CLaNP-5 on PAZ (27) with some modifications (Figure 6.1) as described below. The x direction of the tensor is defined as the unit vector between the two C_α atoms of the attachment site:

$$\vec{x} = \vec{C}_{\alpha 2} - \vec{C}_{\alpha 1} \quad \hat{x} = \frac{\vec{C}_{\alpha 2} - \vec{C}_{\alpha 1}}{|\vec{C}_{\alpha 2} - \vec{C}_{\alpha 1}|}$$

Two unit vectors both perpendicular to \hat{x} are defined to form a plane:

$$\hat{f}_1 = \frac{\vec{x} \times \vec{v}_1}{|\vec{x} \times \vec{v}_1|}$$

$$\hat{f}_2 = \hat{f}_1 \times \hat{x}$$

The average vector between the two $C_\alpha - C_\gamma$ bond vectors, \vec{a} is calculated:

$$\vec{v}_1 = \frac{\vec{C}_{\gamma 1} - \vec{C}_{\alpha 1}}{|\vec{C}_{\gamma 1} - \vec{C}_{\alpha 1}|}$$

$$\vec{v}_2 = \frac{\vec{C}_{\gamma 2} - \vec{C}_{\alpha 2}}{|\vec{C}_{\gamma 2} - \vec{C}_{\alpha 2}|}$$

$$\vec{a} = \frac{1}{2}(\hat{v}_1 + \hat{v}_2)$$

This average vector, \vec{a} , is projected onto the plane created by \hat{f}_1 and \hat{f}_2 :

$$p_1 = \hat{f}_1 \bullet \vec{a}$$

$$p_2 = \hat{f}_2 \bullet \vec{a}$$

These two projections are used to determine the z direction of the tensor:

$$\hat{z} = \frac{(p_1 \hat{f}_1) + (p_2 \hat{f}_2)}{|(p_1 \hat{f}_1) + (p_2 \hat{f}_2)|}$$

The y direction of the tensor is determined by the cross product of \hat{z} and \hat{x} :

$$\hat{y} = \hat{z} \times \hat{x}$$

A point, δ , is calculated, which lies a defined distance, d, from each of the C α atoms of the attachment site and the midpoint of the vector between the two

C α atoms, \vec{b} , is determined :

$$\delta = \sqrt{d^2 - \left(\frac{1}{2}|\vec{x}|\right)^2}$$

$$\vec{b} = \frac{1}{2}(\vec{C}_{\alpha 1} + \vec{C}_{\alpha 2})$$

The position of the paramagnetic centre, L is determined from \hat{z} , δ and \vec{b} , and the initial orientation the magnetic susceptibility tensor, \mathbf{O} , is determined from \hat{x} , \hat{y} and \hat{z} :

$$\vec{L} = \hat{z}\delta + \vec{b}$$

$$\mathbf{O} = \|\hat{x}, \hat{y}, \hat{z}\|$$

If a single-armed probe is used, its position can be user-defined or calculated by PARAssign. The \vec{r}_x vector of the paramagnetic tensor is defined as the direction of the C α -C β bond vector; the \vec{r}_z vector is defined perpendicular to \vec{r}_x and to point away from the protein by using the C α -C γ bond of the residue to provide this direction. The \vec{r}_y vector is derived as the cross product of the \vec{r}_x and \vec{r}_z vectors. The distance between the C α atom of the attachment site and the paramagnetic centre is user-defined.

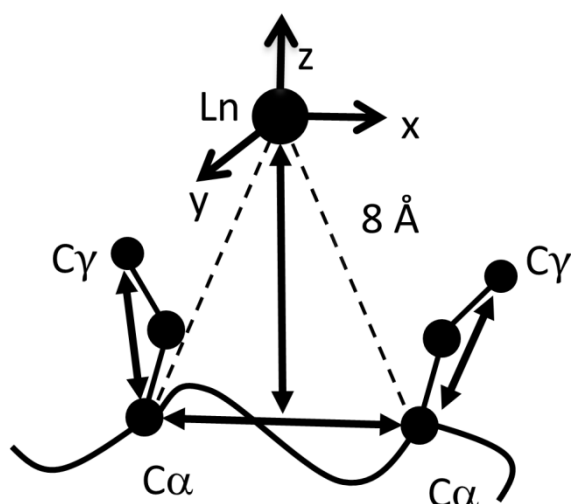


Figure 6.1 Placement of the paramagnetic centres relative to the attachment site for a two-armed probe using a modified method to that previously published (27).

Peaks of nuclei close to the paramagnetic centre are broadened beyond detection as a result of paramagnetic relaxation enhancement and are disregarded. The cutoff distance can be user-specified or calculated by PARAssign, using the equation for Curie spin relaxation (67):

$$r_{\text{cutoff}} = \sqrt[6]{\frac{1}{5\pi PRE} \left(\frac{\mu_0}{4\pi}\right)^2 B_0^2 \gamma_x^2 \frac{(g_e \mu_B)^4 S^2 (S+1)}{(3k_B T)^2} \left(4\tau_r + \frac{3\tau_r}{1 + \omega_x^2 \tau_r^2}\right)} \quad (6.2)$$

As an indication, a PRE of $\approx 200 \text{ s}^{-1}$ will usually lead to broadening beyond detection.

The Hungarian method for minimal cost assignment (217) is used to perform the assignment of the experimental PCSs, using a Q score cost function to populate the assignment matrix:

$$Q = \frac{1}{N} \sum_{i=1}^N \sqrt{\frac{(\delta_{\text{PCS},i}^{\text{pred}} - \delta_{\text{PCS},i}^{\text{exp}})^2}{(\delta_{\text{PCS},i}^{\text{pred}} + \delta_{\text{PCS},i}^{\text{exp}})^2}} \quad (6.3)$$

where N is the number of paramagnetic centres. This algorithm ensures that each experimental PCS is assigned exactly once to a predicted PCS and

therefore, an assignable atom in the protein structure. The Q value is used as a filter when building the assignment matrix. Pairs of predicted and experimental PCSs whose Q score exceeds a user-defined value are excluded from the assignment matrix. In the case of methyl data, the residue type is also used to avoid incorrect assignments. The chemical shifts of methyl groups of different residues are sufficiently different for this to be done by user-defined residue typing. It should also be noted that predicted and experimental PCS with opposite signs are excluded from Q score calculation and therefore cannot be assignment possibility. This avoids any issues that would arise from the summed denominator in equation 6.3.

PARAssign imports peak lists from spectra of both diamagnetic and paramagnetic forms of the protein from either a [$^{13}\text{C},^1\text{H}$] or a [$^{15}\text{N},^1\text{H}$] HSQC spectrum. Using the predicted PCSs, generated from the initial paramagnetic tensor orientations and magnitudes, an algorithm based on Bayesian statistics is used to match the peaks observed for the diamagnetic form of the protein to those for the paramagnetic form of the protein to produce PCSs. For every possible combination of diamagnetic and paramagnetic peaks in the peak lists, the proton and heteroatom pseudocontact shifts are calculated along with the quotient between these two values. The quotients between all predicted proton and heteroatom pseudocontact shifts, based on the initial placement of the paramagnetic centres, are also calculated. A probability score is then calculated for each pair of matched peaks. If the probability score exceeds a threshold level (100 by default) this pairing is added to the list of possible pseudocontact shifts in the data set. From here, a Bayesian matching procedure is carried out using the associated probabilities and this produces a set of pseudocontact shifts for each peak list imported, from which assignment and refinement can be carried out or spectral analysis to identify more pseudocontact shifts can be initiated. This procedure is executed for the data for each paramagnetic centre and the combined sets of PCSs are then used for initial assignment. In this matching algorithm, the PCSs for both the proton and heteroatom dimensions are used, along with their ratios. These PCSs are then used for paramagnetic tensor refinement and protein

assignment. The matching procedure produces on average 85% correct matches based on the initial tensor orientation. These PCSs can be used as a starting point for spectral analysis to identify additional PCSs, which can then be subsequently used for assignment of the protein nuclei.

Using the initial placement of the paramagnetic centre and orientation of the tensors, the protein structure and the user-defined $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ values, a set of predicted PCSs is calculated for each paramagnetic centre. All PCSs that are less than ± 0.02 are excluded from the datasets, since these values could be too small to be accurately measured experimentally. The protein structure is imported as a PDB file using the PDBparser module of Biopython (218, 219) and protonated. Protein structures can also be downloaded from the RCSB within PARAssign. It is possible that the initial tensor calculated by PARAssign may not have an optimal starting orientation and in order to establish whether it can be improved, two sequential searches of the α and β Euler angles are carried out. The signs of the PCSs are an indicator for the orientation of the tensor. To find an approximate orientation the total number of positive PCSs in the experimental and predicted sets is compared. So, using the predicted PCSs, the total number of positive PCSs of all relevant nuclei in the protein for each paramagnetic centre is calculated for the predicted datasets and compared with that of the experimental datasets. The difference between the fractions of positive PCSs in the predicted and experimental datasets is calculated as a cost, c :

$$c = \frac{N_{\text{positive}}^{\text{predicted}}}{N_{\text{all}}^{\text{predicted}}} - \frac{N_{\text{positive}}^{\text{experimental}}}{N_{\text{all}}^{\text{experimental}}} \quad (6.4)$$

This cost is then used as a “goodness of fit” statistic for the sequential searches of the α and β Euler angles. This approach is possible, because the value of c is strongly dependent on the value of the β Euler angle (Figure 6.2). By rotating the β Euler angle over the range $\pm\frac{1}{2}\pi$ from its initial starting value calculated in PARAssign, it is evident that for four proteins of varying size, a similar pattern of sinusoidal variation is observed for all of the tested proteins.

This is a good indication that using the number of positive PCSs is a valid manner in which to refine Euler angles prior to initial protein assignment. Due to the fact that rotation operators do not commute, two searches need to be carried out, one beginning with rotation of the α angle followed by rotation of the β angle (' α/β ') and another beginning with rotation the β angle followed by rotation of the α angle (' β/α '). The ' α/β ' sequential search is performed as follows. Each tensor frame is rotated by a $-\frac{1}{2}\pi$ α angle and then steps of 0.1 radians are performed until a rotation of $\frac{1}{2}\pi$ has been achieved. At each step of the walk, c is used to provide a "goodness of fit" value. The α angle corresponding to the lowest value of c is then selected as the optimal starting orientation for the α angle and denoted α^{opt} . The α angle can take values in the range $-\frac{1}{2}\pi \leq \alpha \leq +\frac{1}{2}\pi$ because the rhombic component of the χ tensor has π radian symmetry and therefore, only half of this angular space needs to be searched.

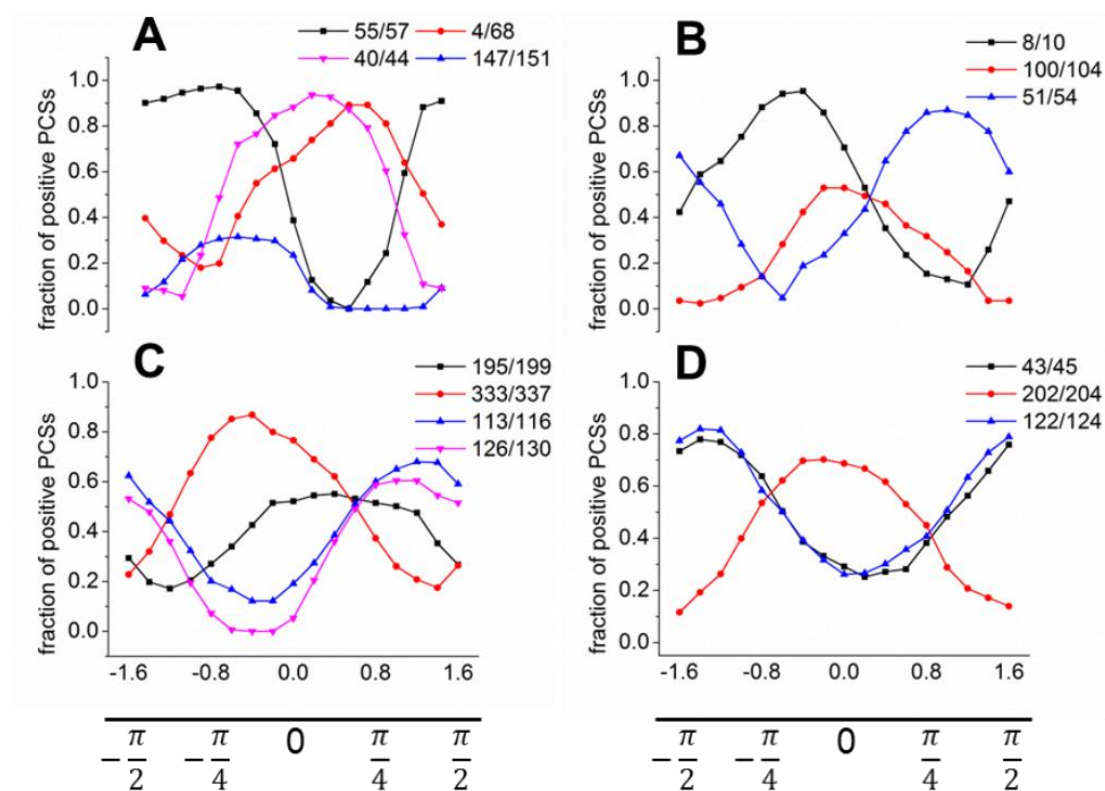


Figure 6.2 Positive PCS. The fraction of predicted positive PCSs is plotted as a function of the β angle (radians) for four proteins: T4 lysozyme (A), PAZ (B),

cytochrome P450cam (C) and Green Fluorescent Protein (D). The lines represent different attachment sites of the paramagnetic centre.

The synthetic datasets for the eight parameter fits were produced using the datasets from the five parameter fitting and random x, y and z positions generated from a grid of $\pm 3 \text{ \AA} \times \pm 3 \text{ \AA} \times \pm 3 \text{ \AA}$ around the initial position for each paramagnetic centre calculated by PARAssign.

The β angle sequential search is then performed analogously, starting from the frame oriented by the angle α^{opt} and producing the angle β^{opt} . The ' β/α ' search is performed in the same way, except the β angle is rotated first. The ' α/β ' and ' β/α ' searches each provide an "optimal" starting orientation and the search that provides the lowest value of c is used to rotate the tensor to the corresponding initial orientation. Only two angles need to be rotated because the orientation of one of the axes is always fixed in this searching procedure. Moreover, if the γ Euler angle were included in the ' β/α ' search, this would be tantamount to performing the last search twice, since both α and γ determine the orientations of the x and y axes of the tensor frame. Using the orientations calculated from the individual sequential searches, the predicted PCSs are recalculated and the initial assignment of the protein is determined. Each paramagnetic tensor is then fitted based on this assignment.

The paramagnetic tensors can be fitted using a constrained or unconstrained five or eight parameter fit. Each paramagnetic tensor is fitted to its assigned dataset individually and the PCSs associated with that tensor are recalculated after application of the resulting fitting parameters. These predictions are then used to populate the assignment matrix required by the Hungarian method and an assignment based on these predictions is carried out. This procedure of fitting and assignment is carried out iteratively until a convergence limit is reached or a user-defined number of minimization steps has been performed. PARAssign then outputs the predicted and experimental values per paramagnetic centre along with the predicted assignment, the deviation between predicted and experimental PCSs for each paramagnetic

centre as well as the average Q score per residue. A PDB file containing the final paramagnetic centre positions and tensor orientations is also written (using Scientific Python 2.8) and, in addition, PARAssign produces a scatter plot showing the fit of predicted to experimental PCSs per paramagnetic centre, along with the Pearson Product Moment Correlation Coefficient (Pearson's r), the linear least squares regression line fitted to the data and a line of $y = x$ for comparison. After the final fitting and assignment step, all predicted PCSs are included in a subsequent assignment, such that all PCSs $< \pm 0.02$ are re-introduced into the dataset. In addition, a user-defined number of alternative assignments are produced by removing current assignments as possibilities from the assignment matrix and performing the Hungarian Algorithm again. This ensures that any alternative assignments will be different from previous assignments and also that the assignment corresponding to the next closest Q score to the previous assignment will be selected as an alternative assignment. This procedure can be carried out as many times as the user determines. All assignments have a reliability indicator associated with them. The trimmed average Q score is calculated using a 5% upper limit for exclusion and this is used to determine assignment reliability; if the moduli of all PCSs used in an assignment of a nucleus are more than 0.02 and the average Q score is less than the trimmed average Q score, this assignment is deemed a highly reliable '***' (two-star) assignment. If only one of the conditions is met, it is deemed a reliable '**' (one-star) assignment and if neither conditions are met, the assignment is deemed unreliable. The complete algorithm is depicted in Figure 6.3.

Two proteins were selected to test the robustness of the PARAssign algorithm. Pseudoazurin (PAZ), a blue copper protein of 13.5 kDa (PDB code 1PY0) (220) with Yb-CLaNP-5 attached at three positions was used to test the backbone amide assignment capabilities and cytochrome P450cam, a 46.6 kDa heme protein (PDB code 1DZ4) (221) with Tm-CLaNP-5 attached at four positions was used to test the methyl assignment capabilities using selective labeling schemes (Ile-C δ 1, 15 residues, Leu-C δ 1, 26 residues, Val-C γ 1, 20 residues, Ala-C β , 23 residues, Leu-C δ 1/Val-C γ 1, 46 residues, Ile-C δ 1/Leu-

C δ 1/Val-C γ 1, 61 residues). Five synthetic datasets and one experimental dataset were used for testing with PAZ and five synthetic datasets were used for testing with P450cam. A dataset represents all sets of PCSs generated from all paramagnetic centres in a certain position, with a certain tensor size and orientation. Two separate tests were carried out, namely five parameter fitting and eight parameter fitting.

The synthetic datasets for five-parameter fitting were generated using a random number generator to produce five sets of α and β Euler angles for each protein with the ranges $-\pi \leq \alpha \leq +\pi$ and $-\frac{1}{2}\pi \leq \beta \leq +\frac{1}{2}\pi$, as well as random $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ values within a range based on published values (27) for Yb-CLaNP-5 ($7 \times 10^{-32} \text{ m}^3 \leq \Delta\chi_{ax} \leq 10 \times 10^{-32} \text{ m}^3$ and $1 \times 10^{-32} \text{ m}^3 \leq \Delta\chi_{rh} \leq 4 \times 10^{-32} \text{ m}^3$) and Tm-CLaNP-5 ($45 \times 10^{-32} \text{ m}^3 \leq \Delta\chi_{ax} \leq 55 \times 10^{-32} \text{ m}^3$ and $9 \times 10^{-32} \text{ m}^3 \leq \Delta\chi_{rh} \leq 11 \times 10^{-32} \text{ m}^3$). The synthetic datasets for the eight parameter fits were produced using the datasets from the five parameter fitting and random x, y and z positions generated from a grid of $\pm 3 \text{ \AA} \times \pm 3 \text{ \AA} \times \pm 3 \text{ \AA}$ around the initial position for each paramagnetic centre calculated by PARAssign.

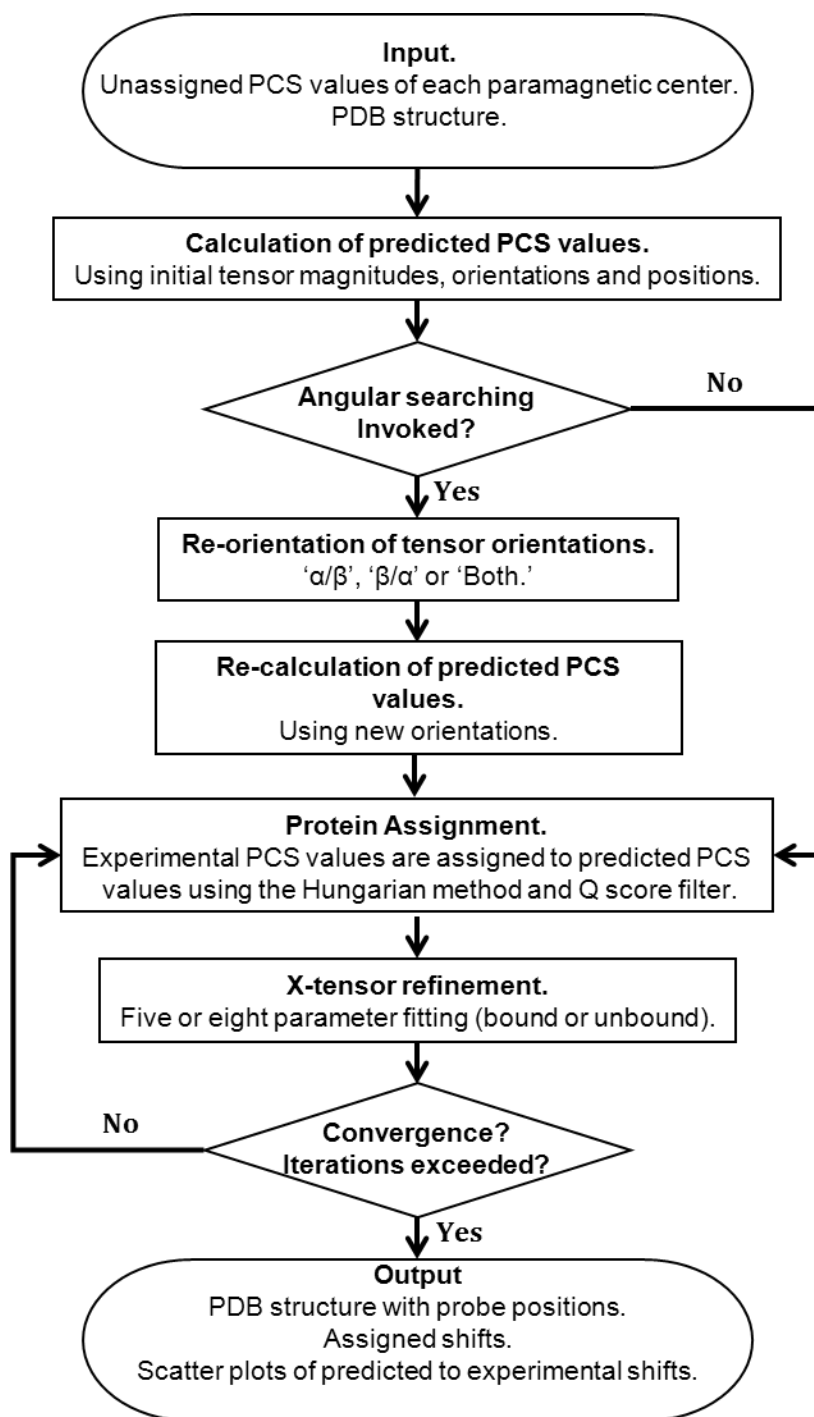


Figure 6.3 Flowchart of the PARAssign Algorithm.

These random values were used to re-orient the tensor and, where necessary, re-position the paramagnetic centre and sets of PCSs were calculated based on the new orientation and position. Noise was also added to each PCS as a random value in the range $-0.01 \leq x \leq 0.01$ ppm to

represent experimental error. This error value was considered appropriate since all data sets used were ^1H PCS data. For heteronuclear data larger error margins may be required.

These datasets were then used as target (“quasi-experimental”) data for PARAssign and the ability PARAssign to refine the tensor and assign each dataset was tested. In all tests, the starting values for $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ were $9.0 \times 10^{-32} \text{ m}^3$ and $2.0 \times 10^{-32} \text{ m}^3$, respectively for Yb-CLaNP and $50.0 \times 10^{-32} \text{ m}^3$ and $10.0 \times 10^{-32} \text{ m}^3$, respectively for Tm-CLaNP. The cutoff distance for excluding residues whose signals would be broadened beyond detection was set to 10 Å for Yb-labeled PAZ and 16 Å for Tm-labeled P450cam and all PCSs that would theoretically be unobservable were excluded from all synthetic datasets. The starting position and orientation were always those obtained using the procedure described above.

Results and Discussion

Synthetic data – PAZ backbone amides

The PARAssign algorithm performs simultaneous χ tensor fitting and assignment of protein nuclei. There are two types of fitting available in PARAssign; namely five parameter fitting, with fixed positions for the paramagnetic centres, and eight parameter fitting. The user can set boundaries for $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ and the metal position if a bound fit is selected. Initially, five parameter fitting and assignment was tested using five synthetic datasets generated from published $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ values and random Euler angles. All initial assignments were generated without applying any angular searching (‘None’), such that the initial position derived from placing the tensor was used to perform the assignment. The Q score filter (Equation 4) was set to 0.25 for all tests and incrementally increased by 0.25 to a maximum of 0.75, if the number of amides assigned in the final output was below the maximal assignable for the protein, taking into account the line broadening caused by the probe (see above). The Q score filter for

assignment is set by the user and is not incremented automatically. In cases where the final assignment was not optimal using the initial tensor placement as the starting position, two sequential Euler angle searches were included separately, namely an ' α/β ' and a ' β/α ' search (see Initial Assignment section) to derive a better initial orientation for assignment and the Q score filter values were incremented as previously stated. If application of neither of the searches to the initial starting orientation resulted in an optimal assignment, both the ' α/β ' and ' β/α ' searches were applied ('Both') in the same assignment run and the search that produced the lowest cost, c (equation 5) was selected to move the tensor to its starting orientation and the Q score filter was incremented; this being the final test for each datasets for which an optimal assignment had not been achieved. The five synthetic datasets yielded diverse results (Table 6.1). For all datasets, the initial orientation of the probe was not sufficient to obtain an optimal assignment, shown by the fact that a search was required in all cases. The starting point for assignment and tensor fitting is a strong determining factor of the final assignment results, as shown in Table 6.1, along with the Q score filter value used for assignment.

Table 6.1 Summary of results from testing of five parameter fitting using PAZ backbone amide synthetic data.

Dataset	Best search	Total assignable amides	Q filter	Total assigned amides	Correct/Assigned		
					2*	1*	0*
1	Both	85	0.75	82/83	51/51	18/18	13/14
2	β/α	85	0.25	78/80	47/47	25/25	6/8
3	Both	85	0.75	83/83	45/45	27/27	11/11
4	α/β	85	0.25	76/78	43/43	26/26	7/9
5	β/α	85	0.50	84/84	43/43	33/33	8/8

The diversity in "Best Search" used to obtain the optimal assignment for each of the five datasets indicates that all search possibilities should be used and a suitable Q filter should accompany each search. In some situations, a low Q

score value of 0.25 was sufficient to obtain a nearly complete and accurate assignment of PAZ amides, however, for other datasets, a higher Q score value was required to enable potentially erroneous assignments to be used to guide the minimization and consequently, result in those erroneous assignments being corrected. An assignment result was deemed optimal based on whether the maximum number of assignable residues had been reached, the value of the target function per paramagnetic centre and the overall fitting statistics for each mutant, i.e. the closeness of the least-squares regression line to the line of $y = x$ and the corresponding Pearson Product Moment Correlation Coefficient (PPMCC). Moreover, the reliability of an individual assignment was determined based on whether the average Q score was less than the trimmed average Q score for that assignment and whether the moduli of all PCSs for that amide were greater than 0.02. An example of the final output from PARAssign, based on experimental PAZ data (*described below*), including the reliability indicators is shown in Table 6.3. The five parameter fitting tests showed that three quarters of the assignments achieved in all cases were in the most reliable category (2*). The 1* assignments were often put in this class due to the fact that one or more of the PCS for that amide were below the threshold for a reliable assignment (0.02). All starred (2* and 1*) assignments were correct for this test and all others reported below. Within the 0* assignments produced by PARAssign at most 20% false positives (wrong assignments) were found in these tests, indicating that 0* assignments are useful suggestions, but need to be checked with other methods. This shows that PARAssign not only produces many reliable assignments of amides on the basis of PCS data, but also that those that are unreliable can be easily identified in the final output. The eight parameter fitting abilities of PARAssign using PAZ synthetic datasets were carried out in the same manner as the five parameter fitting tests. These results are shown in Table 6.2.

Table 6.2 Summary of results from testing of eight parameter fitting using PAZ backbone amide synthetic data.

Dataset	Best search	Total assignable amides	Q filter	Total assigned amides	Correct/Assigned		
					2*	1*	0*
1	Both	85	0.75	82/83	51/51	18/18	13/14
2	$\alpha + \beta$	85	0.5	78/80	44/44	25/25	9/11
3	Both	85	0.75	83/83	45/45	27/27	11/11
4	$\alpha + \beta$	85	0.25	76/78	43/43	26/26	7/9
5	$\beta + \alpha$	85	0.50	84/84	43/43	33/33	8/8

These results show that PARAssign can perform assignment of proteins to an almost identical degree of accuracy whether the positions of the paramagnetic centres are known or unknown.

Experimental data – PAZ backbone amides

A set of previously published experimental PCS data for PAZ (27) was used to test the PARAssign algorithm with a real data set derived from [¹⁵N,¹H]-HSQC spectra. The PCSs used for the assignment were produced both via the peak matching algorithm and from manual identification of the PCSs from the spectra. The PCSs identified by the peak matching algorithm were used as a starting point for building the PCS lists used in assignment and additional PCSs were identified by spectral analysis, providing a comprehensive set of experimental data for the assignment of PAZ amides. The assignment was carried out using a restrained eight-parameter fit. This was performed on a $\pm 3\text{Å} \times \pm 3\text{Å} \times \pm 3\text{Å}$ grid, with the tensor magnitudes restricted to $7 \times 10^{-32} \text{ m}^3 < \Delta\chi_{ax} < 10 \times 10^{-32} \text{ m}^3$ and $1 \times 10^{-32} \text{ m}^3 < \Delta\chi_{rh} < 4 \times 10^{-32} \text{ m}^3$ and the Euler angles restrained using the following limits: $-\pi < \alpha/\gamma < \pi$ and $-\frac{1}{2}\pi < \beta < \frac{1}{2}\pi$. The starting values for $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ were $9.0 \times 10^{-32} \text{ m}^3$ and $2.0 \times 10^{-32} \text{ m}^3$, respectively.

Using this real data, only three amides were mis-assigned out of the 57 assignable amides. The reason that only 57 of the amides were assignable and not 85 as stated for the synthetic data was not related to the PARAssign software, but due to the limited quality of one experimental data sets, such that only a subset of all resonances could be observed. Of the correct assignments, 32 were 2* assignments, 17 were 1* assignments and five had no star. All three incorrect assignments were not starred and therefore deemed unreliable (Table 6.3). It is concluded that the PARAssign algorithm can handle and successfully assign amides using experimental data, as evidenced by the reliability of the assignments obtained and the fitting statistics, Figure 6.4, which show that the fits of experimental to predicted data were excellent.

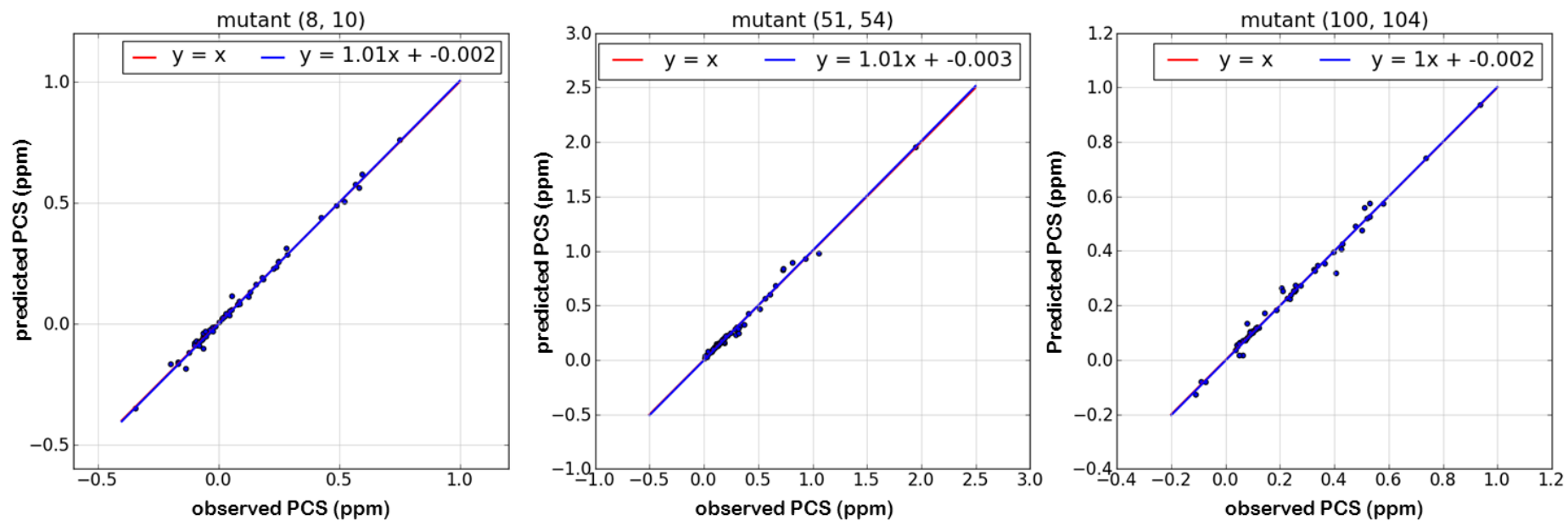


Figure 6.4 Scatter plots of predicted against observed ¹H PCSs for PAZ (27) after fitting and assignment by PARAssign using the structure of PAZ (PDB entry 1PY0) (220).

Table 6.3 Extract from the final output of PAZ experimental assignment, showing three mis-assigned residues as highlighted in bold. In all three cases, the assignments are deemed unreliable by PARAssign as shown by the fact that they are not starred assignments.

Residue number	Residue name	Assigned index	Predicted PCS	Experimental PCS	Deviation (ppm)	Predicted PCS	Experimental PCS	Deviation (ppm)	Predicted PCS	Experimental PCS	Deviation (ppm)	Average Q score	Reliability
1	GLU	1	-0.087	-0.091	0.0041	0.050	0.055	0.0052	0.037	0.035	0.0023	0.0346	**
2	ASN	2	-0.133	-0.187	0.0544	0.048	0.063	0.0148	0.044	0.049	0.0047	0.1178	*
3	ILE	3	-0.164	-0.160	0.0043	0.130	0.136	0.0062	0.080	0.080	0.0001	0.0124	**
4	GLU	4	-0.339	-0.351	0.0123	0.103	0.107	0.0044	0.074	0.071	0.0030	0.0197	**
33	PHE	33	-0.081	-0.092	0.0106	0.133	0.127	0.0061	0.112	0.111	0.0011	0.0300	**
38	LYS	62	0.030	0.028	0.0018	0.033	0.045	0.0115	0.064	0.016	0.0480	0.2594	
59	LYS	59	0.055	0.051	0.0039	0.021	0.024	0.0031	0.102	0.103	0.0010	0.0369	**
64	TYR	64	-0.073	-0.070	0.0034	0.077	0.071	0.0056	0.097	0.097	0.0003	0.0209	**
66	LEU	66	-0.051	-0.054	0.0026	0.187	0.193	0.0055	0.112	0.116	0.0038	0.0186	**
73	ALA	98	0.005	0.005	0.0002	1.065	0.976	0.0892	0.408	0.318	0.0900	0.0611	
93	GLY	94	-0.061	-0.041	0.0195	0.321	0.242	0.0789	0.051	0.016	0.0352	0.2584	
95	SER	95	-0.052	-0.033	0.0192	0.237	0.244	0.0065	-0.107	-0.128	0.0214	0.1099	*
107	LYS	107	0.125	0.110	0.0153	0.098	0.145	0.0468	0.941	0.936	0.0052	0.0867	*
110	VAL	110	0.250	0.256	0.0058	0.078	0.096	0.0175	0.241	0.238	0.0033	0.0395	**

Synthetic data – P450cam

The assignment of methyl groups was also tested using PARAssign. Five and eight parameter fits were tested in the same way as for amide assignment. The results of five parameter fitting using P450cam methyl data showed diverse performance of PARAssign (Table 6.4). All datasets for single residue selective labeling did not converge to a minimum that provided an assignment. In all cases, the number of residues assigned below the Q score filter was below 80% of the total set (*data not shown*).

Table 6.4 Summary of results from testing of five parameter fitting using P450 synthetic data.

Dataset	Total assignable methyl groups	Q filter	Best Search	Total assigned methyl groups	Correct/Assigned			
					2*	1*	0*	
Ile-C δ 1/Leu-C δ 1/Val-C γ 1								
1	63	0.50	β/α	63/63	33/33	29/29	1/1	
2	63	0.75	β/α	63/63	39/39	18/18	6/6	
3	63	0.75	None	60/63	38/38	17/17	5/8	
4	63	0.75	β/α	61/63	37/37	24/24	0/2	
5	63	0.50	α/β	59/63	36/36	23/23	0/4	
Leu-C δ 1/Val-C γ 1								
1	46	0.50	β/α	43/46	19/19	24/24	0/3	
2	46	0.75	β/α	46/46	31/31	15/15	0/0	
3	46	0.75	None	45/46	31/31	13/13	1/2	
4	46	0.75	β/α	44/46	28/28	16/16	0/2	
5	46	0.50	α/β	43/46	26/26	15/15	2/5	

This shows that a simple manner to assign a protein for which the number of residues of a single type is too small, is to combine datasets of several residue types. The fraction of 2* assignments was approximately two-thirds for all except dataset 1 for both the Ile-C δ 1/Leu-C δ 1/Val-C γ 1 and Leu-C δ 1/Val-C γ 1 synthetic data. Furthermore, 1* assignments accounted for most of the remaining methyls assigned

by PARAssign and these were deemed 1* due to the fact that the absolute values of some PCSs in the dataset were less than 0.02. The testing of eight parameter fitting was carried out in an identical manner to the five parameter fitting and a summary of the results is shown in Table 6.5.

Table 6.5 Summary of results from testing of eight parameter fitting using P450 synthetic data.

Dataset	Total assignable methyl groups	Q filter	Best Search	Total assigned methyl groups	Correct/Assigned		
					2*	1*	0*
Ile-C δ 1/Leu-C δ 1/Val-C γ 1							
1	63	0.50	$\beta+\alpha$	63/63	39/39	23/23	1/1
2	63	0.75	$\beta+\alpha$	63/63	38/38	20/20	5/5
3	63	0.75	None	59/63	35/35	22/22	2/6
4	63	0.75	$\beta+\alpha$	61/63	35/35	26/26	0/2
5	63	0.50	$\alpha+\beta$	63/63	40/40	23/23	0/0
Leu-C δ 1/Val-C γ 1							
1	46	0.50	$\beta+\alpha$	46/46	31/31	12/12	3/3
2	46	0.75	$\beta+\alpha$	45/46	30/30	13/13	2/3
3	46	0.75	None	45/46	26/26	17/17	2/3
4	46	0.75	$\beta+\alpha$	44/46	24/24	18/18	2/4
5	46	0.50	$\alpha+\beta$	46/46	29/29	15/15	2/2

Methyl assignment using an eight-parameter fit showed the same issue as for single residue type datasets. However, for both the combined Ile-C δ 1/Leu-C δ 1/Val-C γ 1 and the combined Leu-C δ 1/Val-C γ 1 datasets, a similar degree of success in assignment was achieved as with the five parameter fitting. The reliability of these assignments varied when compared with the reliability of those produced from five parameter fitting. In some cases, the number of correct and reliable assignments increased when eight parameter fitting was employed as opposed to five parameter fitting. This is particularly true for the Leu-C δ 1/Val-C γ 1 datasets. In two cases, datasets one and five, more correct and more 2* assignments were obtained. In addition, more correct

and 2* assignments were obtained for dataset five of the Ile-C δ 1/Leu-C δ 1/Val-C γ 1 series. It is possible that the increase in assignment accuracy is due to the fact that a larger parameter space is searched in eight parameter fitting. These data demonstrate that PARAssign is able to successfully and reliably assign methyls in a large protein with selective labeling, at least using synthetic data. Validation on experimental data sets in progress.

Chapter 7

***Paramagnetic relaxation dispersion of the
K147C/T151C mutant of T4 lysozyme***

Abstract

The use of paramagnetism for studying protein structure and protein-protein interactions has become commonplace over the past two decades, however, its use for investigating protein dynamics is less common. Proteins lacking a metal binding site require the attachment of a lanthanide binding peptide or tag. In order to study protein motions on a specific timescale, it is imperative that an attached paramagnetic probe is not moving on the same timescale. To this end, paramagnetic relaxation dispersion has been applied to T4 lysozyme, a protein for which μs – ms dynamics have reported not to exist for the wild type form. Yb-CLaNP-5 has been attached to the K147C/T151C mutant of T4 lysozyme and dynamics were observed for six resonances in the spectra of the paramagnetically labeled sample. However, this is too small a number to attribute the observed dynamics to probe movement. Consequently, it has been concluded that CLaNP-5 does not undergo μs – ms motions when attached to this mutant of T4 lysozyme and therefore, these are internal protein motions.

Introduction

T4 lysozyme is an 18.7 kDa endoacetyl-muramidase produced in the cells of *E. coli* after infection with bacteriophage T4. This enzyme is composed of a large and a small domain joined by a long helix and the active site cavity is located between the two domains. In the resting state, this cavity appears to be occluded, based on crystallographic data of the wild-type protein. 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 occurs via a hinge bending motion, which consists of a rotation of one domain relative to the other about an axis running through the interface of the two domains. This motion has been demonstrated in solution 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 $\sim 160 \mu\text{s}$ 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). FRET studies have also been used to investigate the conformational dynamics of T4 lysozyme and these have indicated that the formation of the enzyme-substrate complex involves the population of six intermediate conformations and occurs on the ms timescale. A solution NMR structure of a minor state of T4 lysozyme has been determined using chemical shift differences obtained from ^{15}N relaxation dispersion experiments and structure calculations using CS-ROSETTA (123). These studies used the L99A mutant of T4 lysozyme, for which μs -ms dynamics have previously been demonstrated (150) and, furthermore, this dynamic profile was not obtained for the wild-type protein. The absence of μs -ms motions in wild-type T4 lysozyme make it a perfect candidate to test whether an attached paramagnetic probe is moving on this timescale using 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. Moreover, anisotropic chemical shift differences and dipolar couplings can be extracted from relaxation dispersion experiments (104) and these can be of more utility in the determination of structures of lowly populated states (123). Pseudocontact shifts (PCS) resulting from a magnetic susceptibility of paramagnetic centres also contain structural information (131). PCS are determined by the position of the nucleus within a paramagnetic tensor frame of the metal and, therefore, 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 value will be observed for the nuclei in the region. Consequently, 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 this can provide a route to determining the structures of minor states in proteins. Motion detected in the manner may, however, either result from internal motions of the protein itself, or 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 the E51C/E54C mutant of pseudoazurin and the N56C/L58C mutant cytochrome *c* (133). The work of Ubbink and co-workers demonstrated that no movement on the μs - ms timescale of the attached probe was observed when it was attached to the E100C/S104C mutant of pseudoazurin. In this study, CLaNP-5 was attached to the K147C/T151C mutant of T4 lysozyme to determine if, indeed, the probe is moving on the μs - ms timescale using paramagnetic relaxation dispersion.

A method employing lanthanide ions to measure dynamics using $R_1\rho$ relaxation dispersion experiments has also been reported using the N-terminal domain of cardiac troponin C (222)

Methods and Materials

CLaNP labeling

A ^{15}N enriched sample of K147C/T151C T4 lysozyme in 20mM sodium phosphate pH = 7.5 was kindly provided by Anneloes Blok (Universiteit Leiden) and this sample was incubated using 5 mM sodium dithionite on ice for 1 hour to reduce the cysteine residues, after which the sodium dithionite was removed using a PD-10 column (GE Healthcare). Subsequently, the sample was divided into two, each sample was diluted to a concentration of 50 μM and five molar equivalents of CLaNP-5 were added to each, one contained Lu-CLaNP-5 and the other Yb-CLaNP-5. These samples were incubated at 25°C under anaerobic conditions with continuous stirring for 1 hour. CLaNP-labelled protein was purified from excess CLaNP-5 with a HiTrap SP column using 30 column volumes for a 0 to 500 mM NaCl gradient and the samples were concentrated to a volume of 280 μL . These samples were buffer

exchanged three times into 20 mM sodium phosphate, pH = 5.5, 150 mM sodium chloride. The final NMR samples contained 450 μ M T4 lysozyme, 95% H₂O / 5% D₂O, 20mM sodium phosphate, pH = 5.5, 150 mM sodium chloride. CLaNP labelling and the degree of ¹⁵N enrichment were determined by mass spectrometry.

NMR spectroscopy

All NMR spectra were acquired at 25°C on a Bruker Avance III spectrometer equipped with a TCI-Z-GRAD cryoprobe operating at a ¹H proton Larmor frequency of 600 MHz. A [¹⁵N, ¹H]-TROSY-HSQC spectrum was acquired for each sample. For each spectrum 110 and 1024 complex data points were recorded in ¹⁵N and ¹H dimensions, respectively, with 8 scans per FID, separated by a recovery delay of 1.3 s. A B_1 field strength of 6 kHz and 23 kHz were used for all ¹⁵N and ¹H hard pulses. Each spectrum was collected with a sweep width of 2.07 kHz and 8.97 kHz in the ¹⁵N and ¹H dimensions, respectively. Interleaved TROSY ¹H CPMG relaxation dispersion experiments were recorded for each sample (133) with 10 different CPMG periods containing multiples of 16 180° pulses in the forward and reverse INEPT periods and a CPMG period of 20 ms.

Spectral assignment and calculation of the $\Delta\chi$ tensor

All spectra were processed using Topspin 2.1 (Bruker Biospin) and the size of the indirect dimension was doubled using linear prediction. The spectra were analysed using CCPN analysis version 2.1.5 (175). A previously published assignment of T4 lysozyme (223) was used to assign the spectrum of the diamagnetically (Lu-labelled) sample and assignment of the spectrum of the paramagnetically (Yb-labelled) sample was performed in concert with calculation of the $\Delta\chi$ tensor (Equation 7.1) using Numbat (28).

$$\delta_{PCS} = \frac{1}{12\pi r_i^3} \left[\Delta\chi_{ax} (3(\vec{r}_z \cdot \vec{r}_i)^2 - r_i^2) + \frac{3}{2} \Delta\chi_{rh} (\vec{r}_x \cdot \vec{r}_i)^2 - (\vec{r}_y \cdot \vec{r}_i)^2 \right] \quad (7.1)$$

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 \vec{r}_x , \vec{r}_y and \vec{r}_z , represent the unit vectors that determine the orientation of the magnetic susceptibility tensor.

Initially, pseudocontact shift possibilities were determined using a peak matching algorithm supplied with PARAssign (224) and these possibilities were used in initial tensor calculation in Nubat and iterative assignment and tensor refinement were, subsequently, carried out.

Relaxation dispersion analysis

Signal intensities were measured as peak intensities calculated using the box-sum method in CCPN analysis version 2.1.5 and R_2^{eff} values were calculated using the Equation 7.2:

$$R_2^{eff} = \frac{1}{T_{relax}} \ln \frac{I_0}{I} \quad (7.2)$$

and ν_{CPMG} for each spectrum was calculated using Equation 7.3:

$$\nu_{CPMG} = \frac{1}{4\tau_{CPMG}} \quad (7.3)$$

R_2^{eff} was plotted against ν_{CPMG} to determine which signals showed significant dispersion (*data not shown*) and those were fitted using the fast-exchange Carver-Richards Equation (Equation 7.4) to determine the rate of motion, k_{ex} , the chemical shift difference between the two states, $\Delta\omega$ and the maximum R_2^{eff} rate, R_2^0 .

$$R_2^{eff} = R_2^0 + \frac{p_A p_B \Delta\omega^2}{k_{ex}} \left[1 - \frac{4\nu_{CPMG}}{k_{ex}} \tanh\left(\frac{k_{ex}}{4\nu_{CPMG}}\right) \right] \quad (7.4)$$

The populations of the two states, p_A and p_B were fixed to values of 0.5, due to the fact that these cannot be determined using the fast-exchange form of the Carver-

Richards Equation. All data fitting was performed using non-linear least squares fitting as implemented in Origin 8.1.

Results and Discussion

Using a previously published assignment of wild type T4 lysozyme (223) 115 resonances were confidently assigned in the spectrum of the Lu-CLaNP labelled protein sample and using this assignment, 92 resonances in the spectrum of the Yb-CLaNP labelled sample were assigned, 30 in the small domain and 62 in the large domain, to which CLaNP-5 was attached. No minor peaks were observed in the paramagnetically labelled sample, indicated that more than 95% of the protein sample was CLaNP labelled. Mass spectrometry of both samples indicated that the protein samples were 97% ^{15}N enriched and that the probe was attached by both arms. This was confirmed by the observation of single strong paramagnetic shifts in the acquired spectrum of the paramagnetically labelled sample.

Pseudocontact shifts were calculated from 92 assigned paramagnetic resonances and these were used to fit the magnetic susceptibility tensor (Chapter 1). The tensor magnitudes were determined to be $\Delta\chi_{ax} = 10.27 \times 10^{-32} \text{ m}^3$ and $\Delta\chi_{rh} = 6.53 \times 10^{-32} \text{ m}^3$. The fit of predicted to experimental PCS had an R^2 value of 0.995 and a linear least-squares regression line of $y = 0.99x - 0.003$, which shows a good fit between experimental and predicted PCSs (Figure 7.1). Of the 92 assigned resonances assigned in the spectrum of the paramagnetically labelled sample, significant dispersion was observed for six of these, namely G107, T115, S117, G156 W158 and N163 and plots of R_2^{eff} against ν_{CPMG} are shown in Figure 7.2. Moreover, this dispersion was not observed in the diamagnetic spectra, however, G107 is a weak signal in the diamagnetic spectrum and thus confidently ruling out dynamics in the diamagnetic spectrum is not possible. No physiological roles have been reported for these residues, although T115A and S117V mutations have been shown to increase protein stability. The values of k_{ex} , $\Delta\omega$ and R_2^0 for each of these residues obtained

from fitting the data to fast-exchange Carver-Richards equation are shown in Table 7.1. The error values associated with the k_{ex} values are rather high, which is due to the fact that only one or two data points are used in determining this value.

The resonances showing significant dispersion were all within 20 Å of the paramagnetic centre, there are numerous residues within the same region that do not show significant dispersion (Figure 7.3). This indicates that, on the basis of these data, the paramagnetic centre is not undergoing motion on the μ s-ms timescale, which is contrary to previously published studies using this technique and probe, with the exception of the E100C/S104C mutant of pseudoazurin.

The tensor magnitudes determined in this study are larger than have previously been determined for CLaNP-5. The side chain amide of one of the residues that showed significant dispersion, N163, is located 5.5 Å from the probe (Figure 7.4) and W158, another residue for which dispersion was observed, is located on the C-terminal coil. The fact that these residues are located on a loop and that they exhibited motion on the μ s-ms timescale is indicative that this loop may be moving. It is, therefore possible that the side chains of the residues on this loop are interacting with the probe itself. D159 was not assigned in this study; however, the acidic side chain of this residue could form an ionic interaction with the lanthanide ion of CLaNP-5. The increased tensor magnitudes may be due to these interactions and this implies that the local environment of the probe can have an effect on this nature of the paramagnetic field produced by the lanthanide. Further studies would be required to validate this hypothesis.

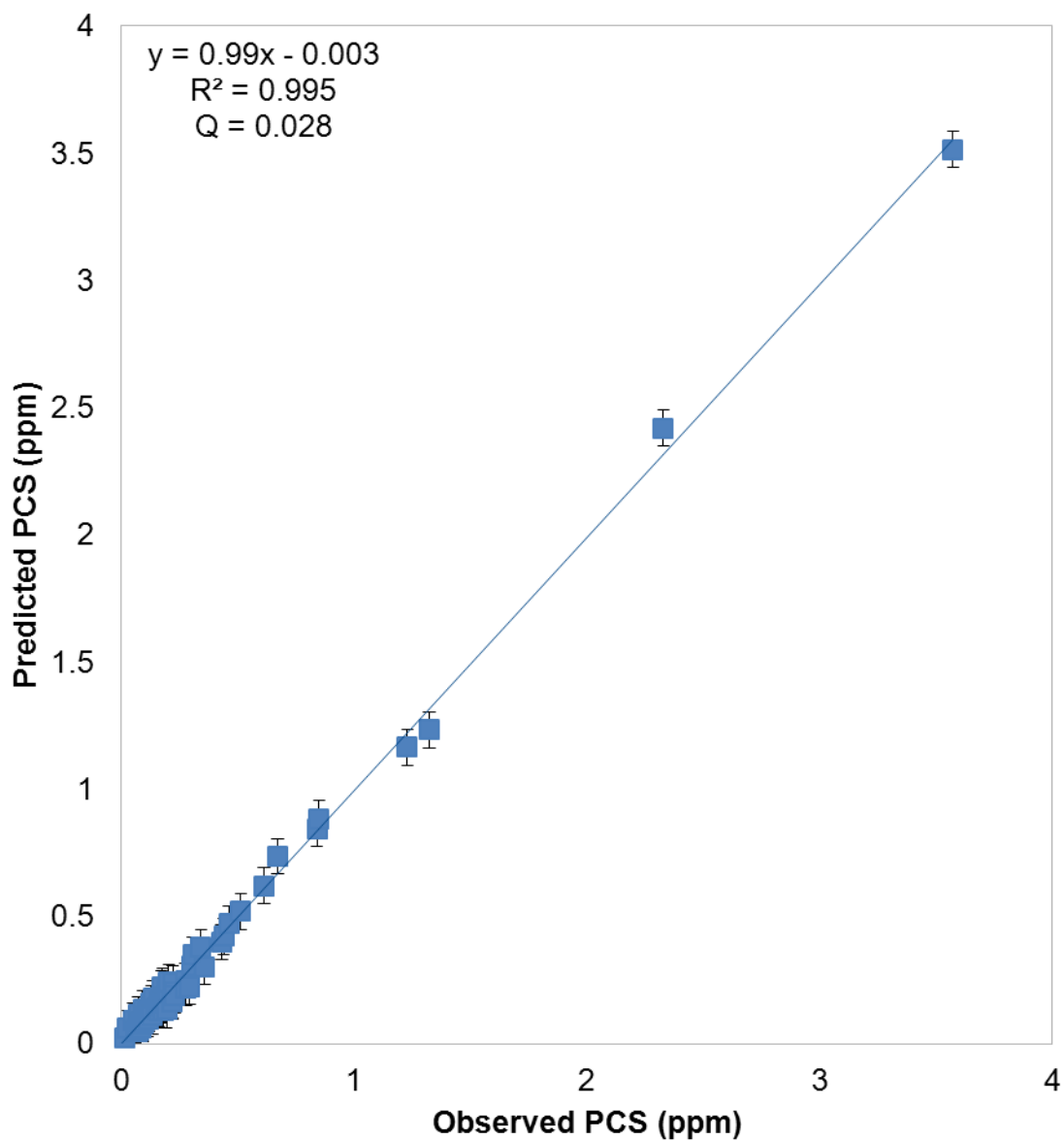


Figure 7.1 Experimentally determined PCS plotted against back-calculated PCS for the K147C/T151C mutant of T4 lysozyme with CLaNP-5 attached. The solid line represents the least-squares linear regression line.

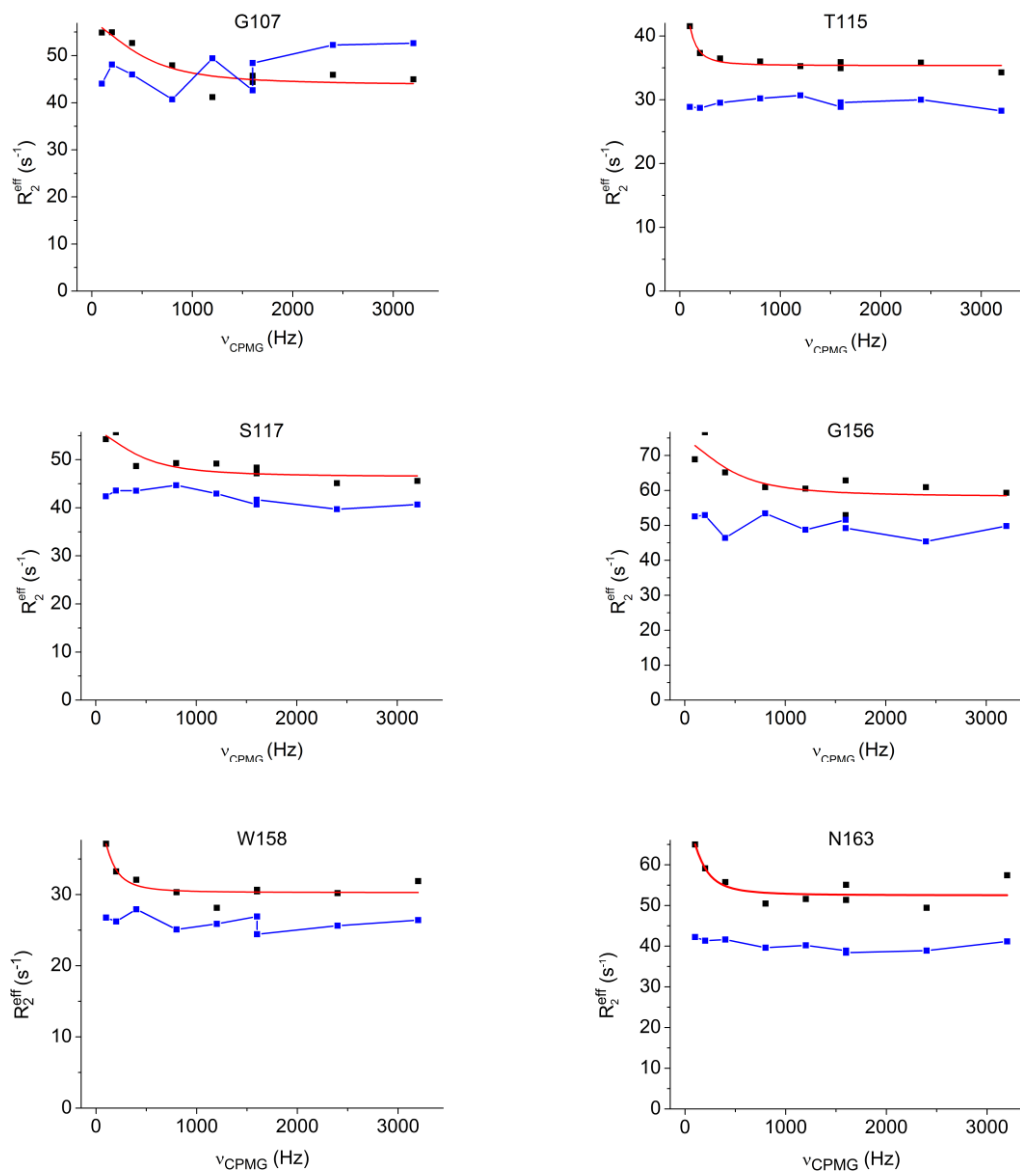


Figure 7.2 Relaxation dispersion curves for the six residues that showed significant dispersion in the 1H CPMG relaxation dispersion measurements of Yb-CLaNP-5 labelled sample of 147C/151C T4 lysozyme. R_2^{eff} values for the diamagnetic and paramagnetic samples are shown in blue and black, respectively. The red lines are the dispersion curves fitted to the paramagnetic data.

Table 7.1 k_{ex} , $\Delta\omega$ and R_2^0 for the six residues that showed significant dispersion in the Yb-CLaNP-5 labelled sample of K147C/T151C T4 lysozyme (errors are shown in parantheses).

Residue	k_{ex} (10^3 s^{-1})	$\Delta\omega$ (Hz)	R_2^0 (s^{-1})
G105	3.4 (1.6)	432 (99)	44 (1.5)
T115	0.5 (0.3)	195 (17)	36 (0.3)
S117	2.9 (1.6)	337 (85)	46 (1.1)
G156	3.0 (2.4)	452 (164)	58 (2.8)
W158	0.8 (0.5)	206 (22)	30 (0.5)
N163	1.0 (0.7)	288 (47)	52 (1.3)

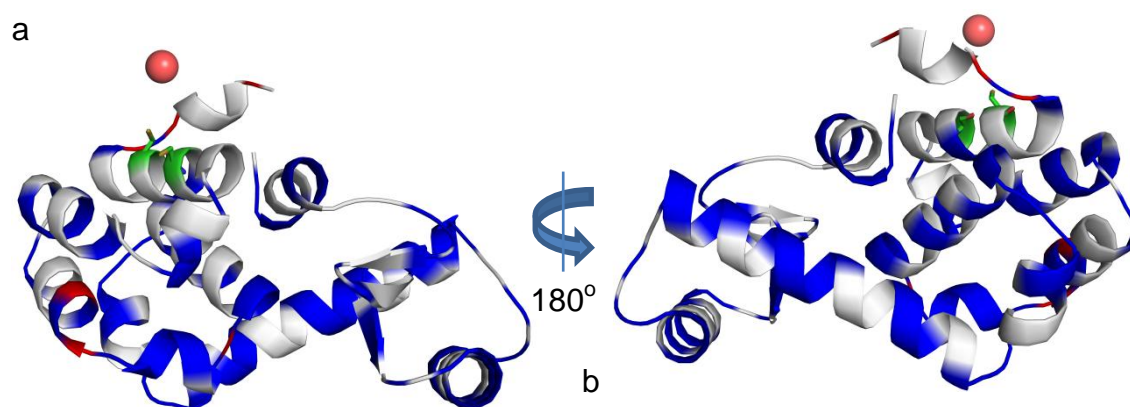


Figure 7.3 Front (a) and back (b) views of T4 lysozyme 147C/151C showing the attachment site (green), the paramagnetic centre as a red sphere, residues showing significant dispersion (red) and residues not showing significant dispersion (blue).

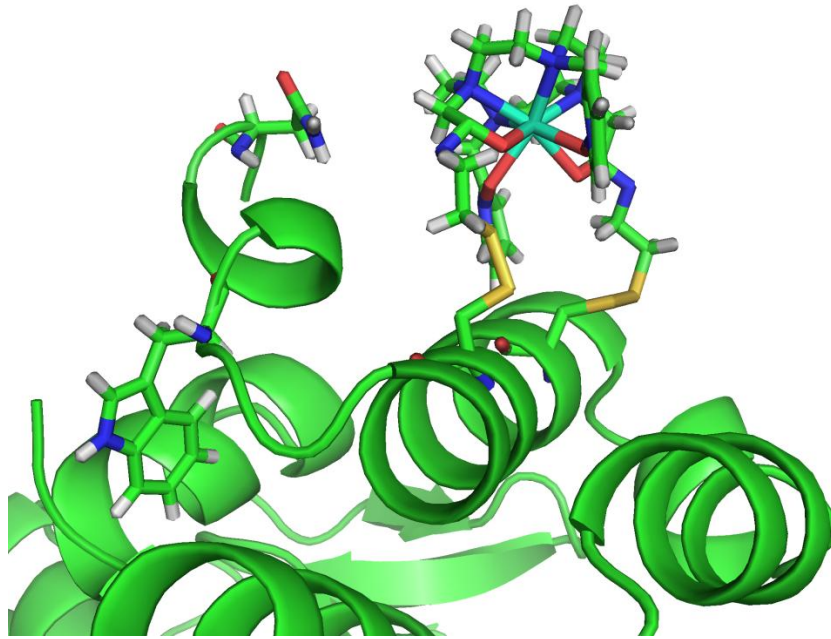


Figure 7.4 CLaNP-5 modelled onto the structure of T4 lysozyme. The attachment site and residues W158 and N163 are shown as sticks.

Chapter 8

General discussion, conclusions and perspectives.

The work presented in this thesis covers a wide range of classical and paramagnetic NMR techniques applied to diverse protein systems. The C-terminal domain of EiAPR studied in chapter 2 is part of a two domain protein that has evolved to possess both reductase and electron transfer components within the same macromolecule. In many redox pathways, electrons are donated from one biomolecule to another; EiAPR, nitrite reductase and cytochrome P450BM3 are exceptions to this rule. P450cam (Chapters 3 and 5) is the archetypal member of the CYP superfamily and as such has been extensively studied over the past five decades. CYPs are terminal acceptors in a variety of redox pathways involved in a plethora of diverse metabolic processes throughout the three kingdoms of life. The precise molecular mechanisms by which CYPs bind to their redox partners and catalyse the hydrogenation of different substrates still remains to be fully elucidated. The data presented in this thesis provide strong evidence that P450cam remains closed when Pdx binds to it. Recent studies (195, 196) have advocated that P450cam opens or adopts an intermediate conformation when Pdx binds. Further investigation is required to determine which of these theories represents the behaviour of P450cam in solution. Avr2 (chapter 4) is a protein produced by a pathogenic fungus that infects tomato plants and causes leaf mould. However, the tomato plants have evolved to possess a level of innate immunity against such an infection, although the precise mechanism, by which this occurs, remains unclear. T4 lysozyme (Chapter 7) has widely been used as a model protein for the development of spin labels for EPR and during the last decade, the L99A mutant of this protein has been used to develop relaxation dispersion techniques (150). This protein is widely used because homologous expression and purification can produce very high yields and it is extremely stable. The diversity of NMR techniques available make this type of spectroscopy ideally suited to unravel the atomic level details of how these different protein systems behave in solution and this thesis outlines but a few of these, applied to the aforementioned protein systems.

The use of classical NMR spectroscopic methods in structural biology

The prerequisite to any detail NMR study is an assignment of the protein nuclei of interest. Nowadays, heteronuclear triple resonance experiments are routinely used to

assign the backbone nuclei of proteins (Chapter 2) and in the case of structure solution, the side chains can be assigned using homo- or heteronuclear experiments (Chapter 4). One of the biggest limitations to the use of classical homo and heteronuclear experiments is the size of the system. As protein size increases, the relaxation rates of the nuclei decrease and in such cases, TROSY techniques can be used overcome this difficulty (Chapter 3). The stability of protein systems can also make assignments challenging to obtain, since classical and TROSY type assignment experiments can last for days, after which time, a protein may have degraded. The introduction of BEST-type pulse sequences (17, 18) has enabled the acquisition times of experiments to be significantly decreased, such that systems such as P450cam can be assigned using BEST-TROSY experiments and a high level of assignment can be obtained (Chapter 3). The assignment of some proteins can still be challenging or even impossible especially if these proteins are especially large and spectra of the uniformly labelled protein would be too crowded for detailed information to be extracted. In such cases, selective labelling of methyl groups (Chapter 3) or the amide nitrogen of a specific residue type (Chapter 5) can yield cleaner spectra. Alternative strategies for the assignment of large proteins have been developed, such as the divide and conquer approach (208) and a method based on systematic mutagenesis (205).

Once an assignment has been obtained, and if the structure of the protein of interest has been solved, detailed atomic level analysis into the behaviour of the protein can commence. Chemical shift perturbation analysis (Chapters 4 & 5) can provide information on how a protein behaves in solution under different conditions at the atomic level. For example when binding to a redox partner or when a specific ligand binds, changes in the chemical shifts of different nuclei are indicative of a change in the chemical environment of those nuclei.

If a structure is not available for the protein of interest, a homology model could be produced, and this could serve as a surrogate until the 3D structure becomes available. Over the past decade software packages, for example TALOS+ (191), have been developed that has the ability to predict secondary structural features on

the basis of $C\alpha$ and $C\beta$ chemical shifts. If a structure can be found that has a high level of sequence homology, predictions of secondary structure can be compared to the structure of the homologous protein and this can provide information on the nature and behaviour of the protein under study (Chapter 2).

If structural information is not available, NMR spectroscopy can be used to solve the solution structure of a protein using NOESY data and software such as ARIA (225) and CYANA (24). However, structures solved in the manner are often *in vacuo* structures and these may be over-constrained and thus not representative of the true solution structure of the protein, and therefore, water refinement of the obtained structure should be carried out and the resultant structural bundle should be validated, for example using CING (190) (Chapter 4).

The applications of classical NMR in structural biology are considerable; however, if the assignment of NMR spectra cannot be obtained for a protein system, the utility of NMR for detailed studies diminishes considerably. In such situations, paramagnetic NMR techniques can provide a solution.

Paramagnetic NMR techniques for the study of proteins and protein-complexes

Paramagnetic NMR is widely used for the study of proteins and protein complexes; however, its use for protein assignment is less widespread. In 2004, Pintacuda *et al.* published the first piece of software for the assignment of proteins on the basis of pseudocontact shifts (66). This software uses a single paramagnetic centre, whose position is known, and its efficacy was demonstrated on selectively labelled samples. However, the larger the protein, the more complex the search space becomes and if the position of a paramagnetic centre is not known, obtaining an assignment from a single paramagnet can be impossible. These difficulties can be overcome by using more than one paramagnetic centre and using a probe for which the approximate position can be consistently and reliably predicted. Each paramagnetic centre yields an independent set of pseudocontact shifts for the same nuclei and an approximate

position can provide a starting point for an assignment algorithm, from which nuclei can be assigned in uniformly and selectively labelled spectral data (Chapter 6).

As simple as using multiple paramagnetic centres may sound, the biochemical work involved in producing numerous mutants to which probes can be attached is laborious and can be extremely time consuming. The development of a methodology whereby fewer mutants are required for an assignment to be obtained would be highly beneficial. The paramagnetic centres produced by different lanthanide ions can be isotropic and anisotropic and the anisotropic fields can have diversely different magnetic susceptibilities. Isotropic paramagnetic lanthanides cause paramagnetic relaxation enhancements in NMR spectra, which provide distance dependent information, and only the position of the paramagnet needs to be fitted. The fact different magnetic susceptibilities are associated with anisotropic paramagnetic fields means that paramagnetic effects can be observed at different distances. For example, using Tm can yield effects up to 60 Å from the paramagnetic centre, whereas Yb can yield effects up to 30 Å (209). The combination of PRE data with short and long range PCS has the potential to provide sufficient information for an assignment to be obtained using a single probe attachment site.

The power of paramagnetism can also be harnessed for the study of protein dynamics. Traditionally, studies of minor states in proteins have utilised relaxation dispersion techniques, however, interpreting the chemical shifts differences obtained from these experiments, in structural terms, can be very difficult. Paramagnetic relaxation dispersion (133) could provide a manner by which the analysis could be simplified (Chapter 7). 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 value will be observed for the nuclei in the region. Fluctuating PCSs can be analysed using relaxation dispersion techniques; thereby the structure of the minor state of a protein could be obtained. A potential pitfall in this method would be if the paramagnetic centre was also undergoing motion on the same time scale as the exchange process being studied and therefore, the source of the paramagnetic field must be rigid for paramagnetic relaxation dispersion to be of

any utility. The development of rigid paramagnetic probes could overcome this limitation and thereby, enable the dynamics of minor states to be investigated. A recently developed paramagnetic probe, CLaNP-7 (61) was designed for increased rigidity, however the dynamic profile of this probe has not been investigated using paramagnetic relaxation dispersion. Paramagnetism could also be using in conjunction with EXSY spectroscopy. Even if two chemical shifts can be observed for an exchange process, the resolution of the spectrum may not be sufficient to identify, unequivocally, the individual autopeaks and cross peaks in an EXSY spectrum. The use of a paramagnetic centre could yield spectra with superior spectral resolution and aid the analysis of processes occurring on the ms-s timescale, where the chemical shift difference between two populations is rather small. The use of saturation transfer for investigating chemical exchange processes has been developed using methyl groups as probes. The utility of this method for studying protein folding has been demonstrated (226).

The significant advances in classical NMR spectroscopy over the last three decades have enabled detailed studies of increasingly complex protein systems. However, the study of some systems is not possible using purely classical techniques. Paramagnetic NMR in isolation can be of considerable utility in the study of biomolecules and this has become increasingly clear over the past two decades. The combination of classical NMR and paramagnetic NMR techniques has the potential to increase the size limit of systems that can be studied at the atomic level and provide solution state information even on the most complex of protein systems.

Reference List

1. Kundu, S., Melton, J. S., Sorensen, D. C., and Phillips, G. N., Jr. (2002) Dynamics of proteins in crystals: comparison of experiment with simple models, *Biophys J* 83, 723-732.
2. Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, 1 ed., Wiley-Interscience, New York.
3. Driscoll, P. C., Clore, G. M., Marion, D., Wingfield, P. T., and Gronenborn, A. M. (1990) Complete resonance assignment for the polypeptide backbone of interleukin 1 beta using three-dimensional heteronuclear NMR spectroscopy, *Biochemistry* 29, 3542-3556.
4. Driscoll, P. C., Gronenborn, A. M., Wingfield, P. T., and Clore, G. M. (1990) Determination of the secondary structure and molecular topology of interleukin-1 beta by use of two- and three-dimensional heteronuclear ¹⁵N-¹H NMR spectroscopy, *Biochemistry* 29, 4668-4682.
5. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., and Clore, G. M. (1989) Overcoming the overlap problem in the assignment of ¹H NMR spectra of larger proteins by use of three-dimensional heteronuclear ¹H-¹⁵N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta, *Biochemistry* 28, 6150-6156.
6. Wijmenga, S. S., Hallenga, K., and Hilbers, C. W. (1989) A three-dimensional heteronuclear multiple-quantum coherence homonuclear hartmann-hahn experiment, *J Mag Reson* 84, 634-642.
7. Fesik, S. W., and Zuiderweg, E. R. P. (1988) Heteronuclear three-dimensional nmr spectroscopy. A strategy for the simplification of homonuclear two-dimensional NMR spectra, *J Magn Reson* 78, 588-593.
8. Shon, K., and Opella, S. J. (1989) Detection of ¹H homonuclear NOE between amide sites in proteins with ¹H-¹⁵N heteronuclear correlation spectroscopy, *J Magn Reson* 82, 193-197.
9. Grzesiek, S., and Bax, A. (1992) An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins, *J Magn Reson* 99, 201-207.
10. Clubb, R. T., Thanabal, V., and Wagner, G. (1992) A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue ¹HN, ¹⁵N, and ¹³C' chemical shifts in ¹⁵N-¹³C-labelled proteins, *J Magn Reson* 97, 213-217.
11. Sattler, M., Schleucher, J. r., and Griesinger, C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients, *Prog Nucl Magn Reson Spectrosc* 34, 93-158.
12. Yamazaki, T., Lee, W., Arrowsmith, C. H., Muhandiram, D. R., and Kay, L. E. (1994) A Suite of Triple Resonance NMR Experiments for the Backbone Assignment of ¹⁵N, ¹³C, ²H Labeled Proteins with High Sensitivity, *J Am Chem Soc* 116, 11655-11666.
13. Kay, L. E., Ikura, M., Tschudin, R., and Bax, A. (1990) Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins, *J Magn Reson* 89, 496-514.
14. Pervushin, K., Riek, R., Wider, G., and Wuthrich, K. (1997) Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution, *Proc Natl Acad Sci U S A* 94, 12366-12371.
15. Keeler, J. (2010) *Understanding NMR spectroscopy*, 2nd ed., Wiley-Blackwell.
16. Fernandez, C., and Wider, G. (2003) TROSY in NMR studies of the structure and function of large biological macromolecules, *Curr Opin Struct Biol* 13, 570-580.

17. Lescop, E., Schanda, P., and Brutscher, B. (2007) A set of BEST triple-resonance experiments for time-optimized protein resonance assignment, *J Magn Reson* 187, 163-169.
18. Schanda, P., Van Melckebeke, H., and Brutscher, B. (2006) Speeding up three-dimensional protein NMR experiments to a few minutes, *J Am Chem Soc* 128, 9042-9043.
19. Pervushin, K., Vogeli, B., and Eletsky, A. (2002) Longitudinal (1)H relaxation optimization in TROSY NMR spectroscopy, *J Am Chem Soc* 124, 12898-12902.
20. Schanda, P., Kupce, E., and Brutscher, B. (2005) SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds, *J Biomol NMR* 33, 199-211.
21. Schanda, P., and Brutscher, B. (2006) Hadamard frequency-encoded SOFAST-HMQC for ultrafast two-dimensional protein NMR, *J Magn Reson* 178, 334-339.
22. Schanda, P., and Brutscher, B. (2005) Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds, *J Am Chem Soc* 127, 8014-8015.
23. Tugarinov, V., Choy, W. Y., Orekhov, V. Y., and Kay, L. E. (2005) Solution NMR-derived global fold of a monomeric 82-kDa enzyme, *Proc Natl Acad Sci U S A* 102, 622-627.
24. Guntert, P. (2004) Automated NMR structure calculation with CYANA, *Methods Mol Biol* 278, 353-378.
25. Herrmann, T., Guntert, P., and Wuthrich, K. (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA, *J Mol Biol* 319, 209-227.
26. Bertini, I., Luchinat, C., and Parigi, G. (2002) Magnetic susceptibility in paramagnetic NMR, *Prog Nucl Magn Reson Spectrosc* 40, 249-273.
27. Keizers, P. H., Saragliadis, A., Hiruma, Y., Overhand, M., and Ubbink, M. (2008) Design, synthesis, and evaluation of a lanthanide chelating protein probe: CLaNP-5 yields predictable paramagnetic effects independent of environment, *J Am Chem Soc* 130, 14802-14812.
28. Schmitz, C., Stanton-Cook, M. J., Su, X. C., Otting, G., and Huber, T. (2008) Numbat: an interactive software tool for fitting Deltachi-tensors to molecular coordinates using pseudocontact shifts, *J Biomol NMR* 41, 179-189.
29. Solomon, I. (1955) Relaxation Processes in a System of Two Spins, *Phys Rev* 99, 559-565.
30. Bloembergen, N., and Morgan, L. O. (1961) Proton Relaxation Times in Paramagnetic Solutions. Effects of Electron Spin Relaxation, *J Chem Phys* 34, 842-850.
31. Volkov, A. N., Worrall, J. A., Holtzmann, E., and Ubbink, M. (2006) Solution structure and dynamics of the complex between cytochrome c and cytochrome c peroxidase determined by paramagnetic NMR, *Proc Natl Acad Sci U S A* 103, 18945-18950.
32. Scanu, S., Forster, J., Finiguerra, M. G., Shabestari, M. H., Huber, M., and Ubbink, M. (2012) The complex of cytochrome f and plastocyanin from *Nostoc* sp. PCC 7119 is highly dynamic, *Chembiochem* 13, 1312-1318.
33. Iwahara, J., Schwieters, C. D., and Clore, G. M. (2004) Ensemble approach for NMR structure refinement against (1)H paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule, *J Am Chem Soc* 126, 5879-5896.
34. Gueron, M. (1975) Nuclear relaxation in macromolecules by paramagnetic ions: a novel mechanism, *J Magn Reson* 19, 58-66.
35. Kowalsky, A. (1965) Nuclear Magnetic Resonance Studies of Cytochrome c. Possible Electron Delocalization*, *Biochemistry* 4, 2382-2388.
36. Wuthrich, K. (1969) High-resolution proton nuclear magnetic resonance spectroscopy of cytochrome, *Proc Natl Acad Sci U S A* 63, 1071-1078.

37. Moura, J. J. G., Xavier, A. n. V., Bruschi, M., and Gall, J. L. (1977) NMR characterization of three forms of ferredoxin from *Desulphovibrio gigas*, a sulphate reducer, *Biochim Biophys Acta* 459, 278-289.
38. Salmeen, I., and Palmer, G. (1972) Contact-shifted NMR of spinach ferredoxin: Additional resonances and partial assignments, *Arch Biochem Biophys* 150, 767-773.
39. Oh, B. H., Westler, W. M., Darba, P., and Markley, J. L. (1988) Protein carbon-13 spin systems by a single two-dimensional nuclear magnetic resonance experiment, *Science* 240, 908-911.
40. Phillips, W. D., Poe, M., Weiher, J. F., McDonald, C. C., and Lovenberg, W. (1970) Proton magnetic resonance, magnetic susceptibility and Mossbauer studies of *Clostridium pasteurianum* rubredoxin, *Nature* 227, 574-577.
41. Ubbink, M., Ejdeback, M., Karlsson, B. G., and Bendall, D. S. (1998) The structure of the complex of plastocyanin and cytochrome f, determined by paramagnetic NMR and restrained rigid-body molecular dynamics, *Structure* 6, 323-335.
42. Salgado, J., Jimenez, H. R., Moratal, J. M., Kroes, S., Warmerdam, G. C. M., and Canters, G. W. (1996) Paramagnetic Cobalt and Nickel Derivatives of *Alcaligenes denitrificans* Azurin and Its M121Q Mutant. A 1H NMR Study., *Biochemistry* 35, 1810-1819.
43. Piccioli, M., Luchinat, C., Mizoguchi, T. J., Ramirez, B. E., Gray, H. B., and Richards, J. H. (1995) Paramagnetic NMR spectroscopy and coordination structure of cobalt(II) Cys112Asp azurin, *Inorg Chem* 34, 737-742.
44. John, M., Pintacuda, G., Park, A. Y., Dixon, N. E., and Otting, G. (2006) Structure determination of protein-ligand complexes by transferred paramagnetic shifts, *J Am Chem Soc* 128, 12910-12916.
45. Feeney, J., Birdsall, B., Bradbury, A. F., Biekofsky, R. R., and Bayley, P. M. (2001) Calmodulin tagging provides a general method of using lanthanide induced magnetic field orientation to observe residual dipolar couplings in proteins in solution, *J Biomol NMR* 21, 41-48.
46. Ma, C., and Opella, S. J. (2000) Lanthanide ions bind specifically to an added "EF-hand" and orient a membrane protein in micelles for solution NMR spectroscopy, *J Magn Reson* 146, 381-384.
47. Johnson, P. E., Brun, E., MacKenzie, L. F., Withers, S. G., and McIntosh, L. P. (1999) The cellulose-binding domains from *Cellulomonas fimi* beta-1, 4-glucanase CenC bind nitroxide spin-labeled cellooligosaccharides in multiple orientations, *J Mol Biol* 287, 609-625.
48. Saio, T., Ogura, K., Yokochi, M., Kobashigawa, Y., and Inagaki, F. (2009) Two-point anchoring of a lanthanide-binding peptide to a target protein enhances the paramagnetic anisotropic effect, *J Biomol NMR* 44, 157-166.
49. Su, X. C., McAndrew, K., Huber, T., and Otting, G. (2008) Lanthanide-binding peptides for NMR measurements of residual dipolar couplings and paramagnetic effects from multiple angles, *J Am Chem Soc* 130, 1681-1687.
50. Gaponenko, V., Dvoretzky, A., Walsby, C., Hoffman, B. M., and Rosevear, P. R. (2000) Calculation of z-coordinates and orientational restraints using a metal binding tag, *Biochemistry* 39, 15217-15224.
51. Barthelmes, K., Reynolds, A. M., Peisach, E., Jonker, H. R. A., DeNunzio, N. J., Allen, K. N., Imperiali, B., and Schwalbe, H. (2011) Engineering Encodable Lanthanide-Binding Tags into Loop Regions of Proteins, *J Am Chem Soc* 133, 808-819.
52. Pintacuda, G., Moshref, A., Leonchiks, A., Sharipo, A., and Otting, G. (2004) Site-specific labelling with a metal chelator for protein-structure refinement, *J Biomol NMR* 29, 351-361.
53. Dvoretzky, A., Gaponenko, V., and Rosevear, P. R. (2002) Derivation of structural restraints using a thiol-reactive chelator, *FEBS Lett* 528, 189-192.

54. Kamen, D. E., Cahill, S. M., and Girvin, M. E. (2007) Multiple alignment of membrane proteins for measuring residual dipolar couplings using lanthanide ions bound to a small metal chelator, *J Am Chem Soc* 129, 1846-1847.
55. Leonov, A., Voigt, B., Rodriguez-Castaneda, F., Sakhaii, P., and Griesinger, C. (2005) Convenient synthesis of multifunctional EDTA-based chiral metal chelates substituted with an S-mesylcysteine, *Chemistry* 11, 3342-3348.
56. Rodriguez-Castaneda, F., Haberz, P., Leonov, A., and Griesinger, C. (2006) Paramagnetic tagging of diamagnetic proteins for solution NMR, *Magn Reson Chem* 44 Spec No, S10-16.
57. Haussinger, D., Huang, J. R., and Grzesiek, S. (2009) DOTA-M8: An extremely rigid, high-affinity lanthanide chelating tag for PCS NMR spectroscopy, *J Am Chem Soc* 131, 14761-14767.
58. Su, X. C., Man, B., Beeren, S., Liang, H., Simonsen, S., Schmitz, C., Huber, T., Messerle, B. A., and Otting, G. (2008) A dipicolinic acid tag for rigid lanthanide tagging of proteins and paramagnetic NMR spectroscopy, *J Am Chem Soc* 130, 10486-10487.
59. Ikegami, T., Verdier, L., Sakhaii, P., Grimme, S., Pescatore, B., Saxena, K., Fiebig, K. M., and Griesinger, C. (2004) Novel techniques for weak alignment of proteins in solution using chemical tags coordinating lanthanide ions, *J Biomol NMR* 29, 339-349.
60. Haberz, P., Rodriguez-Castaneda, F., Junker, J., Becker, S., Leonov, A., and Griesinger, C. (2006) Two new chiral EDTA-based metal chelates for weak alignment of proteins in solution, *Org Lett* 8, 1275-1278.
61. Liu, W. M., Keizers, P. H., Hass, M. A., Blok, A., Timmer, M., Sarris, A. J., Overhand, M., and Ubbink, M. (2012) A pH-Sensitive, Colorful, Lanthanide-Chelating Paramagnetic NMR Probe, *J Am Chem Soc* 134, 17306-17313.
62. Banci, L., Bertini, I., Bren, K. L., Cremonini, M. A., Gray, H. B., Luchinat, C., and Turano, P. (1996) The use of pseudocontact shifts to refine solution structures of paramagnetic metalloproteins: Met80Ala cyano-cytochrome c as an example, *J. Biol. Inorg. Chem.* 1, 117-126.
63. Banci, L., Bertini, I., Savellini, G. G., Romagnoli, A., Turano, P., Cremonini, M. A., Luchinat, C., and Gray, H. B. (1997) Pseudocontact shifts as constraints for energy minimization and molecular dynamics calculations on solution structures of paramagnetic metalloproteins, *Proteins* 29, 68-76.
64. Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Clore, G. M. (2003) The Xplor-NIH NMR molecular structure determination package, *J Magn Reson* 160, 65-73.
65. Banci, L., Bertini, I., Cavallaro, G., Giachetti, A., Luchinat, C., and Parigi, G. (2004) Paramagnetism-based restraints for Xplor-NIH, *J Biomol NMR* 28, 249-261.
66. Pintacuda, G., Keniry, M. A., Huber, T., Park, A. Y., Dixon, N. E., and Otting, G. (2004) Fast structure-based assignment of 15N HSQC spectra of selectively 15N-labeled paramagnetic proteins, *J Am Chem Soc* 126, 2963-2970.
67. Schmitz, C., John, M., Park, A. Y., Dixon, N. E., Otting, G., Pintacuda, G., and Huber, T. (2006) Efficient chi-tensor determination and NH assignment of paramagnetic proteins, *J Biomol NMR* 35, 79-87.
68. John, M., Schmitz, C., Park, A. Y., Dixon, N. E., Huber, T., and Otting, G. (2007) Sequence-specific and stereospecific assignment of methyl groups using paramagnetic lanthanides, *J Am Chem Soc* 129, 13749-13757.
69. Allerhand, A., Doddrell, D., Glushko, V., Cochran, D. W., Wenkert, E., Lawson, P. J., and Gurd, F. R. (1971) Conformation and segmental motion of native and denatured ribonuclease A in solution. Application of natural-abundance carbon-13 partially relaxed Fourier transform nuclear magnetic resonance, *J Am Chem Soc* 93, 544-546.
70. Nigen, A. M., Keim, P., Marshall, R. C., Morrow, J. S., and Gurd, F. R. (1972) Carbon 13 nuclear magnetic resonance spectroscopy of myoglobins and ribonuclease A carboxymethylated with enriched (2- 13 C)bromoacetate, *J Biol Chem* 247, 4100-4102.

71. Glushko, V., Lawson, P. J., and Gurd, F. R. (1972) Conformational states of bovine pancreatic ribonuclease A observed by normal and partially relaxed carbon 13 nuclear magnetic resonance, *J Biol Chem* 247, 3176-3185.
72. Jones, W. C., Rothgeb, T. M., and Gurd, F. R. (1976) Nuclear magnetic resonance studies of sperm whale myoglobin specifically enriched with 13C in the methionine methyl groups, *J Biol Chem* 251, 7452-7460.
73. Campbell, I. D., Dobson, C. M., and Williams, R. J. (1985) The study of conformational states of proteins by nuclear magnetic resonance, *Biochem J* 231, 1-10.
74. Hughes, L. T., Cohen, J. S., Szabo, A., Niu, C., and Matsuura, S. (1984) 13C NMR studies of the molecular dynamics of selectively 13C-enriched ribonuclease complexes, *Biochemistry* 23, 4390-4394.
75. Rozovsky, S., Jogl, G., Tong, L., and McDermott, A. E. (2001) Solution-state NMR investigations of triosephosphate isomerase active site loop motion: ligand release in relation to active site loop dynamics, *J Mol Biol* 310, 271-280.
76. Seifert, M. H., Breitenlechner, C. B., Bossemeyer, D., Huber, R., Holak, T. A., and Engh, R. A. (2002) Phosphorylation and flexibility of cyclic-AMP-dependent protein kinase (PKA) using (31)P NMR spectroscopy, *Biochemistry* 41, 5968-5977.
77. Kay, L. E., Torchia, D. A., and Bax, A. (1989) Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease, *Biochemistry* 28, 8972-8979.
78. Lipari, G., and Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity, *J Am Chem Soc* 104, 4546-4559.
79. Lipari, G., and Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results, *J Am Chem Soc* 104, 4559-4570.
80. Dayie, K. T., and Wagner, G. (1997) Carbonyl Carbon Probe of Local Mobility in 13C,15N-Enriched Proteins Using High-Resolution Nuclear Magnetic Resonance, *J Am Chem Soc* 119, 7797-7806.
81. Yamazaki, T., Muhandiram, R., and Kay, L. E. (1994) NMR Experiments for the Measurement of Carbon Relaxation Properties in Highly Enriched, Uniformly 13C,15N-Labeled Proteins: Application to 13C alpha carbons, *J Am Chem Soc* 116, 8266-8278.
82. Wang, T., Cai, S., and Zuiderweg, E. R. (2003) Temperature dependence of anisotropic protein backbone dynamics, *J Am Chem Soc* 125, 8639-8643.
83. Fischer, M. W., Zeng, L., Majumdar, A., and Zuiderweg, E. R. (1998) Characterizing semilocal motions in proteins by NMR relaxation studies, *Proc Natl Acad Sci U S A* 95, 8016-8019.
84. Pang, Y., Buck, M., and Zuiderweg, E. R. (2002) Backbone dynamics of the ribonuclease binase active site area using multinuclear (15)N and (13)CO NMR relaxation and computational molecular dynamics, *Biochemistry* 41, 2655-2666.
85. Desvaux, H., and Berthault, P. (1999) Study of dynamic processes in liquids using off-resonance rf irradiation, *Prog Nucl Magn Reson Spectrosc* 35, 295-340.
86. Palmer, A. G., 3rd, Kroenke, C. D., and Loria, J. P. (2001) Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules, *Methods Enzymol* 339, 204-238.
87. Deverell, C., Morgan, R. E., and Strange, J. H. (1970) Studies of chemical exchange by nuclear magnetic relaxation in the rotating frame, *Mol Phys* 18, 553-559.
88. Meiboom, S. (1961) Nuclear Magnetic Resonance Study of the Proton Transfer in Water, *J Chem Phys* 34, 375-388.
89. Szyperski, T., Luginbuhl, P., Otting, G., Guntert, P., and Wuthrich, K. (1993) Protein dynamics studied by rotating frame 15N spin relaxation times, *J Biomol NMR* 3, 151-164.

90. Akke, M., and Palmer, A. G. (1996) Monitoring Macromolecular Motions on Microsecond to Millisecond Time Scales by R1 ρ -R1 Constant Relaxation Time NMR Spectroscopy, *J Am Chem Soc* 118, 911-912.
91. Auer, R., Neudecker, P., Muhandiram, D. R., Lundstrom, P., Hansen, D. F., Konrat, R., and Kay, L. E. (2009) Measuring the Signs of ¹H α Chemical Shift Differences Between Ground and Excited Protein States by Off-Resonance Spin-Lock R1 ρ NMR Spectroscopy, *J Am Chem Soc* 131, 10832-10833.
92. Hansen, A. L., Nikolova, E. N., Casiano-Negroni, A., and Al-Hashimi, H. M. (2009) Extending the Range of Microsecond-to-Millisecond Chemical Exchange Detected in Labeled and Unlabeled Nucleic Acids by Selective Carbon R1 ρ NMR Spectroscopy, *J Am Chem Soc* 131, 3818-3819.
93. Korzhnev, D., Orekhov, V., Dahlquist, F., and Kay, L. (2003) Off-resonance R1 ρ relaxation outside of the fast exchange limit: An experimental study of a cavity mutant of T4 lysozyme, *J Biomol NMR* 26, 39-48.
94. Korzhnev, D. M., Orekhov, V. Y., and Kay, L. E. (2005) Off-Resonance R1 ρ NMR Studies of Exchange Dynamics in Proteins with Low Spin-Lock Fields. An Application to a Fyn SH3 Domain, *J Am Chem Soc* 127, 713-721.
95. Lundström, P., and Akke, M. (2005) Off-resonance rotating-frame amide proton spin relaxation experiments measuring microsecond chemical exchange in proteins, *J Biomol NMR* 32, 163-173.
96. Massi, F., Grey, M. J., and Palmer, A. G., 3rd. (2005) Microsecond timescale backbone conformational dynamics in ubiquitin studied with NMR R1 ρ relaxation experiments, *Protein Sci* 14, 735-742.
97. Massi, F., Johnson, E., Wang, C., Rance, M., and Palmer, A. G., 3rd. (2004) NMR R1 ρ rotating-frame relaxation with weak radio frequency fields, *J Am Chem Soc* 126, 2247-2256.
98. Baldwin, A. J., and Kay, L. E. (2013) An R1 ρ expression for a spin in chemical exchange between two sites with unequal transverse relaxation rates, *J Biomol NMR* 55, 211-218.
99. Carr, H. Y., and Purcell, E. M. (1954) Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments, *Phys Rev* 94, 630-638.
100. Luz, Z., and Meiboom, S. (1963) Nuclear Magnetic Resonance Study of the Protolysis of Trimethylammonium Ion in Aqueous Solution---Order of the Reaction with Respect to Solvent, *J Chem Phys* 39, 366-370.
101. Meiboom, S., and Gill, D. (1958) Modified Spin-Echo Method for Measuring Nuclear Relaxation Times, *Rev Sci Instrum* 29, 688-691.
102. Mulder, F. A., Skrynnikov, N. R., Hon, B., Dahlquist, F. W., and Kay, L. E. (2001) Measurement of slow (micros-ms) time scale dynamics in protein side chains by (¹⁵N) relaxation dispersion NMR spectroscopy: application to Asn and Gln residues in a cavity mutant of T4 lysozyme, *J Am Chem Soc* 123, 967-975.
103. Ishima, R., and Torchia, D. A. (2003) Extending the range of amide proton relaxation dispersion experiments in proteins using a constant-time relaxation-compensated CPMG approach, *J Biomol NMR* 25, 243-248.
104. Vallurupalli, P., Hansen, D. F., Stollar, E., Meirovitch, E., and Kay, L. E. (2007) Measurement of bond vector orientations in invisible excited states of proteins, *Proc Natl Acad Sci U S A* 104, 18473-18477.
105. Tugarinov, V., Kanelis, V., and Kay, L. E. (2006) Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy, *Nat Protoc* 1, 749-754.
106. Lundstrom, P., Vallurupalli, P., Hansen, D. F., and Kay, L. E. (2009) Isotope labeling methods for studies of excited protein states by relaxation dispersion NMR spectroscopy, *Nat Protoc* 4, 1641-1648.
107. Kovrigin, E. L., and Loria, J. P. (2006) Enzyme dynamics along the reaction coordinate: critical role of a conserved residue, *Biochemistry* 45, 2636-2647.

108. Fraser, J. S., Clarkson, M. W., Degnan, S. C., Erion, R., Kern, D., and Alber, T. (2009) Hidden alternative structures of proline isomerase essential for catalysis, *Nature* **462**, 669-673.
109. Vallurupalli, P., and Kay, L. E. (2006) Complementarity of ensemble and single-molecule measures of protein motion: a relaxation dispersion NMR study of an enzyme complex, *Proc Natl Acad Sci U S A* **103**, 11910-11915.
110. Beach, H., Cole, R., Gill, M. L., and Loria, J. P. (2005) Conservation of mus-ms enzyme motions in the apo- and substrate-mimicked state, *J Am Chem Soc* **127**, 9167-9176.
111. Sprangers, R., Gribun, A., Hwang, P. M., Houry, W. A., and Kay, L. E. (2005) Quantitative NMR spectroscopy of supramolecular complexes: dynamic side pores in ClpP are important for product release, *Proc Natl Acad Sci U S A* **102**, 16678-16683.
112. Doucet, N., Watt, E. D., and Loria, J. P. (2009) The flexibility of a distant loop modulates active site motion and product release in ribonuclease A, *Biochemistry* **48**, 7160-7168.
113. Korzhnev, D. M., Neudecker, P., Mittermaier, A., Orekhov, V. Y., and Kay, L. E. (2005) Multiple-site exchange in proteins studied with a suite of six NMR relaxation dispersion experiments: an application to the folding of a Fyn SH3 domain mutant, *J Am Chem Soc* **127**, 15602-15611.
114. Schanda, P., Brutscher, B., Konrat, R., and Tollinger, M. (2008) Folding of the KIX domain: characterization of the equilibrium analog of a folding intermediate using ¹⁵N/¹³C relaxation dispersion and fast ¹H/²H amide exchange NMR spectroscopy, *J Mol Biol* **380**, 726-741.
115. Sugase, K., Dyson, H. J., and Wright, P. E. (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein, *Nature* **447**, 1021-1025.
116. Farber, P., Darmawan, H., Sprules, T., and Mittermaier, A. (2010) Analyzing Protein Folding Cooperativity by Differential Scanning Calorimetry and NMR Spectroscopy, *J Am Chem Soc* **132**, 6214-6222.
117. Tzeng, S.-R., and Kalodimos, C. G. (2009) Dynamic activation of an allosteric regulatory protein, *Nature* **462**, 368-372.
118. Wolf-Watz, M., Thai, V., Henzler-Wildman, K., Hadjipavlou, G., Eisenmesser, E. Z., and Kern, D. (2004) Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair, *Nat Struct Mol Biol* **11**, 945-949.
119. Eisenmesser, E. Z., Millet, O., Labeikovsky, W., Korzhnev, D. M., Wolf-Watz, M., Bosco, D. A., Skalicky, J. J., Kay, L. E., and Kern, D. (2005) Intrinsic dynamics of an enzyme underlies catalysis, *Nature* **438**, 117-121.
120. Korzhnev, D. M., and Kay, L. E. (2008) Probing invisible, low-populated States of protein molecules by relaxation dispersion NMR spectroscopy: an application to protein folding, *Acc Chem Res* **41**, 442-451.
121. Korzhnev, D. M., Neudecker, P., Mittermaier, A., Orekhov, V. Y., and Kay, L. E. (2005) Multiple-Site Exchange in Proteins Studied with a Suite of Six NMR Relaxation Dispersion Experiments: An Application to the Folding of a Fyn SH3 Domain Mutant, *J Am Chem Soc* **127**, 15602-15611.
122. Skrynnikov, N. R., Dahlquist, F. W., and Kay, L. E. (2002) Reconstructing NMR Spectra of "Invisible" Excited Protein States Using HSQC and HMQC Experiments, *J Am Chem Soc* **124**, 12352-12360.
123. Bouvignies, G., Vallurupalli, P., Hansen, D. F., Correia, B. E., Lange, O., Bah, A., Vernon, R. M., Dahlquist, F. W., Baker, D., and Kay, L. E. (2011) Solution structure of a minor and transiently formed state of a T4 lysozyme mutant, *Nature* **477**, 111-114.
124. Jeener, J., Meier, B. H., Bachmann, P., and Ernst, R. R. (1979) Investigation of exchange processes by two-dimensional NMR spectroscopy, *J Chem Phys* **71**, 4546-4553.
125. Religa, T. L., Sprangers, R., and Kay, L. E. (2010) Dynamic regulation of archaeal proteasome gate opening as studied by TROSY NMR, *Science* **328**, 98-102.

126. Sahu, D., Clore, G. M., and Iwahara, J. (2007) TROSY-Based z-Exchange Spectroscopy: Application to the Determination of the Activation Energy for Intermolecular Protein Translocation between Specific Sites on Different DNA Molecules, *J Am Chem Soc* 129, 13232-13237.
127. Kuloglu, E. S., McCaslin, D. R., Markley, J. L., and Volkman, B. F. (2002) Structural rearrangement of human lymphotactin, a C chemokine, under physiological solution conditions, *J Biol Chem* 277, 17863-17870.
128. Hwang, P. M., Bishop, R. E., and Kay, L. E. (2004) The integral membrane enzyme PagP alternates between two dynamically distinct states, *Proc Natl Acad Sci U S A* 101, 9618-9623.
129. Nikolaev, Y., and Pervushin, K. (2007) NMR Spin State Exchange Spectroscopy Reveals Equilibrium of Two Distinct Conformations of Leucine Zipper GCN4 in Solution, *J Am Chem Soc* 129, 6461-6469.
130. Li, Y., and Palmer, A. G., 3rd. (2009) TROSY-selected ZZ-exchange experiment for characterizing slow chemical exchange in large proteins, *J Biomol NMR* 45, 357-360.
131. Banci, L., Bertini, I., Cavazza, C., Felli, I. C., and Koulougliotis, D. (1998) Probing the backbone dynamics of oxidized and reduced rat microsomal cytochrome b5 via ¹⁵N rotating frame NMR relaxation measurements: biological implications, *Biochemistry* 37, 12320-12330.
132. Bertini, I., Kursula, P., Luchinat, C., Parigi, G., Vahokoski, J., Wilmanns, M., and Yuan, J. (2009) Accurate solution structures of proteins from X-ray data and a minimal set of NMR data: calmodulin-peptide complexes as examples, *J Am Chem Soc* 131, 5134-5144.
133. Hass, M. A., Keizers, P. H., Blok, A., Hiruma, Y., and Ubbink, M. (2010) Validation of a lanthanide tag for the analysis of protein dynamics by paramagnetic NMR spectroscopy, *J Am Chem Soc* 132, 9952-9953.
134. Omura, T., and Sato, R. (1964) The Carbon Monoxide-Binding Pigment of Liver Microsomes. I. Evidence for Its Hemoprotein Nature, *J Biol Chem* 239, 2370-2378.
135. Nelson, D. R., Koymans, L., Kamataki, T., Stegeman, J. J., Feyereisen, R., Waxman, D. J., Waterman, M. R., Gotoh, O., Coon, M. J., Estabrook, R. W., Gunsalus, I. C., and Nebert, D. W. (1996) P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature, *Pharmacogenetics* 6, 1-42.
136. McLean, K. J., Sabri, M., Marshall, K. R., Lawson, R. J., Lewis, D. G., Clift, D., Balding, P. R., Dunford, A. J., Warman, A. J., McVey, J. P., Quinn, A. M., Sutcliffe, M. J., Scrutton, N. S., and Munro, A. W. (2005) Biodiversity of cytochrome P450 redox systems, *Biochem Soc Trans* 33, 796-801.
137. Williams, P. A., Cosme, J., Ward, A., Angove, H. C., Matak Vinkovic, D., and Jhoti, H. (2003) Crystal structure of human cytochrome P450 2C9 with bound warfarin, *Nature* 424, 464-468.
138. Lewis, D. F. (1999) Homology modelling of human cytochromes P450 involved in xenobiotic metabolism and rationalization of substrate selectivity, *Exp Toxicol Pathol* 51, 369-374.
139. Ortiz de Montellano, P. R. (2005) *Cytochrome P450: Structure, Mechanism and Biochemistry*, 3 ed., Kluwer Academic/Plenum Publishers, New York.
140. Poulos, T. L., Finzel, B. C., Gunsalus, I. C., Wagner, G. C., and Kraut, J. (1985) The 2.6-Å crystal structure of *Pseudomonas putida* cytochrome P-450, *J Biol Chem* 260, 16122-16130.
141. Lee, Y. T., Wilson, R. F., Rupniewski, I., and Goodin, D. B. (2010) P450cam visits an open conformation in the absence of substrate, *Biochemistry* 49, 3412-3419.
142. Ludemann, S. K., Lounnas, V., and Wade, R. C. (2000) How do substrates enter and products exit the buried active site of cytochrome P450cam? 1. Random expulsion molecular dynamics investigation of ligand access channels and mechanisms, *J Mol Biol* 303, 797-811.
143. Winn, P. J., Ludemann, S. K., Gauges, R., Lounnas, V., and Wade, R. C. (2002) Comparison of the dynamics of substrate access channels in three cytochrome

- P450s reveals different opening mechanisms and a novel functional role for a buried arginine, *Proc Natl Acad Sci U S A* 99, 5361-5366.
144. Cojocar, V., Winn, P. J., and Wade, R. C. (2007) The ins and outs of cytochrome P450s, *Biochim Biophys Acta* 1770, 390-401.
 145. Stoll, S., Lee, Y. T., Zhang, M., Wilson, R. F., Britt, R. D., and Goodin, D. B. (2012) Double electron-electron resonance shows cytochrome P450cam undergoes a conformational change in solution upon binding substrate, *Proc Natl Acad Sci U S A* 109, 12888-12893.
 146. Poulos, T. L., Finzel, B. C., and Howard, A. J. (1987) High-resolution crystal structure of cytochrome P450cam, *J Mol Biol* 195, 687-700.
 147. McHaourab, H. S., Oh, K. J., Fang, C. J., and Hubbell, W. L. (1997) Conformation of T4 lysozyme in solution. Hinge-bending motion and the substrate-induced conformational transition studied by site-directed spin labeling, *Biochemistry* 36, 307-316.
 148. Goto, N. K., Skrynnikov, N. R., Dahlquist, F. W., and Kay, L. E. (2001) What is the average conformation of bacteriophage T4 lysozyme in solution? A domain orientation study using dipolar couplings measured by solution NMR, *J Mol Biol* 308, 745-764.
 149. de Groot, B. L., Hayward, S., van Aalten, D. M., Amadei, A., and Berendsen, H. J. (1998) Domain motions in bacteriophage T4 lysozyme: a comparison between molecular dynamics and crystallographic data, *Proteins* 31, 116-127.
 150. Mulder, F. A., Mittermaier, A., Hon, B., Dahlquist, F. W., and Kay, L. E. (2001) Studying excited states of proteins by NMR spectroscopy, *Nat Struct Biol* 8, 932-935.
 151. Yirdaw, R. B., and McHaourab, H. S. (2012) Direct observation of T4 lysozyme hinge-bending motion by fluorescence correlation spectroscopy, *Biophys J* 103, 1525-1536.
 152. Vetter, I. R., Baase, W. A., Heinz, D. W., Xiong, J. P., Snow, S., and Matthews, B. W. (1996) Protein structural plasticity exemplified by insertion and deletion mutants in T4 lysozyme, *Protein Sci* 5, 2399-2415.
 153. Luderer, R., Takken, F. L., de Wit, P. J., and Joosten, M. H. (2002) *Cladosporium fulvum* overcomes Cf-2-mediated resistance by producing truncated AVR2 elicitor proteins, *Mol Microbiol* 45, 875-884.
 154. Dixon, M. S., Jones, D. A., Keddle, J. S., Thomas, C. M., Harrison, K., and Jones, J. D. (1996) The tomato Cf-2 disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins, *Cell* 84, 451-459.
 155. Kruger, J., Thomas, C. M., Golstein, C., Dixon, M. S., Smoker, M., Tang, S., Mulder, L., and Jones, J. D. (2002) A tomato cysteine protease required for Cf-2-dependent disease resistance and suppression of autonecrosis, *Science* 296, 744-747.
 156. Rooney, H. C., Van't Klooster, J. W., van der Hoorn, R. A., Joosten, M. H., Jones, J. D., and de Wit, P. J. (2005) *Cladosporium Avr2* inhibits tomato Rcr3 protease required for Cf-2-dependent disease resistance, *Science* 308, 1783-1786.
 157. Shabab, M., Shindo, T., Gu, C., Kaschani, F., Pansuriya, T., Chinth, R., Harzen, A., Colby, T., Kamoun, S., and van der Hoorn, R. A. (2008) Fungal effector protein AVR2 targets diversifying defense-related cys proteases of tomato, *Plant Cell* 20, 1169-1183.
 158. Axtell, M. J., and Staskawicz, B. J. (2003) Initiation of RPS2-specified disease resistance in *Arabidopsis* is coupled to the AvrRpt2-directed elimination of RIN4, *Cell* 112, 369-377.
 159. Berendt, U., Haverkamp, T., Prior, A., and Schwenn, J. D. (1995) Reaction mechanism of thioredoxin: 3'-phospho-adenylsulfate reductase investigated by site-directed mutagenesis, *Eur J Biochem* 233, 347-356.
 160. Koprivova, A., Meyer, A. J., Schween, G., Herschbach, C., Reski, R., and Kopriva, S. (2002) Functional knockout of the adenosine 5'-phosphosulfate reductase gene in *Physcomitrella patens* revives an old route of sulfate assimilation, *J Biol Chem* 277, 32195-32201.

161. Lillig, C. H., Prior, A., Schwenn, J. D., Aslund, F., Ritz, D., Vlamis-Gardikas, A., and Holmgren, A. (1999) New thioredoxins and glutaredoxins as electron donors of 3'-phosphoadenylylsulfate reductase, *J Biol Chem* 274, 7695-7698.
162. Bick, J. A., Dennis, J. J., Zylstra, G. J., Nowack, J., and Leustek, T. (2000) Identification of a new class of 5'-adenylylsulfate (APS) reductases from sulfate-assimilating bacteria, *J Bacteriol* 182, 135-142.
163. Carroll, K. S., Gao, H., Chen, H., Stout, C. D., Leary, J. A., and Bertozzi, C. R. (2005) A conserved mechanism for sulfonucleotide reduction, *PLoS Biol* 3, e250.
164. Kim, S. K., Rahman, A., Bick, J. A., Conover, R. C., Johnson, M. K., Mason, J. T., Hirasawa, M., Leustek, T., and Knaff, D. B. (2004) Properties of the cysteine residues and iron-sulfur cluster of the assimilatory 5'-adenylyl sulfate reductase from *Pseudomonas aeruginosa*, *Biochemistry* 43, 13478-13486.
165. Kopriva, S., Buchert, T., Fritz, G., Suter, M., Benda, R., Schunemann, V., Koprivova, A., Schurmann, P., Trautwein, A. X., Kroneck, P. M., and Brunold, C. (2002) The presence of an iron-sulfur cluster in adenosine 5'-phosphosulfate reductase separates organisms utilizing adenosine 5'-phosphosulfate and phosphoadenosine 5'-phosphosulfate for sulfate assimilation, *J Biol Chem* 277, 21786-21791.
166. Berndt, C., Lillig, C. H., Wollenberg, M., Bill, E., Mansilla, M. C., de Mendoza, D., Seidler, A., and Schwenn, J. D. (2004) Characterization and reconstitution of a 4Fe-4S adenylyl sulfate/phosphoadenylyl sulfate reductase from *Bacillus subtilis*, *J Biol Chem* 279, 7850-7855.
167. Gao, Y., Schofield, O. M., and Leustek, T. (2000) Characterization of sulfate assimilation in marine algae focusing on the enzyme 5'-adenylylsulfate reductase, *Plant Physiol* 123, 1087-1096.
168. Kim, S. K., Rahman, A., Conover, R. C., Johnson, M. K., Mason, J. T., Gomes, V., Hirasawa, M., Moore, M. L., Leustek, T., and Knaff, D. B. (2006) Properties of the cysteine residues and the iron-sulfur cluster of the assimilatory 5'-adenylyl sulfate reductase from *Enteromorpha intestinalis*, *Biochemistry* 45, 5010-5018.
169. Bick, J. A., Aslund, F., Chen, Y., and Leustek, T. (1998) Glutaredoxin function for the carboxyl-terminal domain of the plant-type 5'-adenylylsulfate reductase, *Proc Natl Acad Sci U S A* 95, 8404-8409.
170. Kim, S. K., Gomes, V., Gao, Y., Chandramouli, K., Johnson, M. K., Knaff, D. B., and Leustek, T. (2007) The two-domain structure of 5'-adenylylsulfate (APS) reductase from *Enteromorpha intestinalis* is a requirement for efficient APS reductase activity, *Biochemistry* 46, 591-601.
171. Sun, Q. A., Su, D., Novoselov, S. V., Carlson, B. A., Hatfield, D. L., and Gladyshev, V. N. (2005) Reaction mechanism and regulation of mammalian thioredoxin/glutathione reductase, *Biochemistry* 44, 14528-14537.
172. Rouhier, N., Gama, F., Wingsle, G., Gelhaye, E., Gans, P., and Jacquot, J. P. (2006) Engineering functional artificial hybrid proteins between poplar peroxiredoxin II and glutaredoxin or thioredoxin, *Biochem Biophys Res Commun* 341, 1300-1308.
173. Li, X., Nield, J., Hayman, D., and Langridge, P. (1995) Thioredoxin activity in the C terminus of Phalaris S protein, *Plant J* 8, 133-138.
174. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *J Biomol NMR* 6, 277-293.
175. Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J., and Laue, E. D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline, *Proteins* 59, 687-696.
176. Shen, Y., and Bax, A. (2012) Identification of helix capping and b-turn motifs from NMR chemical shifts, *J Biomol NMR* 52, 211-232.
177. LeMaster, D. M. (1994) Isotope labeling in solution protein assignment and structural analysis, *Progress in Nuclear Magnetic Resonance Spectroscopy* 26, Part 4, 371-419.

178. Sattler, M., and Fesik, S. W. (1996) Use of deuterium labeling in NMR: overcoming a sizeable problem, *Structure* 4, 1245-1249.
179. Otten, R., Chu, B., Krewulak, K. D., Vogel, H. J., and Mulder, F. A. (2010) Comprehensive and cost-effective NMR spectroscopy of methyl groups in large proteins, *J Am Chem Soc* 132, 2952-2960.
180. Lee, Y. T., Glazer, E. C., Wilson, R. F., Stout, C. D., and Goodin, D. B. (2011) Three clusters of conformational states in p450cam reveal a multistep pathway for closing of the substrate access channel, *Biochemistry* 50, 693-703.
181. Gunsalus, I. C., and Sligar, S. G. (1976) Redox regulation of cytochrome P450cam mixed function oxidation by putidaredoxin and camphor ligation, *Biochimie* 58, 143-147.
182. Fedorov, R., Ghosh, D. K., and Schlichting, I. (2003) Crystal structures of cyanide complexes of P450cam and the oxygenase domain of inducible nitric oxide synthase, structural models of the short-lived oxygen complexes, *Archives of Biochemistry and Biophysics* 409, 25-31.
183. Poulos, T. L., and Howard, A. J. (1987) Crystal structures of metyrapone- and phenylimidazole-inhibited complexes of cytochrome P-450cam, *Biochemistry* 26, 8165-8174.
184. Grzesiek, S., Bax, A., Clore, G. M., Gronenborn, A. M., Hu, J. S., Kaufman, J., Palmer, I., Stahl, S. J., and Wingfield, P. T. (1996) The solution structure of HIV-1 Nef reveals an unexpected fold and permits delineation of the binding surface for the SH3 domain of Hck tyrosine protein kinase, *Nat Struct Biol* 3, 340-345.
185. Crull, G. B., Kennington, J. W., Garber, A. R., Ellis, P. D., and Dawson, J. H. (1989) ¹⁹F nuclear magnetic resonance as a probe of the spatial relationship between the heme iron of cytochrome P-450 and its substrate, *J Biol Chem* 264, 2649-2655.
186. Thomma, B. P., HP, V. A. N. E., Crous, P. W., and PJ, D. E. W. (2005) *Cladosporium fulvum* (syn. *Passalora fulva*), a highly specialized plant pathogen as a model for functional studies on plant pathogenic Mycosphaerellaceae, *Mol Plant Pathol* 6, 379-393.
187. Horger, A. C., and van der Hoorn, R. A. (2013) The structural basis of specific protease-inhibitor interactions at the plant-pathogen interface, *Curr Opin Struct Biol*.
188. Van't Klooster, J. W., Van der Kamp, M. W., Vervoort, J., Beekwilder, J., Boeren, S., Joosten, M. H., Thomma, B. P., and De Wit, P. J. (2011) Affinity of Avr2 for tomato cysteine protease Rcr3 correlates with the Avr2-triggered Cf-2-mediated hypersensitive response, *Mol Plant Pathol* 12, 21-30.
189. Nederveen, A. J., Doreleijers, J. F., Vranken, W., Miller, Z., Spronk, C. A., Nabuurs, S. B., Guntert, P., Livny, M., Markley, J. L., Nilges, M., Ulrich, E. L., Kaptein, R., and Bonvin, A. M. (2005) RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank, *Proteins* 59, 662-672.
190. Doreleijers, J. F., Sousa da Silva, A. W., Krieger, E., Nabuurs, S. B., Spronk, C. A., Stevens, T. J., Vranken, W. F., Vriend, G., and Vuister, G. W. (2012) CING: an integrated residue-based structure validation program suite, *J Biomol NMR* 54, 267-283.
191. Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts, *J Biomol NMR* 44, 213-223.
192. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr D Biol Crystallogr* 54, 905-921.
193. Brunger, A. T. (2007) Version 1.2 of the Crystallography and NMR system, *Nat Protoc* 2, 2728-2733.

194. Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR, *J Biomol NMR* 8, 477-486.
195. Tripathi, S., Li, H., and Poulos, T. L. (2013) Structural basis for effector control and redox partner recognition in cytochrome P450, *Science* 340, 1227-1230.
196. Hiruma, Y., Hass, M. A., Kikui, Y., Liu, W. M., Olmez, B., Skinner, S. P., Blok, A., Kloosterman, A., Koteishi, H., Lohr, F., Schwalbe, H., Nojiri, M., and Ubbink, M. (2013) The Structure of the Cytochrome P450cam-Putidaredoxin Complex Determined by Paramagnetic NMR Spectroscopy and Crystallography, *J Mol Biol*.
197. Sjodin, T., Christian, J. F., Macdonald, I. D., Davydov, R., Unno, M., Sligar, S. G., Hoffman, B. M., and Champion, P. M. (2001) Resonance Raman and EPR investigations of the D251N oxycytochrome P450cam/putidaredoxin complex, *Biochemistry* 40, 6852-6859.
198. Myers, W. K., Lee, Y. T., Britt, R. D., and Goodin, D. B. (2013) The conformation of p450cam in complex with putidaredoxin is dependent on oxidation state, *J Am Chem Soc* 135, 11732-11735.
199. Franke, A., Hartmann, E., Schlichting, I., and van Eldik, R. (2012) A complete volume profile for the reversible binding of camphor to cytochrome P450(cam), *J Biol Inorg Chem* 17, 447-463.
200. Hays, A. M., Dunn, A. R., Chiu, R., Gray, H. B., Stout, C. D., and Goodin, D. B. (2004) Conformational states of cytochrome P450cam revealed by trapping of synthetic molecular wires, *J Mol Biol* 344, 455-469.
201. Dunn, A. R., Dmochowski, I. J., Bilwes, A. M., Gray, H. B., and Crane, B. R. (2001) Probing the open state of cytochrome P450cam with ruthenium-linker substrates, *Proc Natl Acad Sci U S A* 98, 12420-12425.
202. Liu, W.-M., Keizers, P. H. J., Hass, M. A. S., Blok, A., Timmer, M., Sarris, A. J. C., Overhand, M., and Ubbink, M. (2012) A pH-Sensitive, Colorful, Lanthanide-Chelating Paramagnetic NMR Probe, *Journal of the American Chemical Society* 134, 17306-17313.
203. Riek, R., Fiaux, J., Bertelsen, E. B., Horwich, A. L., and Wuthrich, K. (2002) Solution NMR techniques for large molecular and supramolecular structures, *J Am Chem Soc* 124, 12144-12153.
204. Salzmann, M., Pervushin, K., Wider, G., Senn, H., and Wuthrich, K. (1998) TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins, *Proc Natl Acad Sci U S A* 95, 13585-13590.
205. Fiaux, J., Bertelsen, E. B., Horwich, A. L., and Wuthrich, K. (2002) NMR analysis of a 900K GroEL GroES complex, *Nature* 418, 207-211.
206. Kay, L. E. (2011) Solution NMR spectroscopy of supra-molecular systems, why bother? A methyl-TROSY view, *J Magn Reson* 210, 159-170.
207. Amero, C., Asuncion Dura, M., Noirclerc-Savoye, M., Perollier, A., Gallet, B., Plevin, M., Vernet, T., Franzetti, B., and Boisbouvier, J. (2011) A systematic mutagenesis-driven strategy for site-resolved NMR studies of supramolecular assemblies, *J Biomol NMR* 50, 229-236.
208. Sprangers, R., Li, X., Mao, X., Rubinstein, J. L., Schimmer, A. D., and Kay, L. E. (2008) TROSY-based NMR evidence for a novel class of 20S proteasome inhibitors, *Biochemistry* 47, 6727-6734.
209. Keizers, P. H., Mersinli, B., Reinle, W., Donauer, J., Hiruma, Y., Hannemann, F., Overhand, M., Bernhardt, R., and Ubbink, M. (2010) A solution model of the complex formed by adrenodoxin and adrenodoxin reductase determined by paramagnetic NMR spectroscopy, *Biochemistry* 49, 6846-6855.
210. Luchinat, C., Parigi, G., Ravera, E., and Rinaldelli, M. (2012) Solid-State NMR Crystallography through Paramagnetic Restraints, *J Am Chem Soc* 134, 5006-5009.
211. Bertini, I., Bhaumik, A., De Paëpe, G., Griffin, R. G., Lelli, M., Lewandowski, J. z. R., and Luchinat, C. (2009) High-Resolution Solid-State NMR Structure of a 17.6 kDa Protein, *J Am Chem Soc* 132, 1032-1040.

212. Keizers, P. H., Desreux, J. F., Overhand, M., and Ubbink, M. (2007) Increased paramagnetic effect of a lanthanide protein probe by two-point attachment, *J Am Chem Soc* 129, 9292-9293.
213. Jia, X., Yagi, H., Su, X. C., Stanton-Cook, M., Huber, T., and Otting, G. (2011) Engineering [Ln(DPA)₃] 3- binding sites in proteins: a widely applicable method for tagging proteins with lanthanide ions, *J Biomol NMR* 50, 411-420.
214. Bertini, I., Luchinat, C., and Parigi, G. (2002) Magnetic susceptibility in paramagnetic NMR, *Prog Nucl Mag Res Sp* 40, 249-273.
215. Koehler, J., and Meiler, J. (2011) Expanding the utility of NMR restraints with paramagnetic compounds: background and practical aspects, *Prog Nucl Magn Reson Spectrosc* 59, 360-389.
216. Kraft, D. (1988) A software package for sequential quadratic programming, in *Tech. Rep. DFVLR-FB 88-28*, Institute for Flight Mechanics, Koln.
217. Kuhn, H. W. (1955) The Hungarian method for the assignment problem, *Nav Res Logist Q* 2, 83-97.
218. Hamelryck, T., and Manderick, B. (2003) PDB file parser and structure class implemented in Python, *Bioinformatics* 19, 2308-2310.
219. Cock, P. J., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., and de Hoon, M. J. (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics, *Bioinformatics* 25, 1422-1423.
220. Prudencio, M., Rohovec, J., Peters, J. A., Tocheva, E., Boulanger, M. J., Murphy, M. E., Hupkes, H. J., Kosters, W., Impagliazzo, A., and Ubbink, M. (2004) A caged lanthanide complex as a paramagnetic shift agent for protein NMR, *Chemistry* 10, 3252-3260.
221. Schlichting, I., Berendzen, J., Chu, K., Stock, A. M., Maves, S. A., Benson, D. E., Sweet, R. M., Ringe, D., Petsko, G. A., and Sligar, S. G. (2000) The catalytic pathway of cytochrome p450cam at atomic resolution, *Science* 287, 1615-1622.
222. Eichmuller, C., and Skrynnikov, N. R. (2007) Observation of microsecond time-scale protein dynamics in the presence of Ln³⁺ ions: application to the N-terminal domain of cardiac troponin C, *J Biomol NMR* 37, 79-95.
223. McIntosh, L. P., Wand, A. J., Lowry, D. F., Redfield, A. G., and Dahlquist, F. W. (1990) Assignment of the backbone ¹H and ¹⁵N NMR resonances of bacteriophage T4 lysozyme, *Biochemistry* 29, 6341-6362.
224. Skinner, S. P., Moshev, M., Hass, M. A., and Ubbink, M. (2013) PARAssign--paramagnetic NMR assignments of protein nuclei on the basis of pseudocontact shifts, *J Biomol NMR* 55, 379-389.
225. Bardiaux, B., Malliavin, T., and Nilges, M. (2012) ARIA for solution and solid-state NMR, *Methods Mol Biol* 831, 453-483.
226. Bouvignies, G., and Kay, L. E. (2012) A 2D (1)(3)C-CEST experiment for studying slowly exchanging protein systems using methyl probes: an application to protein folding, *J Biomol NMR* 53, 303-310.

Nederlandse Samenvatting

Het gebruik van kernspinresonantie (NMR) spectroscopie in structuurbiologie is toegenomen in de laatste twee decennia, om de structuur van eiwitten te bepalen, voor onderzoek naar de dynamiek van eiwitten en om de relatie tussen structuur, dynamiek en functie te begrijpen. Bij de studies die in dit proefschrift zijn beschreven ligt de focus op de toepassing van zowel klassieke als paramagnetische NMR methoden om voor verschillende eiwitten fundamentele vragen te beantwoorden over het gedrag van deze eiwitten in oplossing. Hoofdstuk 1 geeft een algemene inleiding, waarin de belangrijkste aspecten van de methoden even als de eiwitten die in deze studies gebruikt zijn worden besproken.

In hoofdstuk 2 beschrijven we de toekenning van de resonanties van de backbone van *Enteromorpha intestinalis* adenosine 5'fosfosulfate reductase (EiAPR), met het gebruik van driedimensionale heteronucleaire NMR spectroscopie. Een toekenning van 99% van deze kernen is verkregen en de chemische verschuivingen van de C α en C β kernen zijn gebruikt om de secundaire structuur van dit eiwit te voorspellen. Er is een homolog eiwit gevonden waarvoor de 3D structuur al was bepaald met behulp van NMR spectroscopie: het derde thioredoxine domein van het eiwit disulfide isomerase van *Mus musculus*. De secundaire structuur van dit thioredoxine domein en de voorspelde structuur van EiAPR lijken veel op elkaar wat aangeeft dat EiAPR ook een thioredoxine domein heeft en niet een glutathion-achtig domein. Deze conclusie is in overeenstemming met een model dat David Knaff en zijn medewerkers hebben voorgesteld en geeft verder bewijs dat elektronentransport in dit eiwit is geëvolueerd om vrije thioredoxine in chloroplasten niet te gebruiken.

In hoofdstuk 3, zijn TROSY en BEST-TROSY driedimensionale resonantie NMR experimenten gebruikt om de toekenningen van de HN, N en C α kernen van de 47 kDa monooxygenase, P450cam, geïsoleerd van *Pseudomonas putida*, te verkrijgen. Men neemt aan dat P450cam in oplossing in twee vormen bestaat, een geopende vorm zonder substraat en een gesloten vorm met substraat gebonden. De toekenningen werden gekregen met twee liganden: camphor, het natuurlijke substraat en 1-phenylimidazole (1-PI), een remmer. Aangenomen is dat beide liganden het eiwit in de gesloten vorm houden. De spectra van P450cam met

camphor zijn verkregen in aanwezigheid van kalium cyanide om het eiwit in de lage-spin vorm te houden. Daardoor konden resonanties worden waargenomen, die anders verbreed zouden zijn door paramagnetische relaxatie. Toekenningen van de kernen van de methylwaterstoffen en methylkoolstoffen van P450cam met 1-PI gebonden met het gebruik van een CHD₂-selectief ¹³C TOCSY experiment. Toekenningen van HN en N kernen van P450cam zonder substraat zijn ook verkregen door een camphortitratie met cyanide-gebonden P450cam. Hoewel de RMSD tussen de camphor-gebonden en 1-PI gebonden vormen van P450cam ~0.3 Å is, verschillen de spectra van deze twee vormen aanzienlijk. Dit geeft aan dat de binding van de twee liganden leidt tot magnetische susceptibiliteiten die verschillen in grootte en oriëntatie. Deze verzameling van toekenningen van de verschillende vormen van P450cam kan gebruikt worden om te bepalen of het eiwit geopend of gesloten is, als zijn redoxpartner, putidaredoxin is gebonden (Hoofdstuk 5).

Hoofdstuk 4 beschrijft de bepaling van de structuur van avirulence factor 2 (Avr2) van de pathogene schimmel, *Cladosporium fulvum*, die tomatenplanten infecteert en een overgevoeligheidsreactie veroorzaakt, als Rcr3 aan dit eiwit bindt. De binding van Rcr3 aan Avr2 gebeurt bij pH 5 maar niet bij pH 7. De backbone kernen van Avr2 zijn bij pH 7 toegekend met het gebruik van heteronucleaire twee- en drievoudige resonantie experimenten en de zijketens zijn toegekend met homonucleaire experimenten. Nuclear Overhauser effect (NOE) gegevens en torsiehoeken voorspeld op grond van de chemische verschuivingen van C α en C β zijn gebruikt voor de structuurbepaling. Het blijkt dat er twee structuren bij pH 7 bestaan en beide werden berekend. De RMSDs ten opzichte van de gemiddelde structuur van de backbone is 0.92 ± 0.12 Å en 0.87 ± 0.21 Å. Op basis van protonering van His50 wordt een mechanisme voor de conformatieverandering voorgesteld.

Hoofdstuk 5 beschrijft de invloed van de binding van Pdx op de opening en het sluiten van P450cam in oplossing. De ¹⁵N-selectief verrijkte P450cam variant E195C/A199C is met Yb-CLaNP-7 gelabeld en pseudocontact-verschuivingen zijn gemeten. Daarbij is deze mutant gebonden aan Lu-CLaNP-7 gebruikt als diamagnetische controle. De spectra zijn toegekend met behulp van eerder verkregen toekenningen (Hoofdstuk 3) en verdere toekenningen zijn verkregen op

basis van pseudocontact-verschuivingen en veranderingen van de chemische verschuivingen. Pdx is in beide monsters getitreerd, maar er zijn geen significante veranderingen van pseudocontact-verschuivingen verkregen. Dit levert het bewijs dat P450cam zich in de gesloten vorm bevindt als Pdx bindt, in tegenstelling tot de conclusies van Poulos en zijn medewerkers.

Hoofdstuk 6 beschrijft een nieuwe methode voor de toekenning van kernen in een eiwit op basis van pseudocontact-verschuivingen van meerdere paramagnetische centra met het gebruik van 2D NMR spectra. De software die in dit hoofdstuk wordt gepresenteerd, PARAssign, maakt de identificatie van pseudocontact-verschuivingen in experimenteel verworven piekljsten mogelijk en vervolgens het gebruik van deze gegevens voor het toekennen van de kernen in combinatie met een achtparameter fit van de probe positie en de $\Delta\chi$ grootte en orientatie. Deze software is getest met synthetische en experimentele HN gegevens voor het eiwit pseudozurine en synthetische zijketen methyl gegevens voor het eiwit P450cam. Er wordt aangetoond dat PARAssign betrouwbare resultaten geeft voor beide eiwitten. PARAssign biedt een platform waarop deze benadering verder ontwikkeld kan worden, om andere paramagnetische effecten te benutten en het gebruik van gegevens verkregen uit TROSY experimenten te vergemakkelijken.

In hoofdstuk 7 is beschreven hoe paramagnetische relaxatie dispersie NMR spectroscopie gebruikt is om te bepalen of een paramagnetische centrum, CLaNP-5, beweegt op de μs -ms tijdschaal als hij op een mutant van het eiwit T4 lysozyme wordt bevestigd. Voor zes residuen, wordt een dynamisch proces waargenomen waarvan drie zich dichtbij de bevestigingsplaats bevinden. Er is geen significante beweging van het centrum als het op deze mutant wordt aangesloten.

Hoofdstuk 8 vat de belangrijkste bevindingen en conclusies van dit proefschrift samen en de toekomstperspectieven van het onderzoek.

English Summary

The use of nuclear magnetic resonance (NMR) spectroscopy in structural biology has increased considerably in the past two decades, both for the determination of protein structure and for the investigation of protein dynamics, with a view to understanding the connection between structure, dynamics and function. The studies described in this thesis focus on the application of several classical and paramagnetic NMR methodologies to different protein systems in order to answer fundamental questions regarding how these proteins behave in solution. Chapter 1 is a general introduction outlining the methods used in these studies and the proteins under investigation.

In chapter 2, we described the NMR assignment of the backbone of the C-terminal domain of *Enteromorpha intestinalis* adenosine 5'phosphosulphate reductase (EiAPR), using three-dimensional heteronuclear NMR spectroscopy. 99 % of the backbone was assigned and the C α and C β chemical shifts were used to predict the secondary structure of this protein domain. A homologous protein was found, the third thioredoxin domain of protein disulphide isomerase A4 from *Mus musculus*, for which the 3D structure has been solved by NMR spectroscopy. The secondary structure of this thioredoxin domain and the predicted secondary structure of EiAPR are very similar, which indicated that EiAPR is also a thioredoxin-like domain and not glutathione-like. This conclusion is consistent with the studies of David Knaff and co-workers, and provides further evidence that electron transfer in this system has evolved not to include free thioredoxin present in chloroplasts.

In chapter 3, TROSY and BEST-TROSY triple resonance NMR experiments were used to assign the backbone of the 47 kDa mono-oxygenase, P450cam, isolated from *Pseudomonas putida*. P450cam is believed to exist in an open, substrate-free and a closed, substrate-bound state in solution. The assignments were obtained in the presence of two ligands: camphor, the natural substrate and 1-phenylimidazole, an inhibitor; both ligands are believed to keep the protein in the closed state. The spectra of camphor bound P450cam were recorded in the presence of potassium cyanide, in order to keep the haem in the low spin state and, thereby, enable the acquisition of more resonances, that otherwise would have been broadened out due to paramagnetic relaxation. Side-chain methyl assignments were obtained for the 1-

phenylimidazole bound form using a CHD₂-selective ¹³C TOCSY experiment. An assignment for the camphor free (open) state of P450cam was also obtained via a camphor titration into CN bound P450cam. Even though the RMSD between the camphor and 1-PI bound states of P450cam is ~ 0.3 Å, the spectra of these two states were very different, indicating that 1-PI and camphor cause haem magnetic susceptibilities of different magnitudes and orientations to be observed. This collection of assignments for the different forms of P450cam can be used as probes to determine whether or not the protein is open or closed in the presence of its binding partner, putidaredoxin.

Chapter 4 describes the solution of the structure of Avriulence factor 2 (Avr2) from the pathogenic fungus, *Cladosporium fulvum*, which infects tomato plants and causes a hypersensitive response, when Rcr3 binds to this protein. Rcr3 binding to Avr2 has been reported to occur at pH 5 but not at pH 7. The backbone of Avr2 was assigned at pH 7 using heteronuclear double and triple resonance experiments and the side chains were assigned using homonuclear TOCSY experiments. From these assignments, nuclear Overhauser effect (NOE) data and dihedral angles predictions from the C α and C β shifts were obtained. It was found that two structures exist at pH 7 and both were determined with RMSDs to mean of the backbone of 0.92 ± 0.12 Å and 0.87 ± 0.21 Å. A mechanism for the conformational change was proposed based on protonation of the imidazole ring of H50.

Chapter 5 describes the influence of Pdx binding on the opening and closing of P450cam in solution using ¹⁵N-Leu selectively labelled P450cam. The E195C/A199C mutant of P450cam was tagged using Yb-CLaNP-7 and pseudocontact shifts were observed. A diamagnetic control was produced by tagging this mutant with Lu-CLaNP-7. The spectra were assigned using previously obtained assignments (Chapter 3) and extra assignments were obtained on the basis of pseudocontact shifts and chemical shift perturbations. Pdx was titrated into both samples and no significant changes in pseudocontact shifts were observed. This provides evidence that P450cam is the closed form when Pdx binds in solution, contrary to conclusions of Poulos and co-workers.

Chapter 6 describes a novel methodology for the assignment of protein nuclei on the basis of pseudocontact shifts (PCSs) from multiple paramagnetic centres using two-dimensional NMR spectra. The software presented in this chapter, PARAssign, enables the identification of PCSs from experimentally acquired peaklists and the subsequent use of these data to perform an eight parameter tensor fitting with concomitant assignment of the protein nuclei. This software was tested using experimental HN data for the protein pseudoazurin and synthetic side chain methyl data for the protein, P450cam. It is shown that PARAssign produces reliable results for both proteins and this software represents a platform on which this methodology can be further developed to include other paramagnetic effects and facilitate the use of data acquired from TROSY-based experiments.

In chapter 7, paramagnetic relaxation dispersion NMR spectroscopy was used to determine whether a paramagnetic probe, CLaNP-5, is moving on the μ s-ms timescale when attached to a mutant of T4 lysozyme. It was determined that dynamics were observed for six residues, three of which were near the attachment site. Thus, there is no significant movement of the probe when attached to this mutant of this protein

Chapter 8 summarises the principle themes and conclusions presented in this thesis and outlines future perspectives for the work described herein.

List of Publications

Skinner, S. P., Moshev, M., Hass M. A. S. and Ubbink, M. (2013) PARAssign – Paramagnetic NMR Assignment of protein nuclei on the basis of pseudocontact shifts. *J Biomol NMR* 55:379-389.

Guan, J., Keizers, P. H. J., Liu, W., Löhr, F., **Skinner, S. P.**, Heeneman, E., Schwalbe, H., Ubbink, M. and Siegal, G. (2013) Small molecule binding sites on proteins established by paramagnetic NMR spectroscopy. *J Am Chem Soc* 135:5859-5868.

Hiruma, Y., Hass M. A. S., Kikiu, Y., Liu, W., Ölmez, B., **Skinner, S. P.**, Blok, A., Kloosterman, A., Koteishi, H., Löhr, F., Schwalbe, H., Nojiri, M. and Ubbink, M. The Structure of the Cytochrome P450cam-Putidaredoxin Complex Determined by Paramagnetic NMR Spectroscopy and Crystallography. *J Mol Biol* (*in press*)

Liu W., **Skinner, S. P.**, Timmer, M., Blok, A., Filippov, D., Overhand, M. and Ubbink, M. Thioether linker: A new inert lanthanide-chelating paramagnetic NMR probe. In preparation.

Skinner, S. P., Liu, W. M., Timmer, M., Blok A., de Lint, K., Hass, M. A. S., Hiruma, Y., Löhr, F., Schwalbe, H. and Ubbink, M. P450cam remains closed and in the high spin state when Pdx binds. *In preparation*.

Curriculum Vitae

Simon Skinner was born 30th March 1985 in Barking, East London. He obtained his Bachelor degree in Molecular Medicine and Biochemistry from the University of Essex in 2009. During his Bachelor studies, he completed a year's Industrial placement in 2008 at GlaxoSmithKline in Stevenage working as an ADMET modeller in the Department of Computational and Structural Chemistry. On returning to Essex in 2009, he undertook a Bachelor's internship in Computational Chemistry in the laboratory of Prof Christopher Reynolds. After graduating at the top of his class, he was invited to join the group of Prof. Dr. Marcellus Ubbink at the University of Leiden. During his four years of PhD study, he undertook a variety of different projects including the production of software for NMR assignment using paramagnetic effects and studying the behaviour of P450cam in solution using classical and paramagnetic NMR techniques. He also completed a collaborative project with Prof Dr Renier van der Hoorn to solve the structures of Avr2 at pH 5 and pH 7. In addition to his research, in 2011 he took on the role of ICT coordinator for the Protein Chemistry group in 2011 and became a GTA on the NMR spectroscopy course. Outside the lab, having gained a fair competency in Dutch, he became a volunteer barman for COC Leiden from 2010-2011 and a member of the Science Club Bar Committee in 2012. He is currently waiting for the next chapter of his academic life to begin.