Development of NMR Methodology for the Study of Complex Molecular and Biomolecular Systems



Gogulan Karunanithy

Keble College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary 2019

Acknowledgements

Enormous gratitude is first due to my parents for their constant and unwavering love and support: I have certainly not acknowledged their sacrifices often enough. They have always impressed the importance of education upon me and I hope this thesis represents a worthy conclusion to this part of my life.

I owe a huge thanks to my supervisor Prof Andrew Baldwin. His tireless and infectious enthusiasm for science are a constant source of inspiration. A brilliant scientist, this thesis would be much poorer were it not for his substantial inputs and guidance.

The Baldwin group past and present has been home to many great scientists, who I am proud to call colleagues. Thanks in particular are due to Reid Alderson, an outstanding friend and NMR spectroscopist, for taking the time to proofread this thesis and for many great scientific discussions. Massive thanks are also due to David Dias, Virginia Casablancas-Antras and Richard Wheeler for having the patience to teach a physical chemist the ropes of wet lab molecular biology. I consider myself very lucky to have been part of the wider Robinson and Benesch groups and the many friendships I've made within them. A particular recent highlight has been the re-establishment of the PTCL soup group: a huge thanks to Shane Chandler, Joe Donlan and Joe Gault - I'll treasure the numerous soup lunches plus all the other wonderful social occasions.

I've been blessed with many incredible teachers and mentors along the way who have shaped the scientist I am today: a massive thanks in particular to Andy Baldwin, Stephen Faulkner, Brid Cronin, Harry Anderson, Sarah Jenkinson, Justin Benesch and Mr Moran (who perhaps unwittingly played a huge role in me pursuing an undergraduate chemistry degree).

Away from science, Keble MCR has truly been a home away from home and will forever have a special place in my heart. I've made some lifelong friendships here that I'll always cherish: Leon van Riesen-Haupt: you are a true friend and an absolute legend, Rob Cohen: your exploits are legendary, Charlie Kingston: your cynicism is legendary and Clare Hurst: you are two darts. I don't think a better darts team exists in the world. Thanks also to Antoine Tissier, Zammy Fairhurst-Hunter, Thomas Player and George Potter for their delightful company during my time in the MCR. I'm truly grateful for the Keble friends who have been

there still since the start and are still around: David Johnson, Sarah Worsley, Sasha Salter, Rasheed El-Bouri, Asher Carruthers and Paulin Shek - your friendship means the world to me. And of course, I owe an enormous thanks to my brother for being a constant and willing source of help and advice that I know I can always turn to.

Finally, a massive thank you to my dear Jess. It is truly a wonder that you've stuck with me but I am certainly grateful and should be much more available now that this monkey is off my back! Thanks for all the love and support - I cannot wait for more adventures together!

Abstract

Nuclear magnetic resonance (NMR) is amongst the most powerful and versatile spectroscopic tools for characterising biological systems. A key advantage of NMR techniques is that in addition to being sensitive to structural features of biomolecules through their chemical shift, chemical exchange and relaxation processes allow dynamic molecular properties to be studied. Techniques have also been developed to facilitate the characterisation of translational diffusion with NMR, through the application of pulsed field gradients.

In this thesis we develop and apply NMR methods for studying biomolecular systems. Firstly, in Chapter 2, we introduce an experimental protocol and model for the analysis of cellular systems from the diffusional properties of the water within them. In addition to reporting diffusion coefficients, this analysis also provides information on cell radii, abundance and permeability. A drawback of NMR compared to other spectroscopic techniques is its relative insensitivity. One way of improving this is through chemical exchange saturation transfer (CEST) experiments where the presence of lowly concentrated species is effectively read out through its interactions with a highly concentrated species. In Chapter 3 we show how this experiment can be used to follow changes in a range of solution properties. The diffusion and CEST experiments are combined in Chapter 4, where we introduce the novel diffusion-weighted CEST experiment. Here, the diffusion weighting is used to differentiate between compartments, while the CEST component of the experiment characterises properties of each of the compartments.

In Chapters 5 and 6, modern NMR techniques are used to characterise dynamical features of ^{15}N-labelled proteins. Experiments employed in this section include a fast dynamics analysis as well as CPMG and $R_{1\rho}$ relaxation dispersion experiments. Collectively, these experiments provide important insights into the behaviour of the proteins with atomic-level resolution.

Contents

List of Figures viii

List of Abbreviations xi

1 Introduction 1
- 1.1 Nuclear Spin and Magnetism . . . 2
- 1.2 The Basic NMR Experiment . . . 5
- 1.3 Density Matrices and the Product Operator Formalism . . . 8
- 1.4 J-Coupling . . . 11
- 1.5 Pulse Sequence Elements for Biomolecular NMR . . . 12
  - 1.5.1 INEPT . . . 13
  - 1.5.2 HSQC . . . 15
- 1.6 Relaxation . . . 17
- 1.7 Chemical Exchange . . . 19
  - 1.7.1 Relaxation Dispersion . . . 21
- 1.8 Diffusion . . . 23
  - 1.8.1 Derivation of the Stejskal-Tanner Equation . . . 27
- 1.9 In-Cell NMR . . . 28
- 1.10 Thesis Outline . . . 29

I Development of Label-free NMR Methods for the Characterisation of Biological Systems 31

2 Diffusion in Cellular Systems 32
- 2.1 Introduction . . . 33
- 2.2 Theory . . . 37
  - 2.2.1 Kärger Equations . . . 37
  - 2.2.2 Restricted Diffusion . . . 41
- 2.3 Monte Carlo Simulations . . . 45
  - 2.3.1 Validation of Simulations . . . 46
  - 2.3.2 Non-Uniform Gradients . . . 49
  - 2.3.3 Validation of Theory . . . 49

2.3.4 Monte Carlo Simulation Results . . . 54
2.4 Experimental Validation . . . 58
2.5 Discussion . . . 61
2.6 Methods . . . 63
2.6.1 Monte Carlo Simulation Methods . . . 63
2.6.2 Cell Preparation Methods . . . 66
2.6.3 NMR Experimental Methods . . . 67
2.6.4 Fitting Simulated and Experimental Data to Theoretical Model 67
2.7 Acknowledgements . . . 68

3 Chemical Exchange Saturation Transfer 69
3.1 Introduction . . . 70
3.1.1 Motivation . . . 73
3.2 Theory and Simulations . . . 75
3.2.1 Saturation Pulses . . . 75
3.2.2 Saturation Transfer between Exchanging Spins . . . 77
3.2.3 Direct and Relayed NOE Effects . . . 81
3.3 Results . . . 85
3.3.1 Effect of Viscosity on CEST profiles . . . 86
3.3.2 Effect of Digestion on CEST Profiles . . . 88
3.3.3 Effect of pH on CEST profiles . . . 91
3.3.4 Using CEST to Monitor Protein Aggregation . . . 93
3.4 Discussion . . . 97
3.5 Methods and Materials . . . 100
3.5.1 Simulations of CEST Experiments . . . 100
3.5.2 NMR Experiments . . . 100
3.5.3 Analysis of CEST profiles . . . 102
3.6 Acknowledgements . . . 103

4 Development of DiffCEST Experiment 104
4.1 Introduction . . . 105
4.2 Results and Discussion . . . 110
4.2.1 GUVs: A Model System for Testing DiffCEST . . . 110
4.2.2 Using DiffCEST to Monitor Cell Stress . . . 115
4.2.3 Discussion . . . 121
4.3 Methods and Materials . . . 123
4.3.1 Formation and Imaging of GUVs . . . 123
4.3.2 Cell Sample Preparation . . . 124
4.3.3 NMR Experimental Methods . . . 125
4.3.4 Analysis of diffCEST Experiments . . . 126
4.4 Acknowledgements . . . 127

II Atomic Resolution Studies of Labelled Biomolecules using NMR 128

5 Characterisation of Dynamics in Cyclophilin A 129
5.1 Introduction . . . 130
5.2 Results . . . 133
5.2.1 MD Determined Conformational Ensemble of CypA . . . 133
5.2.2 HSQC Comparison of WT and D66A CypA . . . 141
5.2.3 Fast Dynamics Comparison of WT and D66A CypA . . . 144
5.2.4 Micro-millisecond timescale Dynamics comparison in WT and D66A CypA . . . 151
5.3 Discussion . . . 158
5.4 Methods and Materials . . . 161
5.4.1 MD Methods . . . 161
5.4.2 Sample Preparation . . . 163
5.4.3 NMR Methods . . . 163
5.5 Acknowledgements . . . 166

6 Characterisation of Conformational Exchange in YopH 167
6.1 Introduction . . . 168
6.2 Results and Discussion . . . 170
6.2.1 D106A mutant shows evidence of significant chemical exchange 170
6.2.2 ^{15}N CPMG Relaxation Dispersion Experiments . . . 172
6.2.3 Structural Insights into the YopH^{CBD} excited state . . . 176
6.3 Methods and Materials . . . 178
6.3.1 Sample Preparation . . . 178
6.3.2 NMR Methods . . . 179
6.4 Acknowledgements . . . 180

7 Conclusions 181
7.1 List of Publications . . . 184

References 185

List of Figures

1.1 Illustration of Zeeman splitting 4
1.2 Energy level and vector model description of pulse acquire experiment 7
1.3 INEPT pulse block 14
1.4 HSQC pulse sequence 16
1.5 Exemplar ^{15}N-^{1}H protein NMR spectra 17
1.6 Simulations of two-site chemical exchange 21
1.7 Schematic illustration of the CPMG experiment 23
1.8 NMR Diffusion Experiments 25

2.1 Comparison of Simple models for analysing diffusion data 35
2.2 Overview of INDIANA protocol and model parameters 36
2.3 Simulated effects of exchange on diffusion attenuation 41
2.4 Simulations of restricted diffusion 44
2.5 Schematic showing the calculation of diffusion signal from signal particle trajectories 46
2.6 Comparison of simulations with long time restricted diffusion results 48
2.7 Non-uniform Gradients 50
2.8 Effects of size distributions on displacement distributions 51
2.9 Assessing validity of GPA in presence of non-uniform gradients and size distribution of particles 52
2.10 Assessing GPA fits to simulated data 53
2.11 Error plots comparing fitted parameters with simulation inputs . . . 55
2.12 Diffusion attenuation measured experimentally on cellular systems . 60
2.13 Uncertainty estimates of individual signal intensities 68

3.1 CEST overview figure 71
3.2 Mechanisms of saturation transfer 73
3.3 Stroke Patient CEST data at presentation and 24 hours later 74
3.4 Effect of saturation pulses 76
3.5 Simulated CEST profiles for two spin system 79
3.6 Plots of $J(\omega)$ and cross-relaxation rate 82
3.7 Simulated CEST spectra including three pools 84

3.8 Effect of varying solvent viscosity on CEST profile of BSA models . 87
3.9 Effect of trypsin digestion on BSA CEST profile 89
3.10 Effect of pH on BSA CEST profile 92
3.11 Monitoring aggregation of α-synuclein by 1D NMR and CEST . . . 95

4.1 Schematic comparison of diffCEST and CEST experiments 106
4.2 DiffCEST pulse sequence 107
4.3 Schematic of DiffCEST raw data 108
4.4 Cartoon of GUVs 111
4.5 Images and water diffusion analysis of GUVs 112
4.6 Comparison of watergate and CEST spectra in GUVs 113
4.7 Simulations of sucrose 1D and CEST spectra 114
4.8 DiffCEST of GUVs 114
4.9 Observation of stress droplet formation in modified HeLa cells . . . 116
4.10 Validation of NMR samples by microscopy 118
4.11 Comparison of unstressed and stressed cells using the diffCEST experiment 120
4.12 Normal water CEST experiments on stressed and unstressed HeLa cell suspensions 121
4.13 Summary of diffCEST experimental results 124

5.1 Structure of CypA 133
5.2 Workflow to generate CypA conformational ensembles 134
5.3 MD calculated structural ensemble of WT CypA and experimental validation 136
5.4 MSM for mutated forms of CypA 140
5.5 Comparison of HSQC spectra for WT and D66A CypA 142
5.6 Charcterisation of increased disorder in CypA D66A mutant 143
5.7 R_1, R_2 and hetNOE measurements of WT and D66A CypA 147
5.8 Model-free analysis of WT and D66A CypA 149
5.9 Comparison of experimental and back calculated hetNOE values . . 151
5.10 Experimental CPMG results of WT and D66A CypA 153
5.11 Exemplar $R_{1\rho}$ profiles recorded on D66A mutant and grid searching errors in fitted k_{ex} and P_B values 155
5.12 Comparison of WT and D66A CypA excited states 158
5.13 Comparison of molecular dynamics and NMR results 161

6.1 Representation of Type III secretion system (T3SS) used by toxic effector proteins such as YopH 170
6.2 HSQC and intensity comparison of WT and D106A YopH 171

6.3 Example CPMG relaxation dispersion curves for $YopH^{CBD}$ 173
6.4 Results of fitting CPMG data to global two-site exchange model . . 174
6.5 Schematic illustration showing the relative free energies of the WT and D106A forms of the protein. 175
6.6 $YopH^{CBD}$ structural comparisons 177

List of Abbreviations

1D . . . . . one dimensional

ADC . . . . . apparent diffusion coefficient

aMD . . . . . accelerated molecular dynamics

APT . . . . . amide proton transfer

BSA . . . . . bovine serum albumin

CBD . . . . . core binding domain

CEST . . . . . chemical exchange saturation transfer

CPMG . . . . Carr-Purcell-Meiboom-Gill

CSA . . . . . chemical shift anisotropy

CSP . . . . . chemical shift perturbation

CypA . . . . . Cyclophilin A

diffCEST . . . diffusion-weighted chemical exchange saturation transfer

eNOE . . . . . exact nuclear Overhauser effect

FBS . . . . . foetal bovine serum

FID . . . . . free induction decay

GFP . . . . . green fluorescent protein

GPA . . . . . Gaussian Phase Approximation

GUV . . . . . giant unilamellar vesicle

hetNOE . . . . heteronuclear NOE

HSQC . . . . . heteronuclear single quantum coherence

IDP . . . . . intrinsically disordered protein

INDIANA . . in-cell diffusion analysis

INEPT . . . . insensitive nuclei enhanced by polarisation transfer

ITC . . . . . isothermal titration calorimetry

MD . . . . . molecular dynamics

MRI . . . . . magnetic resonance imaging

MSM . . . . . Markov state modelling

NMR . . . . . nuclear magnetic resonance

NOE . . . . . nuclear Overhauser effect

PBS . . . . . phosphate buffered saline

PDB . . . . . protein data bank

PFG . . . . . pulsed field gradient

PFGSE . . . . pulsed field gradient spin echo

PFGSTE . . . pulsed field gradient stimulated echo

ppm . . . . . parts per million

RDC . . . . . residual dipolar coupling

SAXS . . . . . small angle x-ray scattering

T3SS . . . . . type 3 secretion system

TROSY . . . . transverse relaxation optimised spectroscopy

WT . . . . . wild type

Yop . . . . . Yersinia outer protein

1 Introduction

Nuclear magnetic resonance (NMR) is among the most versatile of all spectroscopic techniques with applications in numerous fields including chemistry, physics, biochemistry, materials science and medicine. Remarkably, for such a ubiquitous and pervasive technique, NMR relies on measuring the seemingly innocuous property of nuclear magnetism. Despite its largely imperceptible effects on the properties of matter, the inherent sensitivity of NMR to local surroundings combined with decades of ingenuity by NMR spectroscopists mean that nuclear magnetism can be harnessed to provide unique and detailed insights into the behaviour of systems. This thesis describes the development and application of solution-state NMR techniques for studying biological systems.

This thesis is split into two parts. The first part is concerned with the development of 'label-free' methods: the NMR experiments developed and performed in this part take advantage of NMR signals associated with endogenous species in the systems, without the need for exogenous agents or chemical labelling. In particular, the experiments here take advantage of the abundant water signal to provide a range of system insights. In the second part of the thesis, we show how NMR can be used to study the structure and dynamics of isotopically labelled proteins with atomic level resolution. As a biophysical technique, NMR is unique

in being able to study dynamic motions over a wide range of timescales, and we highlight this property for different proteins.

This chapter begins with a description of nuclear spin and a brief overview of the theoretical concepts required to describe simple NMR experiments. The approach adopted in these sections borrows heavily from Levitt(*1*), Hore, Jones and Wimperis (*2*, *3*) and Cavanagh, Palmer, Fairbrother, Rance and Skelton(*4*). Following this, a more detailed review of aspects of NMR particularly pertinent to this thesis is presented.

1.1 Nuclear Spin and Magnetism

Like mass or charge, nuclear spin is an intrinsic property of nuclei. Unlike these other properties however, it is purely quantum mechanical, making it difficult to explain accurately by analogy to everyday experience. Although not associated with the rotation of particles, spin is a form of angular momentum. Nuclear spin itself arises from the spin associated with its constituent protons and neutrons, whose own spin come from their constituent quarks.

The size of the nuclear spin is described by the nuclear spin quantum number, I, that can take on integer or half integer values. The nuclear magnetic moment, $\boldsymbol{\mu}$, is related to the nuclear spin through the relationship:

$$\boldsymbol{\mu} = \gamma \boldsymbol{I} \tag{1.1}$$

Where the proportionality constant, γ, is the gyromagnetic ratio that varies between different nuclei. Consequently any nucleus with non-zero nuclear spin will have an associated nuclear magnetic moment whose properties can be potentially measured in an NMR experiment. The total angular momentum of a nucleus is determined by the individual angular momenta of its protons and neutrons according to the standard rules for their coupling(*5*). In general, a nucleus will have several different possibilities for its total nuclear angular momentum, but the energy differences between these different I values is extremely large, many orders of

magnitude above the thermal energy, kT, for any NMR experiment. Consequently, from the perspective of an NMR spectroscopist, only the ground state nuclear angular momentum is important and can be treated as a given property of each nucleus.

Another important quantum number for describing nuclear spin is the projection quantum number, m_I. This quantity takes on 2I+1 values of -I, -I+1,...,I-1,I. For NMR studies spin I=$\frac{1}{2}$ are particularly advantageous as they have spherical charge distributions that have favourable relaxation properties, whereas for spin $I > \frac{1}{2}$ quadrupolar interactions become important leading to faster relaxation and broad linewidths in NMR spectra. In this thesis, our studies are carried out on ^{1}H and ^{15}N nuclei that are both $I = \frac{1}{2}$ so our discussion below is limited to these nuclei.

A spin I=$\frac{1}{2}$ nucleus can have either $m_I = \frac{1}{2}$ or $-\frac{1}{2}$, these are often referred to as the α and β states respectively. In the absence of an external magnetic field these two energy levels are degenerate. However, the degeneracy is broken on application of such a field, described as the Zeeman effect. The Hamiltonian describing this interaction is given by:

$$\hat{H} = -\boldsymbol{\mu} \cdot \boldsymbol{B} \tag{1.2}$$

In this equation $\boldsymbol{B}$ describes the direction of the external magnetic field. In NMR, by convention, the direction of the external field (referred to as B_0) is said to define z-axis. This simplifies the Hamiltonian to $\hat{H} = -\mu_z B_0 = -\gamma I_z B_0$. The corresponding energy eigenvalues are given by:

$$E_I = -m_I \hbar \gamma B_0 \tag{1.3}$$

A spin I= $\frac{1}{2}$ nucleus will thus have energy eigenvalues of $-\frac{1}{2}\hbar\gamma B_0$ and $\frac{1}{2}\hbar\gamma B_0$ for the α and β eigenstates that refer to nuclear magnetisation aligned parallel and antiparallel to the B_0 field respectively. The energy difference between these two states is given by $\Delta E = \hbar\gamma B_0$ and, as illustrated in Fig 1.1, scales linearly with the strength of the B_0 field.

The loss of degeneracy between the two states means that for an ensemble of spins there will be a slight preference for the lower energy α state over the β state,

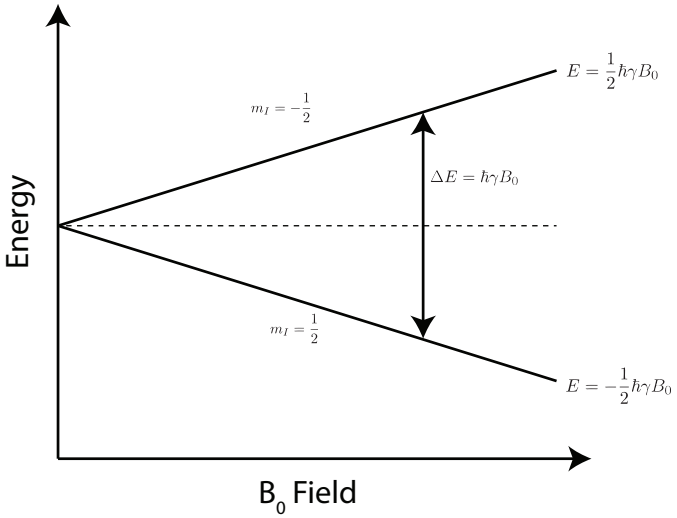

Figure 1.1: Schematic illustration of the Zeeman splitting for spin $I = \frac{1}{2}$ nuclei.

though it should be noted that most nuclear spins will exist in a superposition state rather than one of these eigenstates. The probability of the α and β states in the ensemble (p_α and p_β) can be quantified using the Boltzmann distribution as:

$$\begin{aligned} p_\alpha &= \frac{e^{\frac{\hbar\gamma B_0}{2kT}}}{e^{\frac{\hbar\gamma B_0}{2kT}} + e^{-\frac{\hbar\gamma B_0}{2kT}}} = \frac{e^{\frac{\hbar\gamma B_0}{2kT}}}{2\cosh(\frac{\hbar\gamma B_0}{2kT})} \\ p_\beta &= \frac{e^{-\frac{\hbar\gamma B_0}{2kT}}}{e^{\frac{\hbar\gamma B_0}{2kT}} + e^{-\frac{\hbar\gamma B_0}{2kT}}} = \frac{e^{-\frac{\hbar\gamma B_0}{2kT}}}{2\cosh(\frac{\hbar\gamma B_0}{2kT})} \end{aligned} \tag{1.4}$$

Inserting the value for the gyromagnetic ratio of ^{1}H and assuming a B_0 field strength of 14.1 T (600 MHz) and temperature of 298 K, as is typically employed in this thesis, gives $p_\alpha = 0.50005$ and $p_\beta = 0.49995$. The very tiny preference for the lower energy α state reflects the small energy difference between the energy levels, much smaller than kT, and is also responsible for NMR being an inherently insensitive technique, requiring much higher sample concentrations than other spectroscopies. Nonetheless, this tiny preference is sufficient to confer a net magnetisation or polarisation to the ensemble of spins, aligned with the B_0 field that can be manipulated to give rise to the huge variety of NMR experiments available today.

Measurement of the net magnetisation aligned with the B_0 field (longitudinal magnetisation) itself is largely impractical: its magnetisation is dwarfed by the

strength of the B_0 field and will also be obscured by much larger diamagnetic effects from electrons in the sample. Instead, NMR is measured through the precession of nuclear spins in the transverse plane: the nuclear spins in the magnetic field will experience a torque that causes them to precess around the magnetic field. The cone angle formed by the spins as they precess in the field depends only on the direction in which they are originally pointing. The speed at which the precession occurs depends on the Larmor frequency, ω_0, which can be calculated as:

$$\omega_0 = -\gamma B_0 \tag{1.5}$$

The precession frequency matches the frequency associated with the energy difference for the split Zeeman states ($\Delta E = \hbar\omega_0$). The precession also occurs with a defined direction: spins with positive and negative gyromagnetic ratios precess about the static field clockwise and anticlockwise respectively. When the net magnetisation is aligned with the static magnetic field however, no net precession of the sample exists. The maximum sensitivity is achieved by first rotating the net magnetisation into the transverse plane. The precession of the net nuclear magnetisation then occurs in the transverse plane and induces an oscillating current in a sensitive radio frequency detector perpendicular to the B_0 field. In modern NMR spectrometers the rotation of the magnetisation is achieved using coherent radio frequency pulses (referred to as pulses) as discussed below.

1.2 The Basic NMR Experiment

In principle magnetic resonance can be measured by the varying the frequency of an applied electromagnetic field until the resonance condition, ΔE, is met. This was how continuous wave (CW) NMR spectrometers worked, though often the field of the frequency of the applied field was held constant while the static field strength B_0 was varied. One of the most significant developments in NMR was the development of Fourier transform NMR methods, pioneered by Ernst and Anderson(*6*). Instead of slowly bringing the individual lines into resonance, in this

method short high power radio frequency pulses were used to simultaneously excite all species contributing to the NMR spectrum. The detected signal is referred to as the free induction decay (FID) that is then Fourier transformed to delineate the individual precession frequencies in the sample.

Fourier transform (FT) NMR provides a huge advantage in sensitivity over CW methods. In a CW NMR experiment a considerable amount of time is spent slowly stepping through the spectrum as different spins come into resonance. On the other hand, using the FT method a full spectrum can be acquired after application of a single pulse. After a short delay during which the system relaxes back to thermal equilibrium, the process of pulsing and acquisition is repeated increasing the signal to noise ratio in the spectrum.

The use of pulses to manipulate the nuclear magnetisation also opens up possibilities for designing pulse sequences to probe systems in much greater detail. One of the most important developments in NMR, particularly in the study of biological systems, is multidimensional experiments, first conceived by Jeener(*7*). NMR pulse sequences have also been developed to describe features of molecular behaviour such as diffusion and chemical exchange as described in this thesis.

Turning our attention back to the most basic pulse-acquire NMR experiments, as shown in Fig 1.2**A**, this can be rationalised using a simple energy level model. In a magnetic field the Zeeman effect lifts the degeneracy of the nuclear spin energy levels and we get a slight excess of spins in the lower energy α state (hugely exaggerated here). On application of a 90° pulse the populations of the α and β states are equalised but there is a coherence (correlated phase relationship) between the spins. This coherence between spins leads to detection of the NMR signal in the FID. During and following this time the magnetisation relaxes back to equilibrium, where the coherence between the spins is lost and the Boltzmann populations of spins is re-established.

An alternative but equivalent description of the experiment is provided by the vector model in Fig 1.2**B**. Instead of considering the changes in energy level, this approach arguably offers more intuitive insight by showing the evolution of the

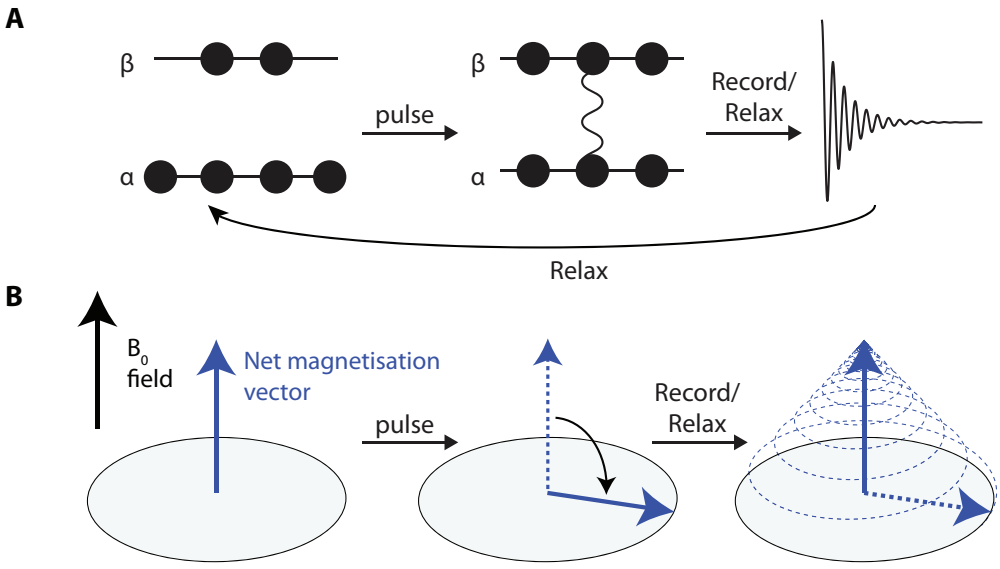

Figure 1.2: **A** Illustration of a pulse acquire NMR experiment from an energy level perspective. **B** An alternative but equivalent view of the experiment provided by the vector model that follows the evolution of the net magnetisation vector.

net magnetisation during the experiment. Initially the nuclear magnetisation is aligned with the static magnetic field. Following the application of a 90° pulse the magnetisation is rotated into the transverse plane. It subsequently precesses in this plane, while relaxing back to equilibrium where it is again aligned with the B_0 field. The effect of relaxation has been mentioned a few times in this section. This is a fundamental topic in magnetic resonance that will be returned to, but here we note that relaxation results in the loss of coherence between spins and the re-establishment of the longitudinal magnetisation according to the Boltzmann distribution. For an uncoupled spin these processes are characterised by the transverse (R_2) and longitudinal (R_1) relaxation rates respectively.

While the vector model provides an excellent visual insight into NMR experiments, it is not sufficient for describing the full complexities of modern NMR experiments. For this we require density matrices and their convenient expression using the product operator formalism, explained in the next section.

1.3 Density Matrices and the Product Operator Formalism

As an intrinsically quantum mechanical technique, quantum mechanics provides the framework with which to describe NMR experiments. Each nuclear spin in the sample can be considered decoupled from the rest of the molecule and will be described by a state vector, $|\psi\rangle$, that evolves according to the time-dependent Schrödinger equation. However, in an NMR experiment we are dealing with an extremely large number of nuclear spins, described by different state vectors meaning that the ensemble can be described as a probabilistic mixture of state vectors. Such an ensemble is referred to as a mixed state and is described using a density matrix. The density matrix, ρ, can be calculated as:

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i| \tag{1.6}$$

where the sum is over all pure states ψ_i that appear with probability p_i and $\sum_i p_i = 1$. A particularly convenient feature of density matrices is that the expectation value of an observable, $\langle A\rangle$, corresponding to operator $\hat{A}$ can be calculated simply as:

$$\langle A\rangle = \mathrm{Tr}(\rho\hat{A}) \tag{1.7}$$

where Tr refers to the calculation of the trace of the matrix product. The evolution of the density matrix under a Hamiltonian, $\hat{H}$, is given by the Liouville-von Neumann equation as:

$$\frac{d\rho}{dt} = -i[\hat{H}, \rho] \tag{1.8}$$

that has the solution:

$$\rho(t) = e^{-i\hat{H}t}\rho(0)e^{i\hat{H}t} \tag{1.9}$$

The density matrix for $I = \frac{1}{2}$ nuclear spins at equilibrium in a static field magnetic field of strength B_0 aligned along the z-axis is given by approximately:

$$\rho_{eq} = \frac{1}{2}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} + \frac{1}{2}B\begin{bmatrix}\frac{1}{2} & 0\\ 0 & -\frac{1}{2}\end{bmatrix} \tag{1.10}$$

where $B = \frac{\hbar\gamma B_0}{kT}$ is a Boltzmann factor accounting for the small polarisation in the sample. In terms of operators this equation can be rewritten as:

$$\rho_{eq} = \frac{1}{2}\hat{E} + \frac{1}{2}B\hat{I}_z \tag{1.11}$$

in which $\hat{E}$ is the identity operator and $\hat{I}_z$ is the z-axis angular momentum operator. The identity operator commutes with all other operators and so will not evolve under any Hamiltonian and therefore will not be considered further. On the other hand, the $\hat{I}_z$ part of the density operator can be easily linked with the vector model for NMR experiments, representing the net nuclear magnetisation aligned with the static magnetic field at equilibrium. This part of the density matrix will also lead to observable signal following the application of pulses. In the discussion below we will remove the $\frac{1}{2}B$ Boltzmann factor for convenience.

A useful device when describing NMR experiments is the rotating frame transformation. This frame rotates with an angular frequency of ω_{RF} usually set to be the middle of the frequency range for the nucleus under study. In this frame, nuclei precess around the z-axis with a reduced offset frequency of $\Omega = \omega - \omega_{RF}$, in which ω is the Larmor frequency of the nucleus in its particular chemical environment. Furthermore, pulses are composed linearly polarised radio frequency fields that oscillate at the frequency ω_{RF}. Consequently, in the rotating frame the pulses can be treated as temporary static fields in the transverse plane, referred to as the B_1 field. In the rotating frame the effect of this field is much greater than the residual B_0 field about which the magnetisation precesses meaning that the effects of precession can be safely neglected during the application of pulses. The Hamiltonian for pulses applied along the x axis is given by:

$$\hat{H} = \omega_1\hat{I}_x \tag{1.12}$$

where $\omega_1 = \gamma B_1$ and $\hat{I}_x$ is operator for the x component of the nuclear angular momentum. For the case when the pulse is applied along the y axis this can be replaced with the $\hat{I}_y$ operator. We can use the Liouville-von Neumann equation to study the effect of this pulse on ρ_{eq}:

$$\begin{bmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{bmatrix} \xrightarrow{\omega_1 \hat{I}_x} \cos(\omega_1 t) \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{bmatrix} - \sin(\omega_1 t) \begin{bmatrix} 0 & -\frac{1}{2}i \\ \frac{1}{2}i & 0 \end{bmatrix} \tag{1.13}$$

or equivalently in operator form:

$$I_z \xrightarrow{\omega_1 \hat{I}_x} \cos(\omega_1 t) I_z - \sin(\omega_1 t) I_y \tag{1.14}$$

The application of the pulse therefore makes the net magnetisation rotate into the yz-plane and the $\hat{I}_y$ operator represents the density matrix when the magnetisation is aligned with the y-axis. Similarly, when the pulse is applied along the y-axis we find:

$$I_z \xrightarrow{\omega_1 \hat{I}_y} \cos(\omega_1 t) I_z + \sin(\omega_1 t) I_x \tag{1.15}$$

where in this case the magnetisation rotates into the xz-plane and the I_x operator represents the density matrix when the magnetisation is aligned with the x-axis.

The operators: E, I_x, I_y and I_z are referred to as product operators and can be used to describe NMR experiments for systems of isolated spin $I = \frac{1}{2}$ nuclei (we remove the operator hats when referring to the product operators). While the product operators refer to density matrix states, they can be used to describe experiments much more simply by following simple rules for the evolution of magnetisation during pulses and periods of free precession. The rules can be summarised as:

$$\begin{array}{lll}
I_z \xrightarrow{\omega_1 \hat{I}_y} \cos(\omega_1 t) I_z + \sin(\omega_1 t) I_x & I_x \xrightarrow{\omega_1 \hat{I}_y} \cos(\omega_1 t) I_x - \sin(\omega_1 t) I_z & I_y \xrightarrow{\omega_1 \hat{I}_y} I_y \\
I_z \xrightarrow{\omega_1 \hat{I}_x} \cos(\omega_1 t) I_z - \sin(\omega_1 t) I_y & I_y \xrightarrow{\omega_1 \hat{I}_x} \cos(\omega_1 t) I_y + \sin(\omega_1 t) I_z & I_x \xrightarrow{\omega_1 \hat{I}_x} I_x \\
I_x \xrightarrow{\Omega \hat{I}_z} \cos(\Omega t) I_x + \sin(\Omega t) I_y & I_y \xrightarrow{\Omega \hat{I}_z} \cos(\Omega t) I_y - \sin(\Omega t) I_x & I_z \xrightarrow{\Omega \hat{I}_z} I_z
\end{array} \tag{1.16}$$

The most common values for $\omega_1 t$ during pulses are $\omega_1 t = \frac{\pi}{2}$ and $\omega_1 t = \pi$, corresponding to 90° and 180° pulses respectively. Most NMR pulse sequences are composed of 90° and 180° pulses separated by judiciously chosen periods of delay or free precession.

A key advantage of the product operators being written in the Cartesian basis is that they can easily be linked to the vector model, giving an intuitive picture

of how the magnetisation evolves during an experiment. An alternative basis that is often useful for describing NMR experiments is the shift basis that is linked to the Cartesian basis by the following relationships:

$$\begin{aligned} \hat{I}_\alpha &= \frac{1}{2}\hat{E} + \hat{I}_z \qquad \hat{I}_\beta = \frac{1}{2}\hat{E} - \hat{I}_z \\ \hat{I}_+ &= \hat{I}_x + i\hat{I}_y \qquad \hat{I}_- = \hat{I}_x - i\hat{I}_y \end{aligned} \tag{1.17}$$

While the shift basis lacks the obvious link to the vector model, its key advantage is that the $\hat{I}_+$ and $\hat{I}_-$ states, which correspond to transverse magnetisation, are eigenstates of the free precession Hamiltonian ($\hat{H} = \Omega\hat{I}_z$).

A common strategy employed in this thesis to analyse the evolution of magnetisation in particular experiments is to write down differential equations for the product operators under the relevant Hamiltonian using the Liouville-von Neumann equation (1.8). This gives us the Bloch equations(*8*). In cases where exchange or diffusion is also important these are extended to give the Bloch-McConnell(*9*) or Bloch-Torrey(*10*) equations respectively. Rather than solving these equations using equation 1.9 we find it more convenient to take the matrix exponential given by $e^{-i[\hat{H},\rho]t}$ that can easily be solved computationally.

1.4 J-Coupling

The J-coupling is an intramolecular coupling between nuclear spins that is modulated by the bonding electrons between them. Even in rapidly tumbling molecules, this interaction has an isotropic component known as the isotropic J-coupling or scalar coupling. The most obvious manifestation of this interaction in NMR spectra is the splitting of peaks it causes: this can be a source of information on bond connectivities as well as torsional angles (particularly three-bond H-H couplings(*11*)).

Of greater importance for the work presented in this thesis however is the ability of J-couplings to be harnessed to control the transfer of magnetisation between nuclear spins. For our purposes, the relevant Hamiltonian for the J-coupling between

two nuclear spins I and S is given by (weak coupling limit):

$$\hat{H} = 2\pi J_{IS} \hat{I}_z \hat{S}_z \tag{1.18}$$

where J_{IS} is the scalar coupling between the two spins.

The product operator formalism is particularly powerful for studying coupled spins: the product operators for the coupled spins are formed by taking the direct products of the single spin product operators yielding sixteen product operators in total for describing the coupled spins. The rules for the evolution under the J-coupling Hamiltonian are:

$$\begin{aligned}
I_x &\xrightarrow{2\pi J_{IS}\hat{I}_z\hat{S}_z} I_x \cos(\pi J_{IS} t) + 2I_y S_z \sin(\pi J_{IS} t) \\
I_y &\xrightarrow{2\pi J_{IS}\hat{I}_z\hat{S}_z} I_y \cos(\pi J_{IS} t) - 2I_x S_z \sin(\pi J_{IS} t) \\
I_z &\xrightarrow{2\pi J_{IS}\hat{I}_z\hat{S}_z} I_z \\
2I_x S_z &\xrightarrow{2\pi J_{IS}\hat{I}_z\hat{S}_z} 2I_x S_z \cos(\pi J_{IS} t) + I_y \sin(\pi J_{IS} t) \\
2I_y S_z &\xrightarrow{2\pi J_{IS}\hat{I}_z\hat{S}_z} 2I_y S_z \cos(\pi J_{IS} t) - I_x \sin(\pi J_{IS} t) \\
2I_z S_z &\xrightarrow{2\pi J_{IS}\hat{I}_z\hat{S}_z} 2I_z S_z
\end{aligned} \tag{1.19}$$

Using appropriate time periods it is therefore possible that after evolving under the J-coupling Hamiltonian in phase magnetisation (e.g. I_x or I_y) can be converted into antiphase magnetisation ($I_x S_z$ or $I_y S_z$) or vice versa. The ability to do this is essential for many NMR pulse sequences as highlighted in the next section.

1.5 Pulse Sequence Elements for Biomolecular NMR

NMR experiments on biomolecular systems are associated with a number of inherent difficulties. Firstly, for large molecules such as proteins, there will be many protons in a range of environments giving rise to a very large degree of overlap in a 1D ^{1}H spectrum. Furthermore, transverse relaxation rates scale with the size of molecules so that for large protein-sized molecules signals may be broadened beyond detection.

One method for reducing the degree of spectral overlap is to perform multidimensional experiments(*7*, *12*) such that the chemical shift dispersion is spread out over multiple frequency dimensions and the overall overlap is reduced. Early 2D ^{1}H-^{1}H experiments were successful at elucidating structural features of relatively small proteins(*13*, *14*), but these approaches become intractable for proteins bigger than approximately 10 kDa due to severe peak overlap. On the one hand, the high gyromagnetic ratio and natural abundance of ^{1}H nuclear spins make them the most useful and widely studied of all spins in NMR experiments. As shown in part I of this thesis through careful design of pulse sequences and analysis tools, ^{1}H-based NMR methods, without any additional isotopic labelling, can still be a very valuable source of information when studying biological systems.

In order to realise the full potential of NMR for the study larger biomolecules at atomic level resolution it is necessary to perform experiments on isotopically labelled samples (*15*), as is done in part II of this thesis. The two nuclei that are most commonly isotopically enriched in biomolecular NMR samples are ^{15}N and ^{13}C. In addition to being elements that appear abundantly in biomolecules both ^{15}N and ^{13}C also benefit from being spin $I = \frac{1}{2}$ with associated favourable relaxation properties. We limit our attention here to ^{15}N labelled proteins that will have this NMR active nuclei at each backbone amide bond as well as amine and -NH containing side chains.

Below we describe the INEPT element, an essential part of most biomolecular pulse sequences, as well as the HSQC experiment used to record 'fingerprint' spectra of proteins.

1.5.1 INEPT

Although ^{15}N-labelled spins give rise to spectra with good dispersion in the ^{15}N-dimension allowing peaks to be easily delineated, a disadvantage of them is that their inherent sensitivity is low. This is a result of the low gyromagnetic ratio of ^{15}N nuclear spins compared to ^{1}H ($\gamma_{1_H} = 267.5 \times 10^6$ rad s^{-1} T^{-1} and $\gamma_{15_N} = -27.1 \times 10^6$ rad s^{-1} T^{-1}). The INEPT (insensitive nuclei enhancement by polarisation transfer)

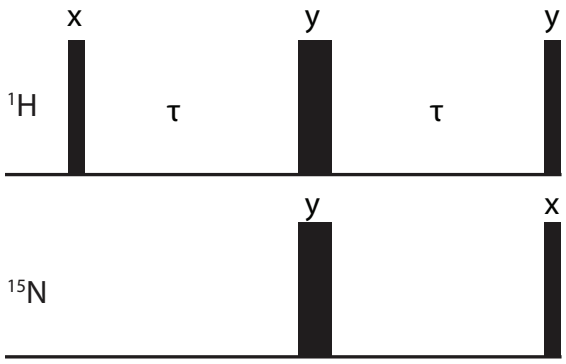

Figure 1.3: The INEPT pulse block: narrow and wide lines represent 90° and 180° pulses respectively. The phase of each pulse is indicated by the pulse.

sequence provides a way of overcoming this difficulty by transferring polarisation from the more sensitive ^{1}H nucleus onto an adjacent ^{15}N(*16*).

The INEPT block is shown in Fig 1.3. Using the product operator formalism, the INEPT sequence can be described as follows (where I refers to the ^{1}H spin and S to the the J-coupled ^{15}N spin):

$$
\begin{aligned}
&I_z \xrightarrow{(\frac{\pi}{2}_x)_I} -I_y \xrightarrow{\Omega\tau} -\cos(\Omega\tau)I_y + \sin(\Omega\tau)I_x \xrightarrow{\pi J\tau} \\
&\quad \cos(\Omega\tau)(-\cos(\pi J\tau)I_y + \sin(\pi J\tau)I_xS_z) + \sin(\Omega\tau)(\cos(\pi J\tau)I_x + \sin(\pi J\tau)I_yS_z) \xrightarrow{(\pi_y)_{I,S}} \\
&\quad \cos(\Omega\tau)(-\cos(\pi J\tau)I_y + \sin(\pi J\tau)I_xS_z) + \sin(\Omega\tau)(-\cos(\pi J\tau)I_x - \sin(\pi J\tau)I_yS_z) \xrightarrow{\Omega\tau} \\
&\quad -\cos(\pi J\tau)I_y + \sin(\pi J\tau)2I_xS_z \xrightarrow{\pi J\tau} \\
&\quad -\cos(2\pi J\tau)I_y + \sin(2\pi J\tau)2I_xS_z \xrightarrow{(\frac{\pi}{2}_y)_I} -\cos(2\pi J\tau)I_y - \sin(2\pi J\tau)2I_zS_z \xrightarrow{(\frac{\pi}{2}_x)_S} \\
&\quad -\cos(2\pi J\tau)I_y + \sin(2\pi J\tau)2I_zS_y
\end{aligned}
\tag{1.20}
$$

Where we have simplified the product operators using the trigonometric relationships: $\cos^2\theta + \sin^2\theta = 1$, $\cos^2\theta - \sin^2\theta = \cos 2\theta$ and $2\sin\theta\cos\theta = \sin 2\theta$. The time period is usually set to $\tau = \frac{1}{4J}$ such that at the end of the INEPT block in-phase magnetisation on spin I is transformed into anti-phase magnetisation on the S spin. Crucially, the polarisation begins on the more sensitive ^{1}H nucleus that is then transferred onto the S spin giving a large sensitivity enhancement, proportional to $\frac{\gamma 1_H}{\gamma 15_N}$. A one bond ^{15}N-^{1}H J coupling has a typical size of approximately 92 Hz corresponding to $\tau = 2.7$ ms. The above analysis also does not take account of

initial S_z magnetisation that can be eliminated by phase cycling or an initial crushing gradient.

The effect of the INEPT block can also be rationalised by studying the overall propagator for this pulse sequence element. From the solution to the Liouville-von Neumann equation (1.9) we can write (ignoring the initial 90° x pulse on spin I):

$$e^{-i\hat{H}t} = e^{-i\frac{\pi}{2}\hat{S}_x}e^{-i\frac{\pi}{2}\hat{I}_y}e^{-i\Omega_I\hat{I}_z\tau}e^{-i\pi J\hat{I}_z\hat{S}_z\tau}e^{-i\pi\hat{I}_y}e^{-i\pi\hat{S}_y}e^{-i\Omega_I\hat{I}_z\tau}e^{-i\pi J\hat{I}_z\hat{S}_z\tau} \tag{1.21}$$

Upon multiple applications of the rotation sandwich property(*1*, *5*):

$$e^{-i\theta\hat{I}_x}e^{-i\beta\hat{I}_y}e^{i\theta\hat{I}_x} = e^{-i\beta(\hat{I}_y\cos\theta+\hat{I}_z\sin\theta)} \tag{1.22}$$

The propagator for the INEPT block can then be simplified to give:

$$e^{-i\hat{H}t} = e^{-i\frac{\pi}{2}\hat{S}_x}e^{-i\frac{\pi}{2}\hat{I}_y}e^{-i2\pi J\hat{I}_z\hat{S}_z\tau}e^{-i\pi\hat{S}_y}e^{-i\pi\hat{I}_y} \tag{1.23}$$

We note that the I spin will initially be described by the $\hat{I}_y$ operator and equilibrium S magnetisation will be suppressed. Consequently, the effects of the two 180° pulses in the propagator can be ignored. The overall effect of this propagator is therefore that the scalar coupling will evolve for a period 2τ followed by two 90° pulses that turn transverse anti-phase magnetisation on spin I into transverse anti-phase magnetisation on spin S. The effects of free precession are refocussed over the INEPT block. Returning to the product operator rules this evolution can be described as:

$$I_z \xrightarrow{(\frac{\pi}{2}{}_x)_I} -I_y \xrightarrow{2\pi J\tau\hat{I}_z\hat{S}_z} 2I_xS_z \xrightarrow{(\frac{\pi}{2}{}_y)_I} -2I_zS_z \xrightarrow{(\frac{\pi}{2}{}_y)_S} 2I_zS_y \tag{1.24}$$

1.5.2 HSQC

The HSQC (heteronuclear single quantum coherence) pulse sequence(*17*) is shown in Fig 1.4. This involves an initial INEPT block to transfer magnetisation from ^{1}H nuclei onto bonded ^{15}N nuclei. This is followed by the t_1 period during which the chemical shift of ^{15}N is recorded. The ^{1}H 180° pulse at the centre of this period refocusses the the scalar coupling. Following this there is a second INEPT block

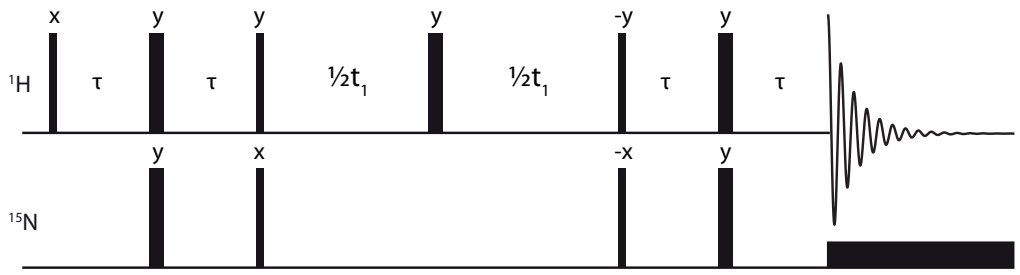

Figure 1.4: HSQC pulse sequence: narrow and wide lines represent 90° and 180° pulses respectively. The phase of each pulse is indicated by the pulse. The black rectangular box during the final detection period represents the decoupler.

allowing anti-phase magnetisation on spin S to be transferred back into in-phase magnetisation on spin I for detection.

In order to perform a two dimensional experiment it is necessary to increment the value of t_1. This yields FIDs in the detection period that are modulated by the frequency of the ^{15}N magnetisation. By performing a double Fourier transform in the two frequency dimensions we get peaks centred at the frequencies of ^{15}N nuclei and ^{1}H nuclei bonded to them. In order to achieve sign discrimination in the frequency of the S spin, the phase of the final 90° pulse on spin S in the first INEPT block is varied giving cosine and sine modulated data in the t_1 period.

The utility of the HSQC pulse sequence is exemplified in Fig 1.5: here we compare 1D and HSQC spectra for two relatively small proteins: α-synuclein (14.5 kDa) and Abp1p SH3 domain (7 kDa). The 1D ^{1}H spectra reveals the presence of a range of functional groups including: aliphatic ($\approx 0-3$ ppm), H^{α} ($\approx 3.5-5.0$ ppm), side chain NH ($\approx 6-7.5$ ppm), aromatic ($\approx 6.5-8$ ppm) and backbone amide ($\approx 6.5-11$ ppm) protons. Even for these relatively small proteins however, the overlap in these spectra is very severe meaning that it is not possible to describe these systems with atomic level resolution details from these spectra. One major difference between the two 1D spectra is that the dispersion in the chemical shifts for amide groups is much smaller in α-synuclein than Abp1p. This reflects the fact that α-synuclein is a disordered protein while the Abp1p SH3 domain is well folded, demonstrating how 1D spectra can elucidate gross structural features of proteins.

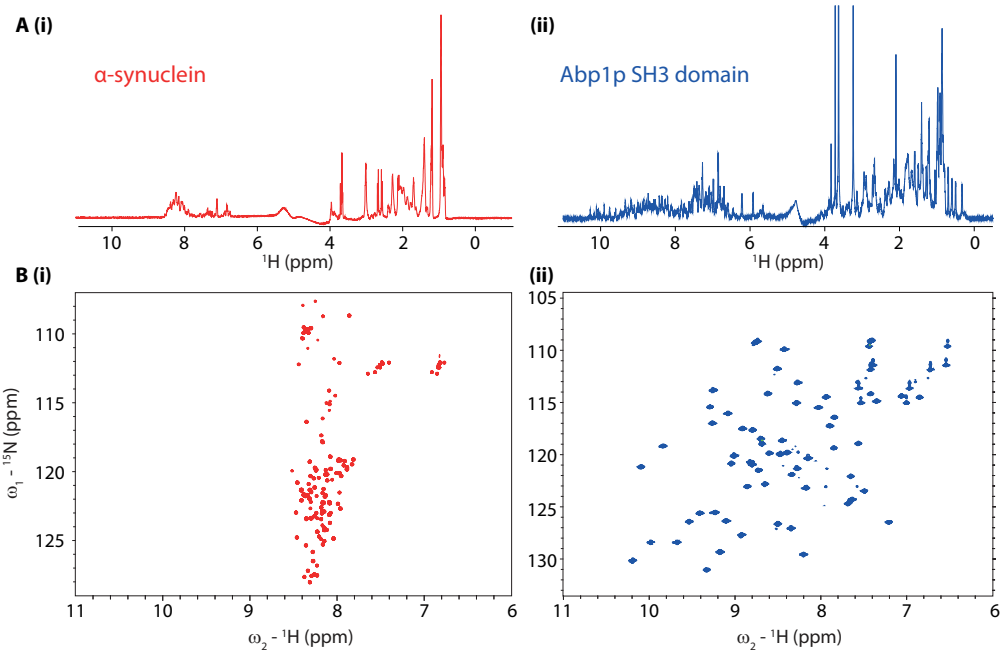

Figure 1.5: Exemplar **A** ^{1}H-1D (watergate (*18*)) and **B** ^{15}N-^{1}H HSQC spectra(*19*) of **i** α-synuclein and **ii** Abp1p SH3 domain. All spectra are recorded on a 14.1 T (600 MHz) spectrometer at 298 K.

By contrast, the 2D HSQC spectra (Fig 1.5**B**) correlates ^{15}N-^{1}H pairs and gives rise to well resolved peaks where each backbone amide (with the exception of proline) gives rise to a peak in the spectrum. Using triply isotopically labelled samples (also with ^{13}C labelling) it is possible to assign each peak to a specific amino acid residue in the protein(*20*). Once an HSQC spectrum of a protein has been assigned it can be used to follow perturbations in the protein structure with residue-level specificity.

1.6 Relaxation

Relaxation is a large and important topic in NMR that has been the subject of investigation since the earliest NMR experiments(*21*). Only a very brief introduction to this topic is provided here, excellent reviews have been written by Goldman(*22*), Fischer, Majmudar and Zuiderweg(*23*) and Cavanagh, Fairbrother, Palmer, Rance and Skelton(*4*).

The rate of a transition, W, for an isolated magnetic dipole by spontaneous emission of a photon with energy $E = \hbar\omega$ is given by(*24*):

$$W = \frac{\mu_0 \gamma^2 \hbar \omega^3}{6\pi c^3} \tag{1.25}$$

Since the difference between energy levels in NMR are very small, unlike in other forms of spectroscopy, spontaneous and stimulated emission provide a negligible contribution to relaxation, due to the very small ω^3 term in the above equation. Instead, relaxation in NMR results from fluctuating local magnetic fields. If these fields fluctuate at the appropriate frequency they will result in transitions between energy levels that will drive the system towards equilibrium. As the fluctuations in local magnetic fields themselves are modulated by molecular motion, NMR relaxation can be studied to probe this motion.

The major source of these fluctuating fields that give rise to relaxation in spin $I = \frac{1}{2}$ nuclei such as ^{1}H and ^{15}N are the dipolar interaction and chemical shift anisotropy or CSA (the latter applies particularly for ^{15}N spins). While these interactions have no effect on the position of peaks in the spectrum, as they average to zero in the liquid state due to rapid molecular tumbling, the have a non-zero effect over short timescales that gives rise to relaxation.

Each nuclear spin in the sample will have an associated local magnetic field. As the vector connecting two dipolar coupled nuclei varies due to molecular tumbling, the local field experienced by the coupled nucleus varies, forming the basis of the dipolar interaction. On the other hand, the CSA results from the variation in the secondary nuclear magnetic fields that result from the motion of electrons induced by the static magnetic field.

The relaxation rate, Γ_{RS}, between operators B_r and B_s is given according to the Bloch, Wangsness and Redfield semi-classical relaxation theory by(*4*, *25*, *26*):

$$\Gamma_{RS} = \frac{1}{2} \sum_{q=-k}^{k} \sum_{p} (\langle B_r | [A_{kp}^{-q}, [A_{kp}^{q}, B_s]] \rangle / \langle B_r | B_r \rangle) j^q(\omega_p^q) \tag{1.26}$$

in which A_k^q is a tensor spin operator from a Hamiltonian that can give rise to relaxation that is then expanded in terms of p basis operators to give A_{kp}^q and q is the rank of the tensor. The function $j^q(\omega_p^q)$ is the associated power spectral density

function that accounts for the spatial dependence of the fluctuating Hamiltonian. A full detailed discussion of the terms in this equation is provided in (*4*). This equation may give the impression that relaxation in NMR shares no features with relaxation mechanisms usually encountered in other spectroscopic methods. However, the rates in equation 1.26 can still be considered as being composed of a selection rule (the double commutator) and the requirement that motion occurs at the appropriate frequency to facilitate transitions between energy levels (through dependence on the spectral density function).

It is interesting to reflect on the synergy between sensitivity and relaxation in NMR. On the one hand, the relative insensitivity of NMR arises from the small energy difference between the Zeeman eigenstates. On the other hand, this leads to NMR having relatively slow relaxation rates, meaning pulse sequences can be designed to manipulate long-lived coherences and allows NMR to be used to study dynamic molecular events over relatively long timescales.

1.7 Chemical Exchange

Chemical exchange is a key concept in NMR that appears throughout this thesis. We begin by describing this phenomenon from the perspective of a single nuclear spin.

The specific resonance frequency of a nuclear spin in a sample is strongly dependent on its chemical environment. Thus, in the absence of scalar couplings, we would normally expect one peak in an NMR spectrum per chemically equivalent spin (in the absence of J-couplings these will also be magnetically equivalent). This overlooks the dynamical nature of many molecules. A single spin may actually exchange between chemical environments as a result of molecular dynamics each of which has a distinct resonance frequency. A spin may also experience a different chemical environment as a consequence of a bimolecular process such as ligand binding. The manifestation of chemical exchange in an NMR spectrum is strongly dependent on both the rate of exchange between the different chemical environments as well as the difference in resonance frequencies between the environments(*27*).

We consider spin I that exchanges between two chemical environments, labelled A and B, with resonance frequencies ω_A and ω_B respectively and forward and reverse exchange rates k_{AB} and k_{BA} respectively. We can follow the evolution of transverse magnetisation using the following Bloch-McConnell matrix(*9*):

$$\frac{d}{dt}\begin{bmatrix} I_+^A \\ I_+^B \end{bmatrix} = \begin{bmatrix} i\omega_A - R_2^A - k_{AB} & +k_{BA} \\ +k_{AB} & i\omega_B - R_2^B - k_{BA} \end{bmatrix}\begin{bmatrix} I_+^A \\ I_+^B \end{bmatrix} \tag{1.27}$$

We note that the populations of the states is given by: $p_A = \frac{k_{BA}}{k_{AB}+k_{BA}}$ and $p_B = \frac{k_{AB}}{k_{AB}+k_{BA}}$. It is also useful to define the variables: $k_{ex} = k_{AB} + k_{BA}$ and $\Delta\omega = \omega_B - \omega_A$. We can take the exponential of the 2×2 matrix in equation 1.27 and multiply it by different time points to simulate the acquisition of an FID. We can then Fourier transform this result to study the effect of chemical exchange on NMR spectra in a number of different regimes. The results of this analysis in the case of both symmetric and asymmetric exchange are shown in Fig 1.6.

The simulated spectra demonstrate the different exchange timescales that are often referred to in NMR. In the slow exchange limit, $k_{ex} \ll |i\Delta\omega + \Delta R_2|$, the two chemical environments behave analogously to two distinct nuclear spins. Each spin resonates independently at its frequency and the intensity of each peak is proportional to population of each environment. As the exchange rate increases whilst remaining in the slow regime both peaks begin to broaden, their linewidths are no longer purely a product of the intrinsic relaxation of the states but also of their exchange dynamics.

Intermediate exchange occurs when $k_{ex} \approx \Delta\omega$. In the case of symmetric exchange, the two peaks broaden to the point where they coalesce into a single peak and very little signal intensity is observed in the spectrum. In the case of asymmetric exchange we see significant broadening in the major state peak while the minor state peak is no longer visible. The additional line broadening experienced by the minor state can be explained as follows: when the minor state exchanges into the major state the lifetime of this state is relatively long and the spins are therefore more likely to lose coherence with the other minor state spins upon returning.

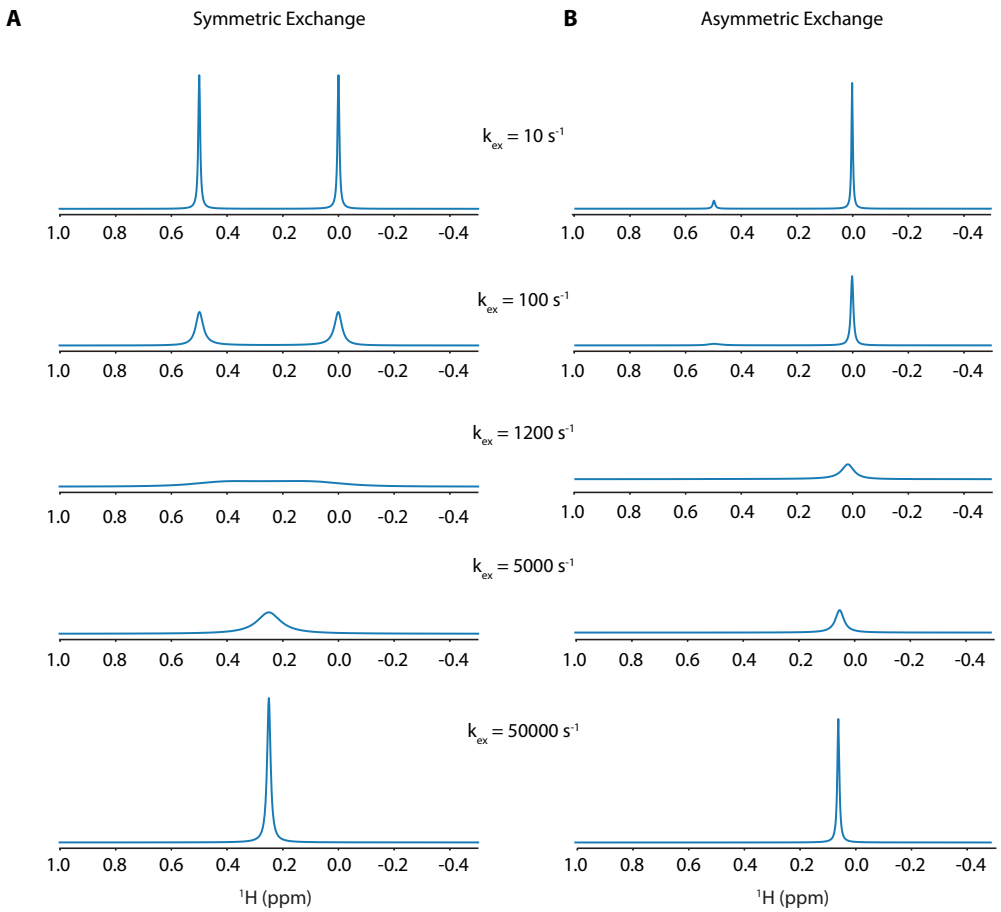

Figure 1.6: Simulated spectra of two-site chemical exchange in the case of **A** symmetric ($p_a = p_b = 0.5$) and **B** asymmetric exchange ($p_a = 0.9$ and $p_b = 0.1$). In both cases $\Delta\omega = 300 \times 2\pi$ rad s^{-1} and the spectrometer frequency is set to 600 MHz. In the different spectra k_{ex} is varied as indicated. In both cases $R_2^A = R_2^B = 10\ s^{-1}$.

Finally, fast exchange occurs where $k_{ex} \gg \Delta\omega$. For fast exchange, the nuclear spins sample the different chemical environments rapidly and therefore a single peak at the population weighted resonance frequency is observed in the spectrum. As the exchange rate continues to increase the linewidth of this peak decreases, converging on the population weighted intrinsic transverse relaxation rate as exchange tends towards the infinitely fast limit.

1.7.1 Relaxation Dispersion

Relaxation dispersion refers to a class of NMR experiments tuned to study $\mu s - ms$ timescale exchange events(*28*). Both CPMG(*29*, *30*) (Carr-Purcell-Meiboom-Gill) and $R_{1\rho}$(*31*) relaxation dispersion experiments are employed in this thesis and

will be explained briefly in this section. The related technique CEST (chemical exchange saturation transfer) forms the basis of the results in Chapters 3 and 4 and is described in detail in Chapter 3.

The relaxation dispersion experiments work on the principle that the effective transverse relaxation due to exchange ($R_{2,eff}$) experienced by a system in exchange can be modulated through pulses applied with a similar frequency to the exchange event itself. By analysing and fitting this dependence it is possible to gain information on the kinetics, thermodynamics and even structural changes underlying the exchange process. Such approaches have proven extremely fruitful when applied to biological systems as they provide a means of probing transiently populated but often functionally important conformational states of biomolecules that are often 'invisible' to other biophysical methods(*32*, *33*).

A schematic illustration of the CPMG pulse sequence is shown in Fig 1.7**A**. In these experiments, referred to as the constant time CPMG, the time T_{relax} is held constant while the number of 180° pulses applied during this block is varied. During the T_{relax} period nuclear spins undergo chemical exchange, that leads to incomplete refocussing of magnetisation and gives an exchange contribution to relaxation. By increasing the number of 180° pulses applied during the relaxation period the exchange contribution to relaxation is gradually quenched. The pulse sequence for the CPMG block is shown in Fig 1.7**B**. A key part of this is the Palmer element in the centre(*34*) that ensures that the magnetisation spends the same amount of time in-phase and anti-phase for all pulsing frequencies so that the $R_{2,eff}$ values are not additionally modulated by varying scalar coupling effects.

The pulse block used to perform off-resonance $R_{1\rho}$ experiments is shown in Fig 1.7**B**(*35*). The $R_{1\rho}$ experiment can be thought as the high pulsing limit of the CPMG where the pulses are applied back to back. An advantage of the experiment is that it is possible to reach much higher effective fields in this experiment as the magnetisation rotates in a tilted field with an additional effect from the residual B_0 field. This approach is therefore generally applied to studying faster exchange events than CPMG experiments are typically sensitive to.

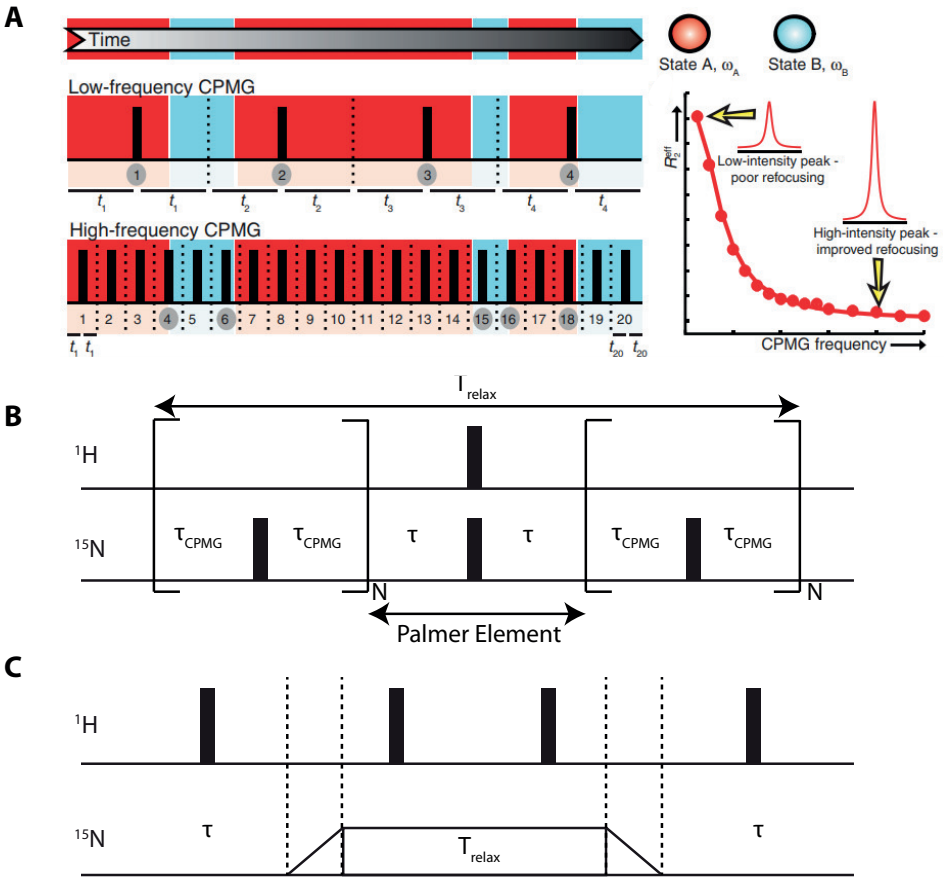

Figure 1.7: **A** Schematic illustration of the CPMG experiment: the relaxation time T_{relax} is kept constant while the number of refocussing pulses is varied. This affects the relaxation contribution to exchange. Figure adapted from (*32*). **B** Pulse sequence diagram of the CPMG block. All black rectangles represent 180° pulses. **C** Pulse block used to perform off-resonance $R_{1\rho}$ experiments: the triangular elements indicate adiabatic sweeps used to align the magnetisation with the effective field and the open rectangle represents the spin lock.

1.8 Diffusion

An important physical property that is central to many of the experiments in part I of this thesis is translational diffusion (herein referred to as diffusion). Diffusion results from the random Brownian motion of particles and is characterised by a diffusion coefficient, D. Mathematically, the diffusion coefficient is defined as:

$$D = \frac{\langle r^2 \rangle}{6t} \tag{1.28}$$

where $\langle r^2 \rangle$ refers to the mean squared displacement of the particle in time t. The diffusion coefficient is an intrinsic property of both the size and shape of a molecule

and its environment. The diffusion coefficient of a spherical particle of radius r is given by the Stokes-Einstein equation as:

$$D = \frac{kT}{6\pi\eta r} \tag{1.29}$$

where η is the viscosity of the solution. Larger diffusion coefficients are therefore associated with smaller molecules, in low viscosity solutions at high temperatures.

NMR experiments have become a powerful tool for determining the diffusion coefficients of molecules in solution(*36*). The ability to perform these experiments is reliant on pulsed field gradients(*37*). During the application of a gradient pulse a linearly varying magnetic field is applied to the sample: modern NMR probes allow for the application of gradients in the x, y and z dimensions, however, for this thesis, we limit our attention to z-axis gradients (the most common variation). Following the application of z gradient, nuclear spins gain a position dependent phase, $\phi(\mathbf{z})$, given by:

$$\phi(\mathbf{z}) = \gamma \mathbf{z} g \delta \tag{1.30}$$

in which $\mathbf{z}$ indicates the position of nuclear spin along the z-axis, g is the strength of the magnetic field gradient (usually in units of G cm^{-1}) and δ is the duration of the gradient.

The principles of NMR diffusion experiments are illustrated in Fig 1.8**A**. The initial gradient pulse encodes a position dependent phase on spins in the sample that can be thought of as imposing a phase helix. Following this there is a period (Δ) during which the spins freely diffuse. After this period a decoding gradient is applied: for spins that have remained in the same z position the effect of the encode and decode gradients cancel out exactly and these spins contribute to the observed intensity in the spectrum. However, when the spins change their position following the initial encode gradient, there is a phase mismatch on application of the decode gradient resulting in signal attenuation in the final spectrum.

The first diffusion NMR experiments were performed using pulsed field gradient spin echo methods(*37*, *38*), as shown in Fig 1.8**B**. The evolution of magnetisation in

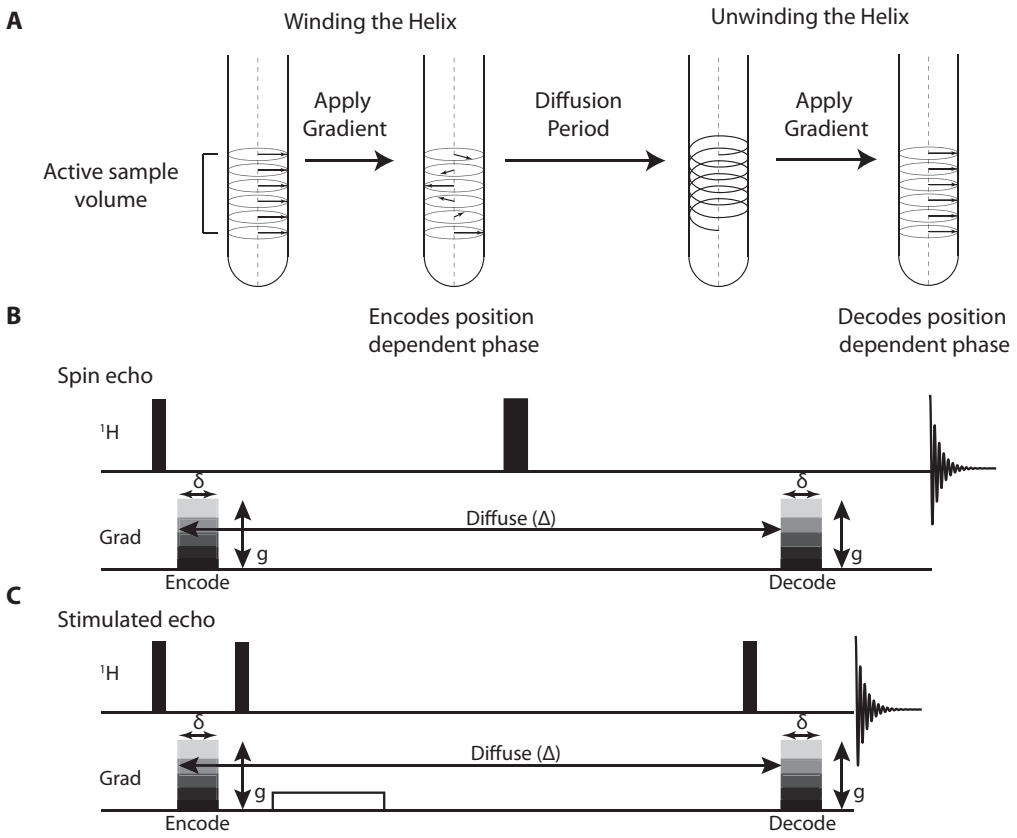

Figure 1.8: **A** Schematic illustration of the principles of an NMR diffusion experiment. The experiments consist of an initial encoding period that can be thought of as imposing a 'phase helix' on the sample followed by a diffusion period and the unwinding of the phase helix. **B** Pulsed field gradient spin echo (PFGSE) experiment for measuring diffusion coefficients. Narrow and wide bars represent 90° and 180° pulses respectively. **C** Related pulsed field gradient stimulated echo (PFGSTE) experiment for measuring diffusion coefficients. In this case the magnetisation is stored longitudinally during the diffusion period. All ^{1}H pulses are 90° and the open rectangle represents a purge gradient.

this sequence can be followed using the product operator formalism (assuming the pulses are applied on resonance so that there is no evolution due to chemical shift) as follows:

$$\begin{aligned} I_z &\xrightarrow{(\frac{\pi}{2}_x)} -I_y \xrightarrow{\gamma g \mathbf{z} \delta} -I_y \cos(\gamma g \delta \mathbf{z}_1) + I_x \sin(\gamma g \delta \mathbf{z}_1) \\ &\xrightarrow{\pi_y} -I_y \cos(\gamma g \delta \mathbf{z}_1) - I_x \sin(\gamma g \delta \mathbf{z}_1) \\ &\xrightarrow{\gamma g \mathbf{z} \delta} -I_y \cos(\gamma g \delta (\mathbf{z}_2 - \mathbf{z}_1)) + I_x \sin(\gamma g \delta (\mathbf{z}_2 - \mathbf{z}_1)) \end{aligned} \tag{1.31}$$

where we have used the trigonometric relations: $\cos(A - B) = \cos A \cos B + \sin A \sin B$ and $\sin(A - B) = \sin A \cos B - \cos A \sin B$. In the above equation $\mathbf{z}_1$ and $\mathbf{z}_2$ are the z positions of spins during the encode and decode gradients

respectively. The final signal is therefore given by:

$$S = \int_{-\infty}^{\infty} P(z, \Delta) e^{ig\gamma\delta z} \, dz \tag{1.32}$$

where $z = \mathbf{z}_2 - \mathbf{z}_1$ and $P(z, \Delta)$ is the probability of a spin undergoing a displacement of $\mathbf{z}$ in time Δ. In the case of isotropic diffusion, $P(z, \Delta)$ is a Gaussian distribution given by(*39*):

$$P(z, \Delta) = e^{-\frac{z^2}{4D\Delta}} \tag{1.33}$$

Upon inserting this into equation 1.32 we almost arrive at the Stejskal-Tanner equation, though this equation neglects the role of diffusion during the application of the gradient(*37*):

$$S = e^{-(\gamma\delta g)^2 D\Delta} \tag{1.34}$$

The form of the displacement probability distribution, $P(z, \Delta)$, for more complex situations is considered in detail in Chapter 2.

Performing the same analysis for pulsed field gradient stimulated echo (PFGSTE) experiment is very similar to its spin echo (PFGSE) counterpart except that the central 180° pulse in the latter experiment is replaced by two 90° pulses(*40*). The result of this is that during the diffusion delay (Δ) the magnetisation is stored longitudinally in the PFGSTE experiment. Since the longitudinal relaxation is less than or equal to the transverse relaxation rate, the PFGSTE approach provides access to longer diffusion delays without the signal suffering excessive attenuation due to relaxation. However, the stimulated echo experiment is not always superior as its maximum intensity is only half that of the spin echo (only half the magnetisation is rotated into the longitudinal axis while the rest is dephased and does not contribute to the final signal).

1.8.1 Derivation of the Stejskal-Tanner Equation

A more detailed analysis of the effect of gradients is required to derive the full Stejskal-Tanner equation(*37*). Ignoring the effects of relaxation, in the presence of a z-axis gradient the evolution of transverse magnetisation, I_+, is given by:

$$\frac{\partial I_+(z)}{\partial t} = -i\gamma B(z) I_+ + D\frac{\partial^2}{\partial z^2} I_+(z) \tag{1.35}$$

where the first term accounts for the effect of the gradients with $B(z)$ the additional magnetic field due to the presence of the gradient. The second term reflects diffusion according to Fick's second law. The following trial equation can be used to solve this equation:

$$I_+(t) = S(t) e^{(-i\gamma \int_0^t B(z,t')\ dt')} \tag{1.36}$$

Where the S(t) term, to be determined, reflects attenuation of magnetisation due to the presence of the gradients. Inserting equation 1.36 into 1.35, simplifying and discarding terms equal to zero we arrive at:

$$\frac{\partial S(t)}{\partial t} = -D\gamma^2 \Big(\int_0^t \frac{\partial}{\partial z} B(z,t')\ dt' \Big)^2 S(t) \tag{1.37}$$

We now note that the gradient causes the the effective magnetic field to vary linearly along the z-axis so that $\frac{\partial}{\partial z} B(z,t) = g(t)$ and g is the gradient strength. Since the gradient is uniform across the sample it is independent of the z position. The solution to equation 1.37 is then given by(*41*):

$$S(t) = e^{-D\gamma^2 \int_0^t \left(\int_0^{t'} g(t)\ dt'' \right)^2\ dt'} \tag{1.38}$$

In order to solve this and derive the Stejskal-Tanner equation it is necessary to split equation 1.38 into three time blocks: application of the first gradient ($t = 0 \rightarrow \delta$), the diffusion delay ($t = \delta \rightarrow \Delta$) and the final gradient ($t = \Delta \rightarrow \Delta + \delta$). Applying this gives(*41*, *42*):

$$\begin{aligned} \int_0^t \Big(\int_0^{t'} g(t)\ dt'' \Big)^2\ dt' = & \int_0^\delta \Big(gt' \Big)^2\ dt' + \int_\delta^\Delta \Big(g\delta \Big)^2\ dt' \\ & + \int_\Delta^{\Delta+\delta} \Big(g\delta - g(t'-\Delta) \Big)^2\ dt' \end{aligned} \tag{1.39}$$

The negative sign in the final term appears because the sign of the final gradient is effectively reversed by the 180° pulse in the PFGSE experiment and two 90° pulses in the PFGSTE experiment. Solving this integral and inserting the result into 1.38 we arrive at the Stejskal-Tanner equation(*37*):

$$S = S_0 e^{-\gamma^2 \delta^2 g^2 (\Delta - \frac{\delta}{3}) D} \tag{1.40}$$

In a normal diffusion experiment a series of spectra are recorded where the gradient strength, g, is varied. By fitting the attenuation of signal to the Stejskal-Tanner equation it is possible to extract the diffusion coefficient for spins at that chemical shift. In Chapter 2 we demonstrate how in the context of cellular systems the diffusion experiment can be used to provide information on intra- and extracellular diffusion coefficients as well as cell radii, permeability and abundance.

1.9 In-Cell NMR

In chapters 2 and 4 of this thesis we carry out a number of NMR experiments on cellular systems. Many early NMR in-cell NMR experiments were performed on erythrocytes due to their relative simplicity and availability(*43*, *44*). Most of these studies focussed on monitoring ^{1}H resonances to investigate aspects of metabolism. One of the disadvantages of using ^{1}H as a probe of these systems is that the lack of spectral dispersion in this dimension can make it difficult to delineate signals from individual species. To overcome this the early pioneers of in-cell NMR experiments also performed ^{13}C and ^{31}P detection based experiments(*45*, *46*). In addition to experiments on cellular systems, a large number of NMR experiments have also been performed on tissue. A landmark study by Damadian found that normal tissue could be differentiated from malignant tumours on the basis of their longitudinal and transverse relaxation rates(*47*), establishing the importance of these measurements for future MRI experiments.

Recently, there has been a significant increase in the in-cell NMR studies. This has been mostly precipitated by the ability to record high-resolution multidimensional spectra of isotopically labelled biomolecules residing in cells: in *E. coli* the

protein of interest can be over-expressed and monitored exclusively over other cellular components(*48*) and methods have also been developed to deliver recombinantly produced protein into mammalian cells by electroporation(*49*).

Within this thesis, our approach takes a greater cue from earlier in-cell NMR experiments. The experimental approaches we develop in chapters 2 and 4 are based purely on endogenous signals, and in particular the water signal is used as the spectroscopic readout.

1.10 Thesis Outline

This thesis is split into two parts: the first part is concerned with developing NMR methods, particularly based on diffusion and chemical exchange saturation transfer, for studying features of biological systems using experiments that do not rely on any isotopic labelling. In the second part, a range of NMR experiments, including relaxation dispersion methods, are applied to two different ^{15}N labelled proteins to characterise their conformational dynamics with atomic-level resolution.

Chapter 2 describes the development of an NMR experimental and analysis protocol for the characterisation of cellular systems. This is achieved by recording variable diffusion delay data that is then globally analysed. The analysis approach we use is tested with extensive Monte Carlo simulations. The analysis provides information on intra- and extra-cellular diffusion coefficients as well as cell radii, permeability and abundance.

In Chapter 3 we introduce the water chemistry exchange saturation transfer experiment. We demonstrate through both simulations and experiments how the CEST signal responds to a range of different environmental factors, information that can be used to inform observations seen in MRI studies. We also demonstrate the ability of the CEST experiment to report on protein aggregation.

Chapter 4 combines the main features of the previous two chapters. By coupling the diffusion and CEST experiments we present the novel diffusion-weighted CEST experiment. This experiment is able to give CEST profiles associated with specific compartments, where compartments are differentiated on the basis of their diffusion

properties. We demonstrate that the experiment works by testing it on model systems with similar dimensions to mammalian cells. The experiment is also shown to be sensitive to cell stress.

In chapters 5 and 6, that form part II of the thesis, there is a switch in focus to using NMR experiment to characterise conformational dynamics in isotopically labelled proteins. Chapter 5 is motivated by a molecular dynamics simulation study of the protein CypA. The simulations predict that mutation of the residue Asp66 will significantly perturb the structure and dynamics of the protein's loop regions around residues 70 and 100. In order to test this hypothesis extensive NMR experiments are carried out on recombinantly produced samples of ^{15}N-labelled WT and D66A CypA. The experiments performed include fast dynamics and relaxation dispersion and a strategy for combining CPMG and $R_{1\rho}$ datasets for characterising millisecond timescale dynamics.

Finally, in Chapter 6 the conformational dynamics of the toxic effector protein YopH is studied. Prior to being translocated into target eukaryotic cells, YopH first binds with its cognate chaperone. We test whether an excited state conformation plays an important role in binding to the chaperone. To study this we perform CPMG experiments on WT YopH as well as a mutant form that shows enhanced binding by ITC measurements.

Part I

Development of Label-free NMR experiments for the Characterisation of Biological Systems

2

Diffusion in Cellular Systems

Abstract

In this chapter, a method and analysis tools for the NMR characterisation of cellular systems via their diffusional properties is presented. The method consists of acquiring variable diffusion time pulsed field gradient stimulated echo (PFGSTE) data and then analysing it globally using a two pool model where the pools correspond to intra and extra-cellular spaces. The model is based upon the Kärger equations but these are substantially modified to account for the effects of (a) differential longitudinal relaxation within the two pools and (b) the effects of restriction within the intra-cellular pool. Extensive Monte Carlo simulations validate the method and demonstrate that it reliably reports on physically meaningful parameters. The errors in fitted parameters are shown to be particularly favourable in regions of parameter space expected for mammalian cells and the analysis is also relatively insensitive to the underlying geometry of the cells, increasing the generality of the method. The experimental approach is applied to the analysis of water within suspensions of yeast and mammalian cells and can report on cell abundance, permeability and radius.

2.1 Introduction

A recent development in NMR is the ability to obtain high quality spectra on systems of living cells allowing for the study of molecules in truly physiological environments at atomic-level resolution(*49–51*). Such experiments have been made possible by continuously evolving hardware that facilitate the detection of species at ever lower concentrations(*49*, *52*) as well as sample preparation methods that allow isotopically labelled molecules to be delivered into cells and be specifically detected amongst otherwise crowded spectra(*53*, *54*). In-cell NMR studies have already yielded significant information on protein folding in cells(*49*), the role of post-translational modifications(*49*) and information on metabolism(*51*). The analysis of diffusional properties of molecules in cell suspensions has the potential to provide significant additional insight into these systems including information on relative viscosities, cell permeabilities and cell sizes and can also potentially act as a non-invasive check that the cell sample retains its integrity during the experiment time.

The analysis of diffusion data within cellular systems is not straightforward however, and requires sophisticated models to extract physically meaningful parameters from the data. Owing to its ubiquity, high concentration and low relaxation rates (giving high sensitivity), water is an excellent probe of cellular systems and is considered exclusively in this chapter, though experiments could in principle be extended to any molecule that is well resolved in the spectrum.

Fig 2.1**A** shows a comparison of the simulated diffusion decay of fast and slow (freely) diffusing populations when they decay independently, and also their combined signal decay when there is no exchange between the populations. The biexponential function used to fit the data is:

$$S = p_a e^{-\boldsymbol{q^2}(\Delta - \frac{\delta}{3})D_a} + p_b e^{-\boldsymbol{q^2}(\Delta - \frac{\delta}{3})D_b} \tag{2.1}$$

Where $(\Delta - \frac{\delta}{3})$ is the effective diffusion time, with Δ the total diffusion time and δ the length of the diffusion gradient pulses, and p_a and p_b are intensity factors associated with the two water pools. The composite parameter $q^2 = (\gamma \delta g)^2$

describes the strength of the gradient pulses where γ is the gyromagnetic ratio and g is the strength of the gradient pulse. At first glance such a function appears a very convenient way to rationalise the behaviour of water in cellular systems(*55*). We have two water pools: an extra-cellular and intra-cellular water population each of which are characterised by a particular diffusion coefficients (D_{out} and D_{in} respectively). We can therefore fit the signal to a biexponential function to yield information on the relative populations of water in the two pools as well as the diffusion coefficient in each pool.

The inadequacy with this analysis is revealed when the analysis is used to attempt to globally analyse data with different diffusion times as in Fig 2.1**B**, which shows actual NMR diffusion data collected on mammalian 3T3 cells. While the cell sample shows evidence of an 'immortal' water population that is not significantly attenuated even at high b-values ($b = q^2\Delta'$ with $\Delta' = (\Delta - \frac{\delta}{3})$), it clearly fails to fully account for the data. A more sophisticated model is therefore required to fit the data and get more physically meaningful insight into the system, as is presented in this chapter (Fig 2.1**C**).

In general, restriction within cavities and exchange between compartments can significantly affect the diffusion attenuation signal of cellular systems. The impact of these effects has been the focus of several studies, particularly in the diffusion MRI literature(*56–58*), owing to the utility of diffusion MRI in clinical diagnosis(*59*). The heterogeneity of living tissue has resulted in a number of models being proposed including two or three pools (where a pool refers to a distinct water population), different aspects of exchange and confinement and with variable experimental protocols to study specific systems including: perfused glial cells(*60*), erythrocytes(*44*, *61*, *62*), rat brain tissue(*63*), blood(*64*), bovine optic nerve(*65*) and cancer tumour tissue(*66*).

In magnetic resonance experiments of cell-suspensions the degree of heterogeneity is significantly lower. Having a single experimental protocol and model to analyse the data that is able to return useful and physically accurate values whilst remaining relatively insensitive to the particular details of the cellular system

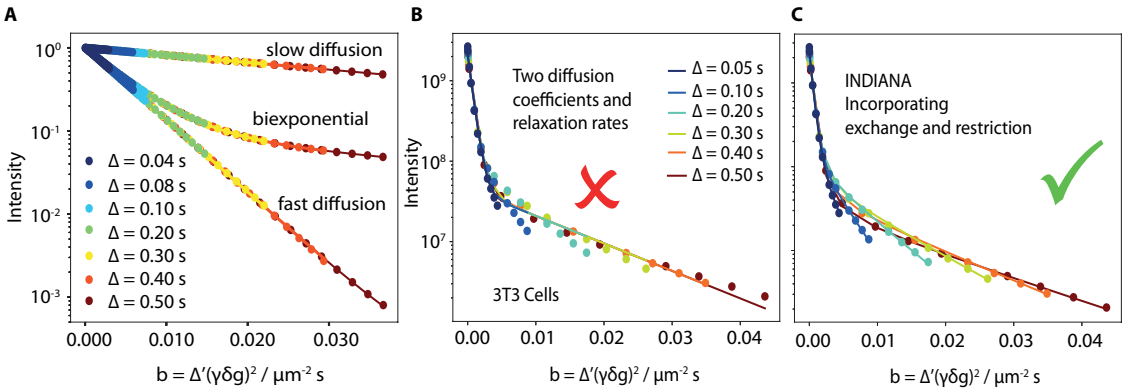

Figure 2.1: **A** Simulated signal intensities from diffusion experiments where populations are freely diffusing. For slow/fast diffusion, diffusion coefficients are set to values of 200/2000 $\mu m^2 s^{-1}$ respectively. In the case of biexponential diffusion slow and fast water components are in the population ratio 1:9 and no exchange occurs. Here the data is fitted to a simple biexponential model (equation 2.1). Details on how simulations are performed are provided in section 2.3. **B** A simple biexponential model such as equation 2.1 is unable to adequately fit real variable time data from a system of suspended 3T3 cells, even when the effects of longitudinal relaxation are accounted for. **C** The INDIANA model presented here is able to fit this data successfully by introducing two new model parameters to account for the effects of exchange and restriction. In this chapter, we show that the method works well for a range of cell types and returns physically meaningful parameters.

under investigation would be of significant utility. The ease associated with a unified approach is likely to make the application of these experiments much more widespread. For example, while as highlighted above, considerable effort has been placed into developing models to rationalise the diffusion behaviour in specific systems, a two-point measurement to yield an apparent diffusion coefficient remains the most commonly used tool in clinical diffusion MRI(*67*).

In order to address this challenge, in this chapter we present an experimental protocol and software package: INDIANA (In-cell DIffusion ANAlysis). The experimental procedure consists of collecting stimulated echo diffusion data where both the gradient strength, g, and diffusion delay, Δ, are varied (Fig 2.2**A**). The data is then analysed using a two pool model including exchange, based on the Kärger equations(*68–70*) but extended to include the differential relaxation in the two compartments, and a restricted intra-cellular environment based on theory developed by Neuman(*71*). The model is able to return intra- and extra-cellular diffusion coefficients, intra- and extra-cellular relaxation rates, cell permeability, cell

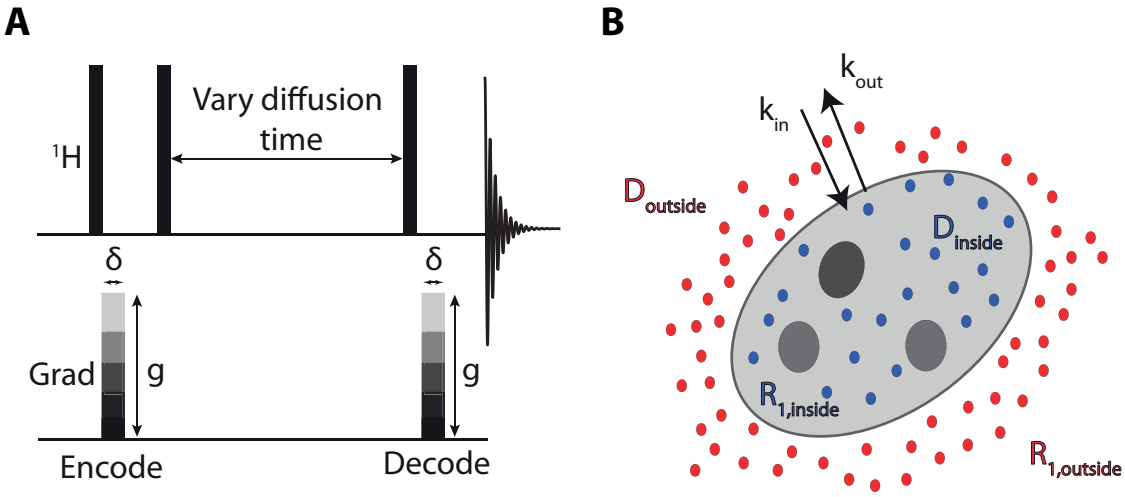

Figure 2.2: **A** Stimulated echo pulse sequence used to record diffusion data: in this analysis data is recorded varying both the diffusion time, Δ, and gradient field strength, g. **B** Schematic of a cell showing the different parameters that are included in the INDIANA analysis.

radius and cell abundance as fitted parameters when used to analyse the diffusion attenuation of the water signal (Fig 2.2**B**).

To validate the approach, extensive Monte Carlo simulations are carried out to create synthetic datasets that are fit to the model where the true values of the parameters are already known. These reveal that the model is generally effective at fitting physically accurate parameters, particularly in regions of parameter space expected for mammalian cells, but also shows where systematic errors in specific parameters are likely to occur. Importantly, the accuracy of fitted parameters is also shown to be relatively insensitive to the specific geometry of the cell under study.

Applications of the method to budding yeast (*Saccharomyces cerevisiae*), 3T3 and HeLa mammalian cells are presented and extracted parameters from the analysis are independently verified. Experimental data can be acquired in as little as five minutes and the experimental parameters required for the analysis should be easily accessible on any modern NMR spectrometer.

This chapter begins with a discussion of the theory required to characterise diffusion in exchanging and confined systems. Subsequently, the Monte Carlo simulation methods are described and the results of the Monte Carlo tests are presented before the experimental applications of the INDIANA method.

2.2 Theory

2.2.1 Kärger Equations

The formalism for describing the effect of exchange between environments on the diffusion attenuation was first described by Kärger(*68–70*). Here we derive the Kärger results for exchange between freely diffusing populations, though presented here slightly differently to Kärger's original form. Since the INDIANA protocol involves variations in diffusion time as well as gradient strength, the signal attenuation due to relaxation also becomes an important factor and so the effects of differential relaxation between compartments are also explicitly included in this model.

Diffusive exchange shares a number of features with chemical exchange and can be thought of as having slow, intermediate and fast regimes as highlighted below.

If we consider the case where we have two populations, a and b, exchanging from A to B and B to A at rates k_{ab} and k_{ba} respectively, characterised by diffusion coefficients, D_a and D_b, signal attenuation S_a and S_b and longitudinal relaxation rates R_{1a} and R_{1b} we can write down the rate of attenuation using a 2×2 matrix as:

$$\frac{d}{dt}\begin{bmatrix} S_a \\ S_b \end{bmatrix} = \begin{bmatrix} -D_a q^2 - k_{ab} - R_{1a} & +k_{ba} \\ +k_{ab} & -D_b q^2 - k_{ba} - R_{1b} \end{bmatrix} \begin{bmatrix} S_a \\ S_b \end{bmatrix} \tag{2.2}$$

In this case we consider the longitudinal relaxation rates, R_1, rather than the transverse ones, R_2, as we want to consider the attenuation during a stimulated echo experiment. A completely analogous approach can however be used for spin echo experiments by using transverse relaxation rates. It is also important to note that the formalism used here assumes two freely diffusing populations in exchange. In the next section we show how the effects of restriction can be additionally included.

In order to solve equation 2.2 we can take the matrix exponential of the propagator matrix, here referred to as $\boldsymbol{\rho}$. In general, $\boldsymbol{\rho}$ is related to the diagonal matrix of its eigenvalues, $\boldsymbol{D}$, by the relation: $\boldsymbol{\rho} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^{-1}$ where the columns of the matrix $\boldsymbol{V}$ represent the eigenvectors of $\boldsymbol{\rho}$. Recasting $\boldsymbol{\rho}$ in this way is useful for calculating the matrix exponential since: $\exp(\boldsymbol{\rho}) = \boldsymbol{V}\exp(\boldsymbol{D})\boldsymbol{V}^{-1}$ and because $\boldsymbol{D}$

is a diagonal matrix its exponential can be found trivially by taking the exponential of each of its elements. Putting this together and adding an additional detection operation that combines the two signals as is the case when they have identical chemical shifts (as is the case in water) we get:

$$S(q,\Delta) = \begin{bmatrix} 1 & 1 \end{bmatrix} \left(\boldsymbol{V} e^{\boldsymbol{D}\Delta} \boldsymbol{V}^{-\mathbf{1}} \right) \begin{bmatrix} p_a \\ p_b \end{bmatrix} S(0,0) \tag{2.3}$$

The eigenvalues of $-\boldsymbol{\rho}$ are:

$$\lambda^{\pm} = \frac{1}{2}(\xi^{+} \pm \phi) \tag{2.4}$$

Expressed in terms of the following substitutions:

$$\xi^{\pm} = D^{\pm} q^2 + R_1^{\pm} + k_{ex}(p_b \pm p_a) \tag{2.5}$$

$$\phi = \sqrt{(\xi^{-})^2 + 4k_{ex}^2 p_a p_b} \tag{2.6}$$

$$D^{\pm} = D_a \pm D_b \tag{2.7}$$

$$R_1^{\pm} = R_{1a} \pm R_{1b} \tag{2.8}$$

$$k_{ex} = k_{ab} + k_{ba} \tag{2.9}$$

The overall exchange rate k_{ex} is related to the individual exchange rates k_{ab} and k_{ba} (themselves the reciprocals of the lifetimes of states A and B) are $k_{ab} = p_b k_{ex}$ and $k_{ba} = p_a k_{ex}$. The matrix of eigenvectors, $\boldsymbol{V}$, is given by:

$$\boldsymbol{V} = \begin{bmatrix} 1 & 1 \\ \frac{-2k_{ex}p_b}{\xi^{-}+\phi} & \frac{-2k_{ex}p_b}{\xi^{-}-\phi} \end{bmatrix} \tag{2.10}$$

And the matrix exponential, $\boldsymbol{V} e^{-\boldsymbol{D}\Delta} \boldsymbol{V}^{-\mathbf{1}}$ is calculated as:

$$\boldsymbol{V} e^{-\boldsymbol{D}\Delta} \boldsymbol{V}^{-\mathbf{1}} = \frac{1}{2\phi} \begin{bmatrix} -(\xi^{-}-\phi)e^{-\lambda^{-}\Delta} + (\xi^{-}+\phi)e^{-\lambda^{+}\Delta} & 2k_{ex}p_a(e^{-\lambda^{-}\Delta} - e^{-\lambda^{+}\Delta}) \\ 2k_{ex}p_b(e^{-\lambda^{-}\Delta} - e^{-\lambda^{+}\Delta}) & (\xi^{-}+\phi)(e^{-\lambda^{-}\Delta} - e^{-\lambda^{+}\Delta}) \end{bmatrix} \tag{2.11}$$

The combined signal of the two populations is given by:

$$\frac{S(q,\Delta)}{S(0,0)} = e^{-\lambda^{-}\Delta} + \left(\phi - k_{ex} + (p_a - p_b)(q^2 D^{-} + R_1^{-}) \right) \frac{e^{-\lambda^{+}\Delta} - e^{-\lambda^{-}\Delta}}{2\phi} \tag{2.12}$$

Where the signals from the two populations are spectroscopically resolvable their individual signal attenuations are given by:

$$\frac{S_a(q,\Delta)}{S(0,0)} = p_a\left(e^{-\lambda^+\Delta} + (\phi + k_{ex} - q^2D^- - R_1^-)\frac{e^{-\lambda^-\Delta} - e^{-\lambda^+\Delta}}{2\phi}\right) \tag{2.13}$$

$$\frac{S_b(q,\Delta)}{S(0,0)} = p_b\left(e^{-\lambda^-\Delta} + (\phi + k_{ex} - q^2D^- - R_1^-)\frac{e^{-\lambda^+\Delta} - e^{-\lambda^-\Delta}}{2\phi}\right) \tag{2.14}$$

From equation 2.12 we can derive the expected limits for fast and slow diffusive exchange. Simulated spectra under different exchange regimes are also shown in Fig 2.3 (the signals are calculated from simulated trajectories as described in section 2.3).

To facilitate the calculation of the fast and slow exchange limits we can rewrite the expression for ϕ by expanding the $(\xi^-)^2$ term and simplifying to give:

$$\phi = \sqrt{(q^2D^- + R_1^-)^2 + 2(q^2D^- + R_1^-)^2k_{ex}(p_b - p_a) + k_{ex}^2} \tag{2.15}$$

In the limit of slow exchange $|q^2D^- + R_1^-| >> k_{ex}$ and in this limit equation 2.15 is well approximated by:

$$\phi_{slow} = (q^2D^- + R_1^-) + k_{ex}(p_b - p_a) \tag{2.16}$$

Inserting this result into the eigenvalues we get:

$$\lambda^+_{slow} = q^2D_a + R_{1a} + k_{ab} \tag{2.17}$$

$$\lambda^-_{slow} = q^2D_b + R_{1b} + k_{ba} \tag{2.18}$$

And the overall signal attenuation is given by:

$$\frac{S_{slow}(q,\Delta)}{S(0,0)} = p_ae^{-\lambda^+_{slow}\Delta} + p_be^{-\lambda^-slow\Delta} + \frac{2k_{ex}p_ap_b(e^{-\lambda^-_{slow}\Delta} - e^{-\lambda^+_{slow}\Delta})}{k_{ex}(p_b - p_a) + (q^2D^- + R_1^-)} \tag{2.19}$$

When $k_{ex} = 0$ we recover the expected population weighted biexponential decay of signal:

$$\frac{S_{slow}(q,\Delta)}{S(0,0)} = p_ae^{-(q^2D_a + R_{1a})\Delta} + p_be^{-(q^2D_b + R_{1b})\Delta} \tag{2.20}$$

By contrast, in the limit of fast exchange $k_{ex} >> |q^2D^- + R_1^-|$ and equation 2.15 can be approximated as:

$$\phi_{fast} = k_{ex} + (q^2 D^- + R_1^-)(p_b - p_a) \tag{2.21}$$

Inserting this into the eigenvalues we get:

$$\lambda^+_{fast} = k_{ex} \tag{2.22}$$

$$\lambda^-_{fast} = p_a(q^2 D_a + R_{1a}) + p_b(q^2 D_b + R_{1b}) = q^2 \bar{D} + \bar{R}_1 \tag{2.23}$$

Where $\bar{D}$ and $\bar{R}_1$ are the population weighted average diffusion coefficient and longitudinal relaxation rate respectively. The total signal decay is then given by:

$$\frac{S_{slow}(q,\Delta)}{S(0,0)} = e^{-\lambda^-_{fast}\Delta} = e^{-(q^2\bar{D}+\bar{R}_1)\Delta} \tag{2.24}$$

Thus, while in slow exchange the diffusive behaviour tends towards the values expected for the isolated species in fast exchange we can only observe average behaviour. The effects of different exchange regimes on variable time diffusion data is explored further in Fig 2.3.

When the exchange between the two environments is slow, the signal retains a strong biexponential character and the signal attenuations at different diffusion times almost overlay. In this case the signal can be well approximated by equation 2.1. When the exchange rate is increased to an intermediate regime however (2.3**B**), a clear bifurcation can be observed between the attenuation curves at different diffusion times. For a specific b-value we see that greater attenuation occurs at higher values of Δ: physically, this corresponds to the slower diffusing population having greater probability of moving into the faster diffusing population where its signal is attenuated more quickly. As the exchange rate continues to increase into the fast exchange regime the signal becomes monoexponential as it decays with the population weighted average diffusion coefficient in line with equation 2.24. The signals associated with different diffusion times also overlay as the lifetime of the individual states is much shorter than the diffusion times. In the fast exchange limit there is also significant correlation between the parameters (the two diffusion coefficients and population factor) meaning extracting precise parameters from this data through fitting procedures becomes intractable.

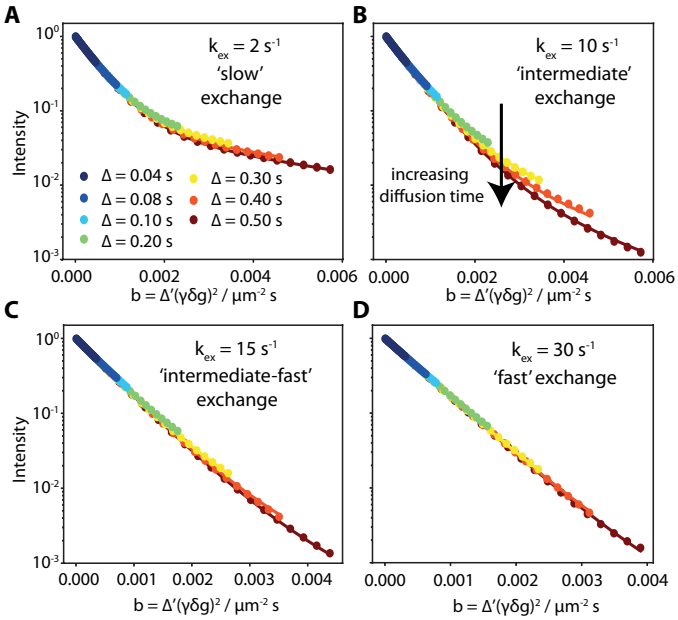

Figure 2.3: Simulated effects of exchange on diffusion attenuation. The slow and fast diffusion coefficients are set to 200 and 2000 μm^2 s^{-1} respectively and weighted in the ratio 1:9. In all cases the diffusion attenuation is plotted as a function of different Δ from 0.04 to 0.5 s, k_{ex} is varied between each sub-figure as stated and relaxation effects are neglected. The simulated data are fully accounted for by equation 2.12. In **A**, in the slow exchange limit the data have biexponential character and curves with different Δ almost overlay. In **B** when the exchange rate is increased to an 'intermediate' exchange regime, there is a maximum bifurication between attenuation curves at different diffusion times. Finally as exchange enters the 'fast' regime in **C-D** the attenuation curves increasingly overlay again and the signal decay becomes monoexponential.

2.2.2 Restricted Diffusion

The approach we use for analysing restricted diffusion in this chapter is the Gaussian Phase Approximation (GPA) developed originally by Neuman(*71*) and Murday and Cotts(*72*). This approach is based on the idea that when a particle undergoes diffusion in a restricted environment its probability displacement distribution remains Gaussian, analogous to free diffusion, but at long times reaches a limiting distribution that is dependent on the geometry of the restricting environment. Since this chapter is associated with diffusion analysis in cellular systems we focus on the analytical results for restricted diffusion in spherical cavities. Other analytical approaches for describing restricted diffusion exist, notably the short gradient pulse (SGP) approach (*39*, *73*), but based on previous studies(*74*) and herein we

show that the GPA approach is able to successfully describe restricted diffusion over experimental ranges of interest.

According to the GPA, the diffusion attenuation for particles within an impermeable spherical cavity of radius r is given by:

$$S(b) = S(0)e^{-bD^{app}} \tag{2.25}$$

Where the apparent diffusion coefficient, D^{app} is given by:

$$D^{app} = \frac{2\Phi(D, \delta, \Delta, r)}{D^2\delta^2\Delta} \tag{2.26}$$

D here refers to the actual diffusion coefficient inside the spherical cavity that is dependent on its internal viscosity, and D^{app} refers to the apparent diffusion coefficient including the effects of restriction. Φ is calculated as:

$$\Phi(D, \delta, \Delta, r) = \sum_{n=1}^{\infty} \frac{1}{\alpha_n^6(r^2\alpha_n^2 - 2)}\Big(2\alpha_n^2 D\delta - 2 + 2\exp(-\alpha_n^2 D\delta) + 2\exp(-\alpha_n^2 D\Delta) - \exp(-\alpha_n^2 D(\Delta - \delta)) - \exp(-\alpha_n^2 D(\Delta + \delta))\Big) \tag{2.27}$$

In this equation, α_n are the roots of the Bessel function equation:

$$(\alpha_n R)J'_{3/2}(\alpha_n R) - \frac{1}{2}J_{3/2}(\alpha_n R) = 0 \tag{2.28}$$

where J is a Bessel function of the first kind and J' is its first derivative. We have developed an efficient numerical procedure for obtaining the roots of equation 2.28. While equation 2.27 includes a sum over infinity, the signal is heavily dependent on the first root and the first twenty roots are sufficient for the sum to get an error of less than 0.1% in applications considered here.

We can quantify the degree of restriction experienced by a molecule in an impermeable sphere using the dimensionless parameter $\zeta = \frac{D\Delta}{r^2}$(*39*). Where $\zeta \ll 1$ the effects of restriction on the particle is small and we expect the diffusive behaviour to converge on the normal unrestricted result, as derived by Stejskal and Tanner(*37*). Conversely, when $\zeta \gg 1$ the restricting geometry becomes the dominant factor in the

probability displacement distribution of the particle and the diffusion attenuation is expected to become independent of the diffusion time and coefficient. The INDIANA approach seeks to gain insight from both these limits by recording diffusion attenuation at a number of times and analysing the data globally: short time attenuations should report mainly on the diffusion coefficient inside the cavity while long time attenuations should contain details on the restricting cavity. A visual representation of this idea is presented in Fig 2.4**A**.

In the limit that $\zeta \ll 1$ the roots of the Bessel equation α_n are small and we fulfil the condition that $\alpha_n^2 Dt \ll 1$. If we do a Taylor expansion of the exponential terms in equation 2.27 up to the third term we find:

$$\Phi(D, \delta, \Delta, r) = \sum_{n=1}^{\infty} \frac{D^3 \delta^2 \Delta}{r^2 \alpha_n^2 - 2} \tag{2.29}$$

We can then apply the Bessel sum rule: $\sum_{n=1}^{\infty} \frac{1}{r^2\alpha_n^2 - 2} = \frac{1}{2}$ where α_n are the roots of equation 2.28. Inserting this result into equation 2.26 gives:

$$D^{app} = D \tag{2.30}$$

And thus the Stejskal-Tanner result for free, unrestricted diffusion is recovered. By contrast, in the long time limit when $\Delta \to \infty$ the exponential terms in equation 2.27 involving Δ tend to zero and so we can rewrite this as:

$$\Phi(D, \delta, \Delta, r) = \sum_{n=1}^{\infty} \frac{1}{\alpha_n^6 (r^2 \alpha_n^2 - 2)} \Big(2\alpha_n^2 D\delta - 2 + 2\exp(-\alpha_n^2 D\delta) \Big) \tag{2.31}$$

Since δ is small, doing a Taylor expansion of the exponential term up to its quadratic term gives:

$$\Phi(D, \delta, \Delta, r) = \sum_{n=1}^{\infty} \frac{D^2 \delta^2}{\alpha_n^6 (r^2 \alpha_n^2 - 2)} \tag{2.32}$$

We can then apply another Bessel sum rule: $\sum_{n=1}^{\infty} = \frac{1}{\alpha_n^2 (r^2\alpha_n^2 - 2)} = \frac{r^2}{10}$. Inserting this result into equation 2.26 gives:

$$D^{app} = \frac{r^2}{5\Delta} \tag{2.33}$$

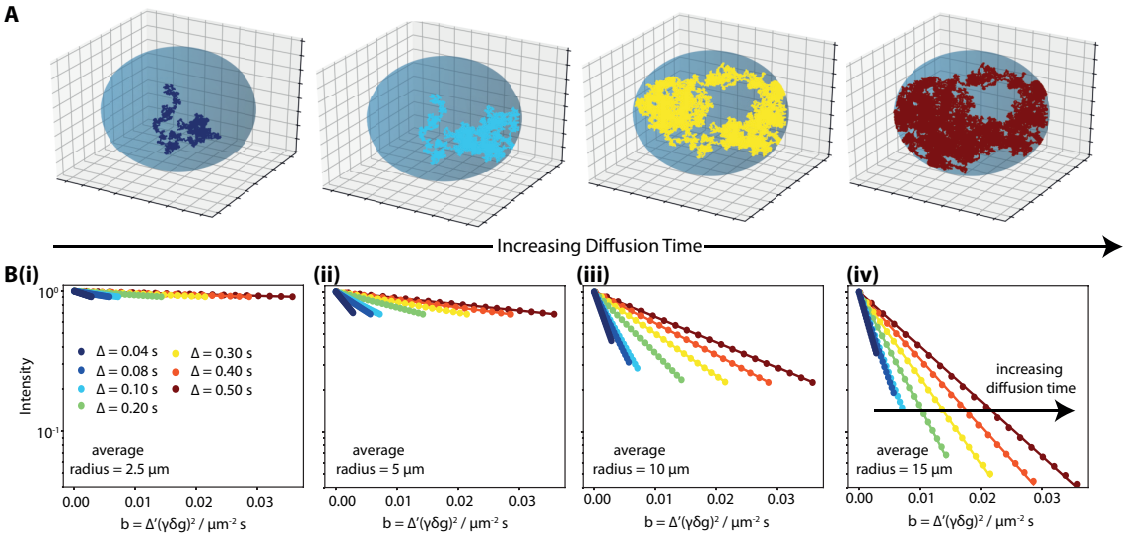

Figure 2.4: **A** Visualisations of single particle diffusion in an impermeable sphere of radius 10 μm and diffusion coefficient of 500 $\mu m^2\ s^{-1}$ at increasing diffusion times. **B** Simulated data showing the effect of diffusion within impermeable spheres of different sizes where the diffusion coefficient is set to 500 $\mu m^2\ s^{-1}$. In each case a distribution of spheres is imposed of stated average radius and standard deviation set to 20% of the average value, reflecting size distributions observed in cellular systems (considered further in section 2.3). Data are fitted to equation 2.25.

And if this in turn is inserted into equation 2.25 gives:

$$S = S(0)e^{-\frac{q^2r^2}{5}} \tag{2.34}$$

A result that was also derived by Stejskal and Tanner(*75*).

Fig 2.4**B** considers the effect of restriction on the diffusion attenuation signal. As the spherical cavities in these simulations are impermeable, as their size increases the diffusion attenuation also increases as the particles are able to diffuse further from their starting points. Similar to diffusive exchange, attenuation curves associated with different diffusion times do not overlay. However, unlike exchange, at a given b-value we see that the curve associated with the highest diffusion time shows the smallest attenuation. The reason for this is that at higher diffusion times the maximum displacement of the particle remains limited by the geometry of the restricting environment, while if instead the gradient pitch is reduced the signal is more sensitive to smaller displacements resulting in greater signal attenuation. Fig 2.4**B** also shows that the Neuman approach for restricted diffusion introduced here

is able to fit the data well, with the caveat that a distribution of particle sizes is imposed. The reasons for this are explored fully in the next section 2.3.

Overall, restriction and exchange can clearly have significant effects on the diffusion signal, though in different ways. In the context of cellular diffusion generally both of these effects will be important. In the INDIANA model, both effects are included by thinking of the system as being composed of intra- and extra-cellular pools in exchange, as described by the Kärger equations (2.12), and treating intra-cellular diffusion as restricted according to the Neuman formalism (equation 2.26). In the next section we use Monte Carlo simulations to demonstrate that this model is robust in a wide range of conditions.

2.3 Monte Carlo Simulations

Monte Carlo simulations are commonly employed in the diffusion NMR and MRI literature both to understand how different features of physical systems affect the diffusion signal(*58*, *76*) and validate analytical models for describing the diffusion signal(*56*, *57*, *77*). The popularity of Monte Carlo simulations in the field results from the simple relationship between the probability displacement distribution $P(z, \Delta)$ and the diffusion signal:

$$S(q, \Delta) = \int_{-\infty}^{\infty} P(z, \Delta) \cos(qz) dz \tag{2.35}$$

Therefore provided sufficient simulations can be carried out to generate the form of $P(z, \Delta)$, they can be used to investigate the diffusion signal in systems of arbitrary complexity. A significant disadvantage of the approach however is that many single trajectories, typically 10^4 to 10^6, are required to give reliable results. Thus while Monte Carlo simulations are ideal for the validation of analytical models, they remain too slow for directly fitting data to, despite recent advances made in this area(*78*).

In this section, Monte Carlo simulations are used to assess the robustness of the model presented in the previous section 2.2 across a wide range of parameters. We begin by first testing the validity of the simulations themselves and then using

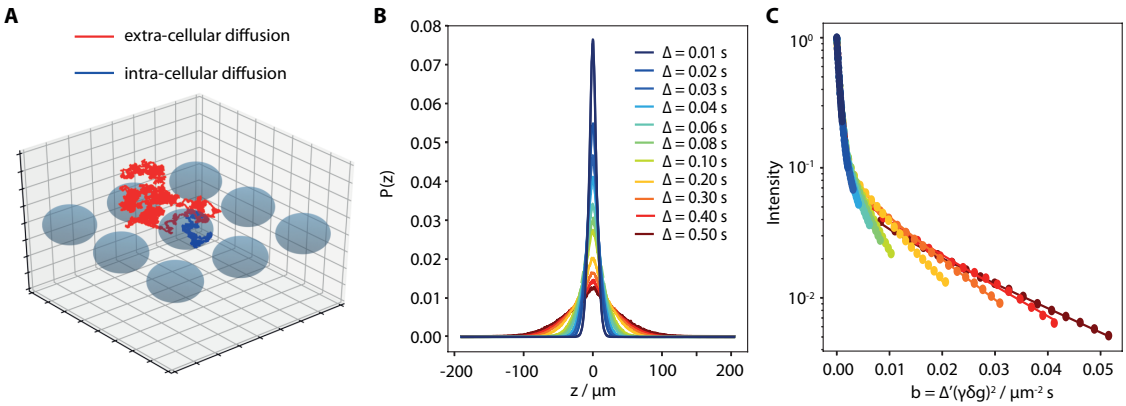

Figure 2.5: Schematic showing the calculation of diffusion signal from single particle trajectories. **A** Example of a single particle trajectory: in general these can include the effects of restriction and exchange with different diffusion coefficients. **B** From a large number single particle trajectories measured for different numbers of times the displacements are recorded to give $P(z, \Delta)$. **C** Using $P(z, \Delta)$ the diffusion attenuation can be calculated with equation 2.35 at given values of gradient strength (g) and time (δ). The data is then fit to an appropriate model and the fitted parameters are compared to the ground truth values.

them to check how well the GPA is able to characterise intra-cellular diffusion and when this approach breaks down. We then simulate a full two pool situation, including the effects of restriction, exchange and relaxation, and systematically evaluate the performance of the model. We aim to keep the simulations as close to experiment as possible and therefore specifically model the effects of cell size heterogeneity and gradient non-uniformity.

The details of how the Monte Carlo simulations are performed are set out in subsection 2.6.1 and a schematic of how the signal is simulated in general is provided in Fig 2.5.

2.3.1 Validation of Simulations

While Monte Carlo simulations have previously proved to be very useful tools in the analysis and validation of NMR diffusion models, before they are used they must be checked rigorously to ensure that they give reliable results. Here, we check that the simulations both provide an accurate distribution of displacements and that the signal attenuation is correctly calculated from the distributions. Fig 2.3 gives a simple example of this, showing that the simulated signal is fully accounted

for by the Kärger Equations. In this subsection we show that the simulations also converge on a more challenging limit, the long-time diffusion attenuation in restricted environments, and in doing so highlight an interesting phenomenon in restricted diffusion, namely diffusive-diffraction(*79*, *80*). The emergence of this effect also demonstrates a failing of the GPA approach.

If we consider restricted diffusion within an impermeable spherical cavity, in the long time limit the final position of particles becomes independent of their initial positions. The z displacement in spherical polar coordinates is given by: $dz = r(\cos(\theta_1) - \cos(\theta_2))$ and the probability of a given displacement will depend on the area: $(r\sin(\theta))^2$. The un-normalised displacement distribution will be:

$$P(z) = \int_0^\pi \int_0^\pi \sin^3\theta_1 \sin^3\theta_2 \delta((r(\cos\theta_1 - \cos\theta_2)) - z)\ d\theta_1\ d\theta_2 \tag{2.36}$$

where the δ represents the Dirac delta function. Changing the variables so that $u = \cos\theta_1$ and $v = \cos\theta_2$ gives:

$$P(z) = \int_{-1}^1 \int_{-1}^1 (1-u^2)(1-v^2)\delta((r(u-v) - z)\ du\ dv \tag{2.37}$$

Following a final change of variable $\kappa = \frac{|z|}{r}$ we arrive at:

$$P(\kappa) = \frac{3}{160}(2-\kappa)^3(4+\kappa(6+\kappa)) \tag{2.38}$$

This can then be converted into an expected NMR signal intensity:

$$S(q,\Delta) = \int_{-\infty}^{\infty} P(z,\Delta)\cos(qz)dz = \frac{9(qr\cos(qr) - sin(qr))^2}{(qr)^6} \tag{2.39}$$

This final result to identical to that previously reported by Tanner(*75*) and Callaghan(*80*) though is here derived in a different fashion and an expression for P(z) in the long time limit is provided.

Fig 2.6 demonstrates that the simulation results do converge on the expected theoretical long time limit results but also highlight a fundamental shortcoming of the GPA approach. With reference to Fig 2.6**A**, while at short times Gaussian fits

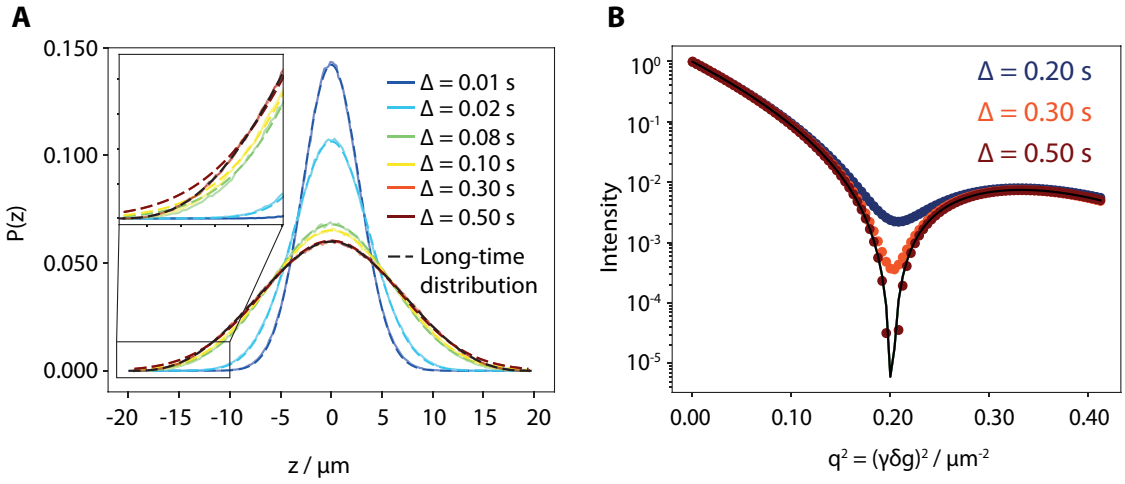

Figure 2.6: **A** Simulated displacement distribution functions (coloured solid lines) at different diffusion times with Gaussian fits (coloured dotted lines) and the expected long-time distribution (black dotted line) for a particle diffusing in an impermeable sphere of radius 10 μm with a diffusion coefficient of 500 $\mu m^2\ s^{-1}$. The inset highlights the discrepancy between the Gaussian fits and the simulated and theoretical results at the tails of the distribution. **B** Calculated signal attenuations from displacement distributions using equation 2.35. The signal attenuation is calculated with $\delta = 3\ ms$ and the 100 quadratically spaced gradient strengths between 4 and 80 $G\ cm^{-1}$. The black line indicates the expected long-time signal attenuation from equation 2.39.

clearly provide a good fit to the observed distributions, at long times there are clear systematic deviations between the Gaussian fits and the simulated displacement distributions that are however in excellent agreement with the theoretical long-time distribution. If we consider for example a displacement of 20 μm: in a sphere of radius 10 μm this can only be achieved by a particle starting on a point at the surface of the sphere and at the end of the diffusion period moving to the diametrically opposed point on the surface of the sphere. The probability of this occurring is vanishingly small, as borne out by the simulations and theoretical result. However, it is clearly overestimated by the Gaussian distribution.

The simulated distributions are able to recapitulate the predicted diffusive-diffraction effects for diffusion in restricted systems, as shown in Fig 2.6**B**. These effects have been observed previously in studies of diffusion in mono-disperse spherical polystyrene beads (*79*), however due to the heterogeneity typically present in systems are often difficult to experimentally observe (*81*). By contrast, the GPA predicts a monotonic decrease in intensity with increasing q^2. The expected

effects and range of validity of the GPA in the context of cellular systems are considered further in section 2.3.3.

2.3.2 Non-Uniform Gradients

We aim in this chapter to make our simulated data match experiment as closely as possible and to this end attempt to include effects commonly encountered in real experiments. One such effect that is commonly overlooked in simulation studies is non-uniform gradients. In a perfect system, the effective gradient field strength would be uniform throughout the active volume of the sample, but in practice all NMR probes suffer from degree of gradient non-uniformity and usually the relative gradient strength at the edges of the active volume is lower than in the centre. While, the effects of gradient non-uniformity are not included in the model presented here (2.2), by including them in the calculation of signal attenuation from simulations we can ensure that this effect does not unduly affect the results.

The effects of non-uniform gradients have been studied extensively by Morris and colleagues(*82*) and we use a gradient probability function based on their work, as shown in Fig 2.7**A**. Fig 2.7**B** illustrates the effect of non-uniform gradients on the appearance on diffusive-diffraction patterns: while minima are still observed in the attenuation profile, they are much less sharply defined and the attenuation more strongly resembles experimental observations of diffusive diffraction (*79*).

2.3.3 Validation of Theory

Having established the reliability of the simulations, in this section we seek to use them to test the validity of the GPA approximation, specifically in the context of cellular systems. The GPA approximation is only valid when $qr \ll 2\pi$. When $qr \geq 2\pi$ the pitch length of the gradients is smaller than the cavity size, meaning signal attenuation will no longer decay monotonically with q. Furthermore, from the previous section we know that the GPA fails to predict the long-time diffusion attenuation that shows diffusive-diffraction effects. By comparing simulated probability distribution functions and signal attenuations to the GPA approximation under

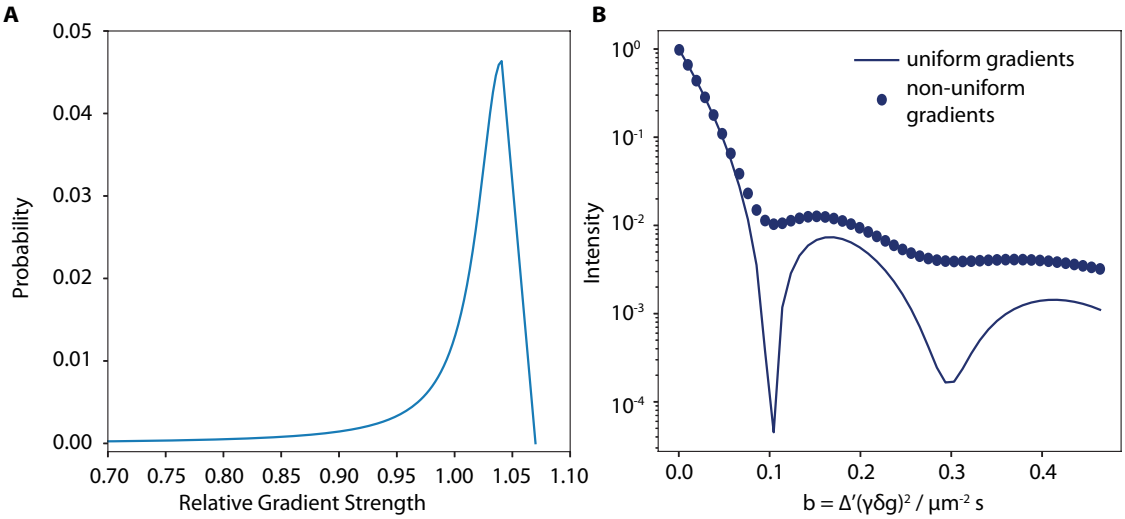

Figure 2.7: **A** Illustration of the range and distribution of gradient strengths used to account for gradient non-uniformity. The distribution is set so that the average relative gradient strength is 1.0. **B** Illustrates the effect of non-uniform gradients on diffusive-diffraction effects for diffusion in an impermeable sphere of radius 10 μm. The signal in the presence of non-uniform gradients is calculated using equation 2.47. For calculation of the diffusion attenuation 50 quadratically spaced gradient strengths between 4 and 120 $G\ cm^{-1}$ are used with $\delta = 3\ ms$ and $\Delta = 0.5\ s$.

a range of conditions that are analogous to cellular systems we can empirically test the validity of the model for such cases.

As mentioned above diffusive-diffraction effects are difficult to realise in heterogeneous systems such as biological cells. The reason for this is that the position of the diffraction dips is dependent on particle size and where there are a distribution of particle sizes these diffraction patterns can interfere and cancel each other out(*81*). For example, HeLa cells are expected to have a standard deviation in radius of $\pm 20\%$ in a typical population (*83*, *84*). Negative curvature effects associated with diffusive-diffraction are not observed in experiments we perform on cellular systems as shown in section 2.4.

Fig 2.8 shows the simulated effect of a distribution of particle sizes on displacement distributions. As shown in Fig 2.8**A** and above, Gaussian fits fail to fully account for the displacement distributions where the particle size is perfectly monodisperse, particularly at the tails of the distribution. Conversely, as shown in Fig 2.8**B** when a normal distribution of radii is included, a situation that is more representative of cell suspensions, the Gaussian fits are able to fit the data well over

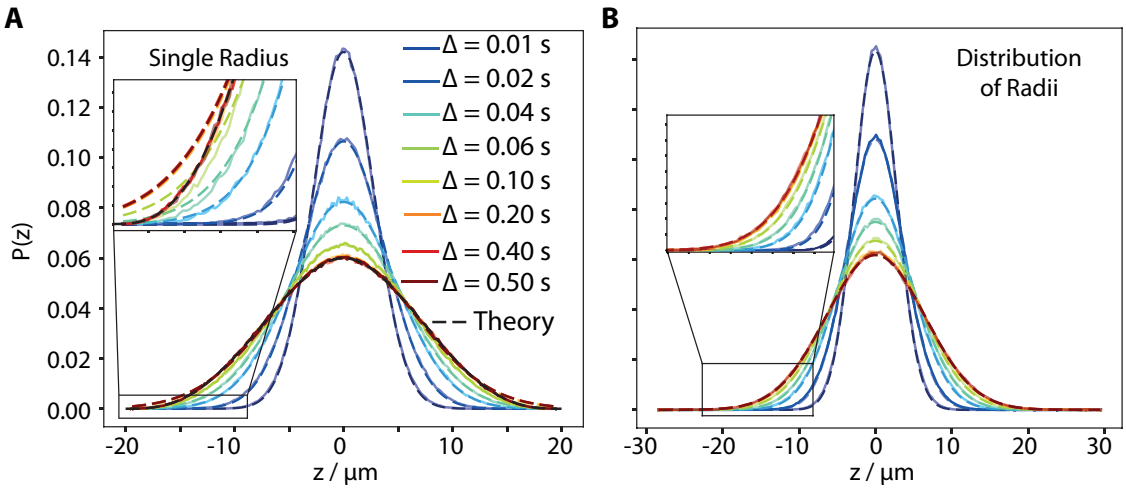

Figure 2.8: **A** Simulated probability displacement distributions for particles diffusing in an impermeable sphere of radius 10 μm with diffusion coefficient 500 $\mu m^2\ s^{-1}$ for stated diffusion times (solid coloured lines). Gaussian fits to the simulations are shown by the coloured dotted lines and the long time distribution (equation 2.38) is indicated by the black line. **B** Simulated probability displacement distributions under identical conditions to **A**, except that the spheres now have an average radii of 10 μm with a standard deviation of 2 μm to reflect size distributions observed in HeLa cells.

a wide range of times and show no evidence of systematic deviations. This opens up the possibility that the range of validity of the GPA is extended in situations where there is a distribution of particle sizes. While this approach is insensitive to the range of sizes present, it could still be used to fit an average radius with good accuracy.

This idea is tested further in Fig 2.9. Using the simulated displacement distributions we can calculate the NMR signal attenuation at high gradient strengths and use this to gain empirical insight into the range of validity of the GPA. From Fig 2.9**A** and **B**, the effects of diffusive-diffraction are still visible and the GPA is only able to account for the data up to $qr < 2$. The discrepancy between the GPA and simulated result is however reduced in the presence of non-uniform gradients. Conversely, in Fig 2.9**C** and **D** when there is a size distribution of spheres, any negative curvature associated with diffusive-diffraction effects are completely obscured(*81*, *85*) and the GPA appears to account for the data up to $qr = 4$. As the q value continues to increase however, the signal decay deviates from the expected GPA attenuation as we approach the $qr = 2\pi$ limit.

Clearly, the approach presented here overlooks much of the complex heterogeneity

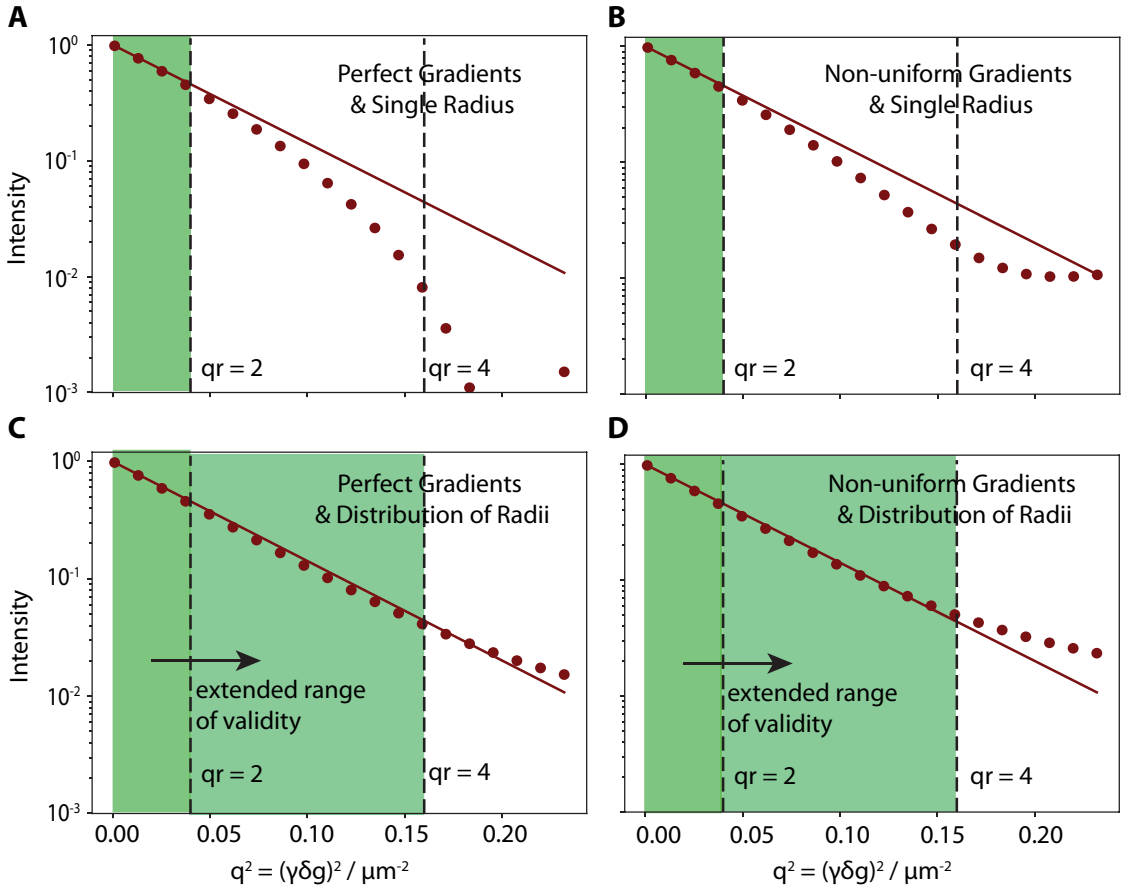

Figure 2.9: Assessment of the validity of the GPA approximation under different conditions. In all cases the NMR signal (dots) is calculated at $\Delta = 0.5\ s$ with $\delta = 3\ ms$ and the average gradient strength is varied from 4 to 80 $G\ cm^{-1}$. In **A** and **B** the signal is calculated assuming all particles diffuse within spheres of radius 10 μm. The solid line shows the predicted signal attenuation from GPA theory (2.25) with the radius fixed to 10 μm. In **C** and **D** the particles diffuse in spheres whose sizes are normally distributed with an average radius of 10 μm and a standard deviation of 2 μm. In **A** and **C** the NMR signal attenuation is calculated assuming uniform gradients, whereas in **B** and **D** the effects of non-uniform gradients (section 2.3.2) are included.

in the size and shape of actual cellular systems. However, given that we (section 2.4) and other studies(*86*) have not observed evidence of more complex signal attenuation in NMR studies of cellular systems, up to $qr \approx 4$ we believe that the approach here provides a good approximation for the underlying complexity in these systems. The extent of validity of the GPA will also be dependent on the range of the size distribution and so in the simulations below and NMR experiments the maximum gradient strength used is $qr \approx 3.3$.

Having established that when a distribution of restricting sizes are present

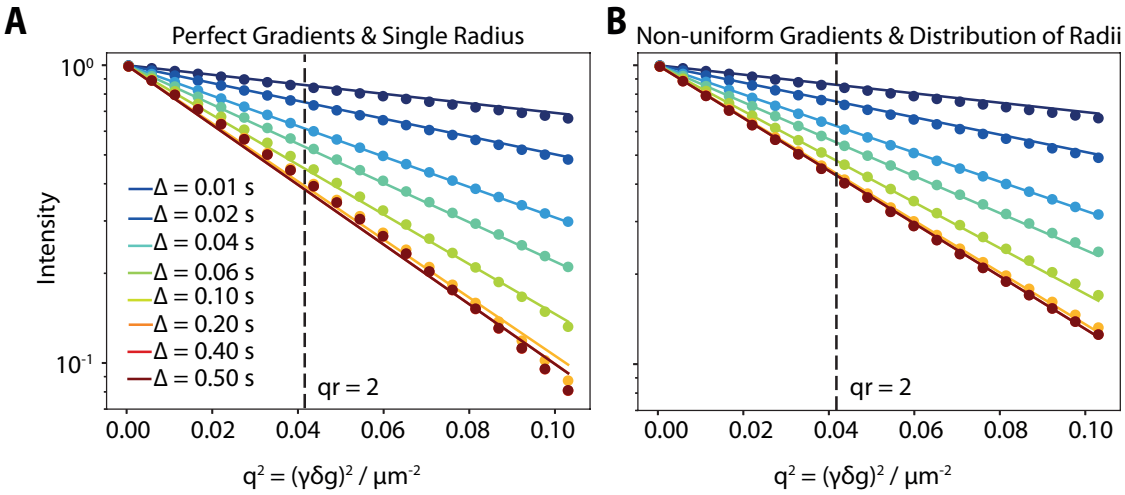

Figure 2.10: Assessing GPA fits to simulated data: in both cases simulated attenuation (dots) are calculated using typical experimental parameters. $\delta = 2\ ms$, Δ is varied from 0.01 to 0.5 s, g is varied from 4 to 60 $G\ cm^{-1}$ and the diffusion coefficient is set to 500 $\mu m^2\ s^{-1}$. The data is globally fitted to the GPA formula (equation 2.25, solid lines). **A** The particles are simulated in impermeable spheres that all have a radius of 10 μm and the signal is calculated assuming perfect gradients. The fitted diffusion coefficient is $483 \pm 7\ \mu m^2\ s^{-1}$ and radius is $10.82 \pm 0.03\ \mu m$. **B** In this case the particles diffuse in spheres with average radius 10 μm and standard deviation 2 μm and the signal is calculated assuming non-uniform gradients. The fitted diffusion coefficient is $480 \pm 10\ \mu m^2\ s^{-1}$ and radius is $10.15 \pm 0.05\ \mu m$. Error bars in parameters are determined by a bootstrapping analysis.

the range of validity of the GPA is extended, we now want to verify that this approach can reliably fit variable time data to give radii and diffusion coefficients where we have a single pool of particles diffusing in impermeable spheres. The results of this are shown in Fig 2.10. When the effects of size heterogeneity and gradient non-uniformity are included, the diffusion attenuation is monoexponential for all times over experimentally relevant gradient values. Moreover, the simulated data can be well fit to the GPA model with only two free parameters: diffusion coefficient and (average) radius. The fitted diffusion coefficients for the single radius vs distribution of radii are very similar, 483 and 480 $\mu m^2\ s^{-1}$ respectively compared to the actual value of 500 $\mu m^2\ s^{-1}$. However, the fitted (average) radii are: 10.85 and 10.15 μm respectively showing that the average radius is fit more accurately when there are a distribution of sizes.

2.3.4 Monte Carlo Simulation Results

We have shown that the GPA approach can be used to accurately parametrise restricted diffusion over experimental ranges of interest and now extend the Monte Carlo simulations to test the robustness of the full model presented in Section 2.2. In these simulations (explained fully in subsection 2.6.1) we aim to capture the main features of diffusion in cellular systems. In the simulations we have two pools corresponding to intra and extra-cellular particles. The intra-cellular particles are in a restricted geometry (usually a sphere but we also test cylindrical geometries) but this is now partially permeable so that exchange occurs between the two environments. In later simulations the effects of relaxation are also included. The simulations are characterised by five parameters: D^{out}, D^{in} (the extra and intra-cellular diffusion coefficients), k_{ex} (exchange parameter), p_{in} (the mole fraction of particles inside a cell) and r (the average radius of the restricting geometry). When the effects of relaxation are introduced, two additional parameters must be considered, R_1^{in} and R_1^{out}.

To quantitatively determine regions of parameter space where the model is able to reliably fit parameters, an extensive set of Monte Carlo simulations were performed where model parameters were varied systematically. The simulated displacement distributions were then used to calculate diffusion attenuations with typical experimental parameters for g, Δ and δ, remaining in the limit where $qr < 4$. The diffusion attenuations were then fit to the INDIANA model and the fitted values were compared to the ground truth values used in the simulations to evaluate the accuracy of the model in that region of parameter space.

In the simulations, the following baseline values were used for parameters when they are not being varied, chosen to reflect typical values expected for the water signal an NMR sample containing mammalian cells: $D^{in} = 500\ \mu m^2\ s^{-1}$, $D^{out} = 2000\ \mu m^2\ s^{-1}$, $k_{ex} = 3.55\ s^{-1}$, $r = 10\ \mu m$ (this is the average radius with the standard deviation set 2 μm) and $p_{in} = 0.15$. The results of this procedure are shown in Fig 2.11.

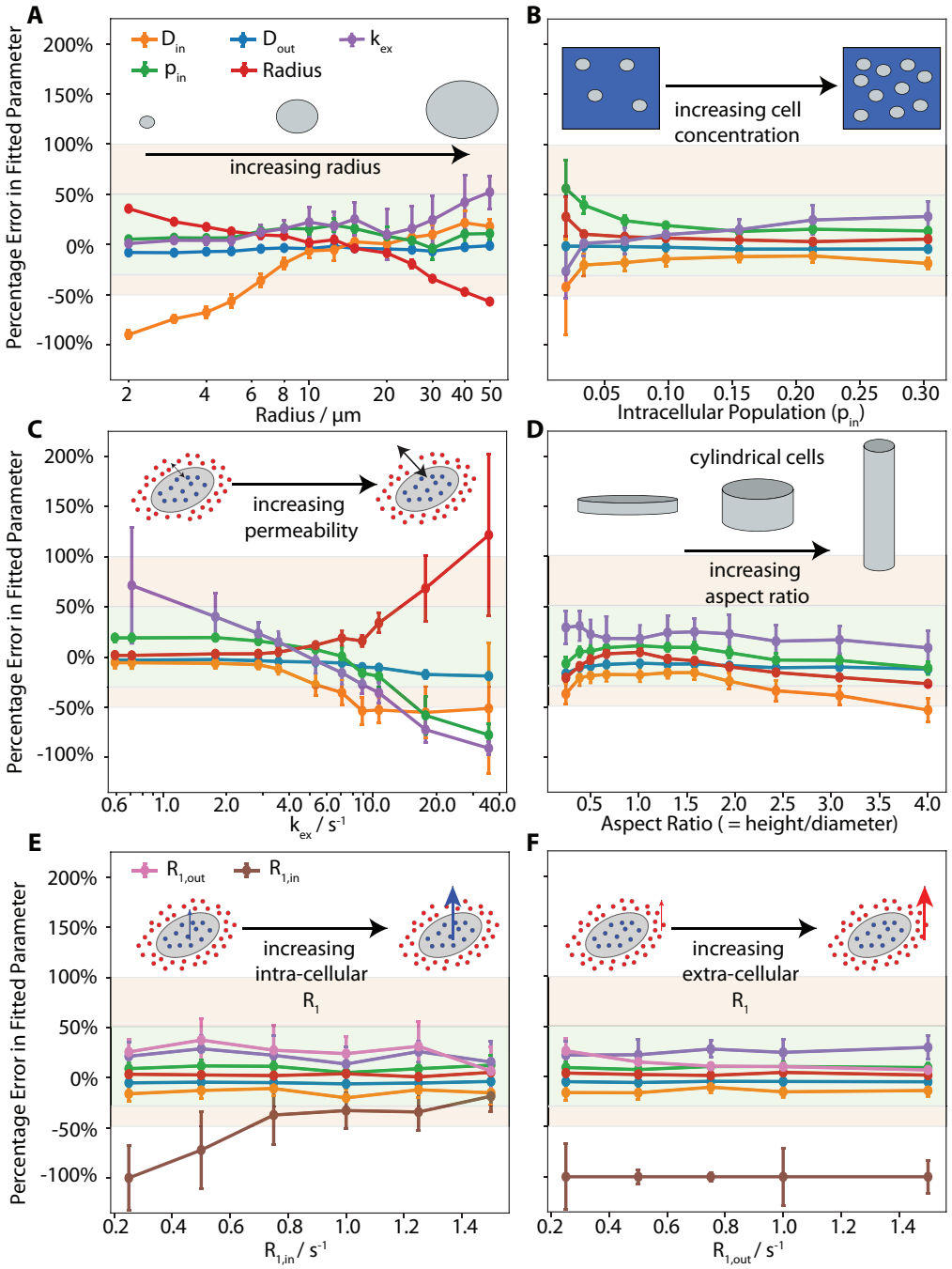

Figure 2.11: Error plots made by comparing the fitted parameters from INDIANA model analysis to ground truth values in simulations. Green regions indicate where a parameter is determined to within a factor of 1.5 and orange regions indicate where it is determined to within a factor of 2. Error bars in model parameters are determined by a bootstrapping analysis, where for each set of parameters a subset of diffusion time curves are fit whose total number is equal to the number of simulated diffusion times but selected at random with replacement, a process repeated 1000 times. In each error plot the value of a parameter is systematically varied as follows: **A** radius **B** p_{in}, **C** k_{ex} **D** aspect ratio **E** R_1^{in}, **F** R_1^{out}. In **A-D** $R_1^{in} = R_1^{out} = 0\ s^{-1}$ whereas in **E-F** the values of $R_1 = 0.25\ s^{-1}$ when they are not being varied. For all other parameters when their value is not varied they are set as follows: $D^{in} = 500\ \mu m^2\ s^{-1}$, $D^{out} = 2000\ \mu m^2\ s^{-1}$, $k_{ex} = 3.55\ s^{-1}$, $r = 10\ \mu m$ (this is the average radius with the standard deviation set $2\ \mu m$) and $p_{in} = 0.15$.

In general, the performance of the model is good, falling into the green area of the figure. This is especially true for the range parameters expected to be encountered in mammalian cells: namely radii of approximately 5 to 20 μm(*84*) and exchange rates of $< 10\ s^{-1}$ as determined in previous studies(*76*, *83*, *87–89*). A key benefit of the Monte Carlo approach for model testing is that it can also elucidate specific instances where parameters may show systematic errors. Optimising q and Δ values for particular systems could significantly improve errors in fitted parameters, and the Monte Carlo simulations could be used as a guide for determining optimal experimental parameters for expected system conditions. The aim here is to show that the INDIANA model is able to robustly fit parameters in many cases for generic choices of q and Δ.

Considering each of the error plots in turn, from 2.11**A**, we see that most parameters are well determined up to radii of approximately 20 μm. At radii below 8 μm the fitted intra-cellular diffusion coefficient is systematically low. This arises because the diffusion times included in the analysis are biased towards $D\Delta \gg r^2$ region and so the effect of the internal diffusion coefficient on the signal attenuation is limited. Conversely, for larger radii, the radius becomes a poorly determined parameter as the diffusion times are not sufficiently long for the particles to experience significant perturbations from the larger restricted environment.

In 2.11**B**, we see that the diffusion parameters are accurately fit over a wide range of p_{in} values chosen to represent the range of cellular concentrations typically observed in in-cell NMR experiments. When p_{in} falls to less than 0.02, the signal associated with the intra-cellular phase becomes smaller and the associated signal becomes noisier at high diffusion attenuation. Consequently, the systematic and random errors of fitted parameters becomes larger.

Fig 2.11**C** shows the effect of increasing k_{ex}, directly proportional to membrane permeability, on error in fitted parameters. The INDIANA model is particularly effective when exchange is relativity slow ($k_{ex} \leq 10\ s^{-1}$). When exchange is very slow the percentage error in k_{ex} is large but the absolute error in k_{ex} is only $\approx 0.5\ s^{-1}$. By contrast, when the membrane permeability becomes very high,

parameters become poorly determined. On the one hand, by reference to Fig 2.3, when exchange is fast there are large correlations between parameters leading to significant random errors in fitted values. Moreover, when the membranes are very permeable restriction has a very limited impact on the diffusion signal as particles usually pass straight through membranes. Thus determining radius from the diffusion data in these systems is inherently challenging.

One assumption of the model presented is that the cells are spherical in shape. While in principal this presents a significant limitation of the method, we test this explicitly in 2.11**D**. For these simulations, the semi-permeable sphere in the simulations is replaced with a cylinder where the aspect ratio (defined here as the ration between the height and diameter of the cylinder) is varied. If the cylinders were aligned, we would expect to see anisotropic diffusion effects, however in these simulations we remove this dependence by averaging over x, y and z displacements. The fitted radius from the INDIANA analysis is compared to the radius of a sphere of the equivalent volume to calculate the percentage error in this parameter. Overall, the model performs very well over a range of aspect ratios: thus while it does not provide insight into the geometric anisotropy in the system, the analysis still provides a useful guide for the overall size of the cells. This further demonstrates the generality of the method presented here, showing that the analysis methodology is relatively insensitive to the specific details of the system under investigation.

In Fig 2.11**E-F** the effects of relaxation on the diffusion signal is considered. Relaxation effects have previously been implicitly ignored in the analysis of variable time diffusion data (*60*), its effects can become important in biological systems and where paramagnetic reagents are present, as has recently been noted in the analysis of FEXSY data (*90*). From 2.11**E** we can see that the intra-cellular longitudinal relaxation rate is poorly determined until its value is above 0.75 s^{-1}. We can rationalise this by considering the effect of the intra-cellular relaxation rate for the given parameters: since $p_{in} = 0.15$ the maximum effect on the signal intensity at $R_1^{in} = 0.25\ s^{-1}$ and $\Delta = 0.5$ will be less than 2% of the total signal attenuation. Due to correlation between parameters, the effects of longitudinal

relaxation can be absorbed into other parameters without introducing significant errors into their fitted values. As the value of R_1^{in} increases it becomes an increasingly well determined parameter as its influence on the diffusion attenuation increases. This is again observed in Fig 2.11**F** where the small value of R_1^{in} means that its value is consistently underfit. Clearly, the parameter errors could be significantly reduced by introducing a minimum value (for example, 0.25 s^{-1} corresponding to the approximate relaxation rate for bulk water). The aim here however is to capture the performance of the model with minimal intervention in the values of fitted parameters, the only restraint being that all parameters must have positive (and therefore physically realisable) values. R_1^{out} by contrast is a consistently well determined parameter reflecting its larger influence on the NMR signal as the majority of water is associated with the extra-cellular space.

2.4 Experimental Validation

Using the Monte Carlo simulations as a guide to the reliability of the model outputs, we now seek to experimentally validate the INDIANA model on cellular systems. As shown in Fig 2.12, the model is tested on four different cell samples: yeast cells (*S. cerevisiae*), mammalian 3T3 cells and two samples of human HeLa cells at different cell concentrations. The samples are prepared as stated in Section 2.6.2 and where possible fitted parameters are compared to independent measurements and literature values to validate the results (Table 2.1 and 2.2).

From Fig 2.12, it is clear that the INDIANA model is able to fit the diffusion attenuation well in a variety of cellular systems. The fitted parameters are also generally in very good agreement with independent verification and literature values, as shown in Tables 2.1 and 2.2. There are a number of parameters whose values are inaccurate but these instances can all be anticipated from the Monte Carlo simulations, highlighting the utility of this approach and the ability of simple simulations to capture the key physical features of the complex, heterogeneous cellular samples.

Sample	D_{in} ($\mu m^2\ s^{-1}$)	D_{out} ($\mu m^2\ s^{-1}$)	k_{ex} (s^{-1})	radius (μm)	p_{in}	R_1^{in} (s^{-1})	R_1^{out} (s^{-1})
Yeast (*S. cerevisiae*)	20± 4	1250± 15	1.23± 0.15	3.57± 0.18	0.092± 0.03	0.86± 0.13	0.81± 0.07
3T3	826± 65	1980± 44	4.45± 0.24	12.42± 0.11	0.086± 0.03	0.00± 0.00	0.29± 0.03
HeLa 1	291± 62	1310± 69	7.39± 0.66	10.38± 0.18	0.169± 0.012	0.00± 0.09	0.51± 0.07
HeLa 2	344± 86	1750± 61	6.43± 0.49	10.69± 0.22	0.046± 0.003	0.00± 0.02	0.28± 0.06

Table 2.1: Fitted parameters from diffusion analysis shown in Fig 2.12. Error bars are derived from 1000 bootstrap runs.

Sample	radius (here)	radius (literature)	k_{IO} (here)	k_{IO} (literature)	cell density (here)	cell density (independent)
Yeast (*S. cerevisiae*)	3.57± 0.18	3.97(*91*), 2.88(*92*)†	1.12± 0.15	1.49(*93*)‡, 1.80(*94*)	$5.3\times10^8\pm 8\times10^7$	≈2.1×10^9§
3T3	12.42± 0.11	5.45 to 14.33	4.07± 0.24	—	$1.1\times10^7\pm 4\times10^5$	—
HeLa 1	10.38± 0.18	10.5± 2.2(*83*, *84*)	6.14± 0.66	8.4(*83*)¶	$3.95\times10^7\pm 6.6\times10^6$	3.5×10^7✠
HeLa 2	10.69± 0.22	10.5± 2.2(*83*, *84*)	6.13± 0.49	8.4(*83*)¶	$8.99\times10^6\pm 1.25\times10^6$	1.1×10^7✠

Table 2.2: Comparing fitted and derived values from 2.1 to literature and independently measured values. †: reported average radii. ‡: Temperature of measurement is not stated. ¶: Measurement is made at 310 K so value is expected to be higher than in this study where measurements are made at 298 K. §: Estimated from OD600 measurement. ✠: Measured using a haemocytometer. Note: $k_{IO} = k_{ex}p_{out}$.

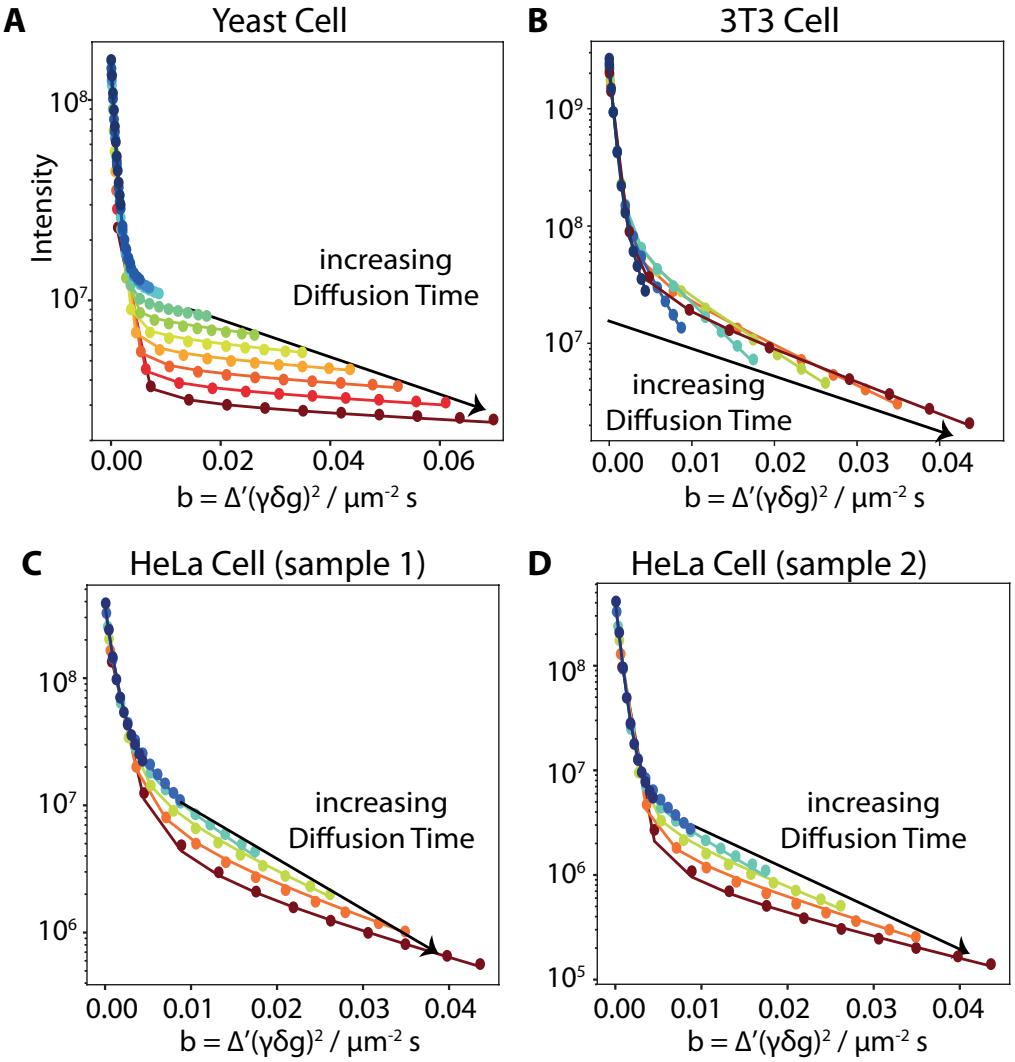

Figure 2.12: Diffusion attenuation measured experimentally on cellular systems of **A** yeast cells (*S. cerevisiae*) **B** mammalian 3T3 cells and **C**, **D** human HeLa cells with different cell concentrations. Dots indicate experimental points with different colours representing different diffusion times and lines showing fits to the data using the INDIANA model presented here. Sample preparation, experimental parameters and data fitting are explained in detail in subsection 2.6. Fitted parameters are given in Table 2.1.

For example, in the yeast sample the intra-cellular diffusion coefficient, D_{in}, is unrealistically low and the fitted intra-cellular longitudinal relaxation rates, R_1^{in}, in the 3T3 and HeLa samples are too low. However, these errors are in keeping with expectations from Fig 2.11. Furthermore, the fitted parameters of radius, exchange rates and intra-cellular populations are in good agreement with independent measures of these properties and are expected to be well determined parameters from simulations, demonstrating the utility of the analysis protocol. A

significant deviation is seen between the cell density measurement for the yeast cell sample. As yeast cells are significantly smaller than mammalian cells, the cell number is very sensitive to the precise fitted values for radius and p_{in}, and the independent measure from the OD600 measurement is a very approximate estimate of cell number.

2.5 Discussion

The INDIANA method can be used to characterise cell density, permeability and size as well as the intrinsic relaxation and diffusion properties of cellular suspensions. This is achieved experimentally by recording pulsed field gradient stimulated echo experiments with variable diffusion times that are subsequently globally analysed. The approach uses a stimulated echo rather than a spin echo sequence and thus takes advantage of the fact that $R_1 \leq R_2$. This allows access to longer diffusion times so that the effects of restriction and permeability can be properly characterised.

While a number of models have been previously proposed for analysing diffusion in confined and/or semi-permeable systems, particularly for diffusion MRI(*64–66*), a shortcoming of these is that due to the complexity or heterogeneity associated with biological tissue, these models tend to be limited in the number systems they can be applied to. Here, we show that when working with cellular suspensions, the model presented is able to work reliably for a wide range of environments. Moreover, while the specific diffusion parameters could be optimised for particular systems, the model is nevertheless able to perform well with generic experimental parameter choices. We show here that with our choices of experimental parameters the model is particularly amenable to the analysis of mammalian cell suspensions.

In the process of validating the model, we have been able to show that the GPA approach for modelling restricted diffusion, surprisingly, has an extended range of validity when accounting for the distribution of sizes in encountered cellular systems compared to a monodisperse system. We have also shown how relatively simple Monte Carlo simulations are able to capture the main features of real data in these complex and heterogeneous systems and can therefore serve as a very useful tool in guiding and interpreting experiments.

There are also several inherent assumptions in the Kärger diffusive exchange model. Notably, in the model all particles in a particular phase have an equal probability of exchanging phases, independent of their position, whereas in cellular systems exchange can only occur at membranes. While we would expect this assumption to place significant limitations on the model, through the extensive simulations and experimental validation presented here we find the it is nonetheless robust over wide parameter ranges.

Another interesting aspect of the model is that the intra- and extra-cellular water diffusion are modelled using single diffusion coefficients. In reality, the water will experience binding and adsorption with membranes, proteins and nucleic acids, particularly in the heterogeneous intra-cellular phase. However, these interactions will typically be on ps-µs timescales(*95*, *96*). The water is therefore in fast exchange between these different environments from the perspective of the diffusion experiment and the reported diffusion coefficients are averaged over these different interactions.

It is interesting to compare the intra-cellular diffusion coefficients measured here to other studies. Persson *et al*(*96*) demonstrated using deuterium relaxation dispersion experiments that 85% of the water in *Escherichia coli* show bulk water like rotational diffusion characteristics while only the remaining 15% of water is significantly motionally slowed. This is inconsistent with the result here where the measured intra-cellular diffusion coefficients are approximately five times smaller than the extra-cellular ones. Two major factors likely contribute to this discrepancy: firstly, as noted in the Persson study, translational diffusion coefficients measure longer-range behaviour and will be significantly affected by crowding or obstruction effects that slow molecular displacement over micrometer length scales, but have minimal impact on measured rotational correlation times. A second factor contributing to the lower measured translational diffusion coefficients may be the effects of further restriction within organelles. As future work, the models here could be extended to include restricted diffusion within organelles (adding greater physical plausibility) and to test whether such a model can provide a good fit to the experimental data.

While in this chapter we have focussed on the water resonance, this analysis can in principle be extended to a resonance from any molecule that can be spectroscopically resolved. Given that ^{1}H NMR spectra of cells are very crowded with many overlapping peaks, when looking at species other than water, this approach will be most powerful when a particular molecule has been isotope edited. This opens up the possibility of doing a background free measurements using an isotope edited pulse sequence.

There has been tremendous recent growth of in-cell NMR studies(*49–51*), with most of these focussing on gaining atomic-level structural and dynamical studies on isotopically enriched molecules in cellular environments. However, with a few notable exceptions (*97*), the study of diffusion has been largely overlooked. The protocol set out in this chapter provides a way to probe these systems in more detail through their diffusional properties. Furthermore, due to the very high concentration of water present in these systems experimental data required for the analysis can be acquired in a few minutes. Therefore, periodically acquiring diffusion data during long in-cell NMR experiments can potentially serve as a real-time probe of the integrity of the cell sample.

2.6 Methods

2.6.1 Monte Carlo Simulation Methods

For all simulations, the displacement of each particle is updated during each step of duration δt in the x, y and z dimensions by amounts:

$$x(t + \delta t) = x(t) + \delta x_{step} \tag{2.40}$$

$$y(t + \delta t) = y(t) + \delta y_{step} \tag{2.41}$$

$$z(t + \delta t) = z(t) + \delta z_{step} \tag{2.42}$$

The displacement step sizes, δx_{step}, δy_{step} and δz_{step} are determined independently by sampling a number drawn randomly from a normal distribution with a standard

deviation $\sqrt{2D\delta t}$ where D is the diffusion coefficient of the particle. When the effects of relaxation are additionally considered an intensity factor is updated by a factor:

$$I(t+\delta t) = I(t)e^{-R\delta t} \tag{2.43}$$

Where R is the relevant relaxation rate.

For Monte Carlo simulations involving restricted geometries space is discretised into cubes with either a semi-permeable sphere or cylinder at the centre of each cube. Identical cubes extend in all directions as shown in Fig 2.5**A**. The volume fractions intra to extra-cellular space are determined by the relative volumes of the cubes and the spheres and cylinders within them. In this case, the particles are initialised at the start of simulations at points in space such that their average initial positions reflect the intra- and extra-cellular mole fractions of the system. The movement and intensity factors also can become dependent on the position of the particle, such that when inside the cavity the particle has a diffusion coffecient of D^{in} and relaxation rate of R^{in} and when outside these become D^{out} and R^{out} respectively.

To account for the effects of exchange over membranes, after each step the new position of the particle is checked. For particles that have moved over a membrane during the previous step we must determine whether this jump will be accepted. The in-to-out P_{IO} and out-to-in P_{OI} jump probabilities are dependent on the permeability (κ) of the membrane and have been previously calculated as(*56*, *98*):

$$P_{IO} = 2s\frac{\kappa}{D_{in}} \tag{2.44}$$

$$P_{OI} = 2s\frac{\kappa}{D_{out}} \tag{2.45}$$

where s is the distance between the particle and cavity surface prior to the jump. The cavity permeability, κ is related to the in to out rate constant, k_{IO}, through the relationship(*87*):

$$\kappa = k_{IO}\frac{V}{S} \tag{2.46}$$

Where V is the volume of the cavity and S is its surface area. In the case of spherical cavities this relationship becomes: $\kappa = k_{IO}\frac{r}{3}$ where r is the radius of the sphere. For all particles that have jumped across the membrane a random number is drawn form a uniform distribution between 0 and 1. When the random number is less than the relevant crossing probability the move is accepted but when it is greater the jump is rejected and the particle is moved back to its previous position. Dealing with wall collisions in this way is both simpler and computationally less expensive than more sophisticated methods, for example specular reflection, but does not result in any loss of accuracy in results as tested here and previously reported(*78*, *99*).

The Monte Carlo simulations were implemented in Python and source code is available on request. For one set of simulation parameters typically 500000 trajectories were calculated with the time step in each simulation set to 10 μs and a typical maximum time of 0.5 s. To account for the variation in sizes of different cells, for each new single particle trajectory the radii of spheres was drawn randomly from a normal distribution where the average radius is set within the simulation and the standard deviation is set to 20% of the average value. Since all single particles are independent of each other the simulations can be run in parallel easily. This is achieved here using the Parallel Python (PP) module. The time for a simulating a single set of parameters (500000 trajectories with a time step 10 μs and a maximum diffusion time of 0.5 s) parallelised over 20 2.4 GHz Intel Xeon CPUs was approximately 50 minutes.

Once the simulations are completed the signal attenuation can be calculated using equation 2.35. The simulation code returns a list of displacements that can then be used to calculate a histogram, $P(z, \Delta)$ (an example of this is shown in Fig 2.5**B**). The values of q($=\gamma\delta g$) are under user control and represent the parameters that could be varied in a real diffusion experiment. Typically, $\delta = 2\ ms$ and g was varied quadratically in the range 4 to 60 G cm^{-1}, mimicking commonly used experimental values. Under conditions where attenuation occurs more quickly the q values are adjusted such that at the highest applied gradient field strength the signal is attenuated by approximately two orders of magnitude.

To model the effects of non-uniform gradients we can adjust equation 2.35 to:

$$S(q,\Delta) = \int_{-\infty}^{\infty} \int_{g_{min}}^{g_{max}} P(z,\Delta)P(g)\cos\left(\gamma\delta(g)z\right)\, dg\, dz \tag{2.47}$$

Where P(g) is the probability of a given gradient strength.

For cases where we additionally consider the effect of relaxation this must be further adjusted. In addition to a displacement term, the simulation code will also return an intensity factor for every particle. We can incorporate this by calculating a two dimensional histogram from which the signal attenuation can be calculated as:

$$S(q,\Delta) = \int_{-\infty}^{\infty} \int_{I_{min}}^{I_{max}} \int_{g_{min}}^{g_{max}} P(z,I,\Delta)I(\Delta)P(g)\cos\left(\gamma\delta(g)z\right)\, dg\, dI\, dz \tag{2.48}$$

Where I is an intensity factor accounting for the effects of relaxation.

2.6.2 Cell Preparation Methods

Yeast cell samples were prepared similarly to previously published protocols (*87*, *94*). Briefly, dried yeast was mixed with tap water in a ratio of approximately 1:3 and the mixture was left for three days during which it was shaken periodically to allow for the release of carbon dioxide bubbles. The mixture was subsequently transferred to a 5 mm standard NMR tube and 5% D_2O was added to provide a lock signal.

For the 3T3 sample, cells were grown in Dulbecco's modified eagle medium (DMEM) in T-175 flasks and detached from the surface by incubation with tyrpsin-EDTA for 5 minutes. Cells were then resuspended in 250 μL Dulbecco's phosphate buffered saline (DPBS) with 5% D_2O and transferred to a 5 mm Shigemi NMR tube for experiments.

For HeLa samples, cells were grown in DMEM supplemented with 10% foetal bovine serum (FBS) in petri dishes and detached using mechanical scraping (using petri dishes gives access to maximal area for scraping cells). Cells were then resuspended in 250 μL FBS with 5% D_2O and transferred to a 5 mm Shigemi NMR tube for experiments.

2.6.3 NMR Experimental Methods

All NMR experiments were carried out at 298 K on a 14.1 T (600 MHz) Agilent DD2 NMR spectrometer equipped with a 5 mm z-axis gradient triple resonance room temperature probe. Gradients are calibrated using the known diffusion coefficient of residual HDO (1902 $\mu m^2\ s^{-1}$) in a sample of D_2O at 298 K(*100*). The spectrometer temperature was calibrated with methanol-d4. A pulsed-field gradient stimulated echo pulse sequence with variable diffusion delays was used. In all experiments $\delta = 2\ ms$ and τ_2, the time for which the magnetisation is transverse, is always set to 10 ms. For the yeast sample, 11 quadratically spaced gradient strengths from 8-60 $G\ cm^{-1}$ were used and Δ = 20, 40, 60 80, 100, 200, 300, 400, 500, 600, 700 and 800 ms. For the 3T3 sample, 11 quadratically spaced gradient strengths were used from 1-60 $G\ cm^{-1}$ and Δ = 50, 100, 200 300, 400 and 500 ms. For the HeLa cell samples, 11 quadratically spaced gradient strengths from 8-60 $G\ cm^{-1}$ were used and Δ = 50, 100, 200 300, 400 and 500 ms. All spectra were Fourier transformed and phased using NMRPipe(*101*). Data are analysed in phase-sensitive mode, with the lowest g value spectrum used determine phase parameters that are then applied to all other spectra. In each spectrum the water peak is integrated using Python scripts, making use of the nmrglue module(*102*), and peak intensities as a function of gradient parameters are carried forward for analysis as described below.

2.6.4 Fitting Simulated and Experimental Data to Theoretical Model

Experimental and simulated NMR data are fitted to the INDIANA model described in section 2.2 (specifically equation 2.12) with the apparent diffusion coefficient in the intra-cellular compartment calculated using equation 2.26. Fitting is performed using the Levenberg-Marquardt algorithm as implemented in the python module Lmfit(*103*). Uncertainties in individual signal intensities were estimated from repeat measurements and found to be approximately a constant percentage over several orders of magnitude, as shown in Fig 2.13. To account for this the fitting is

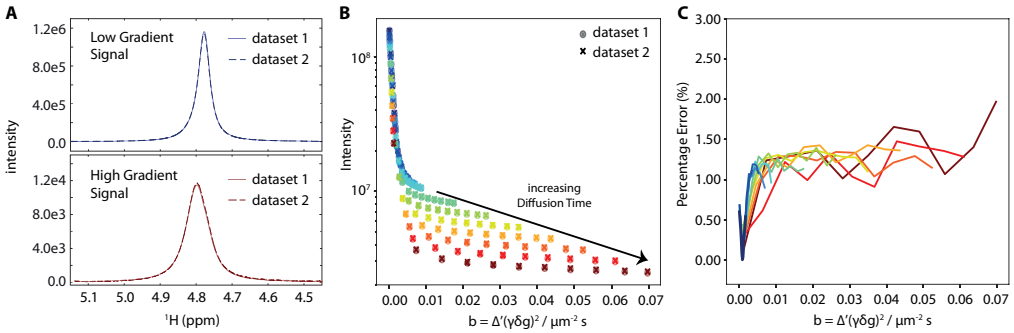

Figure 2.13: **A** Examples of water peaks in spectra of yeast cells under conditions of low and high gradients, highlighting their reproducibility. **B** Overlaid experimental diffusion attenuations carried out on a yeast cell sample under identical conditions (298 K; 12 diffusion times between 20 and 800 ms; $\delta = 2$ ms and 11 quadratically spaced gradient strengths between 8-60 $G\ cm^{-1}$). The signal attenuation is very reproducible even when the attenuation is greater than an order of magnitude. **C** Percentage error in each intensity value calculated by taking the absolute difference between signal intensity under identical conditions and dividing by the average value. The percentage error is slightly smaller at lower b-values but generally relatively constant.

performed on the logarithm of the signal intensity. The only restraint placed on the free parameters is that their values must be ≥ 0 and thus be physically realisable.

Software for performing the INDIANA analysis can be downloaded from: `indiana.chem.ox.ac.uk`.

2.7 Acknowledgements

3T3 cell sample was prepared by Louise R. Tear (Chemistry Research Laboratory, University of Oxford) and HeLa cell samples were prepared by Richard J. Wheeler (Sir William Dunn School of Pathology, University of Oxford).

3

Chemical Exchange Saturation Transfer

Abstract

This chapter describes the water chemical exchange saturation transfer (CEST) experiment. This experiment is particularly useful in MRI studies as it allows system properties to be inferred through the interrogation of the water resonance that is readily measured without the addition of exogenous agents. Here, we use a combination of simulations and experiments to investigate how different factors can affect the CEST signal. We find that while the CEST signal is indeed sensitive to and thus potentially a reporter of a variety of factors this also means that great care must be taken when trying to ascribe changes in the signal to a single particular factor. Furthermore, different saturation frequencies can show distinct reponses to changing conditions meaning that greater physical insight can be gained by studying the CEST profile across a full range of frequency offsets. Finally, we demonstrate how CEST can be employed to monitor a protein aggregation process.

3.1 Introduction

The saturation transfer class of NMR experiments, herein referred to as CEST (chemical exchange saturation transfer), represented an early innovation in NMR that was originally used in studies in the 1960s by Forsen and Hoffman(*104*). In the last decade use of the experiment has increased exponentially in both biomolecular NMR and MRI imaging. In biomolecular NMR, a number of CEST experiments have been developed to characterise the kinetics and thermodynamics of interconversion between the major and minor conformations of proteins as well as the structure of transiently populated states(*105*, *106*). In clinical MRI, CEST experiments have emerged as a potentially useful tool in the diagnosis of various pathologies(*107*, *108*).

The premise of the CEST experiment is simple: consider a species undergoing slow two site exchange where each state has a distinct chemical shift and the exchange is very asymmetric such that there is a clearly identifiable major and minor state. In this situation, assuming no other species are present in the spectrum, the peak associated with the major state will be much more intense and easily identifiable. However, if a long weak pulse (B_1 pulse) is applied at the resonance frequency of the minor state, causing full or partial saturation of the spins, the saturation can subsequently be exchanged and 'read-out' on the more intense major state peak. When the lifetime of the minor state is significantly shorter than the length of the saturation pulse, multiple transfer events can take place leading to a significant build up of saturation on the major state that reports on the exchange kinetics.

A variety of NMR experiments based around the CEST element exist, for example the saturation transfer difference(*109*) and DEST(*110*) experiments. In biomolecular NMR the majority of CEST experiments have been developed for applications in molecules that have been enriched with particular NMR active nuclei, especially ^{13}C and ^{15}N(*105*, *106*, *111*). However, Balaban and coworkers discovered that the CEST experiment had the potential to measure solution pH by simply studying the exchange between labile protons and the solvent water(*112*, *113*).

A major advantage of these water-based CEST experiments, that will be the main focus of this chapter, is that the signal can be derived purely from the

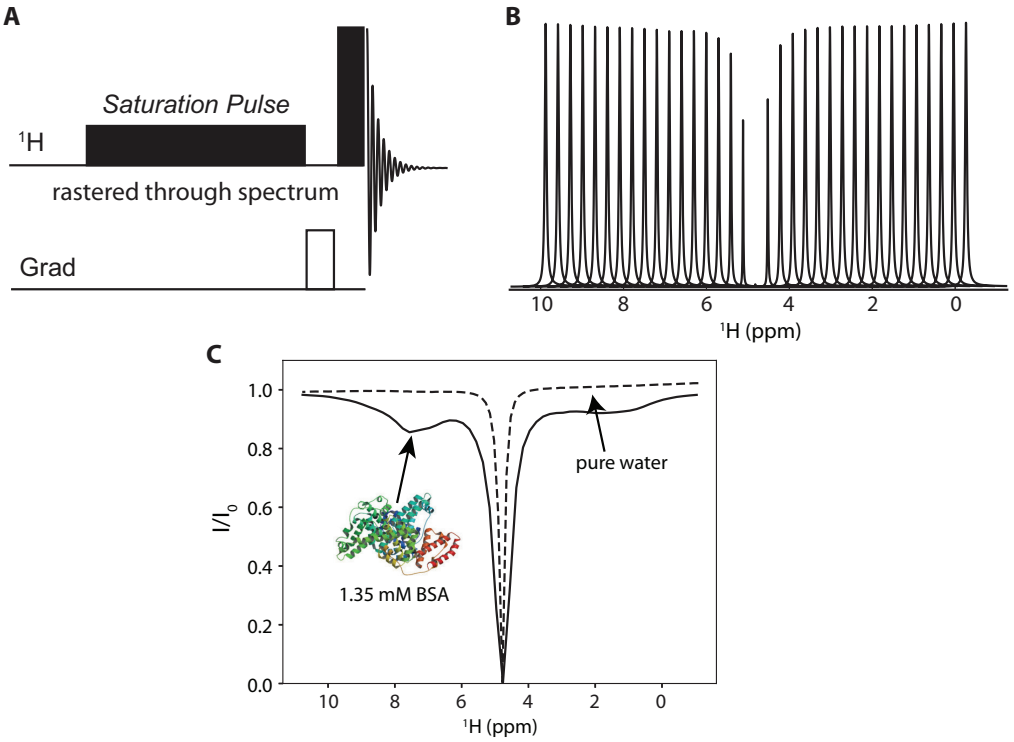

Figure 3.1: **A** Pulse sequence used to record a water CEST experiment. The sequence consists simply of a weak saturation pulse that is rastered through the spectrum and a full power observe pulse. A crushing gradient is employed to eliminate any transverse magnetisation after the saturation pulse, avoiding artefacts in the final signal. **B** At each saturation frequency the intensity of the water signal is measured. When the CEST pulse is applied at the water itself (4.77 ppm at 298 K) its signal is saturated. **C** The CEST spectrum (or profile) is calculated by taking the ratio of the water signal at a given saturation frequency with the signal in the absence of the saturation pulse. While the CEST profile of pure water only shows saturation or CEST effects when the water itself is saturated in a BSA solution these effects are visible at a range of saturation frequencies.

endogenous species in the system, without the need for any chemical labelling. This overcomes a significant hurdle in many MRI modalities that require the addition of an exogenous agent to provide contrast. The potential toxicity of these agents is of topical interest as retained deposits of gadolinium, derived from relaxation contrast agents, have been discovered in tissue following multiple administrations(*114*).

Several *in vivo* and *in vitro* studies of the water CEST experiment have now been carried out(*107*, *115*) and it has emerged as a promising tool in diagnosis, but as discussed in this chapter, fully rationalising the results of these experiments remains challenging(*116*).

The pulse sequence for performing these CEST experiments is particularly

simple, as illustrated in Fig 3.1**A**. Firstly, a low power saturation pulse, with a specified power and time (typically a few seconds), is applied at a certain offset frequency. This is then followed by a crushing gradient that dephases all transverse magnetisation, preventing artefacts in the final signal, and a full power observe pulse. For each offset the saturation pulse is applied at, the water signal intensity is then measured, as shown in Fig 3.1**B**. A CEST spectrum or profile (sometimes referred to in the literature as a z-spectrum), is calculated by plotting the ratio of the water signal in the presence and absence of the saturation pulse as a function of saturation pulse frequency (Fig 3.1**C**). While in pure water there is no mechanism for saturation transfer when the pulse is not directly applied on the water resonance, in the BSA spectrum saturation transfer or CEST effects are clearly visible at a range of offset frequencies.

Analysis of Fig 3.1**C** reveals the the name CEST to be a slight misnomer for these experiments. On the one hand, CEST effects observed downfield of water, can be associated with saturation build up followed by formal chemical exchange of labile protons from -OH, -NH_2 and amide groups leading to an attenuation in the observed water signal. However, as is evident in Fig 3.1**C**, CEST effects can also be observed upfield of the water saturation. In a protein, resonance frequencies in the range 0-2 ppm will be associated with aliphatic groups such as -CH, -CH_2 and -CH_3 groups. Saturation transfers between the protons in these groups and water cannot occur by chemical exchange alone since these protons are non-labile. In this case, the saturation transfer must involve a through space magnetisation transfer, an NOE: this can either be an intra-molecular transfer to a labile proton that subsequently exchanges with water, referred to as a relayed NOE, or a direct inter-molecular transfer between the non-exchangeable protons and water(*116*). The different mechanisms of saturation transfer active in the CEST experiment are summarised in Fig 3.2.

At their core, NOEs are spin relaxation effects whose properties are consequently strongly influenced by the motional properties of the molecules involved. While the CEST literature has largely focussed on the properties of the CEST signal associated

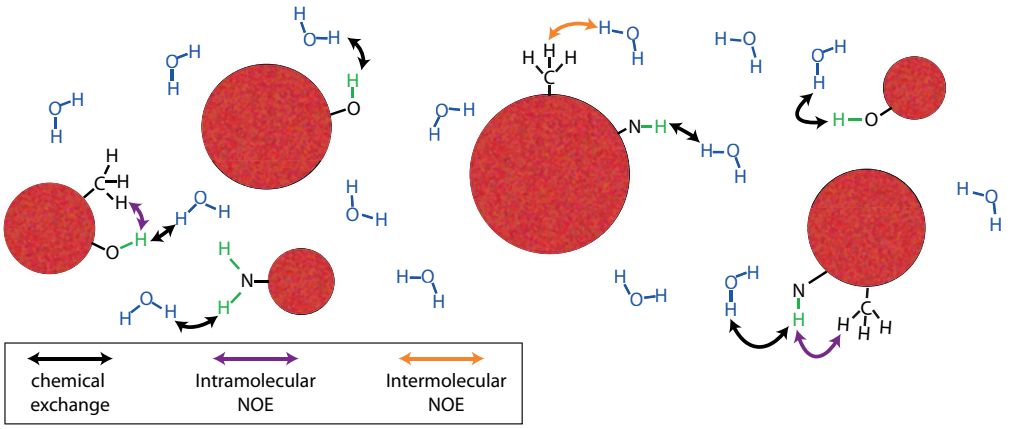

Figure 3.2: Cartoon representation of the different mechanisms of saturation transfer active in water CEST experiments: chemical exchange as well as both relayed and intermolecular NOEs.

with exchanging species resonating downfield of water, in this chapter we seek to demonstrate how the NOE side of the CEST profile, upfield of water, can also be a rich source of information. We probe, both theoretically and experimentally, how a number of different factors affect the form of a CEST profile.

3.1.1 Motivation

Our interest in gaining further insight into the water CEST signal stems from a desire to understand the physical processes underlying changes in CEST effects observed in patients of different pathologies. Highlighting this, in Fig 3.3 we show unpublished CEST data from our collaborators (George Harston and James Kennedy, John Radcliffe Hospital, University of Oxford) for patients who have suffered acute stroke on presentation and 24 hours later. Apparent diffusion coefficients in each of the tissue regions (explained below) are also shown. In all cases the measured CEST effect or ADC is given relative to the value measured on contralateral tissue.

Tissue affected by stroke can be categorised into three groups(*107*): (1) the ischaemic core (IC) represents tissue where irreversible infarction has occurred with no potential for recovery (2) the infarct growth (IG) represents tissue immediately surrounding the IC that will go on to infarct and (3) oligaemic tissue that is slightly

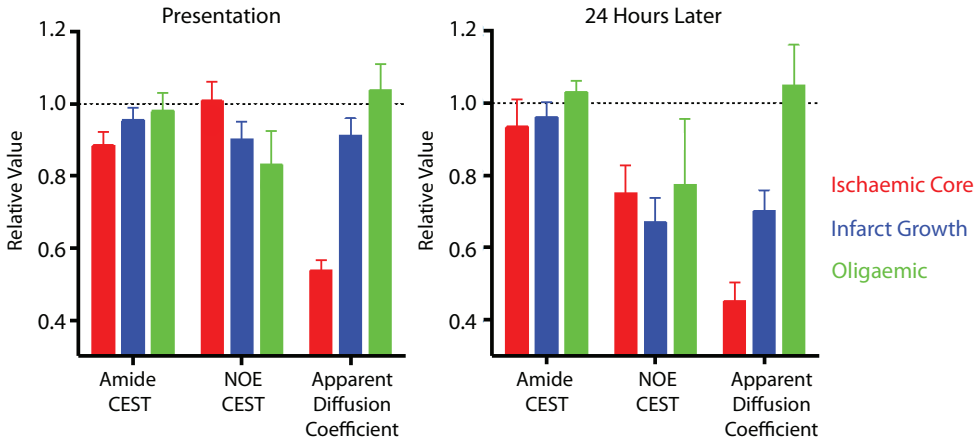

Figure 3.3: Stroke Patient CEST and apparent diffusion coefficient data at presentation and 24 hours later. Data is shown for three categories of tissue: ischaemic core, infarct growth and oligaemic as explained in the text. In all cases, values are shown relative to measurement on contralateral tissue. Figure provided by George Harston (John Radcliffe Hospital, University of Oxford).

hypoperfused but should recover. As shown in Fig 3.3 these regions can be identified on the basis of their differential apparent diffusion coefficients.

However, the CEST signal also shows significant changes for the stroke affected tissue, reflecting a change in the overall molecular environment. Interestingly, the NOE CEST, generally much less studied than the amide CEST, shows bigger changes, particularly after 24 hours. We aim to gain further insights into the features that may drive these changes through simulations and experiments on model systems. While these experiments are clearly far removed from the complexity of a human brain, the systems can be tuned precisely to monitor how the CEST signal responds to a change in a particular variable.

For example, it is generally thought that the major factor that modulates the CEST signal in stroke is pH, since this modulates the rate of proton exchange with the solvent(*107*). This result is however inconsistent with the fact that the NOE CEST signal is generally more sensitive to the effects of ischaemic stroke than the amide CEST. The CEST signal could also be modulated by local viscosity changes, that the ADC results suggest are present, or the degradation of proteins where infarction has occurred. In this chapter we test how these factors can differentially affect the amide and NOE CEST signals.

3.2 Theory and Simulations

3.2.1 Saturation Pulses

We begin this section by considering the effect of a saturation pulse on an isolated spin, temporarily neglecting the role of saturation transfer. When a saturation pulse is applied on-resonance the spin will rotate around the axis about which the pulse is applied in the rotating frame, with the frequency of oscillation dependent on the strength of the pulse. In the absence of relaxation these oscillations would continue indefinitely, however due to its presence in all cases, eventually a steady state magnetisation will be realised. Saturation is achieved when the eigenstates of the Zeeman Hamiltonian are equally populated and there is no coherence between the spins i.e. the macroscopic magnetisation equals zero.

The effect of these pulses can be modelled and simulated conveniently using a Bloch equation formalism. Here we consider the case when the saturation pulse is applied along the x-axis:

$$\frac{d}{dt}\begin{bmatrix} E \\ I_x \\ I_y \\ I_z \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & -R_2 & \Omega_I & 0 \\ 0 & -\Omega_I & -R_2 & \omega_1 \\ R_1 I_{eq} & 0 & -\omega_1 & -R_1 \end{bmatrix} \begin{bmatrix} E \\ I_x \\ I_y \\ I_z \end{bmatrix} \tag{3.1}$$

In this equation R_1 and R_2 refer to the longitudinal and transverse relaxation rates respectively. The identity operator is included to account for the return of magnetisation back to equilibrium. The Ω_I term accounts for precession in the rotating frame and ω_1 is the frequency of the saturation pulse ($\omega_1 = \gamma B_1$ where B_1 is the strength of the saturation pulse in Tesla (T)).

By taking the matrix exponential of 3.1 it is easy to follow the progression of the bulk magnetisation under the effect of a saturation pulse, as illustrated in Fig 3.4. Fig 3.4**A** shows the simple case of an on-resonance saturation pulse. As the pulse is applied on-resonance the effective magnetic field in the rotating frame is aligned with it in the x-axis and the magnetisation oscillates in the yz-plane while undergoing significant attenuation due to relaxation. We can calculate the steady

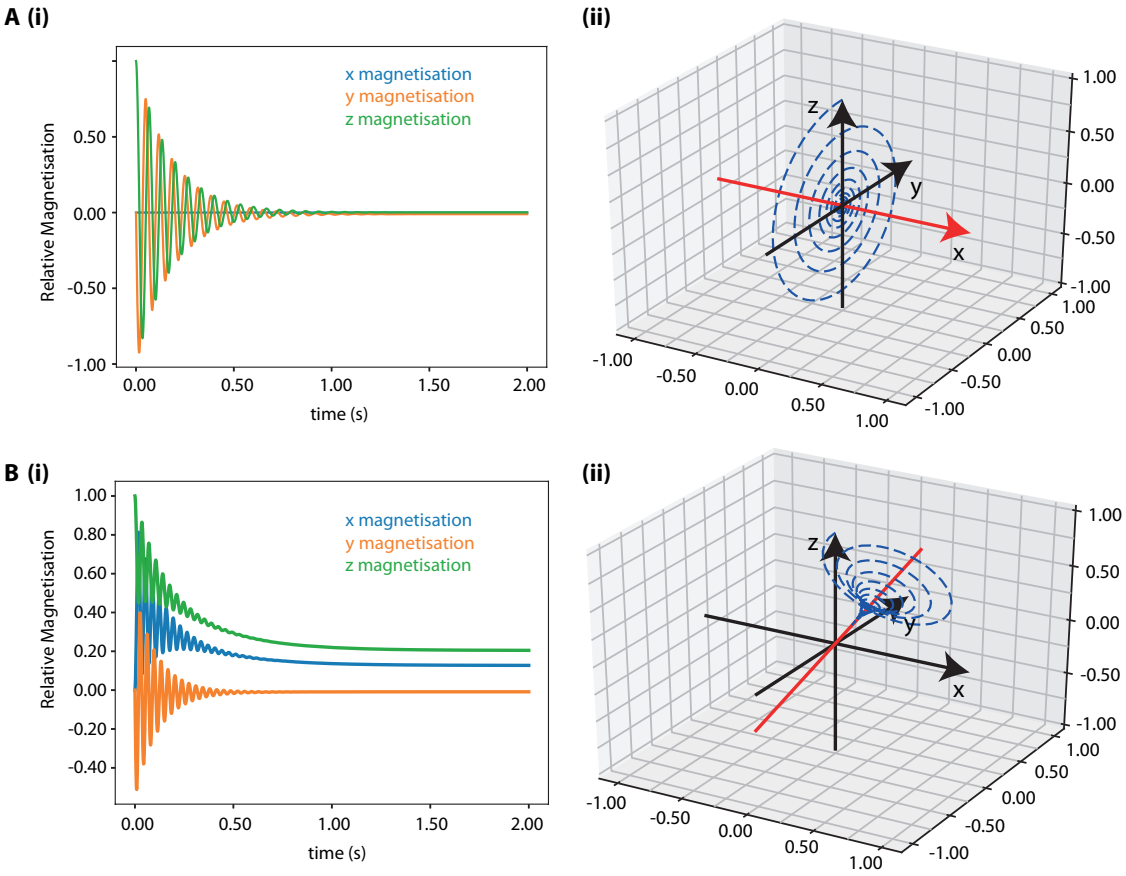

Figure 3.4: Response of bulk magnetisation to on- (**A**) and off-resonance (**B**) saturation pulses applied along the x-axis. In both cases the longitudinal relaxation rate, R_1, is set to 1 s^{-1} and the transverse relaxation rate, R_2, is 10 s^{-1}. The saturation pulse has a frequency of $\omega_1 = 15 \times 2\pi\ s^{-1}$ and is applied along the x-axis in both cases. In **A**, the saturation pulse is on-resonance ($\Omega_I = 0$) meaning that the magnetisation precesses around the x-axis (highlighted in red), while also relaxing. On the other hand, in **B** the saturation pulse is off-resonance, $\Omega_I = 24 \times 2\pi\ s^{-1}$, and consequently the magnetisation precesses about an effective tilted field.

state magnetisation at long times by taking the matrix exponential of equation 3.1 in the limit $t \to \infty$. From this we find:

$$I_x = 0 \tag{3.2}$$

$$I_y = \frac{\omega_1 R_1}{\omega_1^2 + R_1 R_2} \tag{3.3}$$

$$I_z = \frac{R_1 R_2}{\omega_1^2 + R_1 R_2} \tag{3.4}$$

For both y and z magnetisation we observe that when a relatively strong pulse is applied i.e. $\omega_1 \gg R_1 R_2$ then saturation will be achieved at long times and the

steady state magnetisation will tend towards zero. In practice B_1 inhomogeneity will also play a role in depleting the signal: this is caused by small variations in the applied B_1 field across the sample volume that leads to spins precessing at slightly different frequencies and a loss of coherence.

The magnetisation trajectory becomes significantly more complicated when the saturation pulse is off-resonance as shown in Fig 3.4**B**. In this case there is still a residual field along the z-axis in the rotating frame about which the magnetisation can precess. Therefore, on application of the saturation pulse the magnetisation precesses about an axis that is tilted by an angle $\theta = \arctan(\frac{\omega_1}{\Omega_I})$ from the z-axis, into the xz-plane. In general, the stronger the saturation pulse, the more prominent off-resonance (partial) saturation effects become. Thus, in CEST experiments a compromise must be reached between being able to localise saturation effects to specific frequencies and therefore molecular species as well as being strong enough to provide sufficient saturation build up. It may not be possible to have long saturation pulses due to probe limitations and such pulses may be undesirable in any case due to sample heating effects. In the context of MRI experiments, specific absorptions requirements may also become important. In this chapter we therefore only consider relatively weak saturation pulses, $\nu_1 \approx 50$ Hz.

3.2.2 Saturation Transfer between Exchanging Spins

In this section we explore saturation transfer effects when we have two sets of spins in exchange. Although this system represents a gross simplification of typical biological systems studied with water CEST experiments, it can still provide important physical insight into the key determinants of the CEST profile. Equation 3.1 can be expanded to include the effects of chemical exchange(*117*). For two spins I and S undergoing chemical exchange we find:

$$\frac{d}{dt}\begin{bmatrix} E \\ I_x \\ I_y \\ I_z \\ S_x \\ S_y \\ S_z \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -R_2^I - k_{IS} & \Omega_I & 0 & k_{SI} & 0 & 0 \\ 0 & -\Omega_I & -R_2^I - k_{IS} & \omega_1 & 0 & k_{SI} & 0 \\ R_1^I I_{eq} & 0 & -\omega_1 & -R_1^I - k_{IS} & 0 & 0 & k_{SI} \\ 0 & k_{IS} & 0 & 0 & -R_2^S - k_{SI} & \Omega_S & 0 \\ 0 & 0 & k_{IS} & 0 & -\Omega_S & -R_2^S - k_{SI} & \omega_1 \\ R_1^S S_{eq} & 0 & 0 & k_{IS} & 0 & -\omega_1 & -R_1^S - k_{SI} \end{bmatrix} \begin{bmatrix} E \\ I_x \\ I_y \\ I_z \\ S_x \\ S_y \\ S_z \end{bmatrix} \tag{3.5}$$

In this equation $I_{eq} = \frac{[I]}{[I]+[S]}$ and $S_{eq} = \frac{[S]}{[I]+[S]}$. The constants k_{IS} and k_{SI} are the rate constants for exchange between spins I and S and S and I respectively. In these systems the composite exchange variable $k_{ex} = k_{IS} + k_{SI}$ is useful in characterising the exchange.

The above equation can be used to simulate CEST profiles for two spins undergoing exchange as illustrated in Fig 3.5. This is done by taking the exponential of the matrix in equation 3.5 following a CEST pulse of specified time and stepping the saturation pulse through a set of offset frequencies (in the simulations this is achieved by altering Ω_I and Ω_S appropriately). The CEST profile is then calculated by taking the ratio of z-magnetisation of the I spin (assuming $[I] \gg [S]$) in the presence and absence of the saturation pulse as a function of the frequency at which the pulse is applied (the z-magnetisation ratio is taken because in an actual CEST experiment this will be turned into observable signal following application of a 90° pulse).

In Fig 3.5 we demonstrate, by systematic variation, how different factors can affect the form of the CEST profile between exchanging spins. In this case we assume that spin I refers to the water protons that are in very high excess and spin S refers to a pool of labile protons undergoing exchange. The full simulation parameters are given in the figure caption.

In 3.5**A** we see how the exchange regime can profoundly affect the appearance of the CEST profile. In the slow exchange regime ($k_{ex} \ll \Delta\omega$ and $\Delta\omega = |\omega_S - \omega_I|$) we see a characteristic CEST dip at the chemical shift of spin S corresponding to chemical shift saturation transfer. If the exchange rate is increased while remaining within the slow exchange limit the saturation transfer effect becomes larger. However, as the exchange rate continues to increase into the intermediate

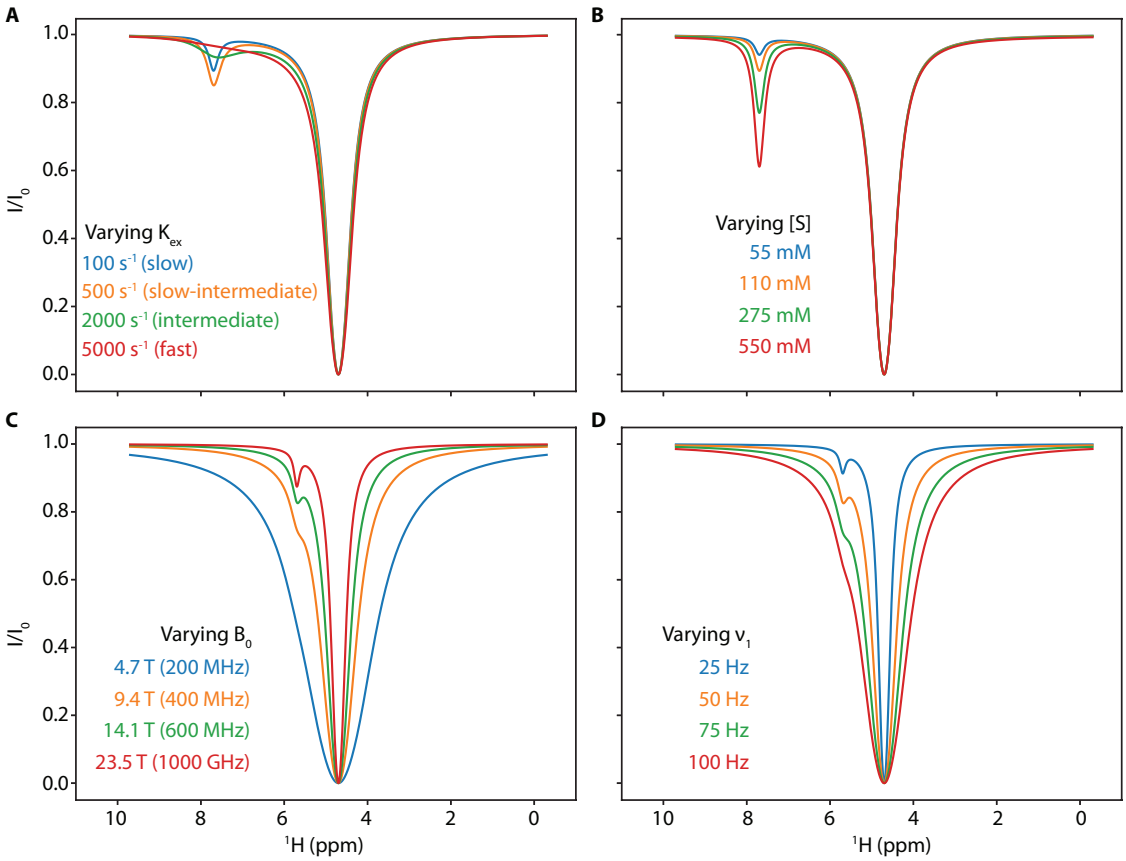

Figure 3.5: Simulated CEST profiles for a two spin system where different parameters are systematically varied. For all simulations the following parameters are fixed: $R_1^I = 0.5\ s^{-1}$, $R_2^I = 10.0\ s^{-1}$, $R_1^S = 1.0\ s^{-1}$, $R_2^S = 100\ s^{-1}$, [I]= 111 M and the saturation pulse is applied for 2 s. We use $\Delta\Omega = 3$ ppm (**A**,**B**) and $\Delta\Omega = 1$ ppm (**C**,**D**), where $\Delta\Omega = \Delta\Omega_S - \Delta\Omega_I$. Where a parameter is not being varied its value is set as follows: $B_0 = 14.1$ T (600 MHz), $k_{ex} = 100\ s^{-1}$, $[S] = 110$ mM and $\nu_1 = 50$ Hz. In each sub-figure the following parameters are varied: **A** k_{ex}, **B** [S], **C** B_0 (the static magnetic field strength) and **D** frequency of the saturation pulse: $\nu_1 = \frac{\omega_1}{2\pi}$.

regime ($k_{ex} \approx \Delta\omega$), the CEST effect becomes more spread out and smaller. If the exchange rate is exchanged further still into the fast region ($k_{ex} \gg \Delta\omega$) the CEST effect at the chemical shift of spin S is completely obscured and instead a slight asymmetric broadening of the water saturation is observed. Consequently, the CEST experiment is most well suited to studying spins in slow exchange, though the effect will be enhanced when more exchange events can occur within this time frame whilst remaining in the slow exchange limit. A further corollary is that any factors that affect the exchange rate will alter the CEST profile, this may include: identity of labile proton (-OH, -COOH, -NH or -NH_2), pH, temperature

and solvent accessibility of labile protons.

Fig 3.5**B** demonstrates how the size of the CEST effect increases linearly with the concentration of exchanging protons as expected (note this is in the slow exchange regime). While the concentration of [S] may appear to be relatively high it is important to note that this refers to the concentration of exchanging protons rather than molecules. If we consider a glucose molecule for example, there are five labile -OH protons per molecule, and in a protein molecule there will potentially be hundreds of exchangeable protons per molecule. This figure also highlights a key feature of the CEST experiment: the ability to amplify signal from a lowly concentrated species by 'reading out' its signal through a highly concentrated species with which it exchanges. For example, the highest concentration of [S] considered in 3.5**B**, 550 mM, is still only 0.5% of the population of [I]. Nevertheless, saturation transfer leads to a decrease in water signal of 40% representing an effective 80 fold amplification of signal.

In Fig 3.5**C** we study how the static magnetic field strength, B_0 can also significantly alter the form of the CEST spectrum(*118*). The changes observed are closely related to the effects chemical exchange. Since $\Delta\omega$ scales with the field strength, going to higher field strengths will push exchange towards the slow limit that is generally more favourable for measuring CEST dips clearly, particularly for those close to the water saturation as seen in the figure. Furthermore, at a fixed ω_1 the bandwidth of a saturation pulse will cover more of the spectrum at lower field strengths making it more difficult to delineate CEST effects between closely spaced residues. On the other hand, at higher field strengths the pulses will require higher power depositions resulting in potential sample heating effects.

Finally, we use the simulations to consider the effect of the B_1 field strength ($\nu_1 = \frac{\gamma B_1}{2\pi}$) on the form of the CEST profile. In general, larger CEST dips are observed at higher B_1 fields, however, as shown in Fig 3.5**D**, increasing the field strength to very high values is undesirable. Firstly, as mentioned previously, there will be hardware limitations on the strength of field that can be applied and strong pulses may result in undesirable sample heating effects. Furthermore, as the B_1

field is increased, the bandwidth of the saturation pulse is also increased causing distinct CEST dips between residues with similar chemical shifts to be obscured.

The results of this subsection can be summarised briefly as follows:

1. Increasing k_{ex} while remaining in the slow exchange regime increases CEST effects. However CEST effects fall as exchange continues to increase into the intermediate and fast regimes.

2. The size of the CEST effect increases linearly with the concentration of the exchanging solute pool.

3. At higher B_0 field strengths it becomes easier to delineate between CEST effects from species that resonante at similar frequencies.

4. CEST effects become larger at higher saturation pulse frequencies ν_1 but again this also causes it to become more difficult to differentiate between CEST effects from spins with similar resonance frequencies.

3.2.3 Direct and Relayed NOE Effects

Exchange mediated saturation transfer is not the only mechanism of saturation transfer active in the CEST experiment. This is made clear by the presence of CEST dips when the saturation pulse is applied upfield of the water resonance. Aliphatic protons, for example -CH, -CH_2 and -CH_3 groups that are unable to exchange with water resonate here. In this case the saturation transfer must involve an NOE (a through space dipolar transfer of magnetisation) either directly to the solvent spins or an intramolecular transfer to a labile proton that subsequently exchanges with the water. These two effects are referred to as direct and relayed NOE effects respectively.

NOE effects are fundamentally spin relaxation effects: thus saturation transfer by this mechanism is intimately linked with molecular properties, particularly the rotational correlation time, τ_c. The cross-relaxation rate, σ_{IS} between two spins I and S is given by(*119*):

$$\sigma_{IS} = d_{IS}^2\left(\frac{6}{10}J(\omega_I + \omega_S) - \frac{1}{10}J(\omega_I - \omega_S)\right) \tag{3.6}$$

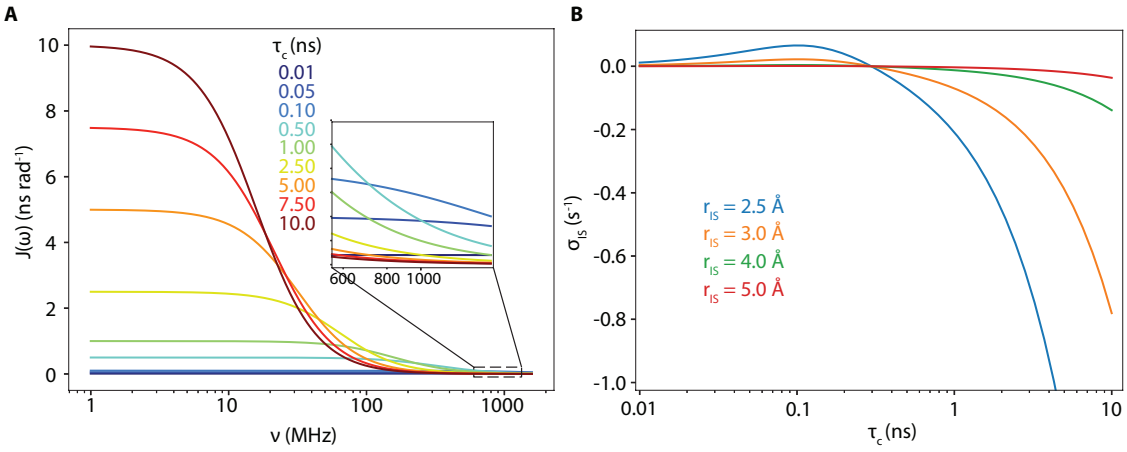

Figure 3.6: **A** Plots of $J(\omega)$ at different rotational correlation times (τ_c). For small molecules we find $J(0) \approx J(2\omega)$, whereas for large molecules the spectral density function decays rapidly and only the $J(0)$ term is important. **B** Variation of cross-relaxation rate with τ_c and at different internuclear distances. For small molecules the NOE is positive and gradually reaches a maximum value. As τ_c continues to increase however its value becomes increasingly negative and in the macromolecular limit $\sigma_{IS} \propto -\tau_c$. Intramolecular NOE transfers are enhanced significantly as the distance between the spins is lowered due to its $\frac{1}{r^6}$ dependence.

In this equation $d = \frac{\mu_0 \hbar \gamma_I \gamma_S}{4\pi r^3}$ where μ_0 is the vacuum permittivity, γ is the gyromagetic ratio and r is the distance between the two spins undergoing cross-relaxation. In this chapter we are interested in homonuclear NOEs between two protons so $\gamma_I = \gamma_S$, $\omega_I + \omega_S \approx 2\omega$ and $\omega_I - \omega_S \approx 0$. The $J(\omega)$ term refers to the spectral density function. For intramolecular NOE effects, the rotational spectral density function, $J(\omega)$, is the most important aspect of motion. For a rigid isotropically tumbling molecule this is given by(*24*):

$$J(\omega) = \frac{\tau_c}{1 + (\omega\tau_c)^2} \tag{3.7}$$

The dependencies of σ_{IS} on different parameters gives rise to several interesting properties, as illustrated in Fig 3.6. Firstly, σ_{IS} is a short ranged interaction (due to its $\frac{1}{r^6}$ dependence) and its effect drops rapidly with increasing separation between spins. Secondly, J(ω) depends strongly on the rotational correlation time (the average time for the molecule to rotate through an angle of one radian) which in turn depends on molecular size. So for small molecules with rotational correlation times on the order of ps, $J(0) \approx J(2\omega)$ and the sign of σ_{IS} is positive. For

species of increasing molecular weight (approximately 1000 Da in water at 298 K) the two terms in equation 3.6 cancel out and no dipolar transfer is observed. Finally, in the macromolecular limit $J(\omega) \approx 0$ and $J(0) = \tau_c$ and consequently $\sigma_{IS} = -\frac{d_{IS}^2 \tau_c}{10}$. Therefore, as molecular weight increases intramolecular NOE transfer to labile protons can occur more efficiently and subsequently transfer onto the solvent resulting in increased CEST effects. Since the intramolecular NOE is a short range effect, the magnetisation may also transfer between several non-labile protons before arriving at a labile proton (referred to as spin diffusion).

In the CEST MRI literature NOE mediated effects are generally split into two categories(*116*, *120*): relayed NOE effects involving soluble proteins where the saturation transfer is relayed via labile protons as described above and so called magnetisation transfer effects(*121*) (MT) involving semi-solid macromolecules where the saturation is transferred through both relayed NOEs and direct saturation transfer between the molecules and bound water molecules(*122*). However, more recent studies have suggested that cross-relaxation effects between a macromolecule and solvent can occur over longer ranges, modulated by the translational spectral density function(*123*, *124*), and are not limited to occurring with directly bound solvent protons as had been assumed(*125*). Nonetheless, in the context of CEST experiments, direct NOE transfers are likely to transfer to water much more slowly than relayed effects except when a water molecule has a particularly high residence time at a protein. In the latter case the spectral density function is adjusted to(*123*):

$$J(\omega) = \frac{\tau_B}{1 + (\omega \tau_B)^2} \tag{3.8}$$

where

$$\frac{1}{\tau_B} = \frac{1}{\tau_c} + \frac{1}{\tau_R} \tag{3.9}$$

In this equation τ_B refers to the effective correlation time of the bound complex and τ_R is the mean residence time of the bound water molecule. The spectral density function here must be summed over all specifically bound spins.

The two pool model used to simulate CEST effects in the previous subsection can be straightforwardly extended to a three pool model that additionally includes

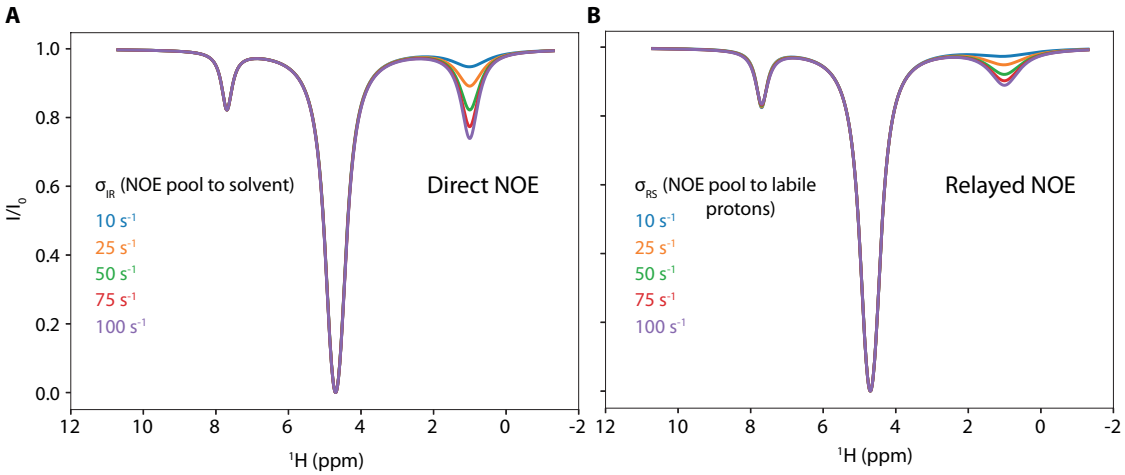

Figure 3.7: Simulated CEST spectra involving three distinct proton pools. In all simulations the following parameters are used: $R_1^I = 0.5\ s^{-1}$, $R_2^I = 10.0\ s^{-1}$, $R_1^S = 1.0\ s^{-1}$, $R_2^S = 200\ s^{-1}$, $R_1^R = 1.0\ s^{-1}$, $R_2^R = 500\ s^{-1}$, [I]= 111 M, [S]= 220 mM, [R]= 440 mM, $\nu_1 = 50$ Hz and the saturation pulse is applied for 2 s. The labile protons spins (S) are located 3.0 ppm downfield of the water resonance, while the NOE pool of spins (R) are 3.7 ppm upfield of it. The exchange between the solvent and labile proton spins $k_{ex}^{IS} = 100\ s^{-1}$ in all cases. **A** and **B** compare direct and relayed saturation transfer: for **A** σ_{IR} characterises saturation transfer between the NOE (R) spins and water while for **B** σ_{RS} characterises intramolecular saturation transfer which is then transferred to the water by normal chemical exchange.

NOE mediated transfer. In line with the discussion above, this can be done in two ways: firstly the so called 'NOE' pool (referred to as 'R' spins below) can transfer saturation directly with the water pool, simulating transfer between long residence time water molecules and proteins (this approach usually used in simulations of CEST experiments(*126*)). Or, we get intra-molecular saturation transfer between non-labile and labile proton pools after which the saturation can be transferred to the water pool by exchange, called a relayed NOE effect. Both of these situations are illustrated in Fig 3.7.

From Fig 3.7 we see the unsurprising result that increasing the direct cross-relaxation rate has a greater effect on the saturation transfer than the intramolecular cross-relaxation rate. However, it is important to note that in practice realising very high direct cross-relaxation rates will be difficult as this will require water protons to have very long residence times with the protein. On the other hand, high intra-molecular cross-relaxation rates are plausible in the context of very

large protein complexes, especially in viscous *in vivo* environments. Importantly however, both of these effects will scale with the rotational correlation time of the protein molecule providing a framework with which to interpret changes in the NOE signal observed in CEST profiles.

Evidence from the literature shows that relayed NOE effects are significantly more important in determining the size of the NOE CEST dip than direct NOE effects(*127*). Therefore, when considering changes in the NOE CEST it is important to take account of changes in the exchange rates of labile protons also.

3.3 Results

In this section we systematically vary system parameters on model systems to experimentally probe how these factors differentially affect the CEST profile. In particular, we are interested in how these different factors can be used to aid the interpretation of CEST signals *in vivo* and also suggest some novel uses of this experiment.

The model systems used in this section are composed of a high concentration BSA sample mixed with different amounts of glycerol to alter the solvent viscosity and pH adjusted to desired levels using aqueous hydrochloric acid or sodium hydroxide. The protein concentration employed in the samples (90 $mg\ ml^{-1}$) was chosen as a reasonable estimation of *in vivo* protein concentration and BSA has previously been employed as a representative protein in phantom samples (and also benefits from being readily available)(*128*). The model systems employed here still represent a gross simplification of any *in vivo* system, however, their advantage is that their properties can be tuned precisely to probe specific effects on the CEST signal.

In each sample we consider the effect of changing composition on the two main minor CEST dips in the spectra: the NOE dip that we record at 1 ppm and the amine/amide dip that we record at 7.6 ppm (aromatic protons also resonate at this chemical shift but contribute less to the CEST dip). CEST dips at amine/amide protons, commonly referred to as APT (amide proton transfer), have attracted particular attention in the CEST MRI literature due to their potential ability to

measure pH non-invasively *in vivo*(*107*, *115*). In all cases, the CEST effect is defined as $(1 - \frac{I}{I_0})$ at the position of the dip.

3.3.1 Effect of Viscosity on CEST profiles

In this subsection we discuss the effect of varying viscosity on the CEST spectra. Increased solvent viscosity will result in reduced translational and rotational diffusion coefficients and therefore longer rotational correlation times. An increased solvent viscosity can be used as a simple way of mimicking the crowded, heterogeneous solvent environment observed in-vivo. Moreover, changes in the apparent diffusion coefficient *in vivo* can be linked with specific pathologies, though it is not necessarily clear whether the change in apparent diffusion coefficient (ADC) is associated with changes in underlying viscosity, tissue morphology causing restriction effects or transmembrane permeability (all discussed in Chapter 2). For example, with reference to the stroke patient data shown earlier (Fig 3.3), differentially affected areas of tissue can be identified on the basis of their ADC values(*129*, *130*). Fully rationalising the CEST data in these systems may help build up a consistent physical picture of the processes underlying the observed changes in signal.

CEST profiles are recorded on samples with varying glycerol concentrations the results of which are shown in Fig 3.8. For each experiment the water diffusion coefficient is also measured and is written in units $\mu m^2\ s^{-1}$ above the bars in Fig 3.8**B** and 3.8**C**. The concentration of BSA in each sample (90 mg ml^{-1}) and the pH (7.0) is kept constant. Full experimental details are given in section 3.5. Control experiments are also performed on just glycerol and water and these indicate that high glycerol concentration has very minimal impact on the form of the CEST profile at 1 ppm and 7.6 ppm and its main impact is to cause an asymmetric broadening of the major water dip due to chemical exchange with the glycerol alcohol groups, which resonate downfield of water. It is important to note that using glycerol that has its own labile protons as a viscogen introduces extra exchange events between the glycerol and water and glycerol and protein, affecting the results.

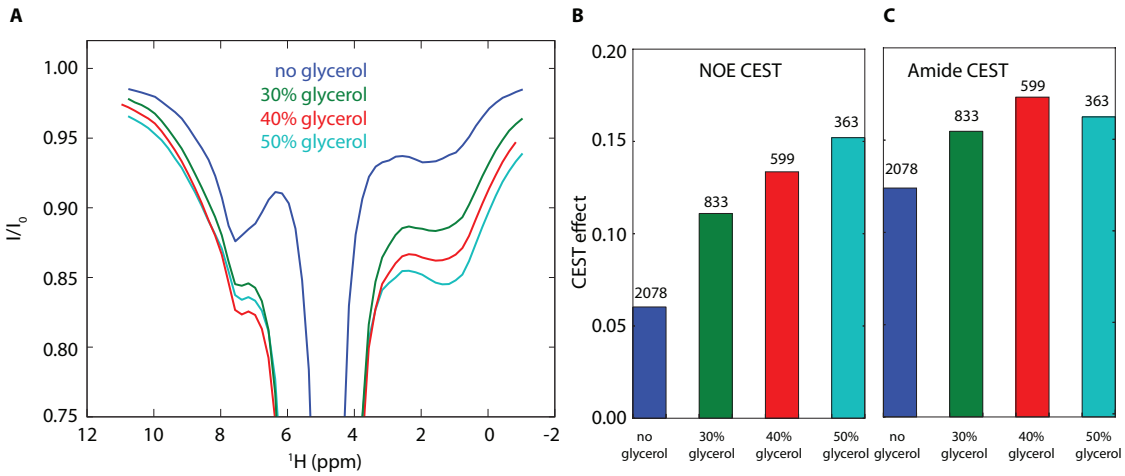

Figure 3.8: Effect of varying solvent viscosity on CEST profile of BSA models. **A** CEST profiles recorded on BSA samples with different percentages of glycerol. In all cases the saturation pulse is applied for 2 s with $\nu_1 = 50$ Hz. **B** NOE CEST effect and **C** Amide CEST effect as a function of glycerol percentage. These values are derived from the CEST spectra in **A** by calculating $(1 - \frac{I}{I_0})$ at the saturation frequency closest to 1.0 and 7.6 ppm respectively. The values above the bars in **B** and **C** are the diffusion coefficients of water (given in units of $\mu m^2\ s^{-1}$) in the sample measured using a stimulated echo diffusion experiment.

To confirm the findings here, the experiment could be repeated with a molecule such as polyethylene glycol (PEG) that has no (or far fewer) exchangeable protons.

It is clear that as the viscosity of the sample is increased the NOE CEST effect (associated with aliphatic protons) also increases. This trend can be rationalised on the basis of the discussion in subsection 3.2.3: as viscosity is increased the tumbling rate of the protein molecules in solution also decreases. Consequently intramolecular cross-relaxation becomes more efficient resulting in the observation of greater relayed-NOE saturation transfer. Moreover, as the viscosity is increased we would also expect solvent residence times on the protein to increase resulting in increased direct NOE transfers to the solvent.

The trend seen in the amide CEST effects is more complex: initially we observe an increase with the CEST effect with increasing viscosity but the effect drops again at the highest tested glycerol concentration. We would expect that the rate of chemical exchange reduces with increasing solvent viscosity. Therefore, the slowest rate of chemical exchange between the water and protein will be associated with the 50% glycerol sample. When the glycerol concentration is reduced to 40% the

exchange rate increases but remains in the slow exchange regime resulting in a larger CEST effect. However, as the glycerol concentration continues to fall, the exchange rate likely moves from a slow towards an intermediate exchange regime and therefore the size of the CEST effect falls again as simulated in Fig 3.5**A**.

Notably, the response to changing viscosity is differential for the two main minor dips in the CEST profile highlighting how they report on different underlying physical processes. We see that up to 40% glycerol concentration the as the NOE CEST effect increases the amide CEST effect also increases, with the latter promoting higher relayed NOE effects. However, when the glycerol concentration increases to 50% the NOE CEST continues to increse while the amide CEST effect falls. This shows that in this case the increased intra-molecular saturation transfer effects outweigh the effects of lowered amide exchange rates.

3.3.2 Effect of Digestion on CEST Profiles

Having established the effect of increasing viscosity on the CEST profiles of our BSA model system, we now seek to study the reverse process. We achieve this by using an enzyme to digest the protein: this means that following digestion the total number of protons in the sample will remain constant as will the functional groups they are associated with. However, the large protein molecules will be broken down into smaller polypeptide chains with significantly higher tumbling rates in solution.

Trypsin is employed as the enzyme to digest the protein: it achieves this by cleaving the C-terminal side of lysine and arginine amino acid residues. A key advantage of this enzyme for the study here is that its activity is optimised at the pH of the samples (7.0)(*131*).

The results of the trypsin digestion on the BSA model system are shown in Fig 3.9. The digestion has a striking effect on a simple 1D ^{1}H spectra of the sample in 3.9**A**: from an NMR perspective BSA is a large soluble protein with a molecular weight of 66.5 kDa, and thus its resonances will have relatively broad linewidths. Upon digestion, the protein is split into smaller polypeptide chains with sharper peaks giving rise to greater intensity in the spectrum (in a simple pulse acquire

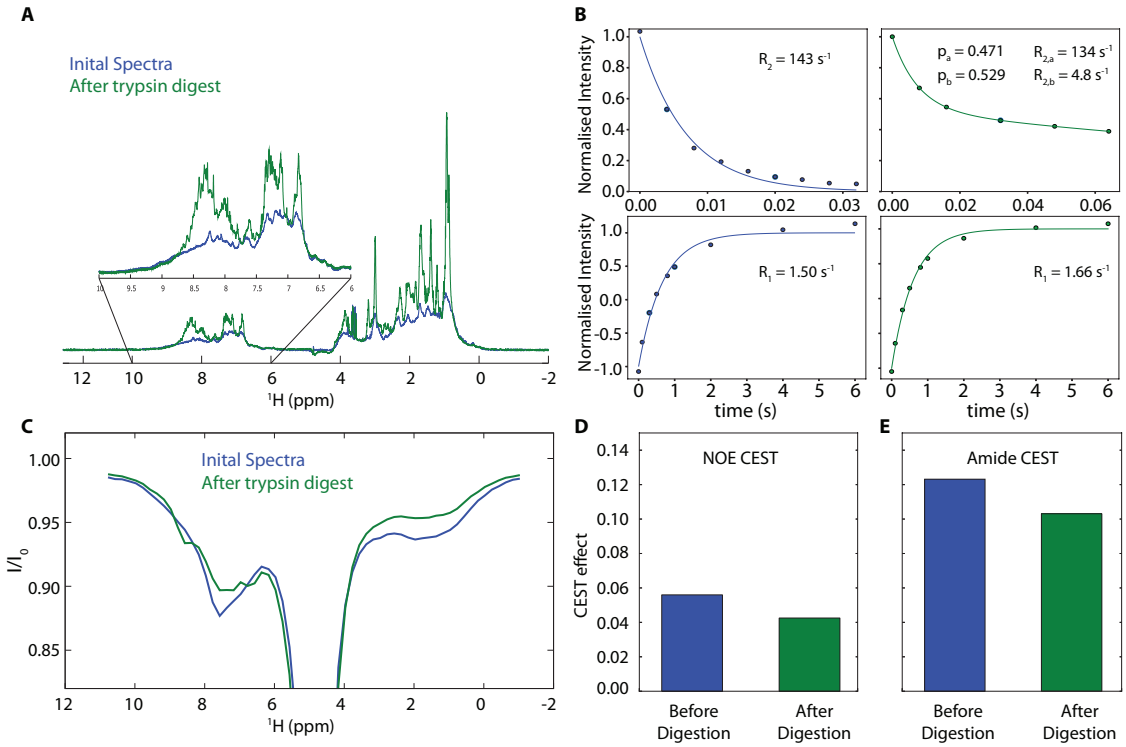

Figure 3.9: **A** Effect of digestion on the ^{1}H watergate spectrum of BSA. The digested protein will have lower transverse relaxation rates leading to sharper signals and increased intensity in the spectrum. **B** Transverse (R_2) and longitudinal (R_1) relaxation rates of the BSA sample measured before and after digestion. The relaxation rates are taken by integrating over the spectral region 0.3-2.0 ppm. While the longitudinal relaxation rate is relatively unaffected by the digestion, a significant change is observed in the transverse relaxation rate requiring a biexponential function to account for the data that suggests around 50% of the protein is digested. **C** Comparison of BSA CEST profiles before and after digestion. After digestion a significant decrease in both the amide and NOE CEST signals is observed. **D** and **E** Bar charts quantifying the minor NOE and amide CEST dips measured at 1.0 and 7.6 ppm respectively. In all cases the saturation pulse is applied for 2 s with $\nu_1 = 50$ Hz.

experiment the integrated intensity in the two spectra would be the same. However as we use a finite time watergate suppresssion pulse sequence faster relaxation results in lower signal intensity for the intact protein).

This point is reinforced in Fig 3.9**B** where the relaxation rates of the protein are measured. Transverse relaxation rates are measured using a CPMG-PROJECT sequence that refocusses homonuclear J-couplings(*132*) while the longitudinal relaxation rates are measured using an inversion recovery experiment. The measured relaxation rates are an aggregate over all resonances between 0.3 and 2.0 ppm. A significant change in R_2 is observed following digestion: the initial BSA sample

can be reasonably well fit to a monoexponential decay with a relaxation rate of 143 s^{-1}. However, following the addition of trypsin the relaxation data requires a biexponential function to be adequately fit. This is comprised of fast relaxing component with a relaxation rate very similar to the initial BSA sample (134 s^{-1}) and a much slower relaxing component (4.8 s^{-1}) in approximately equal fractions. The trypsin digestion results in a complex mixture of peptides with a wide distribution of sizes, that could be more accurately characterised by mass spectrometry. The transverse relaxation rate is particularly sensitive to the correlation times of molecules and in the macromolecular limit is directly proportional to it(*4*). On the basis of the relaxation measurements here, we can conclude that there are still some very large peptides in solution with a transverse relaxation rate approaching that of the undigested protein as well as some much smaller fragments with much slower transverse relaxation.

The CEST results for the sample are summarised in Fig 3.9**C-E**. As expected, following digestion with a significant proportion of molecules tumbling at a much faster rate, the NOE CEST effect falls. This corresponds to slower intra-molecular cross-relaxation making relayed NOE saturation transfer effects less efficient. The remaining NOE CEST effect is likely mainly driven by the large peptides that remain in solution, as suggested by the measured transverse relaxation rates. However, as the drop in NOE CEST is much less than 50%, some of the smaller peptides must also contribute to the observed effect (and therefore must be sufficiently large to give a negative NOE effect).

A more surprising result perhaps is that the CEST effect associated with exchangeable amine and amide protons at 7.6 ppm also decreases upon digestion. A reasonable expectation would be that as the structure of the protein is broken down by the enzyme more exchangeable groups become solvent exposed and therefore the CEST effect should increase. While this likely is the case, an additional important factor is the exchange rate of the process. The more exposed labile protons must undergo faster chemical exchange with the solvent resulting in a reduced observed CEST effect. Interestingly, studying the CEST effect at approximately 8.3 ppm,

a slight increase in the CEST effect is observed after digestion, because these amide resonances move from slow to slow-intermediate exchange with the solvent as they become more solvent exposed.

The experiment in this section is also repeated with a sample in 30% glycerol to mimic viscous in-cell conditions and the same trends in the CEST signal are observed.

3.3.3 Effect of pH on CEST profiles

Since the earliest water CEST experiments were performed considerable interest has been placed in its ability to measure pH(*112*). The potential clinical benefits of measuring pH are significant: for example, tumours in cancer display a drop in extracellular pH (acidosis) and concomitant rise in intra-cellular pH(*133*, *134*). Similarly, during stroke oxygen starved tissue switch from aerobic respiration to anaerobic glycolysis leading to a build up lactic acid and consequent intra-cellular acidification(*135*). However, the use of pH as a diagnostic parameter in the clinic is limited by the lack of readily available, non-invasive tools to measure it. Established methods such as ^{31}P magnetic resonance spectroscopy suffer from poor sensitivity and resolution as well as more specialised hardware requirements(*115*).

In this subsection we study the effect of changing pH on our model BSA system. In particular we are interested in relatively small pH changes that would be of interest in a clinical setting, such as following a stroke (Fig 3.3).

The result of changing pH on the observed CEST is shown in Fig 3.10. From Fig 3.10**A** and **C** there is a clear response in the CEST dips downfield of the water resonance with increased acidification of two pH units. In particular, the amine CEST dip shows a particularly marked increase with decreasing pH. Since this exchange event is known to be base catalysed(*136*), this corresponds to the exchange rate slowing and moving towards the optimal regime for the CEST experiment. While no specific CEST dip is observed at 6 ppm a significant change in the CEST spectrum is also observed here upon acidification of the system. While hydrogen exchange rates in different amino acid residues are known to have differential responses with changing pH (*137*), the change seen in the CEST spectra likely

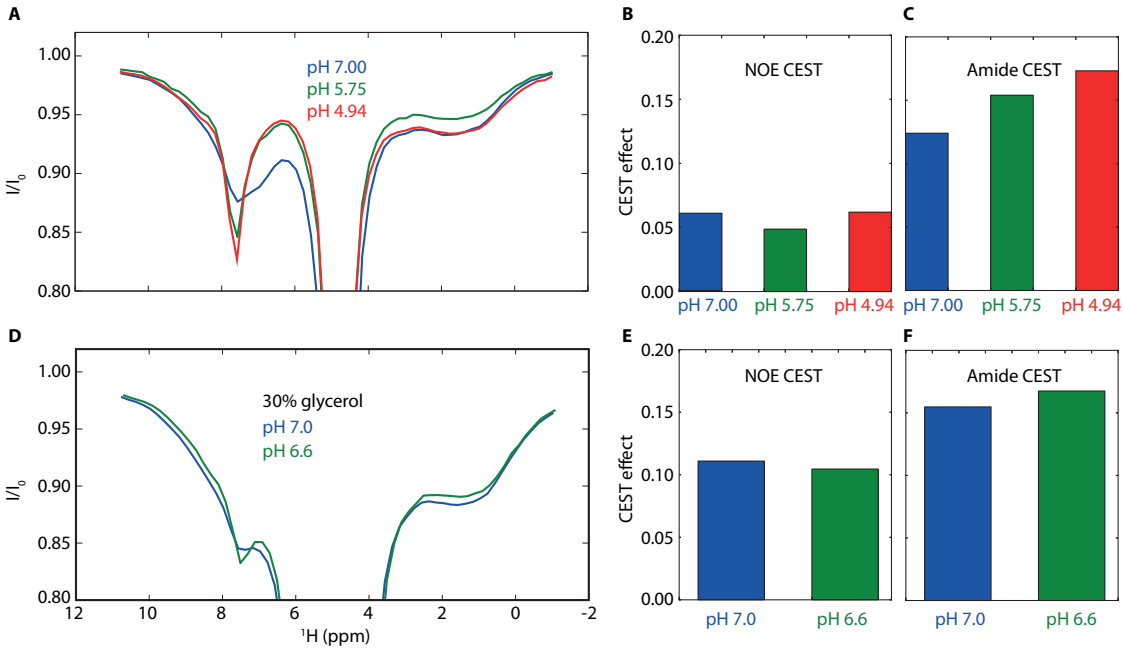

Figure 3.10: **A-C** The effect of pH on the CEST profile of BSA model systems. As the pH is increasingly acidified the amide proton CEST effect clearly increases. On the other hand, the NOE CEST effect shows a more complex response, first rising and then falling with decreasing pH. **D-F** Effect of pH on model BSA systems in the presence of 30% glycerol to mimic physiological viscosities. In this case the effect of a much smaller acidification is investigated, reflecting a typical intracellular acidification that would be expected following a stoke. In all cases the saturation pulse is applied for 2 s with $\nu_1 = 50$ Hz.

represents an aggregate increase in exchange rate for labile protons in alcohol groups at lower pH (as -OH protons are in the fast exchange regime we do not see a specific CEST dip but a narrowing of the major water dip). In general, alcohol groups are much less studied in the CEST literature than amine and amide groups: their exchange rates are generally much higher than amine and amide groups and their chemical shift is much closer to that of water meaning they are more likely to be in the fast exchange regime ($k_{ex} \gg \Delta\omega$) and do not give rise to distinct dips in the CEST spectra (one exception to this is the study of alcohol proton exchange in glycogen and glucose(*138*)). The response of these groups to changes in pH highlights how studying them can nonetheless be of significant utility.

The response of the NOE CEST dip in Fig 3.10**A** and **B** is more complex. Initially, the NOE CEST effect falls with solution acidification, however, upon a further fall in pH the magnitude of the CEST effect returns to its pH 7.0 level.

Notably, the effect of changing pH upfield of the major water dip (NOE side) is much smaller than downfield. We expect that the correlation time of the protein and thus the rate of NOE transfer will be minimally affected by the change of pH. Therefore, in order to rationalise the data we must recall that relayed NOE effects that dominate this signal are also dependent on the exchange rate between labile protons and the solvent. Thus between pH 7.0 and 5.75, though we see an increase in the amine CEST effect, this is outweighed by a loss in CEST effects from alcohol groups. Overall, there is therefore a loss of NOE mediated CEST as the saturation is not as efficiently transferred to the solvent via the labile protons. On the other hand, between pH 5.75 and 4.94, there is no further loss in the alcohol group CEST but the amine CEST dip becomes larger so overall the relayed NOE CEST effect becomes larger. This demonstrates the synergy that exists between the NOE and exchange mediated CEST effects.

As shown in Fig 3.10**D-F** repeating the experiments at 30% glycerol to mimic physiological viscosities and studying smaller (more disease relevant) acidifications show very similar effects to those discussed above. As expected, NOE CEST effects are significantly larger and minor CEST dips are much less sharply defined in the high viscosity conditions, but the general effects of pH on signal are retained. Therefore, in principle, and as noted previously(*108*, *112*), the CEST signal of proteins is able to act as an non-invasive probe of pH. However, the change in CEST signal following pH changes is relatively small and given the huge range of factors that can potentially affect the form of a CEST spectrum, some of which are discussed above, great care must be taken when inferring changes in pH from CEST signals. This is discussed further in section 3.4.

3.3.4 Using CEST to Monitor Protein Aggregation

In subsection 3.3.1 we described how the CEST signal of proteins, especially the NOE CEST, can be affected by the viscosity of the solution due to the effect on the protein's tumbling time. A more direct way in which the tumbling time of a protein would be affected is through oligmerisation. In this subsection we

take advantage of this fact to demonstrate how the CEST effect can be used to follow the process of aggregation.

Here we turn our attention to the protein α-synuclein. α-synuclein is an intrinsically disordered (IDP) pre-synaptic protein composed of 140 amino acid residues with a poorly understood function(*139*, *140*). This protein is of particular interest and significance due to the finding that so called Lewy bodies (abnormal aggregates of proteins that develop in the neurons of sufferers of a range of neurological conditions) are mainly composed of α-synuclein in patients of Parkinson's disease(*141*). Moreover, specific mutations in α-synuclein have been linked to early-onset familial Parkinson's disease(*142*). As a result, α-synuclein and in particular the mechanism and inhibition of its aggregation have been the subject of many recent studies, using a variety of biophysical techniques(*49*, *143–145*).

Within our group methods have been developed to attain reproducible aggregation kinetics of α-synuclein by seeding the process with a small amount of pre-formed α-synuclein aggregates. Methods have also been developed to determine the mechanism of aggregation by following the depletion of signal in a series of 1D NMR experiments as a function of time and at different concentrations(*146*). Below we follow the aggregation of α-synuclein at a single concentration using both 1D NMR as well as the water CEST experiment.

The results of this experiment are shown in Fig 3.11. In this experiment we use ^{15}N labelled α-synuclein and record 1D spectra using ^{15}N edited HSQC experiments that report specifically on protons attached to ^{15}N atoms (namely amine and amide groups) in the protein, shown in 3.11**A**. As α-synuclein is an IDP the chemical shifts of all of these resonances fall within a tight range. As the aggregation of the protein proceeds we see that there is a significant loss of intensity in the 1D spectrum. The reason for this is that the high molecular weight oligomeric species formed during the aggregation process have very large relaxation rates and therefore very broad linewidths, becoming effectively invisible in the spectra. The 1D experiment consequently reports on the remaining concentration of monomeric protein. This is confirmed by performing a full 2D HSQC experiment after aggregation that is

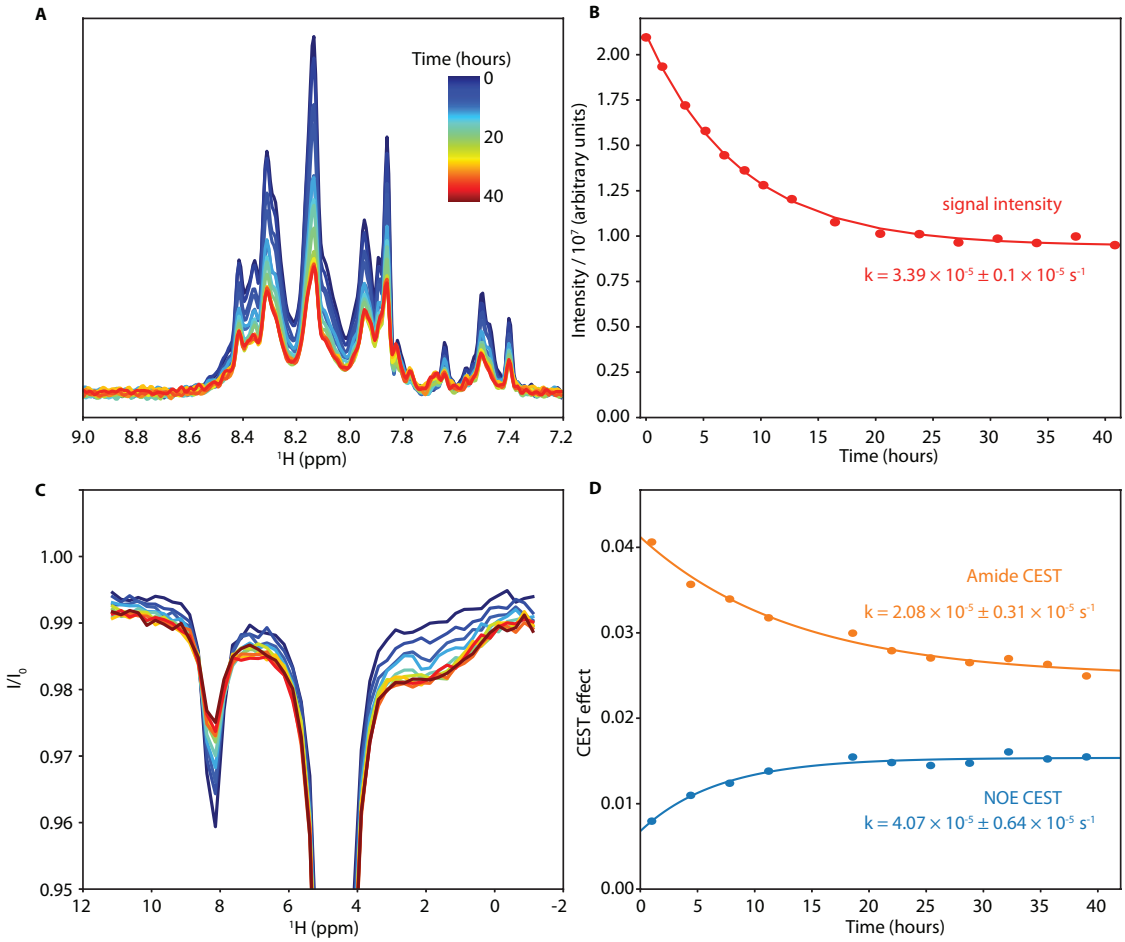

Figure 3.11: Monitoring aggregation of ^{15}N labelled α-synuclein by 1D ^{15}N edited NMR and CEST at pH 7.4. **A** In this experiment we can follow the aggregation of ^{15}N-labelled α-synuclein through the depletion of signal in a series of 1D ^{15}N edited HSQC experiments. **B** Integrated signal shown in **A** as a function of time. The region of integration is 7.2-9.0 ppm. **C** Results of CEST experiments run on the same sample as a function of time. We see changes in both the NOE and amide parts of the spectrum as the sample aggregates. In all cases the saturation pulse is applied for 2 s with $\nu_1 = 50\ Hz$. **D** NOE and amide CEST effects as a function of time. These are calculated from **C** by taking $1 - \frac{I}{I_0}$ at the saturation frequencies closest to 1.0 and 8.15 ppm respectively. In **B** and **D** the time courses are fitted to first order kinetics and the fitted rate constants are given in the figures (errors in fitted values are from the covariance matrix).

indistinguishable from the monomeric protein except for its lower intensity. An important feature of Fig 3.11**A** and **B** is that after approximately 20 hours as the peak intensity levels off, a significant amount of monomeric protein is still present demonstrating that the aggregation process is reversible.

The results of the CEST experiments performed on the same sample during aggregation are shown in Fig 3.11**C** and **D**. It is interested to compare the CEST profile of α-synuclein, an unstructured protein, at time zero, to a larger, structured protein like BSA (for example Fig 3.1**C**). Firstly, while BSA shows minor CEST dips both up and downfield of the water saturation, in α-synuclein a significant CEST effects can only be observed with the amide protons initially, with little evidence of NOE CEST. Each α-synuclein monomer has approximately a quarter the molecular weight of a BSA molecule giving it a much faster tumbling rate than BSA and making cross-relaxation rates slower. Furthermore, as an IDP with a high degree of internal flexibility the effective correlation time of atoms in α-synuclein will be shorter still and cross-relaxation rates slower.

Also, while BSA shows a fairly broad minor CEST dip downfield of the water saturation, with its minimum at 7.6 ppm, in α-synuclein the downfield minor CEST dip is much sharper with its minimum at 8.15 ppm. This feature is also a consequence of the disordered nature of α-synuclein. The sharpness of this CEST dip reflects the fact that the labile amine and amide groups in α-synuclein are restricted to a small chemical shift range. Nearly all of the labile protons will also be solvent accessible leading to a higher average exchange rate with the solvent.

Upon aggregation, two main differences can be observed in the CEST profile of the protein (3.11**C** and **D**). Firstly, there is a significant increase in the NOE CEST as a function of time. This is similar to the increase in NOE CEST that is observed upon increasing solvent viscosity (Fig 3.8). While cross-relaxation effects are very inefficient in monomeric α-synuclein, as it oligomerises forming larger and larger species this becomes more efficient leading to the emergence of an NOE CEST effect. In this sense, the NOE CEST signal therefore provides complementary information to the 1D NMR spectra: while the latter reports on

the residual monomeric concentration, the former is sensitive to the formation of high molecular-weight species. The potential therefore exists to use the CEST experiment in conjunction with 1D spectra to provide additional information on the behaviour of larger oligomeric species that are invisible in the 1D spectra.

The other key difference in the CEST profiles upon aggregation is that the minor amide CEST dip gets smaller. Whereas in monomeric α-synuclein all the amide groups are likely to be solvent exposed, when aggregated some of these groups may be less solvent exposed reducing the amide CEST effect. However, another effect here will be that the large oligomeric species will have very large transverse relaxation rates and consequently their CEST effects will be spread over a large frequency range making the minor dip at 8.15 ppm smaller.

We have also fitted the loss of water intensity and amide and NOE CEST effects to first order kinetics in Fig (3.11). We find that the fitted rate constants for the monomeric signal intensity and NOE CEST are not significantly different (p-value = 0.32). However, the difference in rates between the fall in amide CEST and fall in monomeric signal intensity is significantly different (p-value = 0.004), highlighting the ability of the water CEST experiment to provide additional information on the aggregation process. To make this data more useful however, changes amide and NOE CEST effects with concentrations of individual oligomeric species must be quantified.

As future work, the NMR data acquired here should be compared with dynamic light scattering (the scattered light can be used to quantify the size and population of aggregates in solution) and Thioflavin T (fluoresces upon binding amyloid) assays that are more commonly used to follow aggregation processes (*146*).

3.4 Discussion

The CEST experiment is one of the simplest NMR experiments conceivable, but is nonetheless sensitive to and therefore a potential reporter of a wide range of parameters. A key advantage of the water CEST experiment presented here, is that it takes advantage of only endogenous signals and therefore does not require

any additional, potentially toxic contrast agents to be added. Furthermore, the effective amplification of signal through multiple exchange events allows the indirect detection of species using CEST with MRI experiments where sensitivity and resolution are lower.

The sensitivity of the CEST signal to a range of parameters is also one of its main weaknesses: while in model systems, as above, it is possible to decouple different contributions to the CEST signal, in an *in-vivo* situation confidently ascribing a change in the CEST signal to a particular factor is very difficult. A significant amount of the CEST literature has explored the potential of the amide proton CEST signal to measure pH, even referring to this experiment as pH-weighted imaging (*107*, *126*). However, as described above, only modest changes in the amide CEST signals occur for the small changes in pH that are realisable *in-vivo*. Given that a range of other factors can cause similar magnitude changes in the CEST signal for example a change in aggregate protein oligomerisation, local viscosity or in the labile proton concentration potential changes in these factors must also be accounted for. In the context of the stroke patient data presented at the start of this chapter (Fig 3.3) it is clear that no single factor explored here can explain the variation in CEST effect changes observed, and while pH changes might play a role it is clearly not the most important factor in determining the signal.

A number of different approaches could be used to overcome this issue with the CEST experiment. Firstly, it could be run alongside a diffusion experiment (either a simple ADC measurement or ideally a more detailed analysis analogous to that discussed in Chapter 2) that would allow the CEST results to be viewed in the context of diffusive behaviour. In the next chapter (4) we describe a method that combines the diffusion and CEST experiments and can separate CEST contributions from different compartments. Also, rather than focussing solely on the amide region of the spectrum looking at other regions of the CEST spectrum, including OH groups and the NOE CEST, may also allow CEST results to be interpreted more accurately. Moving towards higher field strengths will aid this as it allows for better delineation between different labile proton groups.

The lack of specificity in the CEST experiment is one reason for the development of paraCEST reagents(*147–149*). These reagents are complexes with a central paramagnetic Lanthanide ion that can induce large chemical shifts changes ($\Delta\omega$) in reversibly binding water molecules. The resonance frequency of the bound water is typically far away from the water resonance (shifted on the order of 10s of ppm) and the molecular exchange rate of water with the complex is typically faster than labile proton exchange. However, as $\Delta\omega$ is very large the system remains in the slow exchange regime allowing the development of significant minor CEST dips and there will be no overlap with endogenous species. Since the molecular exchange rate is sensitive to pH the CEST dips can be used as a measure of pH. However, similar to the endogenous CEST experiments described, the paraCEST effect will also be sensitive to factors such as local viscosity that must be accounted for. Furthermore, since the paraCEST reagent is exogenous care must be taken to ensure that they are non-toxic and an understanding of whether the molecules localise to particular tissue is necessary before they can be used effectively in-vivo.

Another feature of the results in this chapter is the potential utility of the NOE CEST in characterising system properties. Most previous CEST studies have focussed on exchangeable groups downfield of the water saturation (*116*), perhaps somewhat surprisingly given the widespread usage of magnetisation contrast imaging that is analogous to CEST but with the saturation pulses are applied far off resonance(*120*). Here, we have shown how the NOE CEST effect is sensitive to protein oligomerisation, pH and viscosity changes. In particular while exchange effects are not dependent on molecular size, the NOE CEST signal requires relatively fast cross-relaxation and therefore is a protein driven signal. As a further corollary, a common way of analysing CEST data is using an asymmetry analysis where the CEST effect when the saturation pulse is applied at equal spacings either side of the water resonance is measured and the ratio is taken. While this metric has proved useful in specific clinical scenarios, it lacks any rigorous physical basis and can obscure insight into the processes that drive changes in the CEST signal(*128*).

3.5 Methods and Materials

3.5.1 Simulations of CEST Experiments

The limits in equation 3.2 are calculated using a Mathematica script by taking the exponential of the matrix in 3.1 multiplied by time and then taking the limit $t \to \infty$. The visualisations and graphs of the trajectory of the bulk magnetisation during a saturation pulse in Fig 3.4 are produced using a script written in Python.

CEST spectra for two and three pools of spins (as shown in Figs 3.5 and 3.7) were simulated using scripts written in Python. In all simulations the spectrometer frequency is set to 600 MHz and a pulse saturation time of 2 s is employed. The CEST spectra are calculated by taking the exponential of the matrices in equation 3.5 (or the appropriate 10×10 matrix including the effect of an additional spin pool for NOE CEST effects) multiplied by the saturation time. The intensity of the z-magnetisation of the major pool (this gives rise to observable signal) divided by the z-magnetisation of the major pool in the absence of the CEST pulse is then taken to give the CEST signal. For all simulated CEST spectra the signal is calculated at 1000 equally spaced saturation pulse frequencies from -0.7 to 9.7 ppm by adjusting $\Delta\Omega$ terms as appropriate. Cross relaxation rates are calculated using equation 3.6 using the spectral density function for a rigid isotropically rotating molecule (3.7).

3.5.2 NMR Experiments

All NMR experiments are acquired on a 14.1 T (600 MHz) Agilent DD2 NMR spectrometer equipped with a 5 mm z-axis gradient triple resonance room temperature probe. All experiments were performed at 298 K, with the exception of the α-synuclein aggregation (3.11) experiments that were performed at 310 K. All spectra were Fourier transformed and phased using NMRPipe(*101*).

BSA sample preparation

All BSA samples are produced from a 250 $mg\ ml^{-1}$ stock solution in 2x phosphate buffered saline (PBS). Glycerol, deionised water and D_2O are added as appropriate to make up final samples such that every one has a final concentration of 91 $mg\ ml^{-1}$

BSA (corresponds to a concentration of 1.35 mM) and 10% D_2O. The pH of all samples is measured using a pH meter and set to 7.0 (except where otherwise indicated) using small volumes of concentrated hydrochloric acid and sodium hydroxide. For the digestion experiments, the digestion of BSA is initiated following addition of 10 μl of 100 $mg\ ml^{-1}$ trypsin to the sample.

α-synuclein sample preparation

A ^{15}N-labelled α-synuclein sample was prepared according to previously published protocols(*144*). For NMR experiments 500 μM of monomeric protein was mixed with 5% pre-aggregated protein (also with an effective monomer concentration of 500 μM) to catalyse the aggregation process and 5% D_2O to provide the lock signal.

1D NMR experiments

1D ^{1}H spectra of BSA and α-synuclein samples (3.9,3.11) are acquired using a watergate pulse sequence(*18*) for water suppression.

CEST experiments

All CEST experiments are performed using a purpose-written pulse sequence. CEST experiments are performed in a saturation pulse on/off manner such that at every saturation pulse offset frequency the experimet is first performed with CEST pulse (continuous wave power) applied and then with the CEST pulse replaced with a delay of the same length. This allows the ratio of $\frac{I}{I_0}$ to be readily attained at each saturation frequency, accounting for small changes in temperature over time that can affect signal intensity. In all cases the saturation pulse is applied for 2 s and the CEST pulse has a strength of $\nu_1 = 50$ Hz. Experiments employ sixty equally spaced saturation offsets from -1.0 to 11.7 ppm in increments of 0.2 ppm.

The CEST pulse is calibrated by finding the 90° pulse time at 30 dB transmitter power (full power pulses are applied at 59 dB). The low power pulse time is converted into a B_1 field strength in Hz and subsequently scaled to the desired power level for the pulse (50 Hz) by linearly scaling the fine power. Since the water intensity measured in the CEST experiment has an exceptionally high proton concentration (110 M) a shorter pulse time is used for the final observe pulse in the experiment (2 μs corresponding to a flip angle of approximately 20°).

Diffusion experiments

Diffusion experiments are performed on samples with varying concentrations of glycerol as an internal check of the viscosity of samples. All diffusion experiments are performed using a pulsed field gradient stimulated echo pulse sequence with a single diffusion time (Δ) of 100 ms. The gradient length (δ) is set to 2 ms and 10 quadratically spaced gradient field strengths (g) between 8 and 60 $G\ cm^{-1}$ are employed. The water signal attenuation is measured at each gradient field strength and is fit using a Python script to a monoexponential function of the form $S(b) = S(0)e^{-bD}$ where $b = (\gamma\delta g)^2(\Delta - \frac{\delta}{3})$ to find the diffusion coefficient (D).

Relaxation experiments

R_1 and R_2 experiments are performed on the BSA sample before and after trypsin digestion (Fig 3.9). R_1 rates are measured using an inversion recovery experiment. Nine relaxation delays between 600 μs and 6 s are used. The protein signal between 0.3 and 2.0 ppm is integrated and fit to the equation $M(t) = M(0)(1 - 2e^{-R_1 t})$ using a Python script. Transverse relaxation rates are measured using a PROJECT-CPMG(*132*) sequence. The PROJECT element consists of: $delay - 180° - delay - 90° - delay - 180° - delay$ where the additional 90° pulse creates a so called 'perfect echo' that refocusses the effects of homonuclear scalar coupling. Before digestion nine relaxation delays up to 32 ms are used and after digestion six relaxation delays up to 64 ms are employed. Attenuations are fit to either a monoexponential ($I(t) = I(0)e^{-R_2 t}$) or biexponential function ($I(t) = I(0)(p_a e^{-R_{2,a} t} + p_b e^{-R_{2,b} t})$) to determine transverse relaxation rates.

3.5.3 Analysis of CEST profiles

Python scripts have been written to automate the analysis of CEST spectra. The code is able to read in relevant experimental parameters from Varian/Agilent *procpar* files and phased spectra are read in using the nmrglue(*102*) module and subsequently integrated allowing normalised CEST profiles (for example Figs 3.8-3.10) to be rapidly generated. In addition, scripts have been written to automate

the analysis of kinetic 1D watergate and CEST spectra (Fig 3.11). The times are read in from the experimental log files.

3.6 Acknowledgements

Expression and purification of ^{15}N α-synuclein sample was carried out in collaboration with David Dias (Department of Chemistry, University of Oxford). Pre-aggregated α-synuclein 'seeds' were provided by David Dias. Patient stroke data was provided by George Harston and James Kennedy (John Radcliffe Hospital, University of Oxford).

4

Development of DiffCEST Experiment

Abstract

In this chapter we describe the development and applications of a diffusion-weighted CEST (diffCEST) experiment. This experiment combines the diffusion and CEST experiments from the previous chapters to calculate compartment specific CEST profiles, where compartments can be differentiated by their diffusional properties. In this chapter we firstly set out the experimental basis and NMR pulse sequence used in diffCEST. We subsequently show that the experiment is successful at calculating CEST profiles associated with intra- and extra-cavity spaces in giant unilamellar vesicles (GUVs) when the profiles associated with the compartments are known a priori. Finally, the diffCEST experiment is applied to study the formation of stress granules in mammalian cells and is shown to be capable of monitoring cell stress in a non-invasive and label-free manner.

4.1 Introduction

The last two chapters in this thesis have described how diffusion and CEST methods can be used to characterise different aspects of biological systems. In this chapter we describe how these two experiments can be combined to give a diffusion-weighted CEST (diffCEST) experiment. When performed on systems that have compartments with distinct diffusional properties this experiment is capable of distinguishing between their CEST effects, whereas a normal CEST experiment provides an aggregate CEST signal from all compartments.

As extensively explored in Chapter 2 as well as previous studies(*55*, *78*, *150*), the water resonance in cellular systems can be robustly described in terms of a two pool model where we have intra- and extra-cellular signals. Provided exchange between the two environments is slow on the diffusion timescale (Chapter 2), the compartments can be distinguished on the basis of their diffusional properties, a feature that is exploited in the diffCEST experiment. An illustration of the diffCEST methodology is provided in Fig 4.1.

The ability to ascribe CEST effects to particular compartments is of significant potential benefit. Firstly, it opens up the possibility for the CEST signal to be used to identify species from particular compartments or changes in the behaviour or environment within specific compartments. Compartment specific information may also help to rationalise features and changes in CEST spectra so that they can be interpreted in a more rigorous and physically meaningful way. For example, a common assumption in the CEST MRI literature is that the amide proton transfer rate primarily originates from the intra-cellular environment (*108*). Given the high intra-cellular protein concentration within cells this is highly plausible, however, diffCEST provides an opportunity to experimentally test this.

The diffCEST experiment as presented here also retains many of the benefits associated with the ^{1}H water CEST experiment presented in the previous chapter. Firstly by using the water signal as a readout, the signal to noise ratio associated with each measurement point is extremely high meaning data can be acquired quickly. While in this chapter, CEST profiles spanning a wide ppm range are measured, if

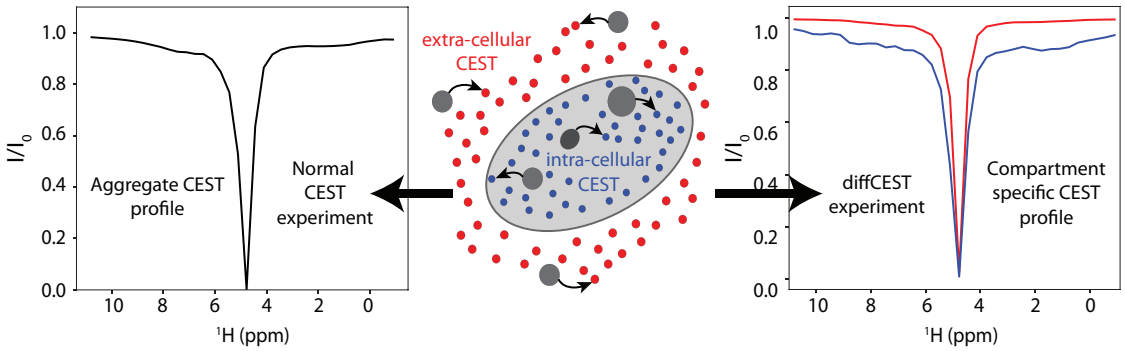

Figure 4.1: Schematic comparison of diffCEST and CEST experiments: when we have compartmentalised systems (for example cells in suspensions) we can have potentially different CEST effects associated with the different compartments. Provided these compartments can be resolved on the basis of their differential diffusion properties the diffCEST experiment is able to calculate CEST profiles associated with each compartment whereas the normal CEST experiment provides an aggregated signal over all compartments. The arrows in the central cartoon represent saturation transfer events between molecules and the water.

only a few specific saturation frequencies are of interest the experimental time can be dramatically reduced. Secondly, the method relies solely on endogenous molecules within the systems under study (effectively label free) and does not require the addition of any potentially toxic or expensive (especially if an isotopically labelled compound is needed) agents to be added to the system. Both of these features mean that the diffCEST experiment should be amenable to MRI studies, in a similar way to CEST experiments.

The pulse sequence used for the diffCEST experiment is introduced in Fig 4.2. The sequence represents a hybrid of the stimulated echo diffusion and CEST sequences used in the previous two chapters. The saturation pulse is placed in the centre of the sequence where magnetisation is held longitudinally. The water intensity is measured then for a range of CEST offset frequencies, and this process is then repeated for several gradient field strength yielding a pseudo-3D dataset. In analogy to the CEST experiment, for each gradient strength a spectrum is also recorded without the CEST saturation pulse allowing a normalised CEST spectrum to be calculated as a function of gradient field strength.

The essence of the diffCEST experiment is that at relatively low gradient field strengths relatively little signal is attenuated by the diffusion filter and therefore

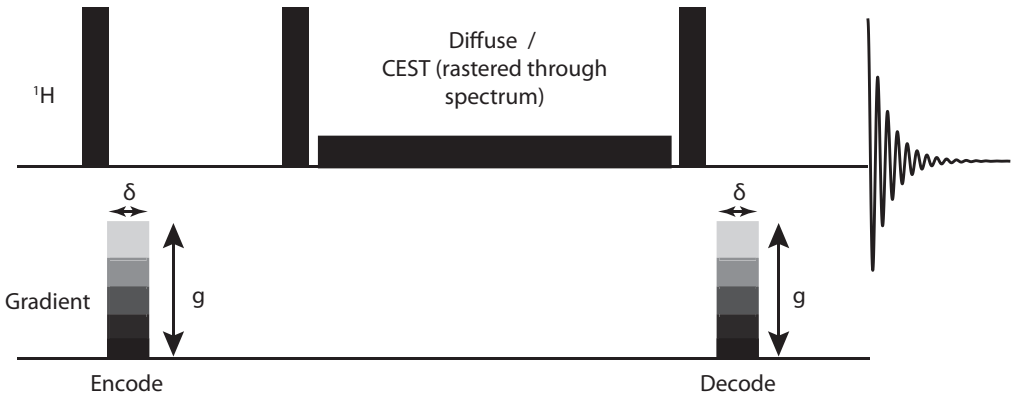

Figure 4.2: Pulse sequence used to carry out diffCEST experiment: this is a straightforward extension of the stimulated echo diffusion experiment where a low frequency saturation pulse is placed in the longitudinal storage time. During the diffCEST experiment, experimental data is recorded both as a function of CEST pulse frequency and gradient field strength yielding a pseudo-3D dataset.

observed CEST effects are from a combination of both fast and slow diffusing water. On the other hand, at high diffusion gradients the only signal left is from water diffusing in restricted or highly viscous compartments and any observed CEST effects will be from the interactions between water and molecules within this environment. In the case where the system can be described in terms of two distinct pools, as is the case with the cellular systems in this chapter, the data ted to a biexponential function allowing for the back-calculation of CEST profiles associated with the different water pools. The biexponential function is of the form:

$$S(b,\tilde{\omega}) = p_a\phi_a(\tilde{\omega})\exp(-bD_a) + p_b\phi_b(\tilde{\omega})\exp(-bD_b) \tag{4.1}$$

Where the b value is an aggregate parameter reflecting the length and effective strength of the pulsed field gradient as well as the diffusion time ($b = (\gamma\delta G)^2(\Delta - \frac{\delta}{3})$). D_a and D_b are the apparent diffusion coefficients of water in pools A and B, p_a and p_b are the apparent mole fractions of the respective pools and the $\phi_a(\tilde{\omega})$ and $\phi_b(\tilde{\omega})$ terms reflect frequency dependent intensity modulations due to the CEST pulses.

It is important to note the difference between the apparent mole fractions and diffusion coefficients referred to in this biexponential model and the actual parameters for the system. The system is complicated by the effects of exchange and restriction: Chapter 2 describes in detail a model and experimental approach for

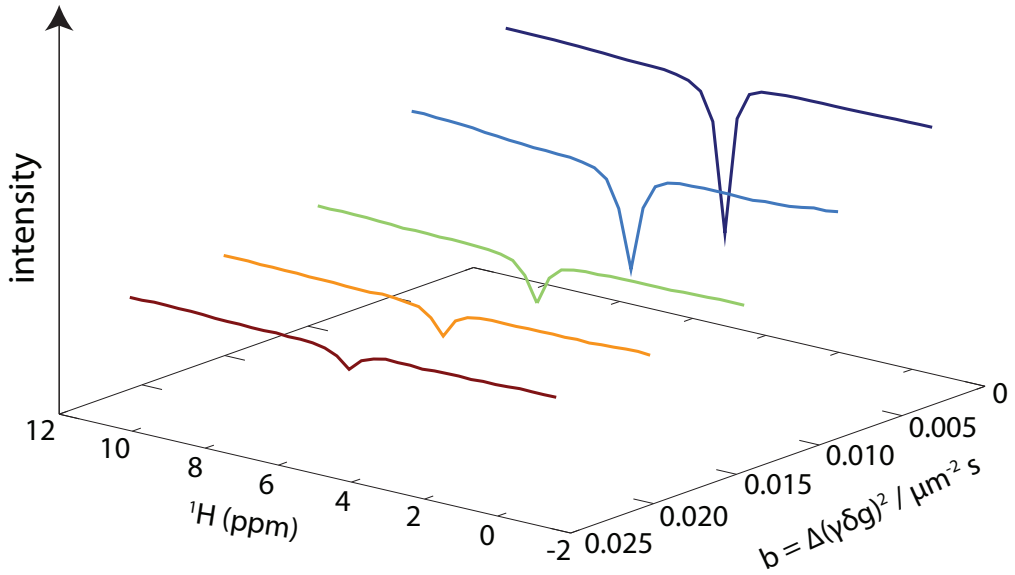

Figure 4.3: In a diffusion-weighted CEST experiment the water intensity is measured both as a function of CEST pulse offset frequency as well as the applied gradient field strength yielding a pseudo-3d dataset. The intensities have been rescaled for clarity.

measuring these parameters robustly in cellular systems but this requires recording variable time diffusion data. A similar approach could be applied here but is made more difficult by the fact that the length of the CEST pulse is coupled to the diffusion delay. The effect of the CEST pulse does not necessarily scale linearly with pulse time and the effect at different offset frequencies may not be consistent, making accurately modelling the data difficult. Furthermore, additionally acquiring data with a range of diffusion/CEST times would turn the experiment into a pseudo-4D version, substantially increasing its length and making it difficult to record on samples with limited stability.

Therefore, we instead adopt an approach where the water intensity is measured at a single diffusion/CEST time and is then fit to equation 4.1. Although the relationship between the true diffusion coefficients and mole fractions and the apparent ones in this equation are not straightforward, provided the two pools still correspond to fast and slow diffusing water compartments this approach is sufficient for the requirements here.

When selecting a diffusion/CEST time in the diffCEST experiment, a balance must be struck between ensuring the CEST pulses are sufficiently long for saturation

transfer effects to build up but also ensuring that there is sufficient signal to noise in the diffusion experiment over a range of gradient strength values. While in the previous chapter CEST pulse times of 2 s were employed, for the diffCEST experiments a compromise time of 500 ms is used to balance the twin requirements of the two experiments. With a diffusion time of 500 ms the water signal attenuatio for systems under study in this chapter at different gradient strengths retain clear biexponential character.

For typical mammalian cell radii (on the order of 10 μm) including the effects restriction, the apparent diffusion coefficient will be approximately an order of magnitude smaller inside the cell than outside. Consequently, at relatively high gradients any signal remaining must have originated from spins that have spent the majority of the diffusion delay inside restricted environments. This view is complicated by exchange: from Chapter 2 and the literature(*83*, *89*, *94*) water exchange rates (k_{ex}) of around 1-10 s^{-1} are expected in cellular systems. But, given that the diffusion coefficient of the extra-cellular domain is much larger, spins that exchange between domains but spend significant periods of time in the fast phase will be attenuated quickly, leading to their association with the fast diffusing phase in the biexponential analysis. Therefore, the apparent slow water mole fraction in the biexponential model will be smaller than the actual value due to exchange but any signal associated with this phase must have resided mostly in a restricted environment. On the other hand, the fast phase will be somewhat 'contaminated' by spins that have exchanged from the slow environment but given that for systems considered here the extra-cellular mole fraction is much greater than the intra-cellular mole fraction this will be a minor effect.

In the next section we demonstrate that the experiment and analysis protcol work successfully in model cellular systems where the intra- and extra-cellular profiles are known a priori. The diffCEST method is then applied to study stress granule formation in a label-free manner in mammalian cells.

4.2 Results and Discussion

4.2.1 GUVs: A Model System for Testing DiffCEST

Giant unilamellar vesicles (GUVs) consist of a spherical amphiphilic lipid bilayer that surrounds an inner aqueous solution. GUVs can be distinguished from other vesicles, namely small and large unilamellar vesicles (SUVs and LUVs, respectively) on the basis of their sizes. Whereas SUVs and LUVs have respective diameters in the range $20 - 50$ nm and $100 - 900$ nm, the size range of GUVs is typically $5 - 50$ μm(*151*). Due to their size and compositional similarity to cell membranes (albeit significantly simplified) GUVs have been used as model systems for studying cell membrane behaviour including aspects of membrane transport(*152*, *153*).

In this subsection, GUVs are used as a model system for validating the diffCEST experimental approach. A key advantage of using GUVs over actual cells, is that the interior composition of cells is a complicated mixture of proteins, metabolites and lipids all of which are likely to contribute to the intra-cellular CEST signal. Therefore, verifying that the diffCEST experiment has reliably distinguished between intra- and extra-cellular CEST effects becomes a difficult proposition (one option may be to lyse cells and record CEST spectra on these samples, but disruption of the membrane environment may affect the CEST signal). Conversely, GUVs can be made to enclose a solution of known composition and placed in an extra-cellular environment that is also of known composition. By comparing the diffCEST derived intra- and extra-cellular CEST profiles to their expected forms the experiment can be rigourously validated.

For experiments here, GUVs are made using an electroformation procedure(*154*) explained in detail in section 4.3. Briefly, the GUVs are formed from 1,2-Dioleoyl-sn-glycero-3-phosphatidylcholine (DOPC) and cholesterol in a 1:3 ratio, within an electroformation chamber with 200 mM sucrose. The GUV rich mixture is then mixed with PBS (1:9 ratio) and allowed to sediment overnight in a falcon tube. The GUVs are then collected from the bottom of the falcon tube and mixed in a 1:1 ratio with a 20% w/v BSA in PBS solution doped with D_2O. This sample is then

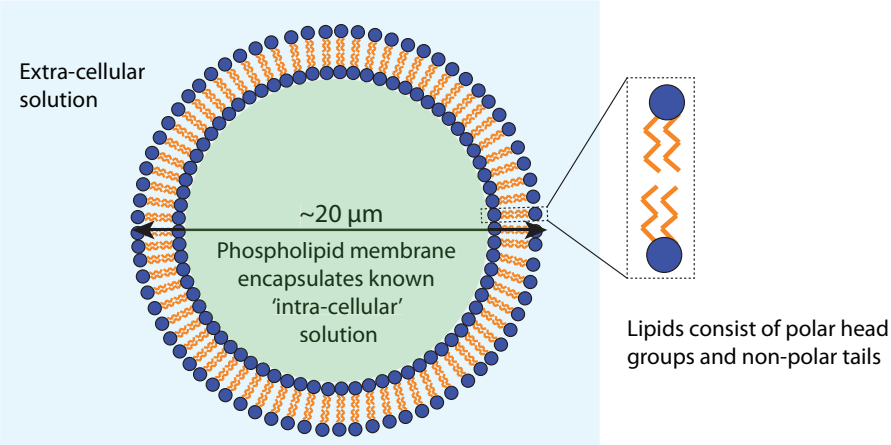

Figure 4.4: Cartoon illustrating the composition of GUVs and the experimental approach used in this subsection. The size of the membrane lipids is greatly increased in this illustration for clarity.

taken forward for NMR analysis. The result of this procedure is that we form GUVs that encapsulate a high concentration sucrose solution, with BSA exclusively in the exterior media of the GUVs. As shown below BSA and sucrose have very distinctive CEST profiles making this an ideal system to test the diffCEST experiment.

Firstly, we demonstrate that the intact GUVs are visible in the NMR sample and give rise to differential water diffusion characteristics necessary for the diffCEST approach to be successful. This is achieved using the variable time diffusion method set out in Chapter 2, using five diffusion times between 0.1 and 0.5 s and explicitly accounting for the effects of restricted diffusion and exchange across the membrane. The results of this analysis are shown in Fig 4.5. The diffusion analysis clearly reveals the the presence of two water pools and the average measured GUV radius on the basis of the water diffusion is in excellent agreement with the microscopy values (10.1 μm and 9.35 μm, respectively) confirming that the slow diffusing pool is associated with the solution encapsulated by the GUVs. The ability to differentiate between the water pools on the basis of their diffusional properties is the key requirement of the diffCEST methodology. The diffusion analysis also estimates that the mole fraction of water inside the GUVs is approximately 5%.

Next, the ^{1}H spectra and CEST spectra of pure BSA, sucrose and the GUV solution are compared to demonstrate the utility of the diffCEST approach. In the

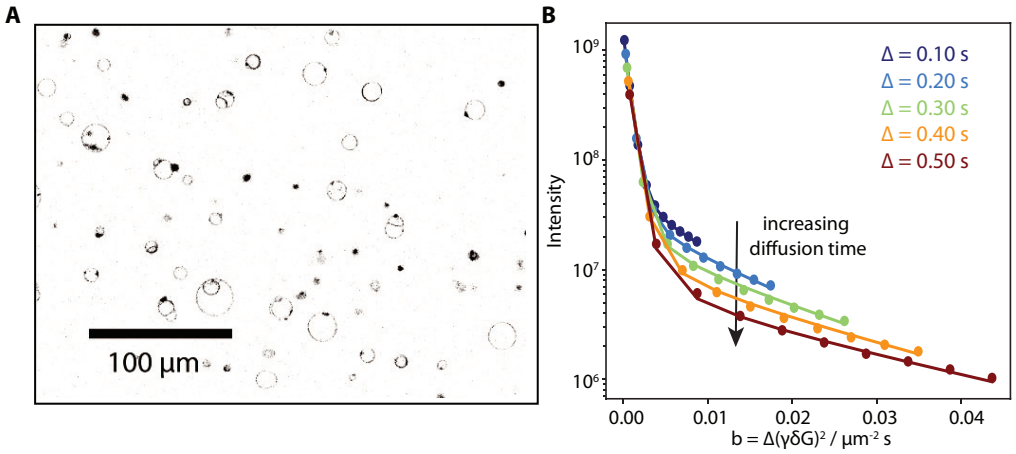

Figure 4.5: **A** Image of a GUV captured by phase-contrast microscopy using a Nikon ECLIPSE Ti inverted microscope. The average GUV radius is estimated to be 9.35 μm, calculated using a purpose-built MATLAB routine. Image captured and cells counted by Valerio Pereno. **B** Diffusion analysis of the water signal in GUV samples reveals clear evidence of a restricted water phase. The estimated GUV radius from NMR is 10.1 μm, in good agreement with the microscopy value.

1D ^{1}H spectra shown in Fig 4.6**A** the sucrose and BSA give rise to very distinct spectra: generally features in the BSA spectra are very broad due to its slow tumbling time and consequent fast relaxation. By contrast, the sucrose, as a small molecule, gives rise to very sharp peaks with the exception of the alcohol protons at around 6 ppm that are broadened due to exchange with water. From this experiment the GUV solution appears as a mixture of the BSA and sucrose spectra. So while the components of the GUV solution can be deduced from these three spectra it provides no information on their respective compartmentalisation.

Similarly, in the comparison of CEST profiles for the three solutions, shown in Fig 4.6**B**, BSA and sucrose show very distinctive CEST profiles. The control sample shows that sucrose does not give rise to any NOE-mediated CEST effects and only shows a single dip in its CEST profile due to the exchange mediated saturation transfer between its alcohol protons and water at 6 ppm. The GUV CEST profile contains elements from the two constituents (sucrose and BSA), though in this case the sharp CEST dip associated with the presence of sucrose is now blurred. Thus while the CEST experiment also provides hints as to the constituents of the GUV solution it also does not provide any information on the

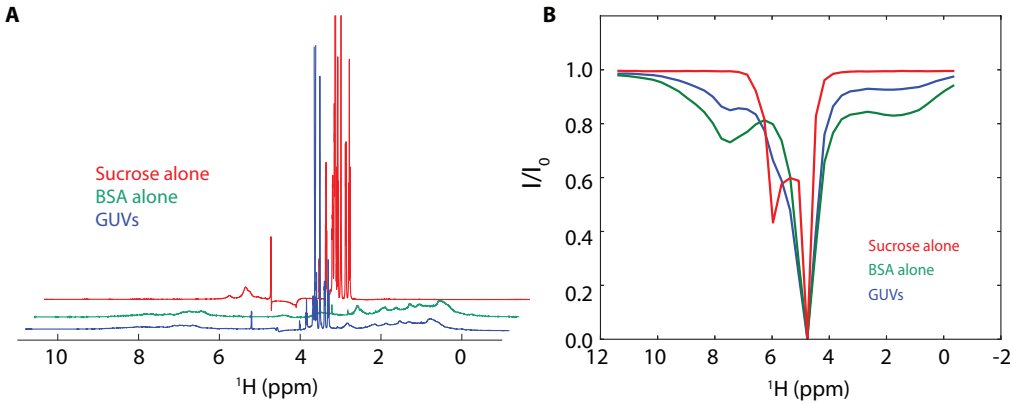

Figure 4.6: **A** ^{1}H water-suppressed NMR spectra comparing solutions of sucrose and BSA alone with the GUV solution. Spectra are offset from each other to aid clarity. **B** Comparison of CEST profiles of solutions of sucrose and BSA alone with the GUV CEST profile, with CEST experiments performed as described in Chapter 3 with a 50 Hz saturation pulse applied for 2 s.

compartmentalisation of those species.

It is somewhat surprising that the alcohol peaks associated with sucrose give rise to such a well defined CEST peak as the protons are expected to be in fast exchange. By simulating CEST spectra with exchange rates of 1000 s^{-1} (Fig 4.7) we see similar CEST effects to those obtained experimentally on the control sucrose sample. The corresponding simulated 1D spectra also show that at 200 mM a small sucrose peak is expected to be visible, whereas this peak is invisible at 20 mM. Very similar CEST effects have been observed in the literature associated with the alcohol groups in glycogen and glucose(*138*).

The diffCEST experiment is now attempted on the GUV sample to test whether it can be used to calculate compartment specific CEST profiles, when their expected forms are known. Fig 4.8**A** shows the normalised diffCEST profiles of the GUVs at increasing gradient field strength. Notably, at low gradient strengths, the CEST profile is dominated by the extra-vesicular water that is in excess and little evidence of CEST with sucrose molecules can be observed. As the gradient strength is increased however, the intra-vesicular restricted water experiences less attenuation and a CEST dip at the expected sucrose position is revealed. We can also observe that the normalised CEST profiles at higher gradient strengths are also more noisy, resulting from the lower signal to noise in these spectra.

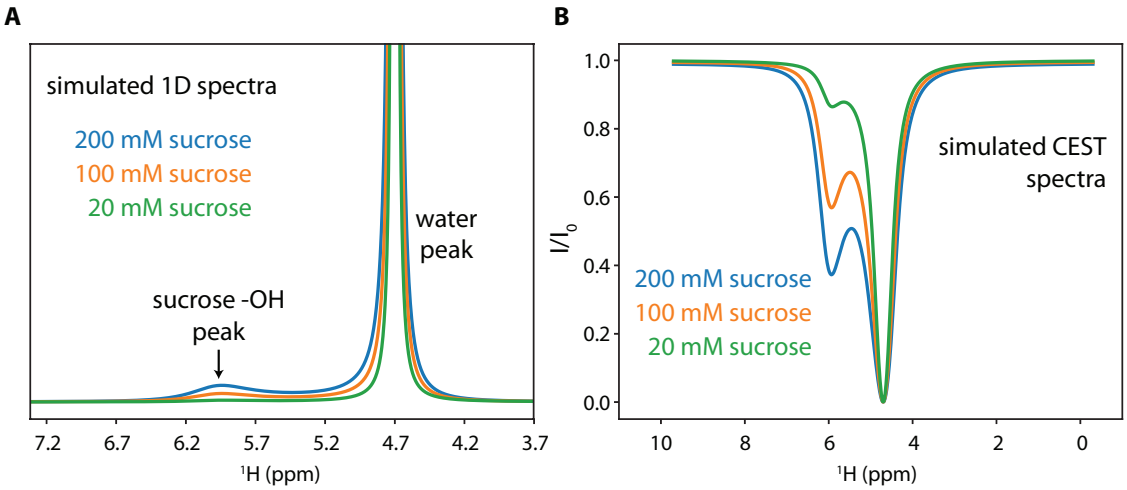

Figure 4.7: **A** Simulated 1D spectra of sucrose -OH peaks in exchange with water (55 M) at different concentrations and **B** corresponding CEST spectra. In all cases the exchange rate between the -OH protons and major water peak is set to 1000 s^{-1}.

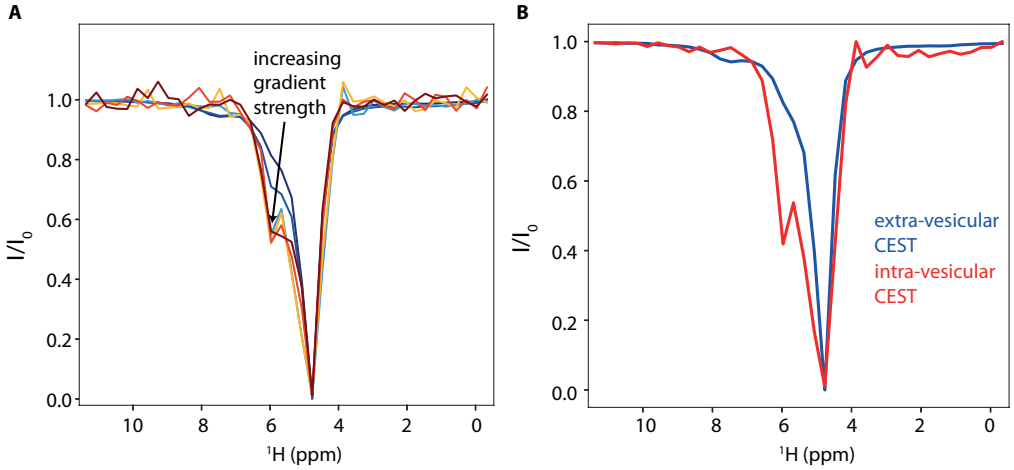

Figure 4.8: **A** Normalised diffCEST profiles at different gradient field strengths. **B** Back calculated compartment specific CEST profile using equation 4.1.

The back calculated CEST profiles for the two water pools are shown in Fig 4.8**B**. The main feature of the back calculated spectra is that the intra-vesicular CEST profile is in excellent agreement with the expected CEST profile for sucrose, and the sucrose CEST dip is missing in the extra-cellular CEST profile. Another important feature is that in the extra-vesicular CEST profile a CEST dip is visible at approximately 8 ppm due to amide proton exchange between the BSA and water, that is missing in the intra-vesicular profile. The combination of these two features provide convincing evidence that the diffCEST sequence is able to robustly

differentiate between CEST effects from different compartments.

There is a discrepancy between the extra-vesicular profile in Fig 4.8**B** and that for BSA in Fig 4.6**B**. This mostly arises because the saturation pulse is much shorter in the diffCEST experiment than in the normal CEST experiment (500 ms vs 2 s), meaning saturation transfer effects are not able to build up to the same extent. On the other hand, the intra-vesicular sucrose CEST effect is still very large, reflecting the very high sucrose concentration inside the GUVs.

While the GUV experiments presented here are an ideal system for testing the diffCEST experiment, they lack the heterogeneity of composition that is present in real biological cells. Moreover, the sucrose inside the GUVs here is at very high concentrations and it is unlikely that a single species in a cell will be at high enough concentration to give a clearly distincitve CEST signature, as is observed for the sucrose in GUVs. Nonetheless, in the next subsection we demonstrate how the experiment can be used to monitor cell stress in a human cell line.

4.2.2 Using DiffCEST to Monitor Cell Stress

An emerging paradigm in cell biology is the importance of liquid-liquid phase separation in the organisation of the cytoplasm(*155*, *156*). While intra-cellular organisation is conventionally thought of as arising from membrane encapsulated organelles, the existence of membraneless organelles has been known for over a century, though an understanding of their liquid droplet like nature and formation from phase separation events is only now emerging(*157*, *158*). Membraneless organelles usually have a concentrated protein and RNA interior from which bulk water is partially excluded, affording these bodies different solvent properties to the cytoplasm(*158*, *159*).

A number of recent studies have pointed to the potential importance of mis-regulated phase transitions in human disease(*156*, *160*). Of particular importance for this subsection is the RNA binding protein FUS, associated with the disease amyotrophic lateral sclerosis (ALS or motor neurone disease). It has been shown that under conditions of cell stress, FUS forms membraneless stress granules and

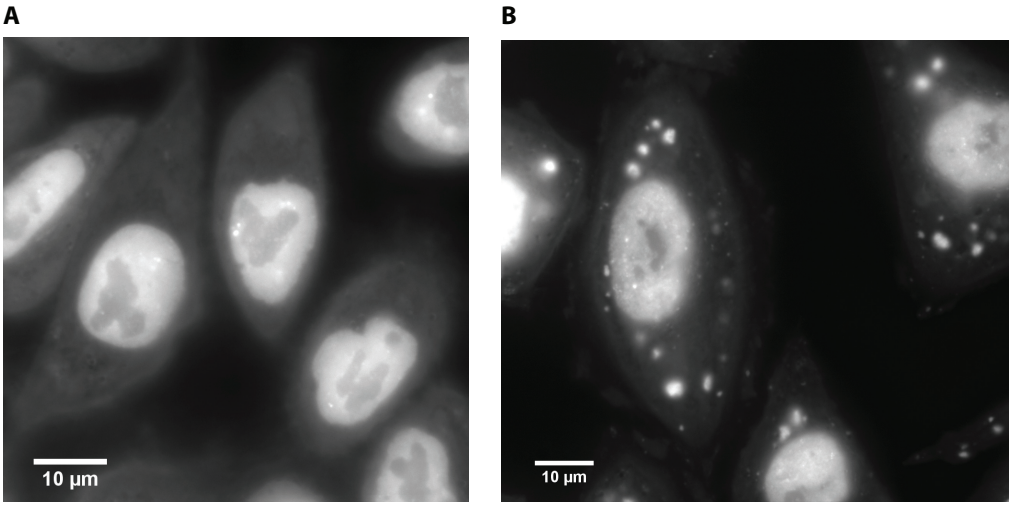

Figure 4.9: Confocal microscope captured images of stress granule formation in modified HeLa cells. **A** In this sample no potassium arsenate is added to the cells and all the GFP-tagged FUS protein remains inside the nucleus of the cell. **B** In a sample of the same cell type that is incubated with 3 mM potassium arsenate for one hour prior to imaging. Image captured by Richard J. Wheeler (Sir William Dunn School of Pathology, University of Oxford).

modulation of these structures is currently being explored as a potential route for treating ALS (*160*, *161*).

Here, we use an *ex vivo* model of HeLa cells expressing green fluorescent protein (GFP) tagged FUS protein at near endogenous levels to test the ability of the diffCEST experiment to monitor cell stress. Given the sensitivity of the CEST experiment to the protein environment (Chapter 3) we expect a CEST response to the formation of concentrated proteinaceous droplets within a cell and the diffCEST experiment provides an opportunity to specifically distinguish effects on the interior of cells where droplet formation occurs (no change is expected to be observed on the exterior of the cells).

Since the FUS protein in the cells is GFP-tagged the presence of stress granules can be conveniently imaged by a confocal microscope. This is illustrated in Fig 4.9 for cell concentrations typically employed in imaging studies, demonstrating the clear presence of stress granules for cells that have been stressed. The cell concentrations used in the NMR experiments are much higher than imaging studies (by around an order of magnitude) so we first test that in the presence of no external

stresses that there is no stress granule formation at higher cell concentrations. For these samples, we then test the cell sample after NMR experiments ($\approx$ 1.5 hour later) to ensure there is still no stress granule formation, as confirmed in Fig 4.10 In this system, we can induce cell stress by adding potassium arsenate to a final concentration of 3 mM to the cells in media and incubating for 1 hour. As shown in Fig 4.10, in the presence of arsenate, stress granule formation is observed and they persist after NMR experiments are completed.

In addition to the diffCEST experiment we also carry out regular water CEST experiments to provide a comparison and diffusion experiments (as described in Chapter 2) to characterise gross features of the samples. In the NMR samples, the HeLa cells are suspended in foetal bovine serum (FBS) doped with D_2O. FBS is a rich media containing a range of proteins and growth factors that cells should be able to comfortably survive in, and therefore attempts to minimise external stresses experienced by the cells (with the exception of the intentional arsenate stress). A consequence of using FBS however is that it introduces a large background signal into the 1D ^{1}H spectrum of the sample making monitoring differences between samples through these spectra unfeasible. Due to variations between biological samples, the experiment is repeated for three samples of unstressed cells and three samples of stressed cells to verify reproducibility. The sample preparation and experimental details are described in detail in subsection 4.3.

The result of the diffCEST experiments on the cellular samples are shown in Fig 4.11, where the intra and extra-cellular CEST profile are back calculated from the raw spectra by fitting to the biexponential model (equation 4.1). The most notable result from this figure is that the stressed cells give rise to consistently higher intra-cellular CEST effects. While these effects are higher throughout the measured CEST profiles, they are particularly marked in the 'methyl' region of the spectrum (between approximately 0.7-2.1 ppm). From a two-sample T-test the intra-cellular CEST effect is significantly increased in the stressed cells with a p-value of 0.016. This result is consistent with the formation of viscous proteinaceous droplets inside the cell that give rise to enhanced saturation transfer with water.

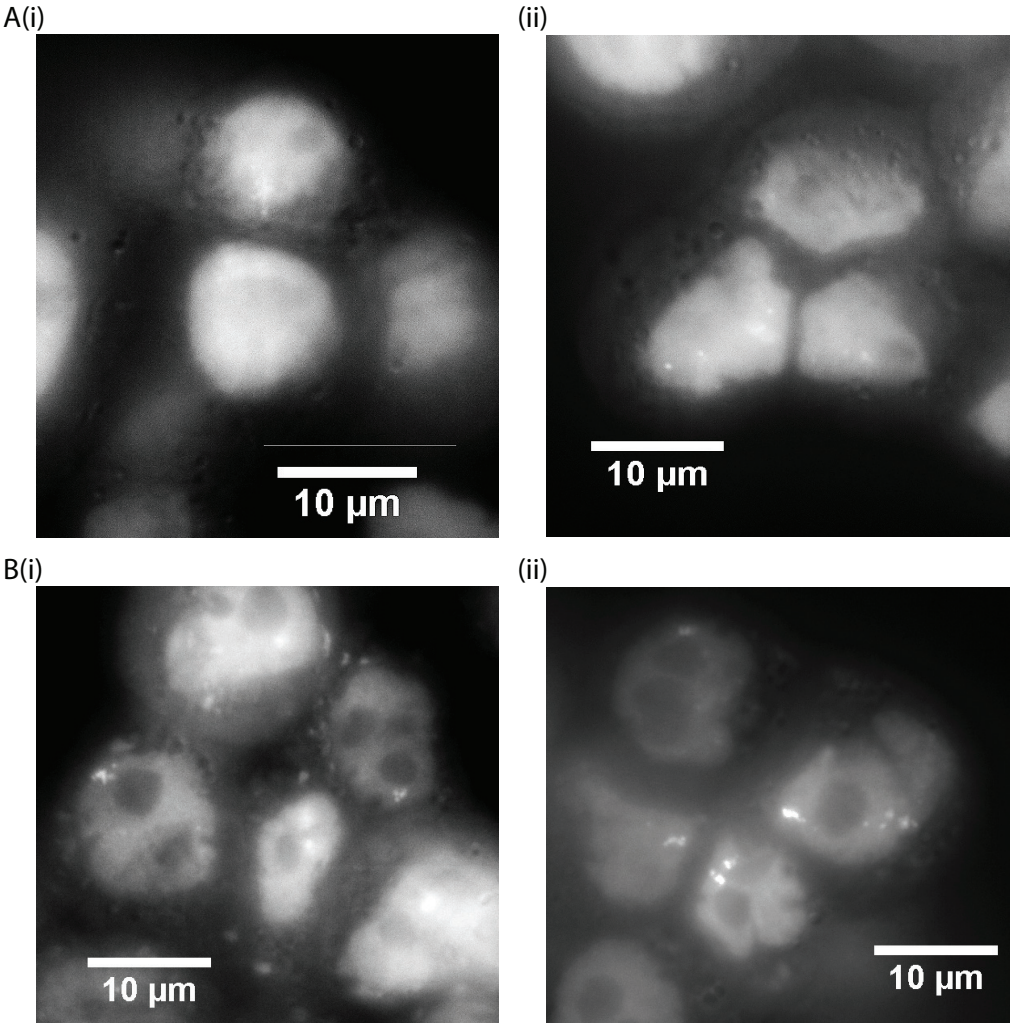

Figure 4.10: Validation of NMR samples by microscopy: since the NMR samples include cells at much higher concentrations than are typically used in imaging and screening studies we check that **A** the unstressed samples show no evidence of stress granule formation over the timescale of NMR experiments and **B** the stress granules in the stressed sample persist over the timescale of the NMR experiments. In **A(i)** the cells are observed before NMR experiments and in **A(ii)** they are observed 1.5 hours later following NMR experiments. **B(i)** stress granule formation is observed in a higher cell concentration sample following incubation with potassium arsenate for 1 hour. In **B(ii)** the granules are still present after NMR experiments. The image quality in these images is reduced compared to Fig 4.9 due to the higher scattering and refraction of light that occurs at higher cell concentrations. Image captured by Richard J. Wheeler (Sir William Dunn School of Pathology, University of Oxford).

Importantly, this does not confirm that the differential CEST effects are due to the formation of the stress granules, since the arsenate may induce some other biochemical change in the cells that is not accounted for here. The increased intra-cellular CEST effects upon cell stress still represents a potentially important finding in the interpretation of CEST profiles.

Another interesting feature common to all the intra-cellular CEST profiles is that they resemble the CEST profile of a protein in viscous solution. This supports the idea that the intra-cellular CEST profile reports predominantly on the behaviour of soluble proteins and thus the signal may be sensitive to the formation of stress granules. It is impossible to confirm this conclusion from the data here however, as the addition of he arsenate poison will likely cause other changes in the cell beyond the formation of stress granules that could drive the observed changes in intra-cellular CEST profiles.

Conversely, the extra-cellular profiles between the six samples is very consistent suggesting that the presence of arsenate has minimal impact on the FBS. It is interesting to compare the extra- and intra-cellular CEST profiles in general. Despite the high concentration of proteins, especially BSA, present in FBS the intra-cellular CEST effects are still substantially higher reflecting the extremely high protein concentration inside cells as well as the higher viscosity inside cells promoting more efficient saturation transfer.

It is interesting to compare the diffCEST result shown in Fig 4.8 to the normal CEST experiment run on the same samples Fig 4.12. The results of the diffusion analysis of these samples are also helpful in interpreting this data and are given in Table 4.1.

Viewed in isolation, the results of Fig 4.12 suggest that the differences between the cell samples under stressed conditions are not significant and seem at odds with the diffusion-weighted results. However when the diffusion analysis results are taken into account we note that the subtle differences in CEST profiles exactly match the trend in cellular concentration: a greater mole fraction of cellular water and therefore higher cell concentration results in bigger CEST effects. This fully supports the

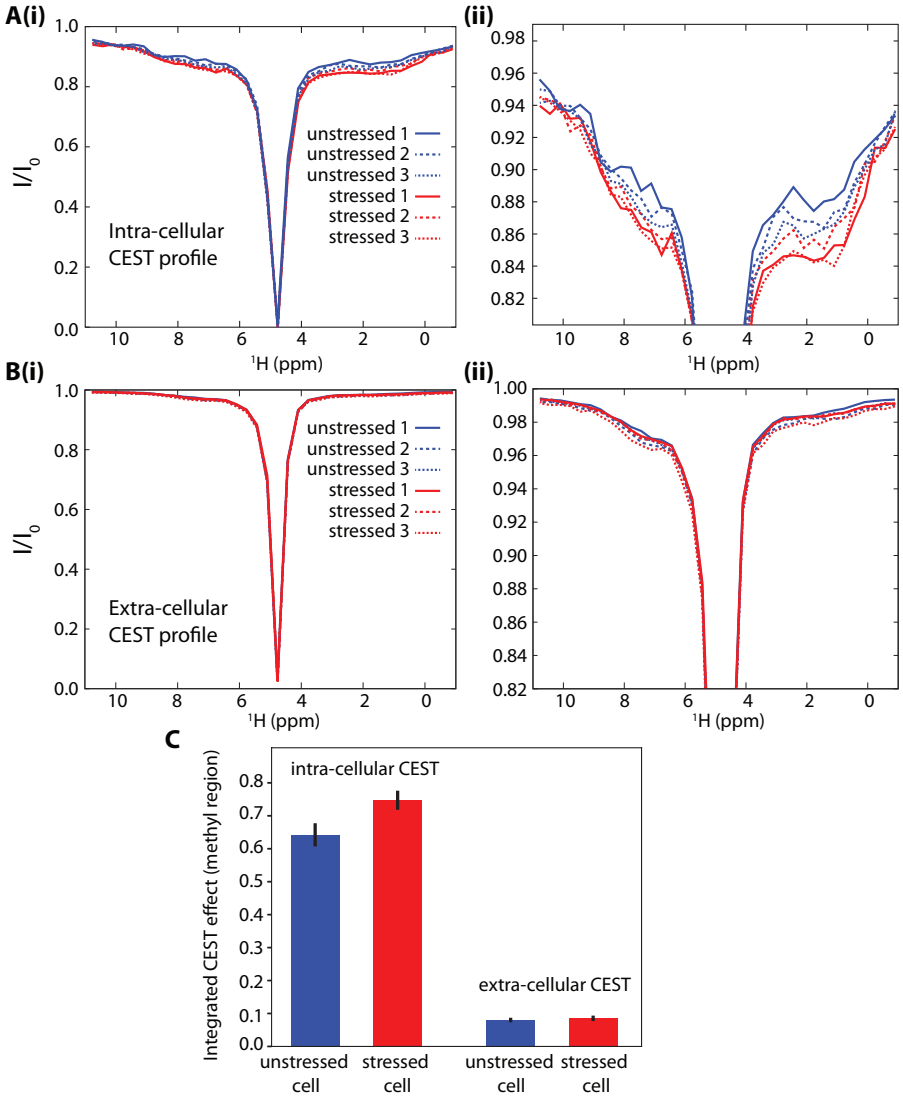

Figure 4.11: Comparison of unstressed and stressed cells using the diffCEST experiment **A** Back calculated intra-cellular CEST profiles for HeLa cells in presence and absence of cell stress (where 3 mM potassium arsenate is used to stress the cells). **ii** zoom-in of back-calculated CEST profile to highlight the differences between the samples. **B** Back calculated extra-cellular CEST profiles with zoom in view of profiles shown in **ii**. **C** Comparison of integrated CEST effects between 0.7-2.1 ppm (methyl region) in stressed and unstressed HeLa cell samples. The stressed cells show a significantly higher intra-cellular CEST effect: from a two-sample T-test we have a p-value of 0.016 for the intra-cellular CEST.

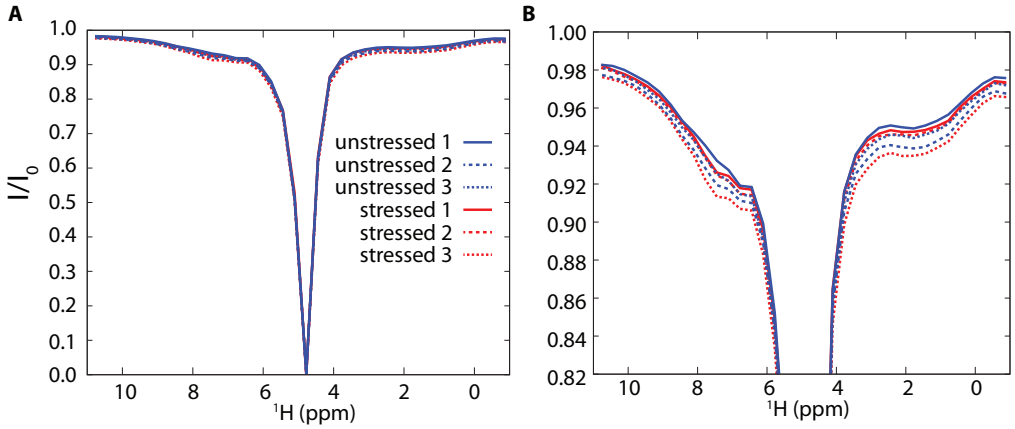

Figure 4.12: Normal water CEST profiles from experiments on stressed and unstressed HeLa cell suspensions. The sample numbers match those for the diffCEST experiments in Fig 4.11. **A** Shows the full CEST profiles (highlighting their similarities) and **B** is a zoom-in view showing small differences between the samples.

conclusion that the intra-cellular CEST profiles show much bigger effects than the extra-cellular ones and also highlights the strength of the diffusion-weighted CEST experiment in being able to unpick compartment specific CEST effects.

The intra-cellular diffusion coefficients obtained from the variable time diffusion analysis in Table 4.1 are also in agreement within error for both stressed and unsressed cell samples. This suggests that the stress granules form a liquid-like rather than gel-like phase, as the latter phase would be expected to have slower intra-cellular diffusion.

4.2.3 Discussion

In this chapter we have presented a novel CEST experiment with diffusion weighting, diffCEST, that allows CEST profiles associated with differentially diffusing water compartments to be measured. By carrying out experiments on model GUV systems with two distinct water pools, we have been able to show that fitting diffCEST data to a phenomenological biexponential model allows CEST profiles for the restricted and unrestricted water to be back calculated. This opens up a range of opportunities for in-cell NMR, as shown in the application to monitoring cell stress.

The diffCEST experiment has several potential uses. Firstly, as a technique in its own right, it offers label-free, non-invasive way of monitoring specific intra- and extra-

Sample	D_{in} ($\mu m^2\ s^{-1}$)	D_{out} ($\mu m^2\ s^{-1}$)	k_{ex} (s^{-1})	radius (μm)	p_{in}	R_1^{in} (s^{-1})	R_1^{out} (s^{-1})
Unstressed 1	249± 101	1750± 85	5.42± 0.97	12.19± 0.60	0.029± 0.004	0.00± 0.00	0.28± 0.08
Unstressed 2	237± 69	1590± 71	5.78± 0.71	12.19± 0.44	0.056± 0.007	0.00± 0.01	0.34± 0.06
Unstressed 3	234± 77	1660± 73	5.49± 0.79	12.45± 0.18	0.041± 0.006	0.00± 0.01	0.31± 0.07
Stressed 1	235± 25	1630± 29	6.63± 0.34	11.19± 0.19	0.035± 0.002	0.00± 0.30	0.26± 0.03
Stressed 2	236± 82	1620± 93	6.15± 0.89	11.98± 0.48	0.040± 0.006	0.00± 0.00	0.23± 0.08
Stressed 3	226± 61	1530± 61	6.84± 0.71	11.14± 0.37	0.058± 0.007	0.00± 0.00	0.32± 0.06

Table 4.1: Fitted parameters from diffusion analysis performed on the same samples as diffCEST experiments in Fig 4.11. Diffusion experiments are carried out using a pulsed field gradient stimulated echo experiment with variable diffusion times and subsequently analysed using the INDIANA approach set out in Chapter 2. Error bars are calculated using 1000 bootstrap runs.

cellular changes in cellular systems. This could be useful in the context of monitoring overall cell health in real-time when it is desirable to perform long NMR experiments on in-cell samples. The label-free nature of the diffCEST experiment also makes it a potentially attractive choice for performing in-cell assays. For example, in the context of the HeLa cell experiments performed above where diffCEST reports on cell stress the effect of drug molecules in modulating this effect could be tested by this method, avoiding the need for a system containing fluorescent tags.

As a label-free method, there is also clear potential for the the diffCEST experiment to be run in pre-clinical and clinical MRI scanners on actual biological tissue. Due to the complexity and heterogeneity of these systems the biexponential model for calculating intra- and extra-cellular CEST profiles may not be appropriate (*63*, *65*). Nevertheless, gradient-edited CEST spectra could still be studied in a 'model-free' sense to compare CEST effects associated with generally faster and slower diffusing water. A potential difficulty of such measurements is that the relatively long diffusion times required to build up sufficient saturation transfer may result in low signal to noise measurements. However, the experimental time of these

experiments could be considerably accelerated by only recording a few saturation frequencies of interest rather than spanning a large frequency range as is done here.

If an appropriate model was found to interpret the data, the ability to extract 'true' intra- and extra-cellular CEST profiles would be of significant utility, as some pathologies are known to exhibit specific intra- and extra-cellular effects. For example, cancerous tissue displays a drop in extracellular pH (acidosis) and a concomitant rise in intra-cellular pH(*133*, *134*). Leaving aside the challenges in measuring pH using CEST approaches (discussed at length in Chapter 3), the diffCEST methodology provides an avenue for delineating these effects. Moreover, there is an assumption in the literature that the observed CEST profile is derived primarily from intra-cellular species(*108*), diffCEST could allow for the relative contributions of intra- and extra-cellular species to be ascertained more quantitatively.

The main experimental findings using the diffCEST experiment in this chapter are summarised in Fig 4.13. As future work, it would be interesting to repeat the cell stress experiments on a cell line where stress granules do not form upon addition of arsenate to further test whether it is the stress granules themselves that drive the increase in saturation transfer effects.

4.3 Methods and Materials

4.3.1 Formation and Imaging of GUVs

GUVs are made using a recently developed electroformation method on stainless steel electrodes(*154*). First, a solution of 1,2-Dioleoyl-sn-glycero-3-phosphatidylcholine (DOPC) and cholesterol (both in chloroform) are mixed in a 1:3 molar ratio. The solution is then pipetted vertically onto the electrodes (25 μl per electrode), with care taken to ensure that the entire surface is coated. The chloroform is allowed to evaporate overnight after which the electrodes are immersed in an electroformation chamber with approximately 3 mL 200 mM sucrose in deionised water. Vesicle swelling is induced by using a signal generator to apply a 5 Vpp, 10 Hz wave signal for 2 hours. Subsequently, the frequency of excitation is then lowered to 5 Hz for 30 minutes to facilitate vesicle detachment. The GUV rich solution is then transferred

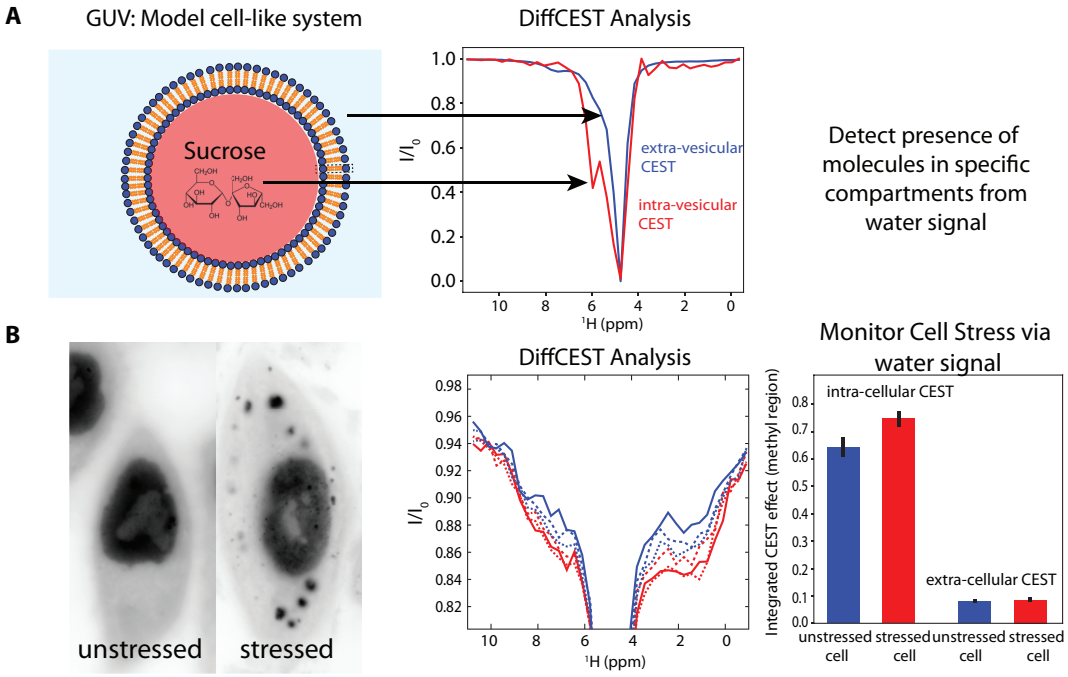

Figure 4.13: Summary of main experimental results using diffCEST shown in this chapter. **A** Using a model system of GUVs we are able to show the successful application of the diffCEST experiment to differentiate between CEST profiles from different compartments. **B** By performing experiments on cells in unstressed and stressed (induced by incubation with arsenate) states we have shown that the intra-cellular CEST profile is sensitive to cell stress and these changes can be captured using the diffCEST experiment.

to a 50 ml falcon tube filled with 27 ml PBS, and is allowed to sediment overnight. Finally, the GUV solution is mixed in a 1:1 ratio with 20% w/v BSA in PBS doped with 10% D_2O and transferred to a standard 5 mm NMR tube for experiments.

The GUVs were imaged using a Zeiss LSM 780 microscope and characterised with a purpose-built MATLAB routine.

4.3.2 Cell Sample Preparation

Stable Kyoto HeLa BAC cell lines expressing FUS with a C-terminal GFP protein were generated using BAC recombineering(*162*), giving a near endogenous expression level of the fusion protein. HeLa cells were grown in Dulbecco's modified eagle medium supplemented with 10% FBS. Cultures were grown in petri dishes and maintained at 37°C and 5% CO_2. A single petri dish of cells is used per NMR sample.

Prior to NMR experiments, for stressed samples, potassium arsenate ($KAsO_2$) is added to the cells to a final concentration of 3 mM for one hour. For both

stressed and unstressed samples the media is then decanted off and 1 ml FBS (with 3 mM $KAsO_2$ for stressed samples) is added to the petri dishes. Cells are then removed from the dishes using mechanical scraping (to minimise disruption to the cell membranes) and transferred to an Eppendorf tube. The cells are then spun down in a centrifuge for 2 minutes at 800× g. The excess liquid is then decanted off and the cells are resuspended in 250 μl FBS (with 3 mM $KAsO_2$ as appropriate) and a further 25 μl of D_2O is added to provide a lock signal. The sample is then transferred to a 5 mm shigemi NMR tube for experiments.

4.3.3 NMR Experimental Methods

All NMR experiments are acquired at 298 K on a 14.1 T (600 MHz) Agilent DD2 NMR spectrometer equipped with a 5 mm z-axis gradient triple resonance room temperature probe. All spectra were Fourier transformed and phased using NMRPipe(*101*).

GUV experiments

1D ^{1}H spectra of sucrose, BSA and GUVs (4.6**A**) are acquired using a watergate pulse sequence(*18*) for water suppression. All 1D ^{1}H spectra are acquired using 32 transients.

CEST experiments shown in Fig 4.6**B** are carried out using a 50 Hz frequency saturation pulse applied for 2 s with 40 evenly spaced saturation frequencies between -0.3 and 11.4 ppm. A single scan per offset frequency is used and for each saturation frequency the experiment is run first with the saturation pulse applied and then with it replaced by a delay of equivalent length to facilitate calculation of $\frac{I}{I_0}$. The CEST spectra are analysed using in-house python scripts as described in Chapter 3, using the nmrglue module(*102*) to integrate the water peak intensity.

Diffusion experiments as shown in Fig 4.5**B** are acquired using a pulsed field gradient stimulated echo experiments with five different diffusion times $\Delta = 100$, 200, 300, 400 and 500 ms. The gradient length, δ is set to 2 ms and 10 quadratically spaced gradient field strengths between $8 - 60\ G\ cm^{-1}$ are employed. The diffusion data are analysed using the INDIANA approach described in Chapter 2.

DiffCEST experiments (shown in Fig 4.8) are acquired using a gradient length, δ, of 1.5 ms and eight quadratically spaced gradient field strengths between $8 - 55$ $G\ cm^{-1}$. The diffusion time is set to 0.5 s. The saturation pulse is applied while magnetisation is stored longitudinally using a 50 Hz pulse at 40 evenly spaced frequencies between -0.3 and 11.4 ppm and is run in a pulse on/off manner as described above. The diffCEST data is analysed using Python scripts as outlined in subsection 4.3.4.

HeLa Cell Experiments

CEST experiments shown in Fig 4.12 are carried out using a 50 Hz frequency saturation pulse applied for 2 s with 36 evenly spaced saturation frequencies between -0.9 and 10.8 ppm and is run in a pulse on/off manner as described above. A single scan per offset frequency is used and the experiment is run in a saturation pulse on/off fashion for normalisation. The CEST spectra are analysed using in-house python scripts as described in Chapter 3.

The diffusion experiments (results in Table 4.1) are acquired using pulsed field gradient stimulated echo experiments with six different diffusion times $\Delta = 50$, 100, 200, 300, 400 and 500 ms. The gradient length, δ is set to 2 ms and 11 quadratically spaced gradient field strengths between $8 - 60$ $G\ cm^{-1}$ are employed. The diffusion data are analysed using the INDIANA approach described in chapter 2.

The DiffCEST experiments (Fig 4.11) are acquired using a gradient length, δ, of 2 ms and five quadratically spaced gradient field strengths between $8 - 44$ $G\ cm^{-1}$. The saturation pulse is applied while magnetisation is stored longitudinally using a 50 Hz pulse at 36 evenly spaced frequencies between -0.9 and 10.6 ppm. The diffCEST data is analysed using Python scripts as outlined in subsection 4.3.4.

4.3.4 Analysis of diffCEST Experiments

As part of the development of the diffCEST experiment Python scripts for automating their analysis have also been written. The code is able to read in relevant experimental parameters from Varian/Agilent *procpar* files and phased spectra are read in using the nmrglue(*102*) module and subsequently integrated allowing

normalised diffCEST profiles (for example Fig 4.8**A**) to be rapidly generated. The analysis script is also able to fit the diffCEST data to a biexponential model and generate plots of back calculated intra- and extra-cellular CEST profiles (for example Fig 4.8**B**). Fitting is performed using the Levenberg-Marquardt algorithm as implemented in the Python module Lmfit(*103*).

4.4 Acknowledgements

Initial diffCEST experiments (not shown here) were carried out in collaboration with Kevin J. Ray (Wellcome Centre for Integrative Neuroimaging, University of Oxford). GUV samples were prepared by Valerio Pereno and Luca Bau (Institute of Biomedical Engineering, University of Oxford) and imaged by Valerio Pereno. HeLa cell samples were prepared with Richard J. Wheeler (Sir William Dunn School of Pathology, University of Oxford) who also captured microscope images.

Part II

Atomic Resolution Studies of Labelled Biomolecules using NMR

5

Characterisation of Dynamics in Cyclophilin A

Abstract

In this chapter a combined molecular dynamics (MD) and NMR approach is used to study structural and dynamical details of the protein Cyclophilin A (CypA) with atomic-level resolution. Initially, MD is used to investigate the energy landscape of the protein, focussing on the loop regions around residues 70 and 100. Accelerated and equilibrium MD methods are combined to provide insight into millisecond timescale protein dynamics. The simulations reveal that the D66 residue plays an important role in determining the loop structure and dynamics and mutating this residue to an alanine is predicted to significantly perturb the loop environments. Following the MD studies, wildtype and D66A forms of the protein are produced recombinantly and studied extensively by NMR, including fast dynamics and relaxation dispersion experiments. Within the chapter, a strategy for characterising millisecond dynamics by combining CPMG and $R_{1\rho}$ datasets is presented. Overall, the NMR results are in good agreement with the MD insights suggesting that the mutation results in an order to disorder transition in the loop region around residue 70 in CypA.

The work presented in this chapter is the result of a collaboration with Jordi Juárez Jiménez, Julien Michel and colleagues in the Department of Chemistry, University of Edinburgh. While the work performed by the author is emphasised here, other parts of the project, particularly the molecular dynamics simulations, are presented for completeness and acknowledged appropriately.

5.1 Introduction

Proteins are typically depicted using a single static structure corresponding to a ground state conformation. While significant insight into behaviour and function can often be provided by such structures, it has long been recognised that this representation is incomplete(*163–165*). A complete description of a protein's behaviour is given by its free energy landscape and includes details of the dynamic perturbations that allow the protein to access different minima in the landscape that correspond to metastable conformations(*166*, *167*). A typical structured protein will spend most of its time in a ground state conformation and will infrequently sample higher energy conformational states (often termed invisible, excited or dark states), though these can be of significant functional importance(*32*). Intrinsically disordered proteins (IDPs) represent an exception to this paradigm as they have no fixed three dimensional structure and instead can be thought of as an ensemble of disordered polypeptide chains that rapidly interconvert(*168*).

As an experimental technique, NMR is uniquely placed for the study of biomolecular dynamics(*4*, *169*). Crucially, spin relaxation is dependent on fluctuations of local magnetic fields at appropriate frequencies. The frequencies of local field fluctuations are themselves a consequence of dynamic motion (mostly rotational motion), making spin relaxation a powerful tool for experimentally probing molecular dynamics with atomic-level resolution(*24*). A large number of spin-relaxation experiments have been designed for probing molecular dynamics on different timescales: the fast dynamics class of experiments (comprising R_1, R_2 and heteronuclear NOE) can be used to probe ps-ns motions(*170*), while relaxation dispersion experiments (CPMG and $R_{1\rho}$) and CEST experiments can be used to probe slower motions on the μs-ms

timescale (over which conformational rearrangements typically occur)(*33*, *34*, *171*). Other techniques available to an NMR spectroscopist for studying dynamics include: residual dipolar couplings, paramagnetic relaxation enhancement, lineshape analysis and hydrogen-deuterium exchange(*172*).

Despite its clear strengths for studying the dynamical behaviour of molecules NMR also has some drawbacks. Firstly, as a relatively insensitive technique, NMR dynamics experiments require high concentrations of isotopically labelled samples. Also, as these experiments typically require hours to days to attain high quality data the sample stability must also be high. Moreover, while it may be possible to produce a high concentration stable sample suitable for NMR experiments, rapid transverse relaxation means that studies of ^{15}N-labelled proteins larger than approximately 50 kDa are unfeasible. Developments in hardware(*52*), pulse sequences(*173*, *174*) and isotope labelling strategies, in particular methyl-TROSY methods(*175*), greatly extend the range of proteins that can be probed by NMR. However, in the latter case the experiment becomes limited to residues in a protein where $^{13}CH_3$ groups can be introduced (typically alanine, methionine, leucine, isoleucine, threonine and valine residues).

Another issue with NMR lies in the difficulty of interpreting the data. Typically following an NMR experiment a quantity such as a chemical shift or relaxation rate is measured: converting this quantity into a visualisable structural change requires robust models. While tremendous progress has been and continues to be made in this area(*171*), when fitting data to the best model is selected by choosing the least complicated, physically reasonable model that fits the experimental data. A lack of resolution in the experimental data, or sensitivity to certain parameters may therefore preclude a complete understanding of the system's behaviour.

Molecular dynamics (MD) simulations provide another potential avenue for studying the free energy landscapes of biomolecules. Unlike NMR that effectively provides windows into different parts of the free energy landscape, the aim of the simulations is to provide a full 'molecular movie' that provides much greater coverage of the structure and dynamical properties of proteins.

However, these simulations are beset by their own issues. Foremost amongst these is that they are ultimately simulations: empirical force fields are used and may lead to inaccurately resolved dynamics and therefore must be robustly tested against experimental findings. Also, in practice MD simulations are extremely computationally intense and usually can only characterise excited states that exchange on a timescale of a few microseconds with the ground state. Methods such as enhanced sampling aim to allow access to more slowly exchanging excited states but this is often accompanied by a loss of temporal resolution(*176*, *177*).

To overcome some of these difficulties associated with particular techniques, in this chapter we use an integrated approach, combining MD and NMR, to describe, with atomic-level resolution, millisecond timescale conformational dynamics in the human protein cylophilin A (CypA). CypA is a cytosolic enzyme that catalyses the cis-trans isomerisation of prolyl peptide bonds(*178*). It is an important drug target, and is particularly known for its binding to the cyclosporin that is widely employed as an immunosuppressant(*179*).

As a 165-residue, 18 kDa protein, CypA is particularly amenable to ^{15}N-relaxation dispersion NMR experiments as have been attempted previously(*178*, *180*). The $\mu s - ms$ timescale dynamics can be parametrised using the chemical exchange contribution to relaxation, R_{ex}(*34*). These studies revealed that most of the residues displaying large R_{ex} values are located in a loop region between Asp66 and Gly75, herein referred to as the 70s loop. A second region showing significant R_{ex} values is located in a loop between Ala101 and Gln 111, herein referred to as the 100s loop. These two regions are highlighted in Fig 5.1. However, a detailed understanding of the structural changes in the protein that underpin its millisecond timescale dynamics has remained elusive, though it has been suggested that its dynamics may be linked to its enzymatic activity(*181*). It is anticipated that this structural information could help facilitate the development of CypA drugs(*182*, *183*).

In this chapter, MD simulations and Markov state modelling (MSM) are combined to compute a structural ensemble of CypA that is validated using a

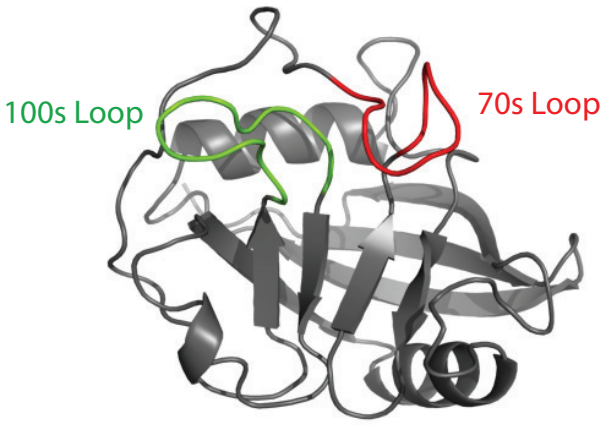

Figure 5.1: Representative structure of CypA (PDB ID: 1OCA) with the positions of the dynamic 70s and 100s loops highlighted

range of experimental measures. Based on insights gained from the structural ensemble of WT CypA, a single-point mutant is designed and shown by simulations to significantly disrupt the energy landscape of CypA. The WT and mutated forms of CypA are subsequently expressed and purified from *E. coli* to allow an experimental validation of the simulation results. A large number of experiments including fast dynamics analysis and relaxation dispersion experiments (CPMG and $R_{1\rho}$) are performed on both forms of the protein that demonstrate its dynamics are significantly perturbed by the mutation.

5.2 Results

5.2.1 MD Determined Conformational Ensemble of CypA

The work in this subsection was performed by Jordi Juárez Jiménez and Julien Michel (Department of Chemistry, University of Edinburgh) and is included here for completeness.

The workflow used to generate the conformational ensemble of CypA is shown in Fig 5.2. Simulations are initialised from an X-ray structure of CypA (PDB ID: 1AK4) and are allowed to equilibrate using standard MD procedures(*184*). An

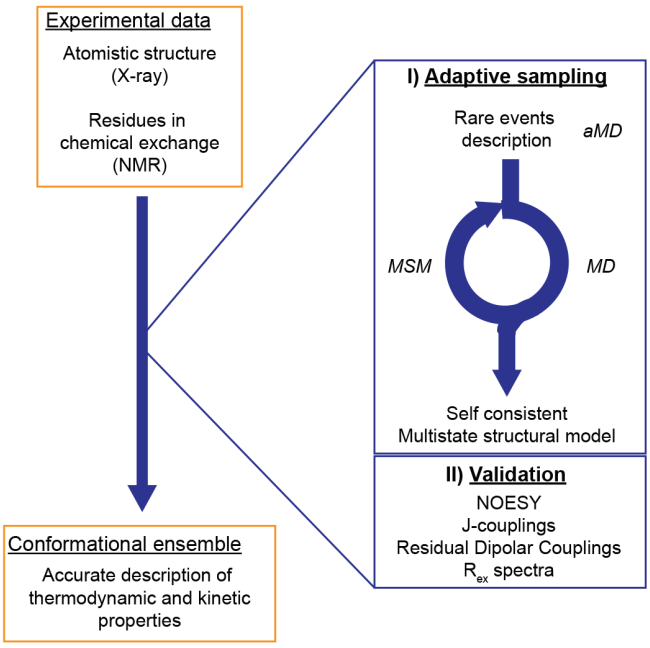

Figure 5.2: Workflow to generate CypA conformational ensembles that can describe millisecond timescale dynamics in the protein. Accelerated MD (aMD) simulations are used to generate rare event loop motions, while MD simulations and Markov State Modelling (MSM) recover equilibrium system properties. Figure adapted from (*184*) and made by Jordi Juárez Jiménez.

accelerated MD (aMD) method is used to explore the conformational landscape of the protein(*185*). aMD methods effectively lower the energy barriers between conformational states in the free energy landscape by allowing dynamics that occur on longer timescales to be accessed more readily. In practice this is achieved by adding a potential energy boost to the molecule when its value falls below a certain threshold. The aMD simulations are found to induce enhanced motions in the 70s and 100s loops, while maintaining the overall folded structure of the protein.

Following the aMD simulations, it is necessary to recover the equilibrium properties of the conformations generated. This is achieved by taking structures sampled from the aMD simulations and using these to initialise equilibrium MD simulations of the protein coupled with MSM construction. The MSMs consist of 100 microstates, themselves obtained from the MD trajectories by k-means clustering of structurally similar conformations, and characterise the transitions between different microstates(*186*). Rounds of equilibrium MD and MSM construction are iterated

until a numerically stable MSM is obtained. This was achieved after three rounds amounting to a cumulative sampling time of 187.5 μs.

After construction of the MSM, it is found that the model can be coarse grained into five macrostates that can aid interpretation of the results. This is indicated in Fig 5.3**A** where each colour represents a different macrostate, the smaller dots denote microstates and the lines between them indicate possible transitions. The five macrostates can be interpreted in the following way: for the most populated macrostate shown in orange in 5.3**A** the protein resembles the conformation observed in the X-ray crystal structure of CypA with the 70s loop in a 'closed' conformation and the 100s loop in an 'open' one. In the orange macrostate the loops undergo low amplitude fluctuations (as measured by the root mean squared positional deviations between microstates). The orange macrostate is able to exchange with the red macrostate in which both the 70s and 100s loops are in a 'closed' conformation. In this macrostate relatively small positional fluctuations are observed in loops also. The cyan macrostate in the centre of the MSM serves as a gateway that all other macrostates are able to exchange with. In this conformation both loops are in intermediate positions and show average positional fluctuations compared to the other macrostates. Finally, in the magenta and blue macrostates the 70s loop is in an 'open' conformation while the 100s loop is in an 'open' and 'closed' conformation, respectively. However, while the dynamics of the 100s loop is comparable to the other macrostates, the 70s loop shows much larger amplitude dynamics in the magneta and blue macrostates, suggesting that the loop is disordered in these states.

Further insight into the nature of the order to disorder transition of the 70s loop can be gained by studying the predominant atomic interactions in the different macrostates as shown in Fig 5.3**B**. This shows that the transition is controlled by a network of hydrogen bonds between the backbone nitrogen atoms from residues Thr68 to Gly74 and the side chain of residue Asp66. In the orange macrostate the hydrogen bonding network is intact giving rise to the relativity rigid structure of the loop in its 'closed' form. In the cyan intermediate macrostate there is a breakage in the hydrogen bonds between residues Gly72-Gly75 resulting in a partial unwinding of

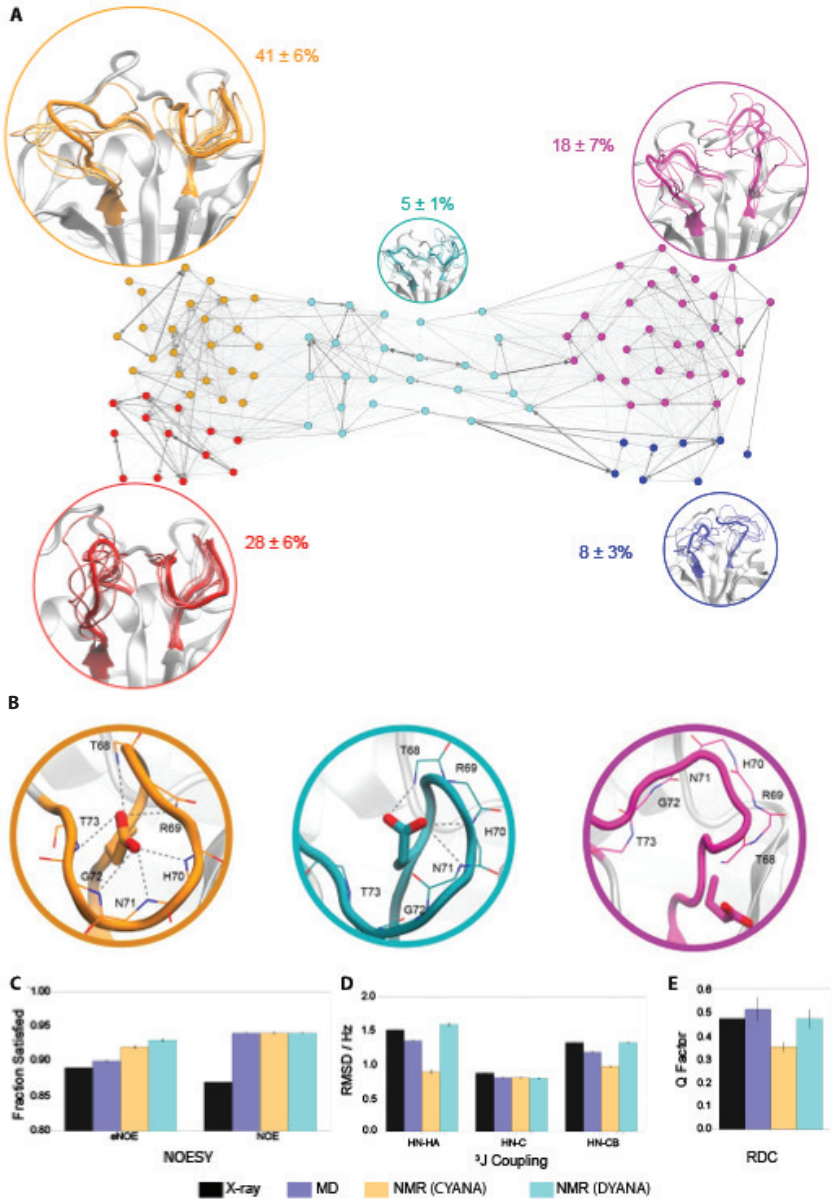

Figure 5.3: MD calculated structural ensemble of WT CypA and experimental validation. **A** Visualisation of MSM: each small dot depicts a protein microstate that can be further coarse-grained into five macrostates indicated by the different colours. Lines between the dots indicate possible transitions between microstates. The populations of the macrostates are shown with the errors calculated from a bootstrap analysis. **B** Detailed images of interactions between the Asp66 residue and 70s loop in the orange, cyan and magenta macrostates. **C-E** Validation of calculated ensemble against different NMR observables and comparison with two other published NMR ensembles of the structure and the X-ray structure of CypA. **C** Comparing agreement of measured and backcalculated NOE and eNOEs. **D** RMSD between measured and back calculated ^{3}J scalar couplings. **E** Back calculated Q-factor value comparison for NH residual dipolar couplings (RDCs). Figure adapted from (*184*) and made by Jordi Juárez Jiménez.

the loop and increased dyanmics. Finally, in the open form of the loop, as exemplified by the magenta macrostate, no interactions occur between Asp66 and the backbone residues of the loop allowing it to adopt a much more flexible disordered structure.

While the discussion above demonstrates the detailed atomistic information that can be provided by the MD and MSM approach, it is essential that the computed structural ensemble is experimentally validated. In order to do this the NMR observables: NOEs, exact NOEs (eNOEs)(*187*), three-bond scalar couplings (^{3}J) and NH residual dipolar couplings (RDCs)(*188*) are back calculated from the MD ensemble and compared to experimentally measured values (Fig 5.3**C-E**). It is important to note that the simulations are not refined against these observables and they therefore can provide an independent measure of the quality of the ensemble. The performance of the MD ensemble is also compared to the X-ray structure of CypA(*189*) as well as two previously determined NMR structural ensembles calculated using the software packages DYANA, composed of 20 structures(*190*), and CYANA, composed of 40 structures(*188*), respectively. The CYANA ensemble is refined using the NMR observables shown here.

NOEs result from dipolar relaxation between spins that are located close in space and can thus be used as structural restraints(*191*). While conventional NOEs are usually interpreted in a semi-quantitative way, eNOEs are obtained from NOE build-up curves and can take account of spin diffusion effects to provide much more precise upper and lower limits for distances between spins(*192*). In the present case, we find that the MD ensemble is able to satisfy 90% and 95% of eNOEs and NOEs, respectively. This is similar to the degree of agreement observed with the two NMR ensembles and better than the X-ray structure.

Three-bond scalar couplings are known to be sensitive reporters of protein backbone torsion angles(*193*). Dihedral angles were calculated from the MD ensemble, from which three-bond scalar couplings were computed in a population weighted manner using the Karplus equation(*11*) for backbone HN-HA, HN-C and HN-CB pairs. For HN-HA couplings, the discrepancy between measured and computed scalar couplings are similar for the X-ray and DYANA ensemble and

very slightly improved in the MD ensemble. The agreement is much better for the CYANA ensemble that is refined against these values. On the other hand, for HN-C couplings, the accuracy of the MD and NMR ensembles is very similar while it is slightly worse in the X-ray structure. Finally, the HN-CB couplings show similar behaviour to the HN-HA pairs, with the CYANA ensemble able to reproduce the experimental values with higher accuracy.

The final experimental parameter used to assess the computed ensemble is the residual dipolar coupling (RDC). Unlike NOEs and J-coupling that report on local structural features, the RDC is a very useful property in NMR structural determination or validation as it can report on more global features of a protein's 3D structure, for example the relative alignment between two domains(*194*). Usually, rapid molecular tumbling results in dipolar couplings averaging to zero. However, in the presence of an anisotropic environment, such as an oriented liquid crystal phase, provided the alignment is kept sufficiently weak (approximately 1 in 1000 solute molecules is aligned), a series of RDCs can be measured that provide structural information while maintaining the overall simplicity of the spectra. The Q-factor, shown in Fig 5.3**E**, measures the deviation between experimental and back calculated RDCs (similar to a χ^2). For the ensembles considered here, the CYANA ensemble that is refined against the RDCs shows better agreement with the experimental RDCs while the other ensembles that are not refined against them show very similar levels of accuracy.

However, in all cases the Q factors are relatively high demonstrating that none of the structural ensembles fully capture the solution-state behaviour of CypA. This emphasises the importance of using an independent method to validate the predictions of the ensembles. It should also been noted in the literature that relatively small changes in the alignment angle can considerably alter the measured Q factors(*195*).

Overall, the quality of the computed MD ensemble is deemed promising according the experimental metrics considered here. The MD ensemble shows similar levels of accuracy to previously determined NMR structural ensembles and generally

provides better levels of agreement to experimental parameters than the X-ray structure. On the basis of this validation, the detailed description of the loop dynamics provided by the simulations is considered credible.

The simulations reveal that the Asp66 residue plays a pivotal role in controlling the dynamics of the 70s loop. This led to the hypothesis that mutating this residue to an Alanine that lacks a polar side group would result in a shift in the major macrostate of the protein towards the cyan intermediate form. As an initial probe of this, the same MD and MSM protocol described above is applied to D66A mutated CypA. As an additional control, the protocol is also applied to the H70A mutant: while the His70 residue is located close to Asp66, it has been previously studied by NMR experiments and shown to cause only minimal perturbations to the 70s loop dynamics(*178*).

The results of this analysis are shown in Figs 5.4**A** and **B** and are generally within expectations from insights gained from studying the WT loop dynamics. The loss of the hydrogen bonding interactions from the sidechain of the Asp66 residue to the 70s loop means that its 'closed' conformation in the D66A mutated CypA is significantly destabilised. The orange macrostate goes from being the most abundant in the WT form of the protein (40% population) to a relatively minor state in the D66A mutant (14% population). On the other hand, the cyan macrostate where the loops are in intermediate positions is considerably stabilised in the D66A mutant with its population increased 7-8 fold relative to the WT protein. Overall the 70s loop is in the 'closed' conformation 70% of the time in the WT but only 25% of the time in D66A mutated protein.

By contrast, the MSM for the H70A mutant is much more similar to the WT protein. Here, the orange macrostate remains the most populated are there is a slight stabilisation of the 'closed' conformation of the 70s loop overall (80% 'closed'). This is expected as there is little disruption to the hydrogen bonded array in the loop as a result of this mutation.

In Fig 5.4**C** we see the mean passage times between the different macrostates. This shows that transitions between macrostates for the WT and H70A proteins

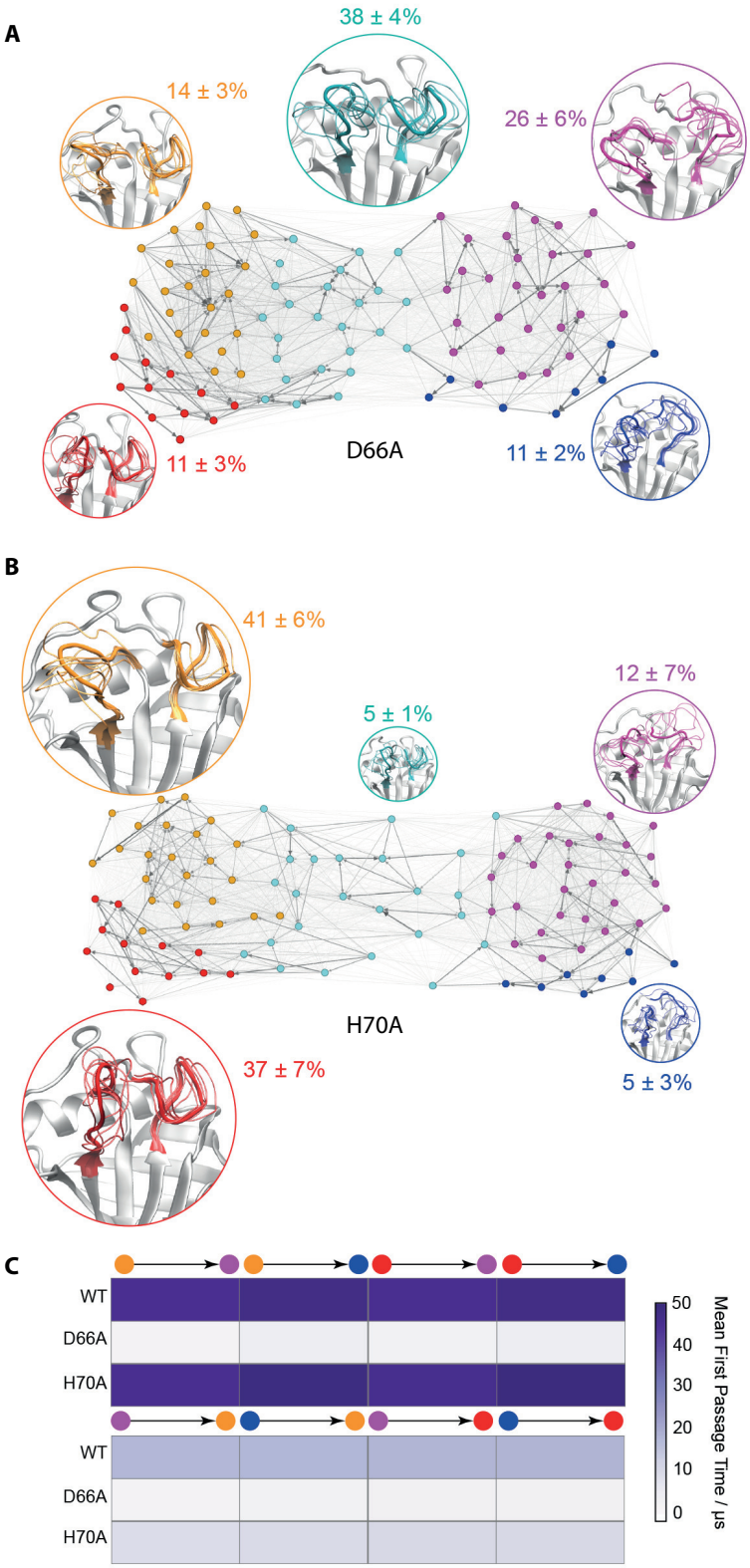

Figure 5.4: MSM for (**A**) D66A (**B**) and H70A mutated forms of CypA. Models have the same form as Fig 5.3**A**. **C** Mean passage times between macrostates for WT and mutated forms of CypA. Figure adapted from (*184*) and made by Jordi Juárez Jiménez.

occur on very similar timescales. In the D66A mutant however, all transitions occur much more quickly meaning that the 70s loop moves between ordered and disordered states more rapidly.

On the basis of the MD and MSM performed above we conclude that the rationally designed D66A mutation in CypA results in a dramatic perturbation in the energy landscape of the 70s loop. In order to experimentally probe these findings isotopically labelled samples of WT and D66A CypA are expressed in and purified from *E. coli* cells allowing the two forms of the protein to be compared by NMR.

5.2.2 HSQC Comparison of WT and D66A CypA

All experiments presented from this point in the chapter were performed by the author.

One of the key predictions of the MSM models is that the ground state conformation of the protein is shifted in the D66A mutant compared to the WT protein. We would expect this to be reflected in the ^{15}N-^{1}H HSQC spectra of the proteins, shown in Fig 5.5**A** as the chemical shift report predominantly on properties of the ground state. From inspection of the spectra, it is seen that CypA has a well dispersed spectrum reflective of a folded protein. The majority of residues in the D66A mutant resonate at identical frequencies to the WT protein indicating that the mutated protein largely retains its native structure. However, as shown, a significant proportion of residues also show very large chemical shift perturbations (CSPs), these are mostly clustered around the 70s loop, for example: Gly65, Thr68, Gly72 and Gly74. Indeed, initially in this study it was though that assignment of the D66A protein could be performed by transferral from the WT protein, but the presence of these large CSPs meant that this approach was abandoned and the D66A mutant was assigned from scratch using a double-labelled ^{15}N-^{13}C sample(*4*).

A per residue comparison of CSPs in D66A and H70A (taken from a previous study(*178*)) is shown in Fig 5.5**B**. This demonstrates how in the D66A mutant the largest CSPs are observed in the 70s loop while smaller perturbations occur around residues 50 and 110 that flank this region. The plot also shows that the CSPs observed in D66A are much higher than in the H70A mutant despite these residues

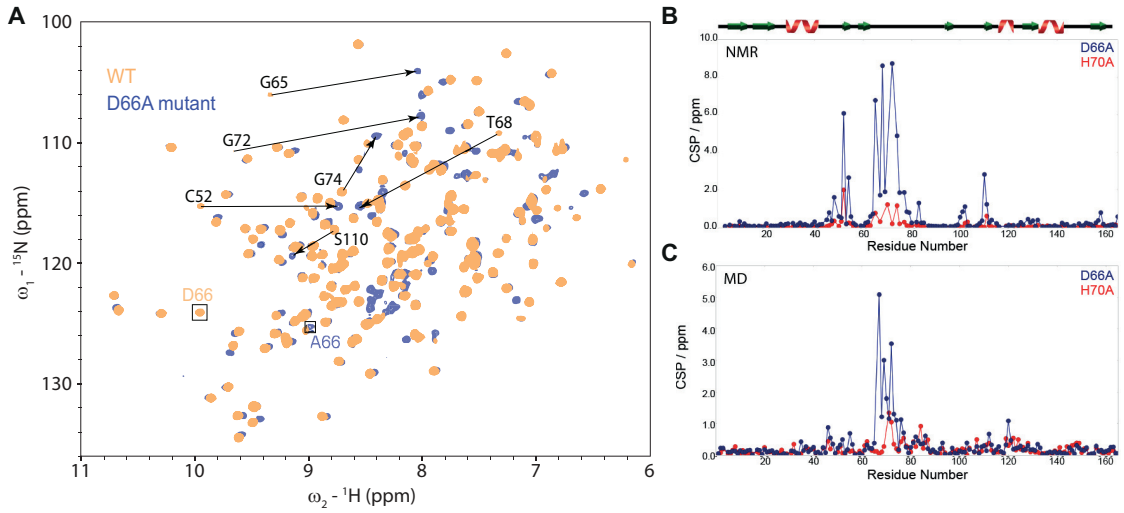

Figure 5.5: **A** Comparison of ^{15}N-^{1}H HSQC spectra for WT and D66A CypA. Spectra are acquired at 298 K on a 14.1 T (600 MHz) spectrometer. Residues showing particularly large chemical shift perturbations (CSPs) are highlighted. **B** Comparison of per residue CSPs for D66A and H70A mutants compared to WT CypA as calculated from NMR experiments. Cartoon indicates positions of secondary structure (α-helices, β-sheets and loops) within the protein structure. **C** Comparison of per residue CSPs for D66A and H70A mutants compared to WT CypA as calculated from MD structural ensembles.

being in close proximity. This highlights the potential importance of the Asp66 residue in determining the 70s loop dynamics as suggested by the MD simulations. Furthermore, other previously studied mutants R55A and K82A are found to induce CSPs of a very similar magnitude to H70A(*178*).

We also back calculate chemical shifts from MD structural ensembles, as shown in Fig 5.5**C**. The qualitative agreement with the actual CSPs is good: the largest CSPs are concentrated in the region of the 70s loop and much larger CSPs are observed for the D66A mutant that the H70A one. Complete quantitative agreement between the metrics is not present however and this is not unexpected: we have already established above that the MD ensemble is an incomplete description of the system, unable to reproduce all experimental features (Fig 5.3**C-E**). Furthermore, even with such an ensemble in hand, chemical shift predictors are known to not be 100% accurate(*196*).

A further insight from the MSMs is that the 70s loop in the D66A protein adopts a more disordered form. In order to test whether this is the case the measured CSPs between WT and D66A protein are compared to expected chemical shifts

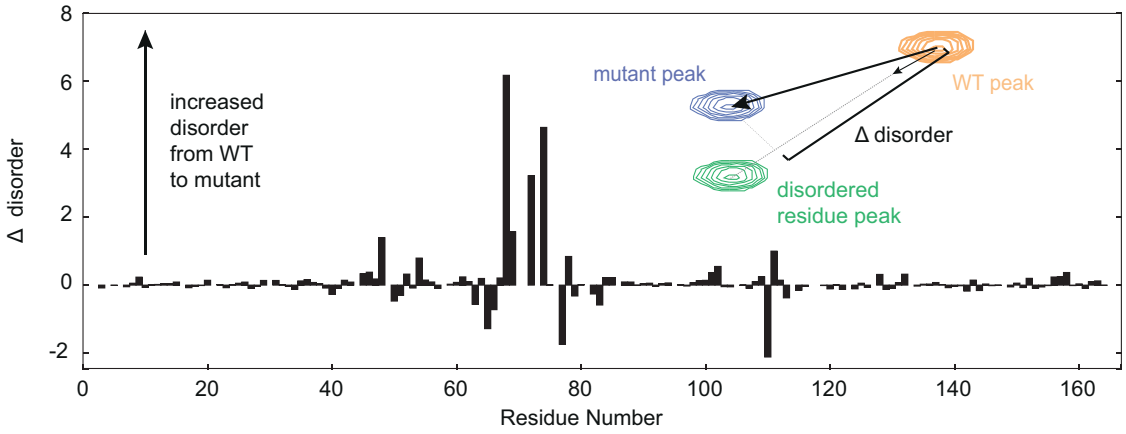

Figure 5.6: Charcterisation of increased disorder in CypA going from WT to D66A mutant. The level of disorder is parametrised by the variable Δdisorder that measures the projection of CSPs between WT and D66A on calculated CSPs between WT and disordered protein residues.

for an intrinsically disordered protein. Expected ^{15}N and ^{1}H chemical shifts for a disordered protein with the same amino acid sequence as CypA are calculated using the neighbour corrected IDP library(*197*). The relative change in order/disorder is then characterised by the projection of the actual CSP between WT and D66A mutated protein onto the calculated CSP between the WT and the predicted disordered resonance, as shown in Fig 5.6. The projection (labelled Δdisorder) is a signed value such that a positive value indicates a move towards the disordered residue, a negative value is a move away from the disordered residue and a zero value indicates there is no CSP or that the CSP occurs orthogonally to the position of the disordered peak. As shown in Fig 5.6, a number of residues around the 70s loop undergo large CSPs upon the D66A mutation towards more disordered chemical shifts, consistent with the insight from the MSMs.

The analysis of the HSQC spectra shown here shows that the D66A mutant induces large changes in the structure of the 70s loop of CypA, as predicted by the MD simulations and in contrast to other mutations that have only caused much smaller structural changes. The MD simulations also predict that this mutation significantly alters the dynamics of the 70s loops: in the next two subsections NMR experiments are carried out to probe different aspects of the protein's dynamics.

5.2.3 Fast Dynamics Comparison of WT and D66A CypA

Extensive work has been done in NMR to develop methods to characterise fast dynamics in molecules, defined as motions on the ps-ns timescale(*170*, *172*). A very brief introduction to this topic is presented before moving onto the results for CypA.

Analysis of fast dynamics involves the measurement of ^{15}N R_1 and R_2 rates as well as the steady state heteronuclear NOE (hetNOE) effect for each residue. The aim of a backbone fast dynamics analysis is to characterise both the overall rotational dynamics of the protein as well as the internal dynamics associated with each ^{15}N-^{1}H bond vector. In order to do this it is necessary to find the spectral density function associated with each bond vector that parametrises its rotational motion. The R_1, R_2 and hetNOE effect are in turn related to the spectral density functions through the following equations(*24*):

$$R_1 = \frac{d^2}{10}\Big(J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)\Big) + \frac{2c^2}{5}J(\omega_N) \tag{5.1}$$

$$\begin{aligned} R_2 = \frac{d^2}{20}\Big(4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)\Big) \\ + \frac{c^2}{15}\Big(4J(0) + 3J(\omega_N)\Big) + R_{ex} \end{aligned} \tag{5.2}$$

$$NOE = 1 + \frac{d^2}{10R_1}\frac{\gamma_H}{\gamma_N}\Big(6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)\Big) \tag{5.3}$$

In these equations $d = \frac{\mu_0 h \gamma_H \gamma_N}{8\pi^2 < r_{NH}^3 >}$ and $c = \frac{\omega_n \Delta\sigma}{\sqrt{3}}$ where r_{NH} is the nitrogen-hydrogen bond length and $\Delta\sigma$ is the chemical shift anistropy of the ^{15}N nuclei. Terms associated with the constant d represent relaxation due to dipolar couplings while terms associated with the constant c are due to relaxation from chemical shift anisotropy. In addition, R_2 relaxation also has a contribution from chemical exchange processes on the $\mu s - ms$ timescale, parametrised by the variable R_{ex}. In order to accurately extract this value it is necessary to record datasets at different magnetic field strengths due to the dependence of R_{ex} on the square of the static field strength(*198*) (this is also useful for providing further data points for fitting).

In total, the fast dynamics experiments give access to the value of the spectral density at five different frequencies: $J(0)$, $J(\omega_N)$, $J(\omega_H+\omega_N)$, $J(\omega_H)$ and $J(\omega_H-\omega_N)$.

Based on the dependencies of the different measurements, measured R_1 and R_2 rates and hetNOE effects may be interpreted in a purely phenomenological manner(*172*). This is particularly useful in the application here as we consider the same protein with a single point mutation and are most interested in how the respective measures change as a consequence of the mutation. This approach also avoids the assumptions and difficulties (particularly in model selection) inherent in more sophisticated analysis protocols as discussed below.

The main pitfall of such a phenomenological approach is that the relaxation rates and hetNOE effects in themselves lack specific and quantitative structural insight. In order to do this it is necessary to fit the data to a motional model, as expressed through the form of the spectral density function. The most common way of doing this is through a model-free analysis, originally proposed by Lipari and Szabo(*199*, *200*). This does not attempt to fit a specific structural model to the data (hence being referred to as model-free) but instead assumes that the spectral density function can be decoupled into independent contributions from (1) the overall tumbling of the molecule (this can be isotropic or anisotropic) and (2) independent internal motions that are faster than the overall rotational diffusion(*170*). Consequently, for isotropic tumbling, the spectral density function has the form(*199*, *200*):

$$J(\omega) = \frac{S^2\tau_c}{1+\omega^2\tau_c^2} + \frac{(1-S^2)\tau}{1+\omega^2\tau^2} \tag{5.4}$$

In this equation S^2 is the square of the generalised order parameter that parametrises the amplitude of internal motions, τ_c is the isotropic overall correlation time and τ_e is the effective internal correlation time with $\tau = (\frac{1}{\tau_c} + \frac{1}{\tau_e})^{-1}$. The S^2 variable takes on values between 0 and 1 can be considered in terms of the cone angle formed by the rotating bond vector: a completely rigid bond vector will have an S^2 value of 1 and more rigid bond vectors will be associated with smaller cone angles whereas a completely flexible bond vector will have an S^2 value of 0 and regions of lower S^2 will be associated with larger cone angles. The spectral density function can also be extended to the case where there are two independent internal motions as follows(*201*):

$$J(\omega) = \frac{S^2\tau_c}{1+\omega^2\tau_c^2} + \frac{(1-S_f^2)\tau_f'}{1+\omega^2\tau_f'^2} + \frac{(S_f^2-S^2)\tau_s'}{1+\omega^2\tau_s^2} \tag{5.5}$$

In this case $\tau_f' = \frac{\tau_f\tau_c}{\tau_f+\tau_c}$ and $\tau_s' = \frac{\tau_s\tau_c}{\tau_s+\tau_c}$ where τ_f and τ_s are effective correlation time for fast and slow internal motions respectively. Similarly, S_f^2 and S_s^2 are the squares of the generalised order parameters for fast and slow internal motions respectively with $S^2 = S_f^2 S_s^2$.

The forms of the spectral density functions clearly involve a large number of parameters. In a typical fast dynamics experiment the R_1, R_2s and hetNOE of each residue will be measured at a minimum of two static magnetic field strengths giving six data points per residue against which different forms of the spectral density function may be fit. A key challenge in the model-free analysis process lies in model selection: each residue as well as the molecule as a whole must be fit to a model that can adequately reproduce the relaxation data without overfitting it. In practice this is done by first finding the most appropriate diffusion tensor to characterise the overall rotation of the molecules and then fitting each residue to the spectral density function in equation 5.5 in different limits (for example, $S^2 = S_f^2$ or $\tau_f \to 0, \tau_e = \tau_s$). The importance of the R_{ex} term in fitting the relaxation data for each residue can also be considered. After fitting the data to different models, statistical tests (usually F-tests or the Aikaike information criterion) are used to assess which of these models provides the best statistical fit (taking into account the number of free parameters in each model). A number of software packages have been written to automate this fitting and model selection process(*202*, *203*).

The results of R_1, R_2 and hetNOE measurements made on WT and D66A CypA at 14.1 T (600 MHz) are shown in Fig 5.7 (these experiments are also carried out at 11.7 T (500 MHz) for the model-free analysis below). In general, the values seen across all three metrics are similar in the two forms of the protein, with most large changes restricted to the vicinity of the 70s loop. This is consistent with the HSQC spectra (Fig 5.5) of the proteins that show the protein retains its native fold following the mutation and that most large CSPs are located in the 70s loop. This is also consistent with the MSM insights (Figs 5.3 and 5.4) that

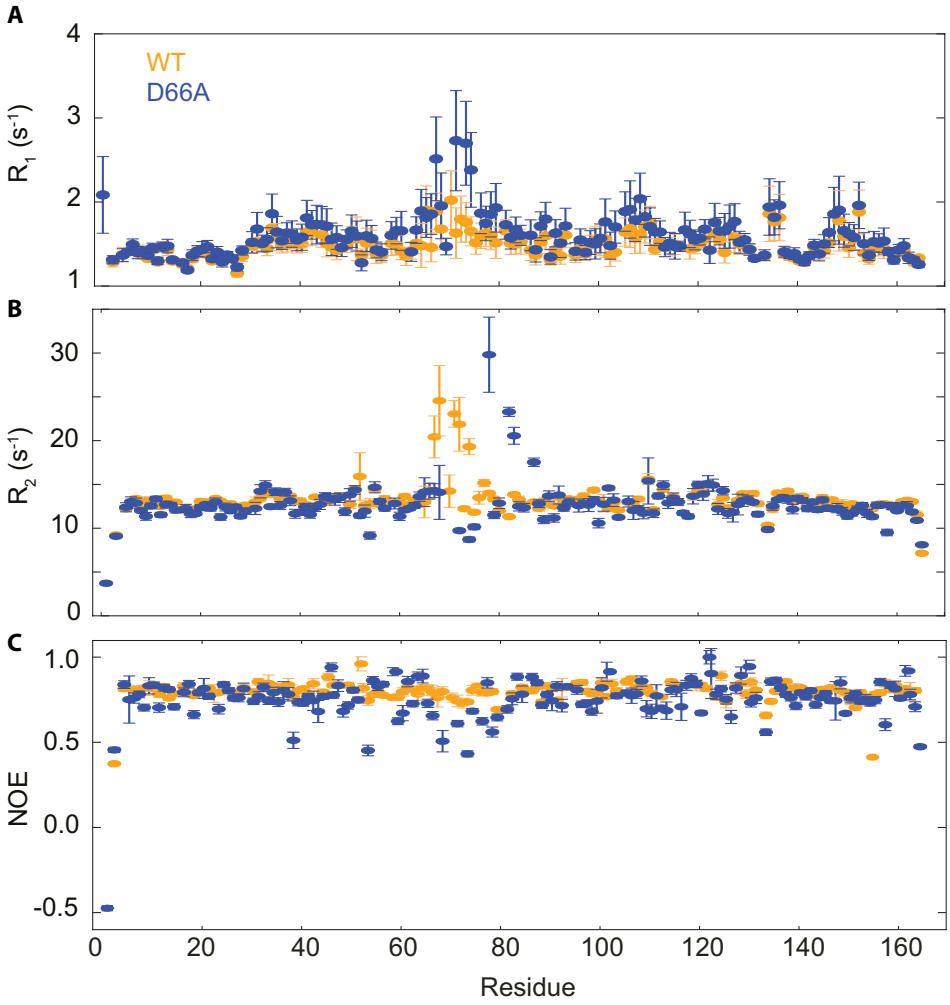

Figure 5.7: R_1 (**A**), R_2 (**B**) and hetNOE (**C**) measurements of WT and D66A CypA. All measurements are made at 298 K and results shown were performed at 14.1 T (600 MHz), but all experiments were also performed at 11.7 T (500 MHz). Experiments were performed using previously reported pulse sequences(*204*).

predict a significant perturbation in the dynamics of the 70s loop whilst retaining the overall folded integrity of the protein.

From the R_1 relaxation rates (5.7**A**), both forms of CypA show elevated longitudinal relaxation rates around the 70s loop, though this is to a greater extent in the D66A mutant. A secondary smaller rise in R_1 rates is also visible around the 100s loop. Since the R_1 rate is sensitive to the value of the spectral density function at non-zero frequencies (namely $J(\omega_N)$, $J(\omega_H - \omega_N)$) and $J(\omega_H + \omega_N)$), higher R_1 rates are indicative of greater motional flexibility. This highlights that (a) as expected the 70s and 100s loop regions are among the most flexible in the

protein and (b) the 70s loop region appears to be more flexible in the D66A mutant than the WT protein. This is consistent with the mean passage time prediction from the MSMs of the different protein constructs (Fig 5.4**C**).

Stark differences in the R_2 relaxation rates are observed for the two protein forms in the 70s loop region. In the WT form, many residues show significantly increased relaxation rates while in the D66A mutant this is not observed and some residues in this region actually show slightly depleted R_2 rates. Residues adjacent to the 70s loop (around Gly80) however do show significantly raised R_2 rates in the D66A mutant. Based on the above discussion that the 70s loop represents a particularly flexible region of the protein the observed rates are not a consequence of intrinsic relaxation effects but rather due to contributions to relaxation from μs-ms timescale motions. These motions have been previously recorded and characterised for WT CypA(*178*, *181*) and are interrogated in detail in the next subsection. The absence of elevated R_2 rates in the 70s loop for the D66A mutant suggests that these motions have been quenched reflecting the dramatic perturbation made to the energy landscape of the protein. The depressed R_2 rates are also consistent with the CSP analysis above that suggests that the D66A mutation increases disorder in the 70s loop. There is however evidence of the emergence of μs-ms timescale motions around the Gly80 residue in the mutated protein, that is not observed in the WT.

The hetNOE values (5.7**C**) paint a similar picture to the discussion for the R_1 rates. It has been shown previously(*205*) that for large molecules such as proteins, the hetNOE effect can be considered to be a function of a single frequency, $J(0.87\omega_H)$, giving: $NOE = 1 + \frac{d^2}{10R_1}\frac{\gamma_H}{\gamma_N}5J(0.87\omega_H)$. Given that the hetNOE is uniquely sensitive to high frequency motions it is especially sensitive to motions on the ps-ns timescale. Since the gyromagnetic ratios of ^{1}H and ^{15}N nuclei have the opposite sign, a smaller value for the hetNOE effect implies a more flexible residue. The results show that in both forms of the protein there are a cluster of residues in the 70s loop that show decreased NOE values, though this effect is more exaggerated in the D66A mutant. Like the R_1, this trend is consistent

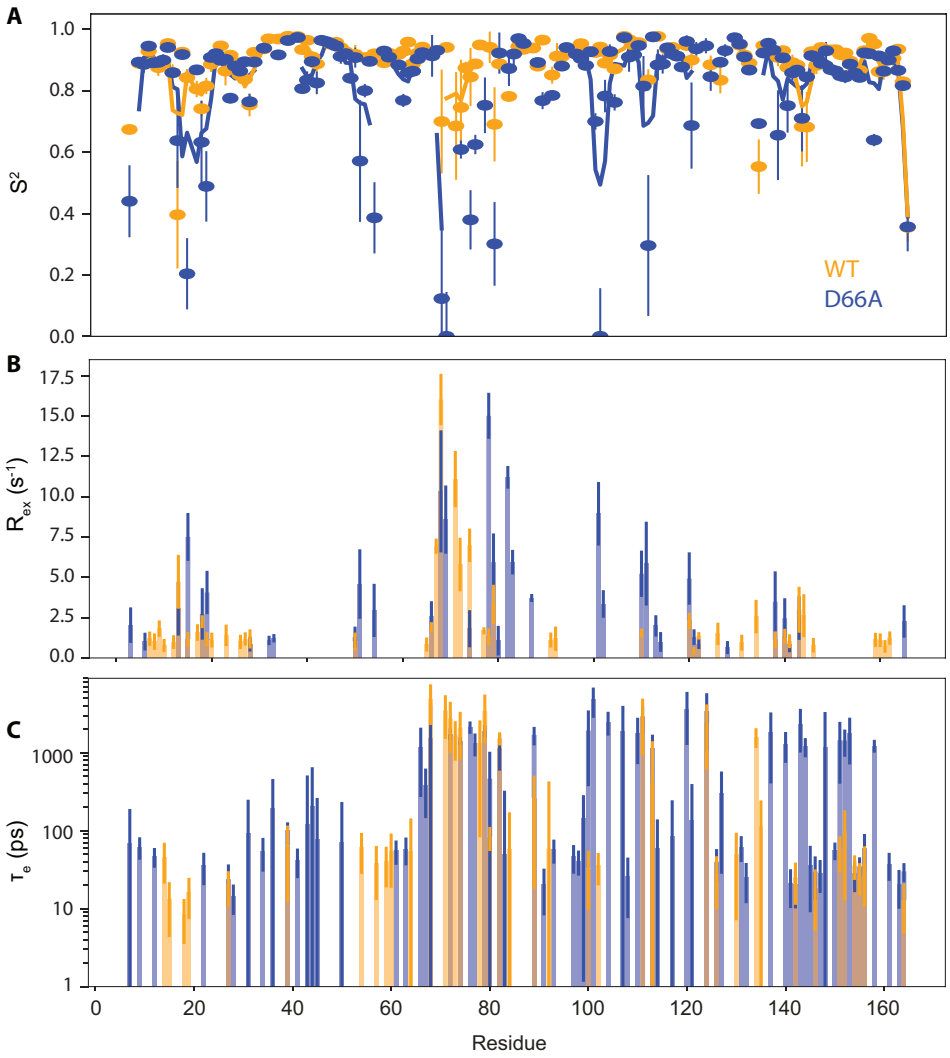

Figure 5.8: Results of model-free analysis of WT and D66A CypA carried out using the software package relax(*203*, *206*). Plots show fitted values for (**A**) the generalised order parameter, S^2, (**B**) chemical exchange contribution to relaxation, R_{ex}, and (**C**) effective correlation time for internal motions, τ_e.

with the MSMs that indicate a more disordered structure in the mutated protein with greater positional fluctuations.

Using the R_1, R_2 and hetNOE measurements a model-free analysis is carried out on the two proteins, the results of which are shown in Fig 5.8. The model-free analysis is carried out using the software package relax(*203*, *206*). The overall correlation times of the two protein forms are found to be very similar at 9.1 ns and 9.2 ns for the WT and D66A protein respectively. However, while the global diffusion of the WT protein is best characterised using an ellipsoid model with

three distinct rotational diffusion axes the mutant is found to diffuse as a spheroid with only two distinct rotational diffusion axes.

Fitted S^2 values from the model free analysis are shown in Fig 5.8**A**. The trends shown are in line with the raw relaxation rates, with the 70s loop identified as a particularly flexible region in both forms of the protein but the degree of flexibility greater in the D66A mutant. The fitted S^2 values, particularly for the mutant, show very large scatter however. This highlights an issue with the model-free analysis: in practice we expect that adjacent residues will show some degree of correlated dynamics such that more flexible and more rigid residues will form clusters. Correlated motion is not accounted for in the model-free analysis process, as each residue is treated independently, and this can lead to a physically implausible degree of scatter in the S^2 plot.

Fitted R_{ex} values are largely in line with expectations (Fig 5.8**B**). The largest R_{ex} values for the WT protein are observed in 70s loop and the D66A mutant shows smaller or quenched exchange contributions in this region. Significant R_{ex} contributions are however present for a cluster of residues around Gly80 in line with expectations from the R_2 rates.

The timescales of internal motions (τ_e) shown in Fig 5.8**C** show that there is a large diversity in the fitted values from 10s of ps to several ns. Very fast motions likely reflect fast random motions, while slower motions are mostly confined to loop regions, reflecting more substantial loop movements. It is important to note that a number of residues showing τ_e values approaching the overall tumbling time of the protein, τ_c, are also associated with very low S^2 values. The correlation between these two parameters makes the interpretation of S^2 values in terms of protein flexibility more difficult and also explains some of the scatter obtained in the S^2 plots. Between residues 1 and 60, where the fitted values of τ_e are two orders of magnitude lower than the overall correlation time, a large amount of scatter is still observed in the S^2 plot showing that the correlation between these two parameters is not the only problem in the model-free analysis here. Measurement of the R_1, R_2 and heteronuclear NOE rates for the WT and D66A mutant at a further static

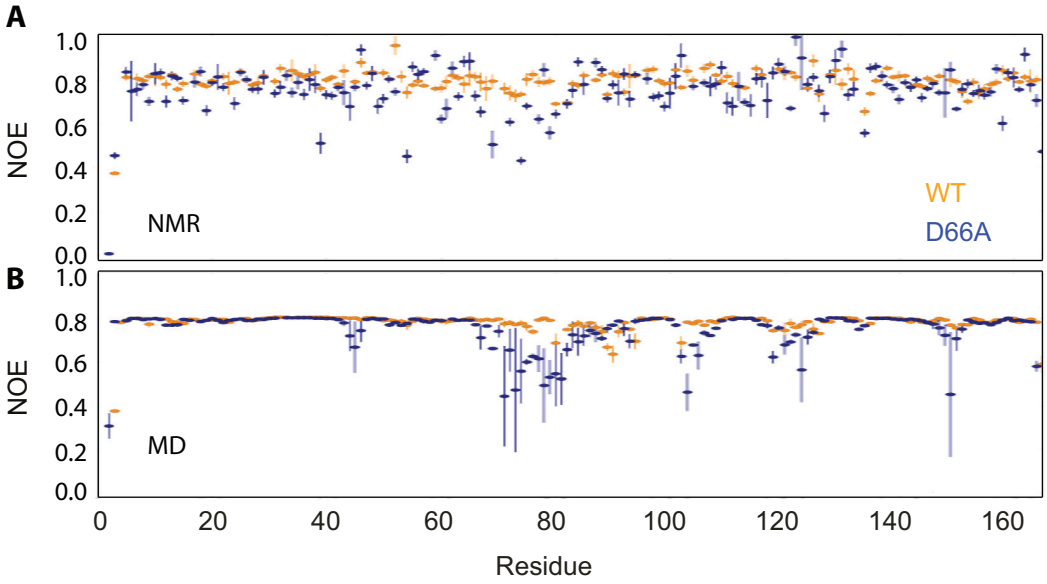

Figure 5.9: Comparison of (**A**) experimentally measured and (**B**) back calculated hetNOE values from MD simulations for WT and D66A CypA. Figure adapted from (*184*).

magnetic field strength and reanalysis of the data is suggested as future work to improve the quality of the model-free analysis.

Consequently, while it is most common to compare ps-ns dynamics through the square of the generalised order parameter, here we resort to back-calculation of the hetNOE values from the MD simulations as this is the most sensitive experimental metric for rapid motions. The comparison of experimental and back calculated hetNOEs is shown in Fig 5.9. There is good agreement between these two measures, with the hetNOE values generally consistent through most of the protein with the main difference occurring in the 70s loop where the more disordered and dynamic structure for the D66A mutant gives rise to smaller NOE values in this region.

5.2.4 Micro-millisecond timescale Dynamics comparison in WT and D66A CypA

Having established that the D66A mutant significantly alters the ground state structure and dynamics in the 70s loop of CypA, we now seek to discover its effect on longer timescale $\mu s - ms$ dynamics. In terms of the MSM constructions, dynamics on this timescale are expected to intersect with transitions between

different macrostates (Fig 5.3 and 5.4). The presence of exchange events on these timescales have already been revealed in the fast dynamics analysis shown above, however here we employ relaxation dispersion techniques to probe these events in much greater detail. This has been attempted previously for WT CypA, and this was actually one of the early examples of the use of these NMR techniques to provide detailed insight of such dynamical events(*178*).

Dynamics on $\mu s - ms$ timescales are investigated initially using ^{15}N CPMG experiments on the WT and D66A forms of CypA. An important aspect of the CPMG experiment however is that it is only able to characterise exchange rates when k_{ex} is in the approximate range 200-3000 s^{-1} (though this can be extended in some circumstances(*207*, *208*)) and the population of the excited state exceeds 0.5% (for two site exchange k_{ex} is the sum of the forward and backward rates)(*34*, *209*). In order to bring the dynamics present in WT CypA into this CPMG-active range, it is necessary to perform the CPMG experiments at 283 K, as was done previously(*178*).

The results of the ^{15}N CPMG experiments performed on a 14.1 T (600 MHz) spectrometer are shown in Fig 5.10. For fitting the CPMG data to a two state model (discussed below) the experiments are also performed at 11.7 T (500 MHz). The contribution of micro-millisecond timescale dynamics to the R_2 relaxation rate can be characterised using the approximate measure R_{ex}^* as shown in Fig 5.10**A**. This is calculated by taking the difference between the measured $R_{2,eff}$ rates at a CPMG pulsing frequency of 50 and 1000 Hz (maximum frequency) and assumes that the exchange contribution to the relaxation rate is 'pulsed out' at the highest CPMG pulsing frequency. The results show that in the WT protein very large R_{ex}^* values are observed in the 70s loop but are not observed in the corresponding residues in the D66A. This is further highlighted in some of the exemplar dispersion profiles shown in Fig 5.10**B**. For the residues Asp66, Gly74 and Ser77 for example that show very large dispersions in the WT protein the effect is almost quenched in the mutant. On the other hand, the residues Gly80, Asn87 and Glu120 reveal dispersions of a similar magnitude for both forms of the protein (and larger in the

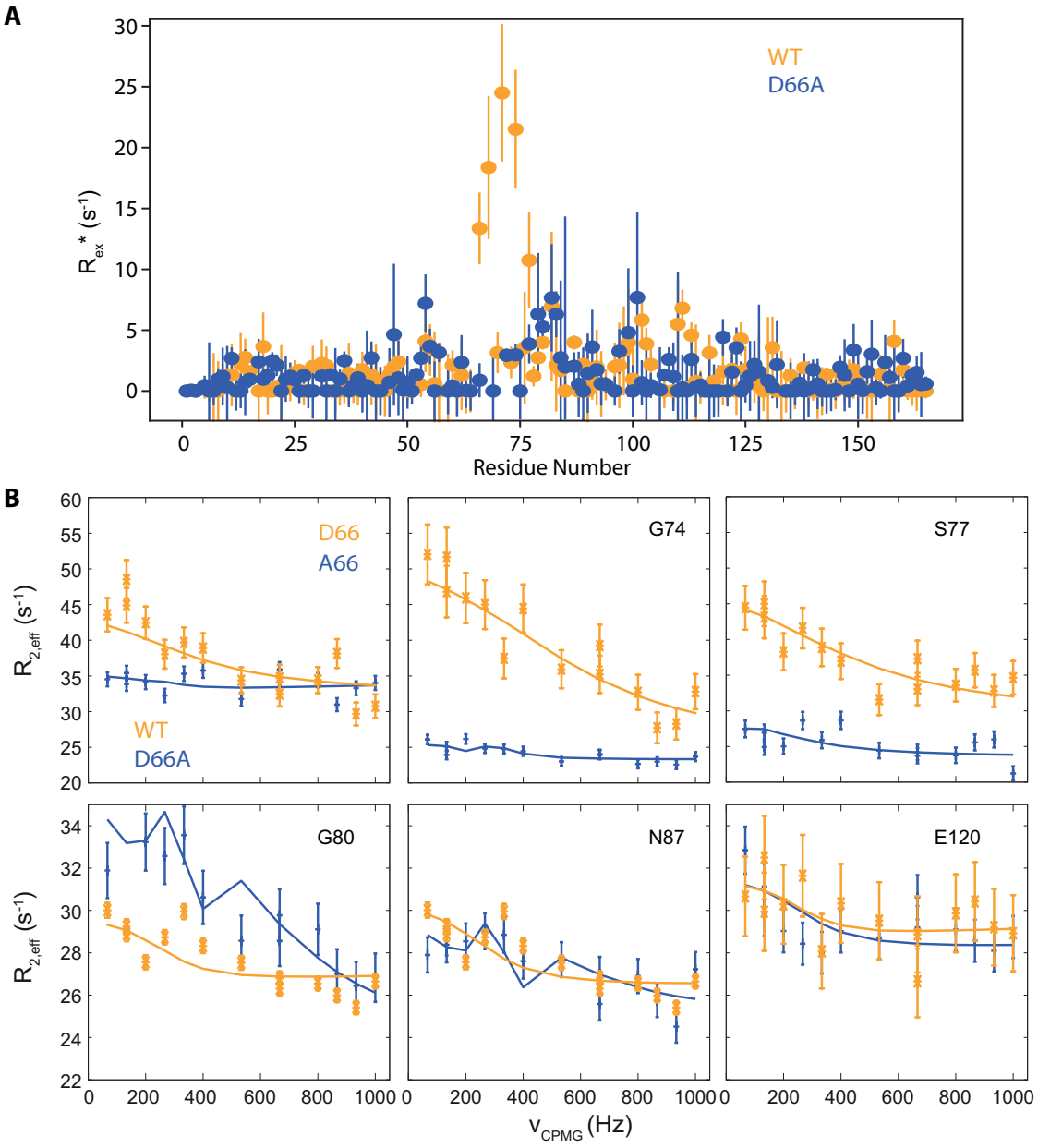

Figure 5.10: **A** Calculated R_{ex}^{*} from CPMG relaxation dispersion profiles. The parameter R_{ex}^{*} estimates the exchange contribution to relaxation by taking the difference between $R_{2,eff}$ measured at CPMG pulsing frequencies of 50 and 1000 Hz. **B** Exemplar CPMG dispersion profiles for WT and D66A CypA recorded at 283 K on a 14.1 T (600 MHz) spectrometer.

mutant for the Gly80 resiude). This demonstrates that CPMG timescale sensitive motions are still present in the mutated protein to some extent.

The discussion above is useful for providing qualitative insight into micro-millisecond timescale dynamics the key benefit of recording full CPMG dispersion profiles is that they can be fit to exchange models that provide much greater quantitative insight. Here we find that the relaxation dispersion data can be globally fit to a two-site exchange model: the free parameters in such a model are k_{ex} and P_B (population of excited state) parameters that are common to all exchanging residues as well as a per residue $|\Delta\omega_N|$ value that gives the absolute difference in ^{15}N chemical shift between the ground and excited states. While several closed form equations exist for fitting CPMG data(*210*, *211*), the approach employed here involves fitting the CPMG data based on numerical simulation of the Bloch-McConnell equations as implemented in the software package CATIA(*212*). The advantage of this approach is that additional relevant spin physics including off-resonance pulse effects and scalar couplings are included in the calculation of CPMG profiles, allowing fitted parameters to be determined more accurately. In addition to CPMG dispersion curves, in-phase (R_{Nz}) and antiphase (R_{NzHz}) longitudinal relaxation rates are measured at 14.1 T and used as inputs into the fitting process.

We found that for WT CypA, CPMG relaxation dispersion data recorded at field strengths of 14.1 T and 11.7 T could be satisfactorily fit to a global two state exchange model. The fitted k_{ex} is 2150$\pm$90 s^{-1} with an excited state population (P_B) of $2.1 \pm 0.1\%$. The fitted exchange rates and excited state populations are in broad agreement with a previous ^{15}N CPMG on the WT protein study by Eisenmesser et al(*178*). However, in this study it is proposed that the residues should be split into two clusters that are analysed independently. These, clusters termed Group 1 and Group are found to have k_{ex} rates of 1140 ± 200 and 2260 ± 300 s^{-1}(*178*). We note that given the errors associated with the fitted k_{ex} the differences in exchange rates between these two groups is small. Also, there are two significant differences between the data and analysis processes used previously and herein. Firstly, the dispersion curves in the Eisenmesser study are fitted using the Carver-Richards formula(*210*).

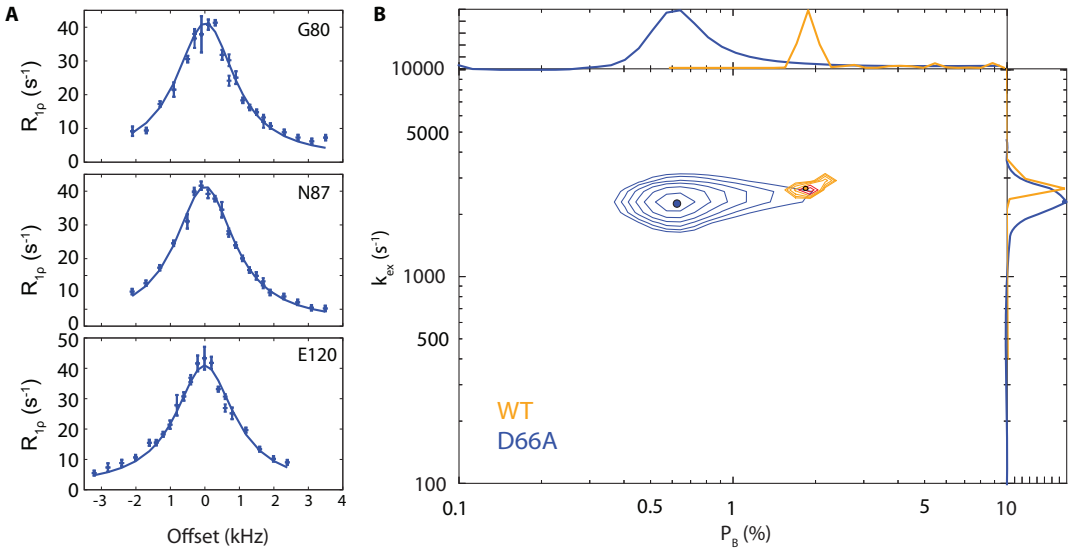

Figure 5.11: **A** Exemplar $R_{1\rho}$ profiles recorded on D66A mutant on a 14.1 T spectrometer at 283 K. Solid lines represent fits to the data from numerical simulation of Bloch-McConnell equations for two site exchange in an applied B_1 field. **B** Grid searching errors in k_{ex} and P_B from global fits of relaxation dispersion data for WT and D66A mutant.

As mentioned above, this approach ignores the effects of off-resonance pulses and scalar couplings that will likely have a small but potentially significant effect on the quality of extracted parameters. More importantly, the results in the Eisenmesser study are based on CPMG experiments performed at a single static field strength (14.1 T) and it is known that significant correlations between fitted parameters can occur when doing this (particularly k_{ex} and $|\Delta\omega_N|$)(*171*). By contrast, the results presented here are based on fitting the data to a model that includes all the relevant spin physics and the data is collected at two field strengths, giving us confidence in the extracted parameters. Moreover, grid searching the k_{ex} and P_B parameter space around the best fit value, as shown in Fig 5.11**B** reveals the presence of a primary minima in the error surface, centred on the positions of the extracted parameters.

While the WT CypA relaxation dispersion data can be fit to a two-state global exchange model, attempting this on the D66A mutant protein results in very large errors in fitted values for k_{ex} and P_B (not shown here). As observed in Fig 5.10, although some dispersions are clearly present in the D66A mutant the magnitude of these is generally much lower than in the WT protein making reliable extraction

of model parameters more difficult. In order to address this we seek an additional experimental restraint that can be used in the fitting process.

Here, we find that off-resonance $R_{1\rho}$ experiments can be used in complement to CPMG, and these datasets can be analysed globally to determine model parameters more precisely when CPMG data alone is insufficient to provide precise fits. Usually, $R_{1\rho}$ experiments are used to characterise exchange events on faster timescales than are accessible to CPMG methods(*172*). This is because it is possible to reach much higher effective applied fields in the $R_{1\rho}$ experiments that will modulate more rapid exchange processes, whereas probe limitations and sample heating considerations limit the pulsing frequency in CPMG experiments. Previous studies have shown that off-resonance $R_{1\rho}$ using relatively weak B_1 field strengths can be used to probe exchange events in slower regimes (and within the CPMG-sensitive window)(*213*). As shown here, data from the two experimental methods can be analysed simultaneously to parametrise challenging exchange dynamics more precisely.

Exemplar $R_{1\rho}$ relaxation profiles recorded on the D66A mutant are shown in Fig 5.11**A**. The $R_{1\rho}$ data are recorded at 283 K on a 14.1 T spectrometer as described previously, with adiabatic pulses employed to rotate the ^{15}N magnetisation from the z-axis to the appropriate effective field (further details provided in subsection 5.4.3)(*214*). In analogy to the CPMG analysis, a number of closed form equations exist for fitting $R_{1\rho}$ data to exchange models(*215*, *216*) but these all involve some level of approximation. Thus, as with the analysis of CPMG data, we therefore use a fitting protocol that numerically simulates the Bloch-McConnell equations(*9*) for two site exchange in the presence of an applied field (further described in subsection 5.4.3). This is implemented as an additional module in the CATIA software allowing the $R_{1\rho}$ profiles to be fit simultaneously with the two-field CPMG data.

The combined $R_{1\rho}$ and CPMG for the D66A can be fit to a global two state exchange model yielding $k_{ex} = 2200 \pm 350\ s^{-1}$ and $P_B = 0.7 \pm 0.1\%$. Grid searching errors in the k_{ex}-P_B plane, shown in Fig 5.11**B**, confirms that these parameters are well determined and the utility of the combined CPMG and $R_{1\rho}$

fitting approach. To the best of our knowledge, this represents the first example of characterising an exchange process through the simultaneous fitting of CPMG and off-resonance $R_{1\rho}$ datasets.

The extracted exchange rates and excited state populations for the WT and mutated protein (shown in Fig 5.11**B**) may give the impression that the exchange event being reported on is similar for the two proteins but it is important to emphasise that this is not the case. Firstly, the number and identity of residues that show significant dispersions is very different for the two forms: far more large R_{ex}^{*} values are seen in the WT protein and those residues that do show dispersions in the D66A mutant tend to be located adjacent to the 70s loop rather than in it, as is the case for WT protein.

This is further emphasised by comparison of the the $\Delta\omega_N$ values from the fits of the relaxation dispersion data, that gives the difference in ^{15}N chemical shifts between the ground and excited states. This parameter is particularly useful for assessing what structural changes take place during the exchange event. While CPMG experiments are only sensitive to the magnitude of the difference in chemical shift, another advantage of the $R_{1\rho}$ experiments is that they are also sensitive to the sign of the chemical shift. The positions of the residues showing the significant R_{ex}^{*} effects sorted by the size of the fitted $\Delta\omega_N$ values for WT and D66A mutated protein are shown in Fig 5.12**A** and **B** respectively. This highlights the significant difference in the position, number of and size of $\Delta\omega_N$ values between the two forms of the protein that reflect the different excited state sampled by the two proteins.

The fitted values for $\Delta\omega_N$ also provide an opportunity to interrogate the structure of the excited protein state in greater detail. The earlier comparison of the HSQC spectra of WT and D66A protein (Fig 5.5) revealed significant difference in their ground state structures. In Fig 5.12**C**, we compare the absolute differences in ^{15}N chemical shifts in WT and D66A mutant from the HSQC spectra to fitted $|\Delta\omega_N|$ from CPMG relaxation dispersion. While the correlation between these parameters is clearly imperfect, this analysis is consistent with the ground to excited state WT transition representing an ordered to more disordered structural rearrangement

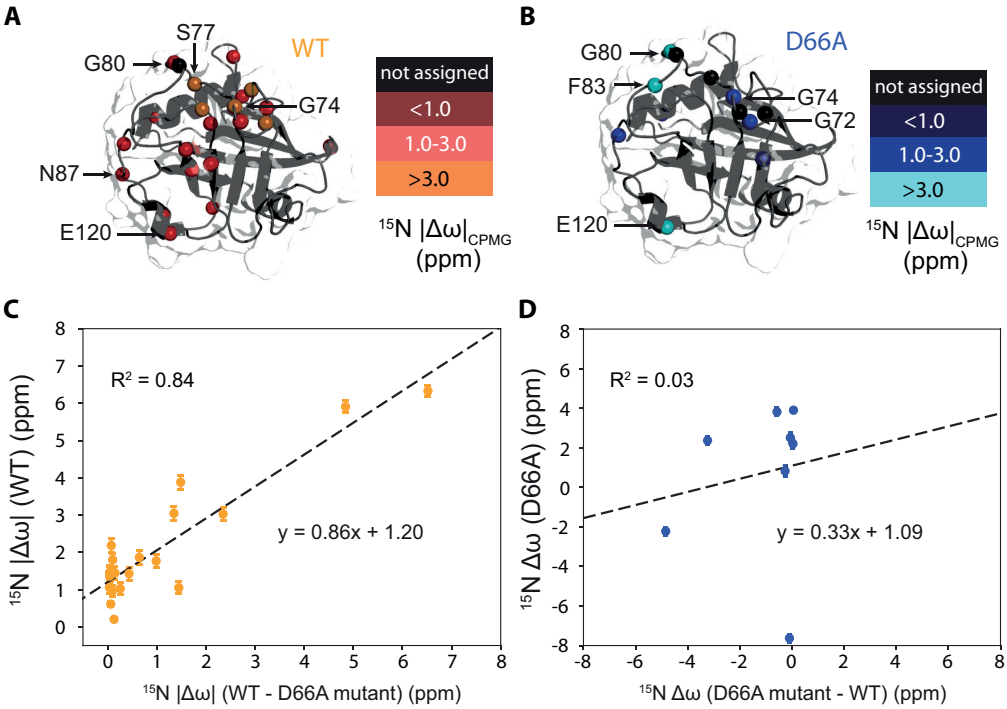

Figure 5.12: **A** Location of residues showing significant dispersions in the WT protein and indication of the size of $\Delta\omega_N$, reflecting the magnitude of structural change experienced by the residue. **B** Same as **A** but for D66A mutant. **C** Comparison of $|\Delta\omega_N|$ between WT and D66A measured from HSQC spectra at 283 K and $|\Delta\omega_N|$ determined from fits to CPMG data for WT protein. **D** Comparison of $\Delta\omega_N$ between WT and D66A measured from HSQC spectra at 283 K and $\Delta\omega_N$ obtained from fits of $R_{1\rho}$ and CPMG data for D66A.

in the 70s loop region. By contrast, no such correlation is observed between the fitted $\Delta\omega_N$ values for the D66A mutant and the difference in observed HSQC chemical shifts, suggesting that the exchange process in this case is unrelated to an order-disorder transition (for the mutant protein the $R_{1\rho}$ experiments give access to the sign of $\Delta\omega_N$).

5.3 Discussion

Taken together, the results in this chapter demonstrate how the rationally designed D66A mutant significantly perturbs both the structure and dynamics of the 70s loop region in CypA. The NMR results are generally in good agreement with the MD insights, demonstrating that the mutation causes the loop to adopt a more flexible and disordered structure. The results also suggest that the millisecond timescale

motions in WT CypA that can be probed using relaxation dispersion techniques are largely consistent with an order to disorder transition in the 70s loop region. The ability to gain insight into processes of this timescale using MD simulations, without the use of extremely specialised hardware, represents a significant development for this technique(*217*, *218*) and opens up greater possibilities for the design of drugs that bind specific conformations of proteins(*219*).

It is important to note the specific role of the Asp66 residue in controlling the 70s loop structure and dynamics as highlighted by the MD study. A previous study by Eisenmesser et al(*178*) probed the effects of R55A, H70A and K82A mutants on the 70s loop's structure and intrinsic dynamics. The CSPs caused by these mutations were all found to be far smaller than for the D66A mutant studied here (Fig 5.5**B**). Indeed, MD simulations and MSM of the H70A suggest its structure and dynamics are very similar to the WT protein (Fig 5.4**B**). Furthermore, relaxation dispersion experiments on these mutants indicate chemical exchange occurs at similar sets of residues in all cases with very similar values for k_{ex}, but with the amplitude of dispersions altered by nearly constant values for all residues in each mutant. Put together, this suggests that in all cases the CPMG results in the Eisenmesser et al study report on the same excited state structure, but that the population of this state is altered between mutants. Conversely, the evidence presented here for the D66A mutant, strongly suggests that it undergoes an entirely different millisecond timescale motion to the order to disorder transition likely experienced by the WT protein.

A large discrepancy is however observed between the in the MD and NMR results in the excited state populations. While NMR relaxation dispersion data is usually fit to a phenomenological two or three state model, as required to adequately fit the data(*220*), MSMs constructed from MD simulations typically use a much larger set of states to describe protein dynamics(*186*). In order to reconcile these two approaches we can combine the populations from the WT MSM where the 70s loop is in a 'closed' conformation to give the ground state populations and all those where the 70s loop is in an 'open' conformation to give an excited state. Using this

approach we expect the populations of ground and excited state to be approximately 75% an 25% respectively compared to the 2% excited state population from the CPMG experiments. Clearly, the level of quantitative agreement between the simulations and NMR data is very poor. In a best case scenario, full quantitative agreement is not expected due to technical limitations in the simulations. For the simulated structural ensemble of CypA used here there are discrepancies between back calculated and measured experimental observables that severely limits its ability to provide quantitative insight into the loop dynamics.

However, other factors may also contribute to this apparent discrepancy. Different 'states' in NMR are defined according their relative $\Delta\omega$ values compared to the exchange rate, and are thus a function of the static field strength(*198*). Inspection of chemical shift distributions from snapshots sampled from the different macrostates show that in many cases there is significant overlap in chemical shift values of residues between different macrostates. Re-analysis of the MSMs, under the assumption that open 70s loop conformations with chemical shifts similar to closed 70s loop conformations cannot be distinguished from the ground state lead to populations for the ground and excited state of 91% and 9% respectively, in much closer agreement to the NMR results. This demonstrates how CPMG analyses may underestimate the populations of minor states owing to a degree of degeneracy of chemical shifts between different protein conformations. A comparison of the molecular dynamics and NMR results for WT and D66A protein is shown in Fig 5.13.

Overall, the results in this chapter demonstrate the potential of a combined NMR and MD approach to characterise the structure and dynamics of proteins with atomic level resolution. We have also shown how a single point mutation made at the appropriate residue can result in the dramatic alteration to the protein's free energy landscape. Such insights should be of significant utility in understanding protein structure-dynamics-function relationships as well as opening up further opportunities in bioengineering. However, the results here also demonstrate that cutting edge simulation methods still remain a long way from reliably providing quantitative

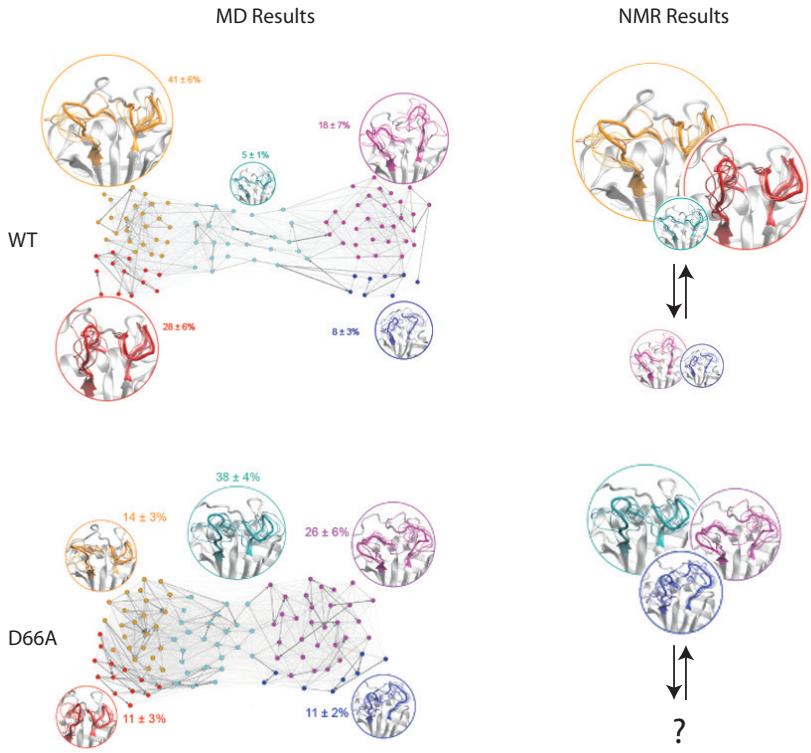

Figure 5.13: Figure giving a comparison of results for WT and D66A mutated protein from molecular dynamics simulations and NMR.

information on the dynamics of large biomolecules on milli-microsecond timescales and the importance of experimentally testing predictions of these simulations.

5.4 Methods and Materials

5.4.1 MD Methods

Only brief details of the MD simulations are provided here, the reader is referred to (*184*) for more detailed methods. Initial exploration of the protein's energy landscape is carried out using aMD simulations implemented in the PMEMD package(*185*). This provides an energy boost to the potential energy of the protein when it is above a certain threshold allowing barriers in the free energy surface to be traversed more easily. Structures sampled from the aMD trajectory were then used as initial structures to seed equilibrium MD trajectories. These were set up and conducted using appropriate modules of the GROMACS5.0 software package(*221*).

The MSMs were then built from the set of equilibrium MD trajectories using the pyEMMA 2.3.0 software package(*222*).

For calculation of NOE and eNOE effects from MD simulations the distance between atoms involved in NOE and eNOE signals was calculated for each structure in the ensemble using MDTRAJ(*223*). A histogram weighted by the probability of each microstate was then built and the average distance between atoms i and j was calculated as: $r_{i,j}^{avg} =< r_{i,j}^{-6} >^{-\frac{1}{6}}$. eNOEs are considered fulfilled if the predicted value was within 0.5 Å of the experimental limits while NOEs are fulfilled if they lie below the experimentally determined upper limit (Fig 5.3**C**).

For ^3J coupling constants, backbone ψ/ϕ angles are calculated with MDTRAJ for all structures in the ensemble. Karplus equations(*11*) are used to calculate ^3J couplings for each structure that are then weighted by their probability to give average predicted ^3J values (Fig 5.3**D**).

The best fit alignment tensor that provides the best fit of experimental residual dipolar couplings, for a subset of 10000 structures from the MD ensemble is fit using the software package PALES(*224*). The Q value from the fits of each of the structures is then calculated that is then weighted according to the probability of each microstate, giving an ensemble Q value (Fig 5.3**E**).

Predicted chemical shifts are calculated for individual microstates for both WT and D66A CypA. The MSM transition probabilities are converted into a rate matrix and the predicted chemical shifts used as inputs to calculate average chemical shifts for the two proteins. These can then be used to predict CSPs from the MD ensembles (Fig 5.5**C**).

Heteronuclear NOEs were calculated for simulated WT and D66A CypA ensembles using a previously published protocol (*225*). Briefly, for each protein a 2.5 μs MD simulation is carried out and snapshots are saved every 5 ps. The simulation is split into two 1.25 μs trajectories that are analysed separately to estimate the uncertainty in the calculated heteronuclear NOEs. The autocorrelation function ($C(t)$) is split into contributions from global tumbling ($C_{glob}(t)$) and internal

motions (C_{int}). To calculate C_{int} the MD snapshots are aligned to a reference frame (removing contributions from global tumbling) and calculated as:

$$C_{int} =< P_2[\mu(t) \cdot \mu(t+\tau)] > \tag{5.6}$$

Where $P_2(x)$ is the second Legendre polynomial, μ is the relevant bond vector scaled to unit magnitude, the maximum of time of τ is 100 ns and the angular brackets indicate averaging over time t. The total autocorrelation function, $C(t)$, is calculated by multiplying C_{int} by C_{glob} where $C_{glob} = \exp(-t/\tau_c)$ and τ_c is determined from the NMR relaxation measurements (9.1 ns and 9.2 ns for WT and D66A mutant respectively). The spectral density function $J(\omega)$ is then obtained from the autocorrelation functions by Fourier transform:

$$J(\omega) = \frac{2}{5} Re[\int_{-\infty}^{+\infty} C(t) e^{-i\omega t} dt] \tag{5.7}$$

The heteronuclear NOE is then calculated using equation 5.3 where ω_H and ω_N are the proton and nitrogen resonance frequencies on a 600 MHz spectrometer. Error bars are given by the standard error of the mean of the heteronuclear NOE values from the analysis of the two 1.25 μs trajectories.

5.4.2 Sample Preparation

WT CypA was prepared according to previously published protocols(*226*). For the D66A mutant, a directed CypA D66A mutant gene construct was synthesized by GeneArt (ThermoFisher Scientific) and was subsequently subcloned into a modified pET-15b vector with N-terminal hexa-histidine tag and a thrombin protease cleavage site. The construct is subsequently transformed into BL21 (DE3) pLysS host cells and the same expression and purification procedure as the WT protein is followed.

5.4.3 NMR Methods

^{15}N-^{1}H HSQC Experiments

^{15}N-^{1}H HSQC experiments were performed using a gradient enhanced pulse se-

quence(*19*). The following formula was used to calculate CSPs ($\Delta\omega$):

$$\Delta\omega = \sqrt{(5\Delta^1 H)^2 + (\Delta^{15} N)^2} \tag{5.8}$$

In this equation the factor of five accounts for the greater variation in chemical shift observed in the ^{15}N dimension.

^{15}N Backbone Relaxation Experiments

Backbone ^{15}N heteronuculear nuclear Overhauser enhancements (hetNOE) and R_1 and R_2 relaxation rate measurements were carried out using previously published pulse sequences(*204*) on 11.7 T (500 MHz) and 14.1 T (600 MHz) spectrometers equipped with a 5 mm z-axis gradient triple resonance room temperature probes at 298 K. For R_1 experiments seven relaxation times were employed in total: 0, 150, 250, 400, 500, 750 and 1000 ms with a repeat point at 250 ms for error estimation. For R_2 measurements seven relaxation times were also used: 0, 30, 60, 80, 90, 120 and 150 ms with a repeat point at 30 ms for error estimation. R_1 and R_2 relaxation times were found by fitting signal intensity as a function of time to an exponentially decaying function of the form $I = I_0 e^{-Rt}$. For hetNOE experiments, values were calculated by taking the ratio of peak intensities $\frac{I_{on}}{I_{off}}$ from experiments performed with and without ^{1}H saturation. All spectra were processed using NMRpipe(*101*) and cross peaks were fitted to determine signal intensities using FuDa(*227*).

Model-free Analysis

Model-free analysis was performed using R_1, R_2 and hetNOE values for WT and D66A CypA at 11.7 T and 14.1 T with the software package relax(*203*). Within this analysis, a sophisticated model selection protocol is used where the data is fit against a range of global diffusion models and a global tumbling time (τ_c) and individual residues are fit to models including the square of the generalized order parameter (S^2), effective internal correlation time (τ_e) and chemical exchange relaxation parameter (R_{ex}) as required to explain the data without overfitting it. For all residues the ^{15}N CSA is assumed to be -160 ppm and the ^{15}N-^{1}H bond length is set to 1.02 Å. The PDB coordinates of wt-CypA were utilised for diffusion tensor determination for both WT and D66A variants.

^{15}N Backbone Relaxation Dispersion Experiments

Relaxation compensated ^{15}N backbone CPMG experiments were carried out on 1.0-1.3mM ^{15}N-labelled wild-type and D66A CypA variant protein samples. Data were collected at 283 K using magnetic field strengths of 11.7 T (500 MHz) and 14.1 T (600 MHz). The constant time relaxation delay (T_{rel}) was set to 32 ms and spectra with 12 values of ν_{cpmg} were recorded with two additional repeat points acquired for error estimation as well as a reference plane where $T_{rel} = 0$. Cross peaks were fitted and peak intensities quantified using FuDa. Peak intensities were then converted into effective relaxation rates using $R_{2,eff} = -\frac{1}{T_{rel}} ln \frac{I_{\nu,CPMG}}{I_0}$ where I_0 is the peak intensity in the reference frame and $I_{\nu,CPMG}$ is the peak intensity at a particular CPMG pulsing frequency. Relaxation dispersion profiles can then be plotted with $R_{2,eff}$ as a function of ν_{cpmg}. Individual residue relaxation dispersion curves were evaluated for statistically significant dispersions for two-site exchange versus no exchange model (flat horizontal line). Residues that displayed appreciable exchange contributions were identified and clustered for global analysis using the numerical simulation software CATIA, which fits k_{ex} and P_B globally and a $\Delta\omega$ for each exchanging residue. R_{Nz} and R_{HzNz} rates were measured for the proteins and used as inputs for CATIA to determine the differential relaxation rates of in-phase and anti-phase magnetization. Individual residue peak positions were also used as inputs for CATIA so that the effects of off-resonance pulses are fully accounted for. Off-resonance ^{15}N $R_{1\rho}$ relaxation dispersion experiments on D66A CypA were recorded at 14.1 T (600 MHz) field strength at 283 K. The spin-lock field strength is set to 1042 Hz and 21 field offsets from -3000 to 3000 Hz from the ^{15}N carrier frequency are employed with three additional repeat points acquired for error estimation. The spin lock is applied for times 16, 48 and 80 ms. For each residue peaks are fit and intensities quantified using FuDa as a function of offset frequency and intensities are fit as a function of spin lock time to a single exponential to yield $R_{2,eff}$ as a function of offset frequency. $R_{1\rho}$ data is fit simultaneously with CPMG data for the D66A using a bespoke module written in CATIA to give robust characterization of the millisecond dynamics in this protein.

For two site exchange between a ground (G) and excited (E) state, the $R_{1\rho}$ analysis involves numerically simulating the following equation(*216*):

$$\frac{d}{dt}\begin{bmatrix} E \\ M_x^G \\ M_y^G \\ M_z^G \\ M_x^E \\ M_y^E \\ M_z^E \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -R_2 - k_{GE} & -\delta_G & 0 & k_{EG} & 0 & 0 \\ 0 & \delta_G & -R_2 - k_{GE} & -\omega_1 & 0 & k_{EG} & 0 \\ R_1 M_{eq}^G & 0 & \omega_1 & -R_1 - k_{GE} & 0 & 0 & k_{EG} \\ 0 & k_{GE} & 0 & 0 & -R_2 - k_{EG} & -\delta_G - \Delta\omega & 0 \\ 0 & 0 & k_{GE} & 0 & \delta_G + \Delta\omega & -R_2 - k_{EG} & -\omega_1 \\ R_1 M_{eq}^E & 0 & 0 & k_{GE} & 0 & \omega_1 & -R_1 - k_{EG} \end{bmatrix} \begin{bmatrix} E \\ M_x^G \\ M_y^G \\ M_z^G \\ M_x^E \\ M_y^E \\ M_z^E \end{bmatrix} \tag{5.9}$$

Where $\delta_j = \omega_j - \omega_{RF}, j \in \{G, E\}$, $\Delta\omega = \omega_E - \omega_G$, ω_1 is the strength of the applied field and k_{GE} and k_{EG} are the forward and backward rates of interconversion respectively.

5.5 Acknowledgements

All MD work in this chapter was performed by Jordi Juárez Jiménez and Julien Michel in the Department of Chemistry, University of Edinburgh. WT and D66A CypA samples were expressed and purified by Arun Gupta in the Department of Chemistry, University of Edinburgh. WT CypA assignments were provided by Xavier Hanouelle and Guy Lippens. Beat Vögeli provided eNOE, NOE, ^3J and RDC NMR observables. Assignment of D66A CypA was carried out by Arun Gupta. All other NMR experiments and analyses presented in this chapter were performed by the author.

6

Characterisation of Conformational Exchange in YopH

Abstract

YopH is a toxic effector protein employed by *Yersinia* species that is translocated into target eukaryotic cells via the type III secretion system. Prior to translocation the YopH protein first binds its cognate chaperone SycH. In this chapter, we use relaxation dispersion experiments to characterise the intrinsic conformational dynamics in YopH. A mutant is identified that stabilises the excited state conformation of YopH and facilitates tighter binding the chaperone, suggesting that the excited state is an important gateway conformation for chaperone binding.

The work presented in this chapter is the result of a collaboration with Arun Gupta, Alexander Schug, Magnus Wolf-Watz and colleagues in Karlruhe institute of Technology and Umeå Univeristy that has been published(228). The results presented herein focus on the author's contribution, namely, characterisation of intrinsic conformational exchange in the protein YopH.

6.1 Introduction

A key weapon in the arsenal of pathogenic bacteria to elicit a toxic response in host cells or disarm their immune response is through the export of effector proteins. In a number of gram-negative bacteria this is achieved through a multi-protein nanomachine known as the type III secretion system (T3SS) or injectisome(*229*). This system is associated with diseases including typhoid fever and bubonic plague(*230*).

The structure of the T3SS has been the subject of several investigations, particularly using the techniques of solid state NMR and cryo-electron microscopy(*231–233*). It consists of a basal structure made up of concentric ring proteins that span both the inner and outer membrane of the bacteria. This is connected to a stiff needle component that is capable of penetrating the membranes of infected host cells and provides a channel through which effector proteins can enter the cytoplasm of the host. A key aspect of the needle component is that is has an internal diameter of 20-30 Å meaning that effector proteins can only be transported through it in a fully or partially unfolded state(*231*).

In the *Yersinia* genus of bacteria (including *Yersinia psuedotuberculosis* responsible for Far East scarlet-like fever and *Yersinia pestidis* responsible for bubonic plague) effector proteins are referred to as *Yersinia* outer proteins or Yops. Before being presented to the T3SS complex, the core binding domain (CBD) of Yops form specific high-affinity complexes with their respective Yop chaperone protein, referred to as Sycs(*234*). A previous study, that serves as an important benchmark for the work here, considered the interaction between YopE and its cognate chaperone SycE. Monomeric YopE has an unfolded structure in isolation but is found to

form a 'wrap-around' structure in complex with dimeric SycE(*235*, *236*). The 'wrap-around' structures have been found to be conserved among T3SS substrates and chaperones, with the effector proteins forming extended polypeptides that are stabilised by hydrophobic patches on the surface of the chaperone dimer(*237*).

The Yop protein of particular interest in this chapter is YopH. In common with other Yop proteins, YopH is composed of three domains: an N-terminal secretion signal sequence(*238*), a CBD and a C-terminal catalytic domain. *In vivo*, the catalytic domain of YopH has been found to have phosphatase activity, associated with anti-phagocytic functions, while the CBD has been shown to be important in the recognition and binding of targets prior to dephosphorylation(*239*). Unlike YopE that is unfolded in its monomeric state however, monomeric YopH is known to adopt a compact globally folded structure(*240*). Crystallographic studies of a peptide from YopH^{CBD} comprising residues 21-63 suggest that YopH forms a wrap around complex with SycH, in line with other T3SS substrates, despite its folded structure. This is further confirmed by a low-resolution SAXS structure between YopH^{CBD} and SycH(*228*). We know that in order to be translocated into target eukaryotic cells YopH must undergo an unfolding event to fit through the injectisome, a process that in principle could occur before or after its interaction with its cognate chaperone. Fig 6.1 highlights the initial part of the translocation process that is under study in this chapter.

To provide more insight into this process we use NMR spectroscopy to study the monomeric YopH^{CBD} protein in greater detail. We focus specifically on the CBD as it is responsible for binding the Syc chaperones and as a 128-residue domain with a 15 kDa molecular weight it is well suited for ^{15}N NMR studies (as opposed to the full length protein that is composed of 468 residues). We identify that the YopH^{CBD} undergoes millisecond timescale exchange with an excited state that we interrogate using CPMG relaxation dispersion experiments. A D106A mutant is also identified in which the excited state conformation is stabilised relative to the WT protein. Isothermal titration calorimetry (ITC) is used to compare the

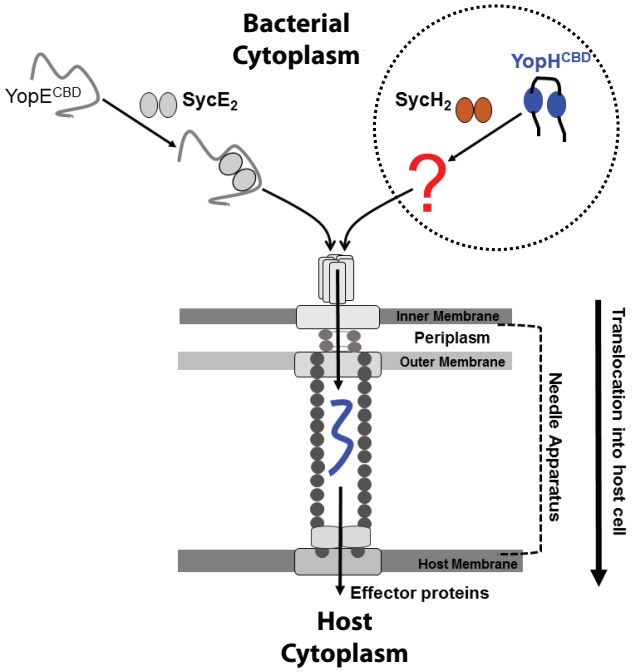

Figure 6.1: Representation of Type III secretion system (T3SS) used by toxic effector proteins such as YopH and YopE. Yop proteins bind with specific Syc chaperones prior to their presentation to the secretion system. The YopH system under investigation here is circled. Figure adapted from (*228*).

binding to SycH of the mutant and WT proteins that suggests the excited state conformation provides a route for binding the chaperone.

6.2 Results and Discussion

6.2.1 D106A mutant shows evidence of significant chemical exchange

Due to the globally folded structure of YopH^{CBD}, we hypothesise that it is in exchange with a globally expanded conformational state that facilitates its interaction with the SycH chaperone. In order to test this hypothesis a mutant form of the protein is sought that will modulate the chemical exchange process between the ground and excited states in YopH^{CBD}, following which the binding with SycH can be compared for the two protein forms. Mutation of the Asp106 residue to an alanine (D106A) is proposed to facilitate this as the residue is involved in a hydrogen bond with the Ile44 residue that is thought to stabilise the ground state structure.

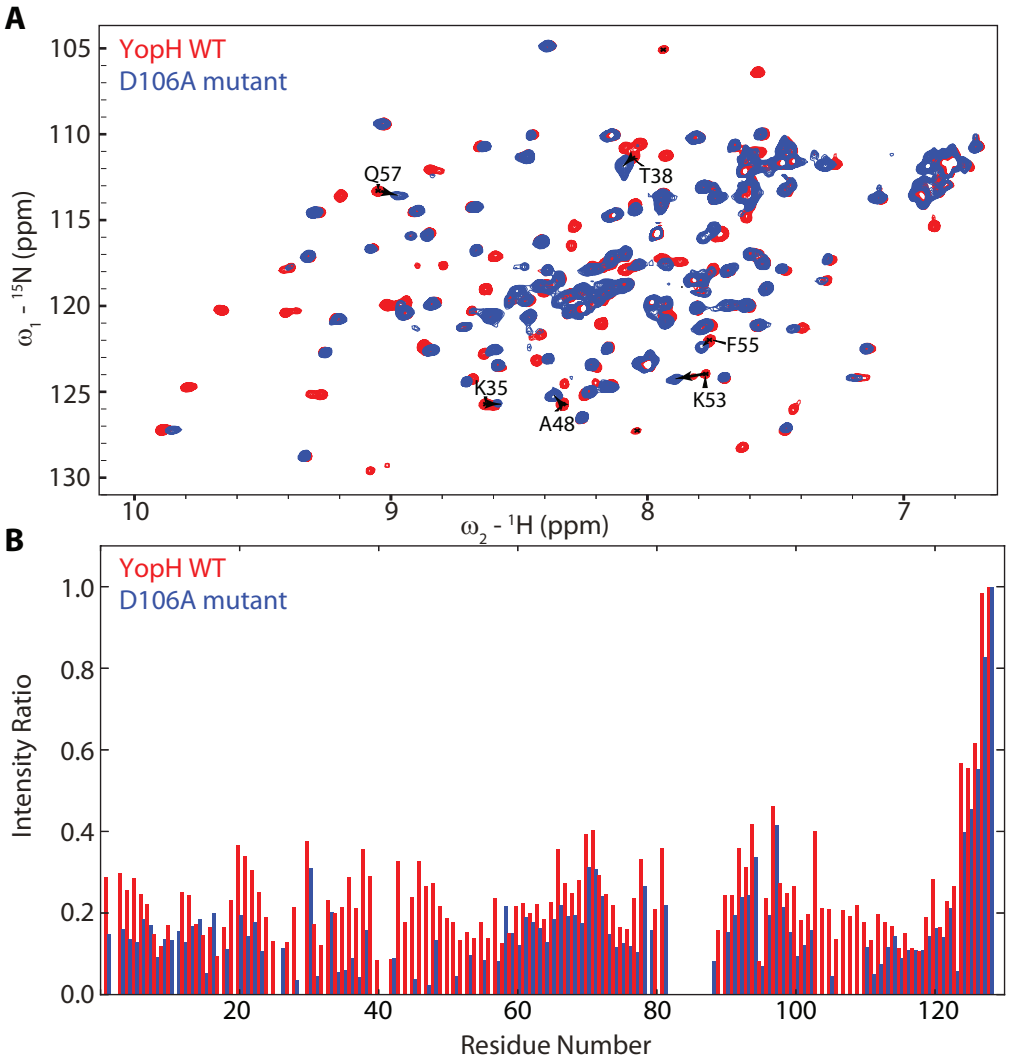

Figure 6.2: **A** ^{15}N-^{1}H sensitivity enhanced(*19*) HSQC spectra of WT and D106A YopH^{CBD} acquired on a 14.1 T spectrometer at 298 K. **B** Residue specific relative intensity of peaks in both WT and D106A YopH^{CBD}. In both cases, the intensities are normalised to the C-terminal residue (Arg128) that is the most intense in the spectrum.

The effect of the D106A mutation on the ^{15}N-^{1}H HSQC spectra of the protein is shown in Fig 6.2**A**. As expected for a well folded protein YopH^{CBD} shows a good deal of chemical shift dispersion in both frequency dimensions. While some small chemical shift perturbations are apparent between the WT and D106A mutant forms of the protein (as highlighted in the figure), the peaks are largely in very similar positions for the two forms suggesting it adopts very similar ground state conformations. The major difference between the two spectra however is that a large number of peaks that are present in the WT protein are broadened beyond

detection in the D106A mutant. This is very diagnostic of the presence of chemical exchange on the intermediate timescale, highlighting how the D106A mutant has significantly shifted the conformational dynamics in the protein.

A residue level comparison of relative intensities from the ^{15}N-^{1}H HSQC spectra is shown in Fig 6.2**B**. We observe here that although only a single point mutation is made to the protein, the WT form gives rise to more intense peaks throughout the protein. This suggests that the protein undergoes a global conformational change, at least in the D106A form. To interrogate this exchange event more quantitatively, we turn to ^{15}N CPMG relaxation dispersion experiments.

6.2.2 ^{15}N CPMG Relaxation Dispersion Experiments

Example relaxation dispersion profiles from ^{15}N-CPMG experiments are shown in Fig 6.3**A**. While dispersions are present in both forms of the protein, they are clearly much larger for the D106A mutant, as exemplified by the dispersions for residues Ala33 and Thr38. Furthermore, several residues that show large dispersions in the mutant protein, for example Ala48 and Ser51, show no evidence of dispersions in the WT protein.

This is further highlighted by the R^*_{ex} plot shown in Fig 6.3**B**. The parameter R^*_{ex} can be used as a proxy for the exchange contribution to relaxation by assuming that the exchange is fully quenched at the highest CPMG pulsing frequency. In the present case, we see that for most residues in the WT protein $R^*_{ex} = 0$, except for two small clusters around Thr37 and Asp106. By constrast, in D106A, large R^*_{ex} values are observed throughout the protein, consistent with the depleted signal intensities observed in the HSQC spectra of the mutant.

Fitting CPMG data can provide information on the kinetics (rates of interconversion), thermodynamics (populations) and structural changes (chemical shifts) that accompany a conformational exchange process(*28*). Here we find that both the WT and D106A mutants can be fit to a two-site global exchange model, as shown in Fig 6.4. The fits are performed using the numerical simulation software CATIA (*171*). For WT YopH^{CBD} the minor state population is $1.4 \pm 0.2\%$ and

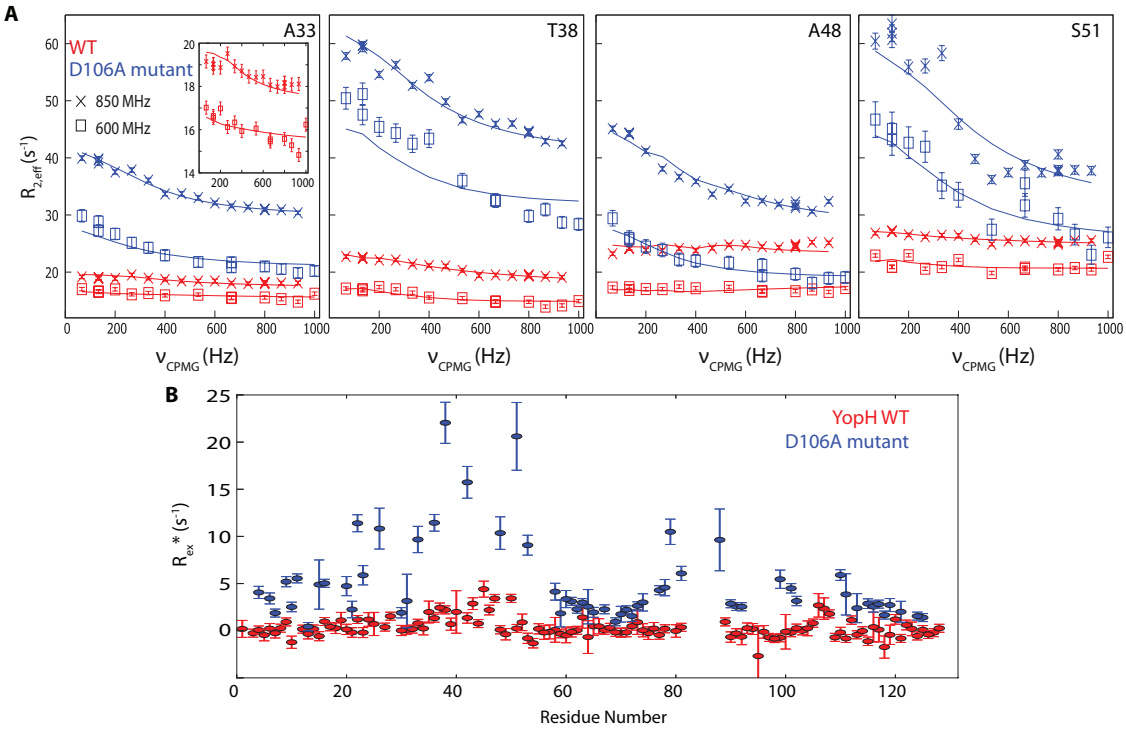

Figure 6.3: Example CPMG relaxation dispersion curves for WT and D106A YopH^{CBD}. All experiments are performed at 298 K on both 14.1 T (600 MHz) and 19.9 T (850 MHz) spectrometers for both forms of the protein and additionally on a 11.7 T spectrometer for WT only (not shown). **B** Calculated R_{ex}^{*} per residue at 14.1 T. This parameter is calculated by taking the difference of $R_{2,eff}$ at the lowest and highest CPMG pulsing frequencies.

the sum of the forward and reverse exchange rates, k_{ex}, is $3490 \pm 398\ s^{-1}$. On the other hand, in the D106A mutant the minor state population is $2.4 \pm 0.1\%$ and $k_{ex} = 1772 \pm 233\ s^{-1}$. The mutation thus has the effect of enriching the minor state population and slowing down the exchange process in the protein. Fig 6.4**B** shows the magnitude of the fitted $|\Delta\omega|$ values through the two forms of the proteins and in Figs 6.4**B** and **C** these are projected onto structure of the protein (PDB:1huf). Overall, we see that the exchanging residues all cluster to the top face of the WT protein (as drawn in 6.4**B**), while in the D106A mutant the largest $|\Delta\omega|$ values occur on this top face with smaller values occurring further down the protein. This suggests that the largest structural change occurs in this region of the protein with smaller rearrangements propagating through the structure.

Further insights into the effect of the mutation on the exchange process can be gained by calculating the specific forward and reverse exchange rates. For the WT,

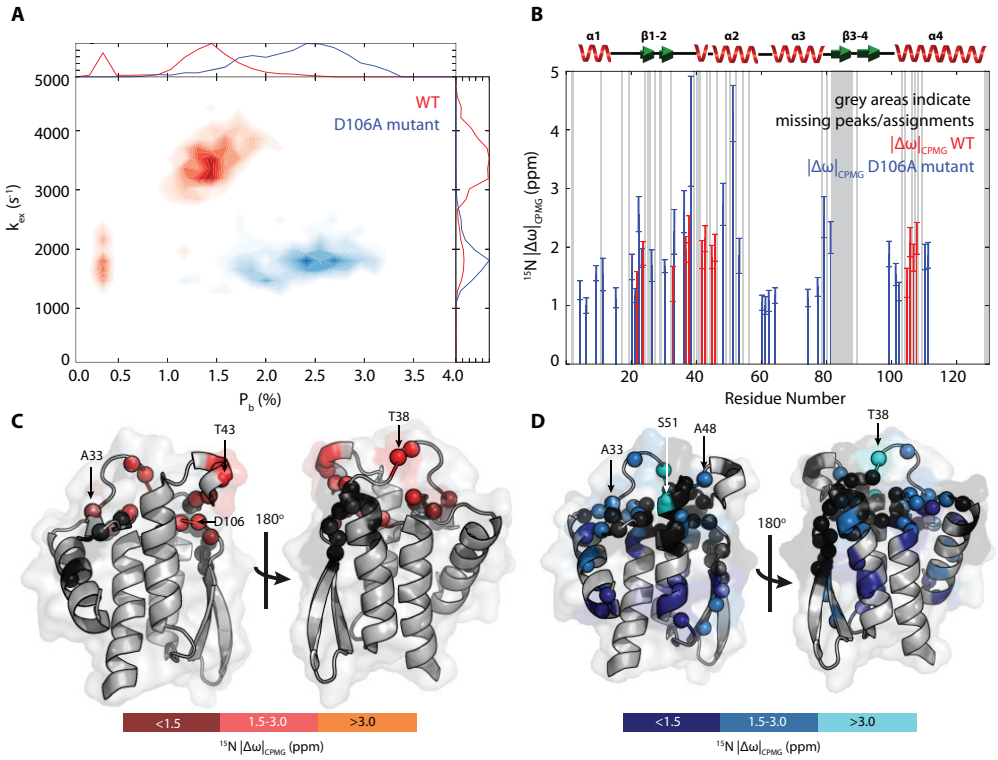

Figure 6.4: Results of fitting CPMG data to global two-site exchange model **A** Histogram results of bootstrap analysis performed on both CPMG datasets to obtain uncertainty estimates in the fitted parameters. **B** Fitted $|\Delta\omega|$ values across the protein, with cartoon illustration of secondary structure elements. **C** and **D** Fitted $|\Delta\omega|$ values for WT and D106A respectively projected onto structure of YopH^{CBD} (PDB:1huf).

the forward reaction rate is 49 s^{-1} and the reverse reaction rate is 3440 s^{-1} while for the D106A mutant the rates are found to be 42 s^{-1} and 1730 s^{-1} respectively. The similarity in the forward exchange rate and two fold reduction in the reverse rate show that the mutation results in a specific stabilisation of the excited state relative to the ground state, while the energy barrier for the forward reaction remains roughly constant. This is illustrated schematically in Fig 6.5.

The CPMG experiments reveal that the D106A mutation results in a relative stabilisation and elevated population of the excited state in YopH^{CBD}. The equilibrium dissociation constant between YopH^{CBD} and SycH can help determine whether the excited state conformation is important for chaperone binding. Apparent equilibrium dissociation constants (K_d^{app}) for the two forms of the protein with SycH were determined using isothermal titration calorimetry (ITC). K_d^{app} for YopH^{CBD} WT and D106A mutant are found to be 76 ± 9.3 nM and 41 ± 26 nM respectively.

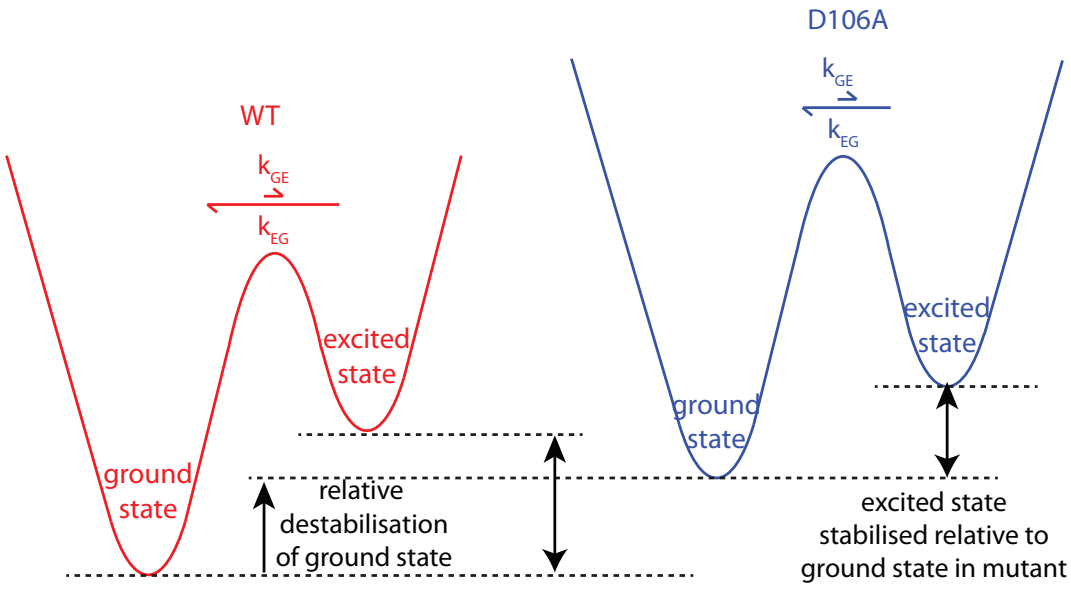

Figure 6.5: Schematic illustration showing the relative free energies of the WT and D106A forms of the protein. In the mutant, the excited state is stabilised relative to the ground state resulting in a slower rate of the reverse reaction. However, the most important factor in the observed binding results is that the mutant ground state is stabilised relative to the WT ground state causing tighter chaperone binding in the former case.

The nM affinities reflect the tight complexes formed between the $YopH^{CBD}$ and chaperone in both cases. While the error bars are relatively large, there is some evidence of increased binding in the D106A. However, from the relaxation dispersion experiments results above there is a stabilisation of the excited state relative to the ground state in the D106A mutant. If the ground state energy of the mutated protein remained unchanged relative to the WT protein there would be an overall stabilisation of the unbound mutated protein and the dissociation constant would increase. Since the opposite trend is observed, there must be a destabilisation of the ground state in the D106A mutant relative to WT protein, as indicated in Fig 6.5, consistent with the loss of a hydrogen bond.

One possibility is that the excited state of $YopH^{CBD}$ serves as an important gateway conformation in the formation of the complex with SycH according to the following scheme:

$$1.\ YopH^{CBD} \rightleftharpoons YopH^{CBD*}$$

$$2.\ YopH^{CBD*} + SycH \rightleftharpoons YopH^{CBD}{:}SycH$$

where $YopH^{CBD}$* represents the excited conformational state and $YopH^{CBD}$:SycH the bound complex.

However, given the tight nM affinities between $YopH^{CBD}$ and SycH the lowly populated excited states (1-2%) within $YopH^{CBD}$ are unlikely to have an important effect on the binding of the protein. The main factor promoting increased binding in the D106A mutant is therefore the destabilisation of the ground state.

6.2.3 Structural Insights into the $YopH^{CBD}$ excited state

In this section we seek greater structural insight into the excited state conformation of $YopH^{CBD}$, using chemical shifts as a guide. Fig 6.6**A** shows the crystal structure of $YopH^{CBD}$ along with an example model structure for the YopH:SycH based on SAXS data(*228*). For this model it is important to note that the SAXS data is low resolution and consequently an ensemble of similar conformations is able to fit the data. Thus while SAXS data can be used to rule out structures that are inconsistent with the data, it has insufficient resolution to determine the correct structure of a complex. In this case, rigid body modelling is inconsistent with the SAXS data, suggesting that the $YopH^{CBD}$ must undergo a significant structural change for binding to occur. Moreover, NMR spectra of the $YopH^{CBD}$:SycH complex while unassigned, suggest that a large number of $YopH^{CBD}$ are significantly disordered in the complex due to their lack of dispersion in the ^{1}H dimension(*228*).

The correlation between predicted ^{15}N chemical shifts for the crystal structure of $YopH^{CBD}$ and the experimental ^{15}N shifts for WT and D106A protein is shown in Fig 6.6**B**. Chemical shifts are predicted using the SHIFTX2 web server(*196*). The RMSD in chemical shifts for both within the error of the prediction (1.1169 ppm for ^{15}N shifts) showing that the solution state NMR structure is in good agreement with the crystal structure for both forms of the protein. This result is also consistent with the similar HSQC spectra observed for the WT and D106A forms of the protein, with some residues being broadened beyond detection in the mutant.

Next, the experimental ^{15}N chemical shifts of WT and D106A mutated $YopH^{CBD}$ are compared to predicted ^{15}N chemical shifts of the SAXS-based model for $YopH^{CBD}$

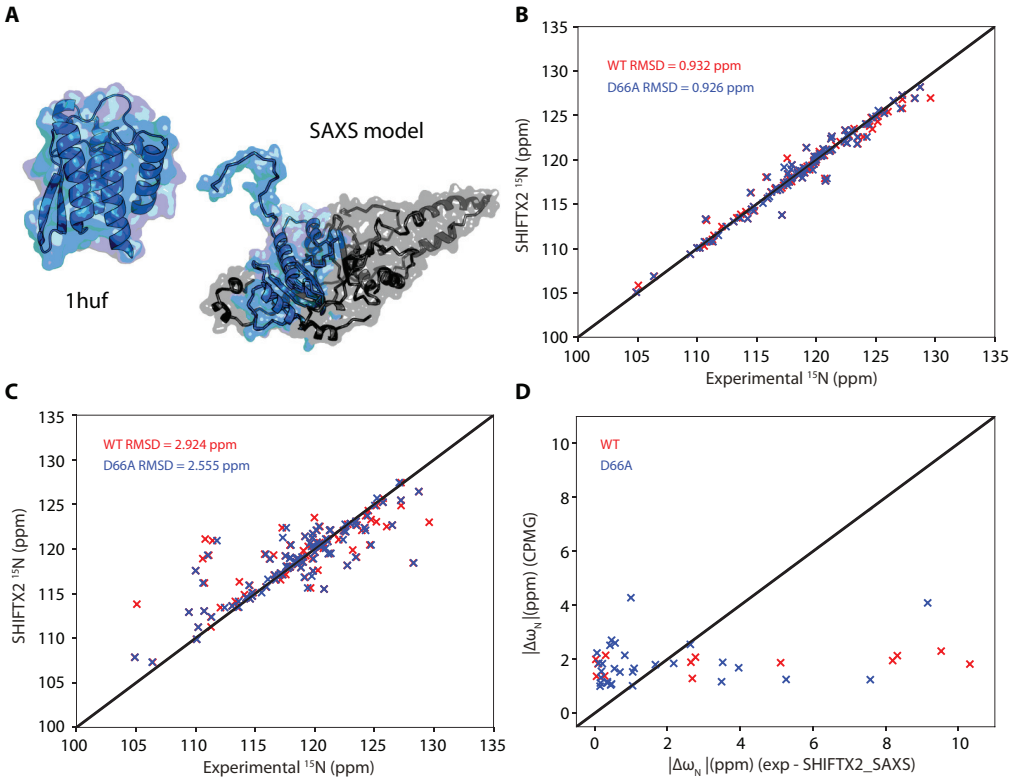

Figure 6.6: **A** Crystal structure of YopH^{CBD} (PDB: 1huf) and example SAXS-based structure of YopH^{CBD}:SycH complex. **B** Comparison of experimental ^{15}N chemical shifts for WT and D106A YopH^{CBD} with predicted shifts for crystal structure. **C** Comparison of Comparison of experimental ^{15}N chemical shifts for WT and D106A YopH^{CBD} with predicted shifts for the SAXS structure. **D** Correlation between $|\Delta\omega_N|$ values from CPMG experiments on WT and D106A YopH^{CBD} and difference between calculated $|\Delta\omega_N|$ values between experimental ^{15}N chemical shifts and calculated shifts for SAXS structure.

in the chaperone bound complex (6.6**C**). The chemical shifts for the SAXS-based models are predicted for four representative structures and the average is taken for the comparison in the figure. The correlation for both WT and mutated protein is materially worse in this case and far exceeds the expected error in predicted chemical shifts. This provides further evidence that rigid body modelling is inconsistent with the observed SAXS data.

We also seek to establish the extent of structural similarity between the excited state and the protein in its bound state to see whether there is evidence that the intrinsic dynamics of the protein prime it for binding. In particular, in the SAXS data suggests that in the bound state a large number of residues are unfolded and so it seems possible that the excited state conformation is partially unfolded,

facilitating the binding process.

In order to test this we compare $|\Delta\omega_N|$ values for residues showing significant dispersions from CPMG experiments on WT and D106A YopH^{CBD} to $|\Delta\omega_N|$ from the difference between experimental and predicted ^{15}N chemical shifts from the SAXS model of the bound protein. Since the CPMG experiments are only sensitive to the absolute difference in chemical shifts it is necessary to use the absolute differences in chemical shift for these comparisons. As shown in Fig 6.6**D**, the correlation between these two parameters is very poor showing that the excited state of YopH^{CBD} is not locally unfolded. Essentially the unfolding process is predicted to give rise to very large $|\Delta\omega_N|$ for some residues that are not observed in the CPMG analysis, while a number of residues with $|\Delta\omega_N|$ values on the order of 2 ppm are not expected to show CSPs upon binding. This observation adds weight to the idea that the excited state does not play an important role in binding and that the increased binding observed in the D106A mutant is driven by destabilisation of the ground state in the mutant.

The conserved structure of the YopH^{CBD}:SycH complex and the requirement for YopH to partially unfold to fit through the T3SS raises the question of why, unlike YopE, it adopts a globular folded monomeric form. The likely explanation for this is that the ground conformation is functionally important and following translocation YopH refolds into this form.

6.3 Methods and Materials

6.3.1 Sample Preparation

WT YopH^{CBD} samples were prepared as described previously(*228*). For the D106A mutant, gene constructs were synthesised by Eurofins MWG Operon and were subsequently subcloned into a modified pET-His-1a vector. The clone constructs were confirmed by DNA sequencing. An identical expression and purification protocol to the WT protein is followed.

6.3.2 NMR Methods

All NMR experiments were performed at 298 K and processed using NMRPipe(*101*). Spectra were subsequently visualised using Sparky(*241*).

HSQC Experiments

^{15}N-^{1}H HSQC experiments were performed on a 14.1 T (600 MHz) spectrometer using a gradient enhanced pulse sequence(*19*).

^{15}N CPMG Experiments

^{15}N backbone CPMG experiments were carried out on 1.0-1.3 mM perdeuteurated ^{15}N-labelled WT and D106A YopH^{CBD} protein samples. Data were collected at multiple magnetic field strengths of 11.7 T (500 MHz) (WT only), 14.1 T (600 MHz) and 19.9 T (850 MHz). The constant time relaxation delay was set to 30 ms. CPMG experiments were carried out using 14 values of the CPMG field strength ν_{cpmg}, varying from 67 to 1000 Hz in an interleaved fashion. In each experiment, two values of ν_{cpmg} were repeated for error estimation. Peak intensities as a function of CPMG frequency were quantified using FuDa(*242*), from which relaxation dispersion curves for every residue were produced. Each dispersion profile was tested for a statically significant dispersion using an F-test that compares the fit provided by a two-site exchanging model to a no exchange model (flat horizontal line). Residues that showed significant dispersions were carried forward for global analysis using the numerical simulation software CATIA(*171*), which fits k_{ex} and P_B globally and a $|\Delta\omega_N|$ for each exchanging residue. R_{Nz} and R_{HzNz} rates were measured for the proteins and used as inputs for CATIA to determine the differential relaxation rates of in-phase and anti-phase magnetization. For bootstrap analyses (6.4**A**) synthetic datasets were produced by sampling CPMG dispersion curves for individual residues with replacement such that the total number of residues was a constant. These were then fitted to the 2-state model, and the exchange parameters were recorded, and represented as a histogram.

6.4 Acknowledgements

Preparation of WT and D106A YopH^{CBD} samples as well as ITC expierments and 19.9 T (850 MHz) CPMG data acquisition was performed Arun Gupta and Magnus Wolf-Watz (Umeå University). All other NMR experiments and data analyses presented here were carried out by the author.

7
Conclusions

Usually, the main advantage of using NMR to study biomolecular systems is its ability to provide detailed atomic level resolution information. But, as highlighted in chapters 2-4 NMR is also able to describe more macroscopic features of such systems. However, gaining this insight requires creative thinking about the nature of the signal in NMR experiments and robust models that can be used to interpret the results.

For example, the results in Chapter 2 are dependent on the idea that the diffusion attenuation signal is sensitive to the distribution of displacements. For a molecule in a confined environment the displacement distribution will be limited by the size of the confining environment. Also, in the context of a cellular system, the displacement distribution is further complicated by the possibility of exchange with the extra-cellular environment and the intrinsic differential diffusional and relaxation properties of the two environments. By recording diffusion data with different delays and analysing the data globally we have shown that it is possible to disentangle these different effects and have written software to facilitate this analysis.

In Chapter 3 we show how the water CEST experiments can be used in principle to report on changes in a range of solution properties such as viscosity and pH due to the way in which these changes affect water-solute interactions via chemical exchange and NOEs. We demonstrate that insights gained from studying CEST effects in model systems can help elucidate the physical processes underlying

observed changes in the CEST signal *in vivo*. We have also shown how the CEST experiment can be used to follow the process of protein aggregation and provides complementary insight to 1D experiments.

Chapter 4 combines elements of the previous two chapters to introduce the novel diffusion-weighted CEST (diffCEST) experiment. The diffusion weighting in this experiment can be used to differentiate between different compartments, while the properties of the compartments themselves are probed using the CEST pulses. The successful performance of the experiment is demonstrated using a model system of GUVs. We also show that this experiment is able to monitor cell stress: remarkably this is achieved in a label-free manner and via the measurement of the water resonance exclusively.

Exploiting the full potential of NMR to study the structure and dynamics of biomolecules requires isotopically labelled samples, as highlighted in chapters 5-6. In Chapter 5, NMR is utilised to verify predictions from molecular dynamics simulations that the D66A mutation in the protein CypA significantly perturbs its loop motions and dynamics. While MD simulations possess enormous potential for studying biomolecular systems, they remain inexact, relying on empirical force fields, and therefore require rigorous experimental validation. Overall, the predictions of the simulations are in good agreement with the NMR results, suggesting that the loop structure undergoes an order to disorder transition as a result of the mutation and that the ground state of the mutant protein resembles the excited state of the WT. These experiments also highlight the differences in approaches for describing conformational dynamics: in NMR this is usually done using two or three state models, whereas MD simulations typically employ a much larger number of states. Forming an overarching physical picture consistent with both of these views can be difficult but is an important challenge given the requirement to use multiple techniques to characterise challenging biological target systems.

The unique ability of NMR to report on conformational dynamics and features of transiently populated states of proteins is further highlighted in Chapter 6. In this case the intrinsic conformational dynamics of the protein YopH were investigated.

The findings demonstrate how subtle changes in the energy landscape of a protein can affect binding and the utility of NMR for investigating these phenomena.

Overall, this thesis demonstrates the enormous versatility and capability of NMR experiments to characterise biological systems. Most modern biomolecular NMR experiments take advantage of isotope labelling strategies to probe the structure and dynamics of biomolecules with exquisite detail. However, as demonstrated here, there remains significant scope for the development of approaches based purely on endogenous species, to yield tremendous insight into these systems.

7.1 List of Publications

A list of the author's publications is provided below. This includes collaborative work on the development of ^{19}F labelling methods that is not included in this thesis as well as a project describing homonuclear relaxation interference effects in oligomeric porphyrin systems that was commenced during the author's Masters project and completed during the start of the DPhil.

1. G Karunanithy, RJ Wheeler, LR Tear, NJ Farrer, S Faulkner, AJ Baldwin: An in-cell diffusion method to characterize the size, abundance and permeability of cells, Journal of Magnetic Resonance, In Press

2. AA Gupta, I Reinartz, G Karunanithy, A Spilotros, VR Jonna, A Hofer, DI Svergun, AJ Baldwin, A Schug, M Wolf-Watz: Formation of a secretion competent protein complex by a dynamic wrap-around binding mechanism, J. Mol. Biol., 2018, 430, 3157-3169

3. M Imiołek, G Karunanithy, WL Ng, AJ Baldwin, VE Gouverneur, BG Davis: Selective radical trifluoromethylation of native residues in proteins, J. Am. Chem. Soc., 2018, 140, 1568-1571

4. RL Leung, MD Robinson, AA Ajabali, G Karunanithy, B Lyons, R Raj, S Raoufmoghaddam, S Mohammed, TD Claridge, AJ Baldwin, BG Davis: Monitoring the Disassembly of Virus-like Particles by 19F-NMR, J. Am. Chem. Soc., 2017, 139, 5277-5280

5. KJ Ray, JR Larkin, YK Tee, AA Khrapitchev, G Karunanithy, M Barber, AJ Baldwin, MA Chappell, NR Sibson: Determination of an optimally sensitive and specific Chemical Exchange Saturation Transfer MRI quantification metric in relevant biological phantoms, NMR in Biomedicine, 2016, 29, 1624-1633

6. G Karunanithy, A Cnossen, H Muller, MD Peeks, NH Rees, TDW Claridge, HL Anderson, AJ Baldwin: Harnessing NMR relaxation interference effects to characterise supramolecular assemblies, Chem Commun., 2016, 52, 7450-7453

References

(*1*) Levitt, M. H., *Spin Dynamics : Basics of Nuclear Magnetic Resonance*; John Wiley & Sons, Ltd: 2008, p 740.

(*2*) Hore, P. J., *Nuclear magnetic Resonance*, Second Edi; Oxford University Press: 2015, p 112.

(*3*) Hore, P. J., Jones, J. A.J. A., and Wimperis, S., *NMR the Toolkit: How Pulse Sequences Work*, Second Edi; Oxford University Press: 2015, p 120.

(*4*) Cavanagh, J., Fairbrother, W. J., Palmer III, A. G., Rance, M., and Skelton, N. J., *Protein NMR Spectroscopy: Principles and Practice*; Academic Press: 2007; Chapter 885, p 885.

(*5*) Rose, M. E., *Elementary theory of angular momentum*; Dover: 1995, p 248.

(*6*) Ernst, R. R., and Anderson, W. A. (1966). Application of Fourier Transform Spectroscopy to Magnetic Resonance. *Review of Scientific Instruments 37*, 93–102.

(*7*) Jeener, J In, Basko Polje, 1971.

(*8*) Bloch, F. (1946). Nuclear Induction. *Physical Review 70*, 460–474.

(*9*) McConnell, H. M. (1958). Reaction Rates by Nuclear Magnetic Resonance. *The Journal of Chemical Physics 28*, 430–431.

(*10*) Torrey, H. C. (1956). Bloch Equations with Diffusion Terms. *Physical Review 104*, 563–565.

(*11*) Karplus, M. (1963). Vicinal Proton Coupling in Nuclear Magnetic Resonance. *Journal of the American Chemical Society 85*, 2870–2871.

(*12*) Aue, W. P., Bartholdi, E., and Ernst, R. R. (1976). Two-dimensional spectroscopy. Application to nuclear magnetic resonance. *The Journal of Chemical Physics 64*, 2229–2246.

(*13*) Wagner, G., and Wüthrich, K. (1982). Sequential resonance assignments in protein 1H nuclear magnetic resonance spectra: Basic pancreatic trypsin inhibitor. *Journal of Molecular Biology 155*, 347–366.

(*14*) Williamson, M. P., Havel, T. F., and Wüthrich, K. (1985). Solution conformation of proteinase inhibitor IIA from bull seminal plasma by1H nuclear magnetic resonance and distance geometry. *Journal of Molecular Biology 182*, 295–315.

(*15*) Kay, L. E., Ikura, M., Tschudin, R., and Bax, A. (1990). Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *Journal of Magnetic Resonance (1969) 89*, 496–514.

(*16*) Morris, G. A. (1980). Sensitivity enhancement in nitrogen-15 NMR: polarization transfer using the INEPT pulse sequence. *Journal of the American Chemical Society 102*, 428–429.

(*17*) Bodenhausen, G., and Ruben, D. J. (1980). Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chemical Physics Letters 69*, 185–189.

(*18*) Piotto, M., Saudek, V., and Sklenář, V. (1992). Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *Journal of Biomolecular NMR 2*, 661–665.

(*19*) Kay, L., Keifer, P., and Saarinen, T. (1992). Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *Journal of the American Chemical Society 114*, 10663–10665.

(*20*) Ikura, M, Kay, L. E., and Bax, A (1990). A novel approach for sequential assignment of 1H, 13C, and 15N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry 29*, 4659–67.

(*21*) Bloembergen, N, Purcell, E., and Pound, R. (1948). Relaxation effects in nuclear magnetic resonance absorption. *Physical Review 73*, 679–715.

(*22*) Goldman, M. (2001). Advances in magnetic resonance: Formal theory of spin-lattice relaxation. *Journal of Magnetic Resonance 149*, 160–187.

(*23*) Fischer, M. W. F., Majumdar, A., and Zuiderweg, E. R. P. (1998). Protein NMR relaxation: theory, applications and outlook. *Progress in Nuclear Magnetic Resonance Spectroscopy 33*, 207–272.

(*24*) Abragam, A., *The Principles of Nuclear Magnetism*; Clarendon, Ed., Oxford, 1961.

(*25*) Wangsness, R. K., and Bloch, F. (1953). The Dynamical Theory of Nuclear Induction. *Physical Review 89*, 728–739.

(*26*) Redfield, A. G. On the Theory of Relaxation Processes., 1957.

(*27*) Bain, A. D. (2003). Chemical exchange in NMR. *Progress in Nuclear Magnetic Resonance Spectroscopy 43*, 63–103.

(*28*) Sauerwein, A. C., and Flemming Hansen, D. In *Protein NMR: Modern Techniques and Biomedical Applications*; Springer, Boston, MA: 2015, pp 75–132.

(*29*) Carr, H. Y., and Purcell, E. M. (1954). Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Physical Review 94*, 630–638.

(*30*) Meiboom, S., and Gill, D. (1958). Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Review of Scientific Instruments 29*, 688.

(*31*) Deverell, C., Morgan, R. E., and Strange, J. H. (1970). Studies of chemical exchange by nuclear magnetic relaxation in the rotating frame. *Molecular Physics 18*, 553–559.

(*32*) Baldwin, A. J., and Kay, L. E. (2009). NMR spectroscopy brings invisible protein states into focus. *Nature Chemical Biology 5*, 808–814.

(*33*) Anthis, N. J., and Clore, G. M. (2015). Visualizing transient dark states by NMR spectroscopy. *Quarterly Reviews of Biophysics 48*, 35–116.

(*34*) Palmer, A. G., Kroenke, C. D., and Loria, J. P., *Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules*; 1997; Elsevier Masson SAS: 2001; Vol. 339, pp 204–238.

(*35*) Massi, F., and Palmer, A. G. (2006). Characterization of the Dynamics of Biomacromolecules Using Rotating-Frame Spin Relaxation NMR Spectroscopy. *Chem Rev 106*, 1700–1719.

(*36*) Johnson, C. (1999). Diffusion ordered nuclear magnetic resonance spectroscopy: principles and applications. *Progress in Nuclear Magnetic Resonance Spectroscopy 34*, 203–256.

(*37*) Stejskal, E. O., and Tanner, J. E. (1965). Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient. *The Journal of Chemical Physics 42*, 288.

(*38*) Hahn, E. L. (1950). Spin echoes. *Physical Review 80*, 580–594.

(*39*) W.S. Price, *NMR Studies of Translational Motion*; Cambridge University Press: 2009.

(*40*) Tanner, J. E., and Annert, J. E. T. (1970). Restricted Self-Diffusion of Protons in Colloidal Systems by the Pulsed-Gradient. *Restricted Diffusion and Flow The Journal of Chemical Physics 52*, 3597.

(*41*) Loening, N. M., Keeler, J., and Morris, G. A. (2001). One-dimensional DOSY. *Journal of Magnetic Resonance 153*, 103–112.

(*42*) Sørland, G. H., *Dynamic Pulsed-Field-Gradient NMR*; Springer Series in Chemical Physics, Vol. 110; Springer Berlin Heidelberg: Berlin, Heidelberg, 2014.

(*43*) Brown, F. F., Campbell, I. D., Kuchel, P. W., and Rabenstein, D. C. (1977). Human erythrocyte metabolism studies by 1H spin echo NMR. *FEBS Letters 82*, 12–16.

(*44*) Andrasko, J. (1976). Water Diffusion Permeability of Human Erythrocytes Studied By a Pulsed Field Gradient NMR Technique. *Biochimica et Biophysica Acta 428*, 304–311.

(*45*) London, R. E., Gregg, C. T., and Matwiyoff, N. A. (1975). Nuclear magnetic resonance of rotational mobility of mouse hemoglobin labeled with (2-13C)histidine. *Science 188*, 266–268.

(*46*) Salhany, J. M., Yamane, T, Shulman, R. G., Ogawa, S, Circello, B., Griffin, B. M., Labeda, D. P., and Metcalf, W. W. High resolution 31P nuclear magnetic resonance studies of intact yeast cells., tech. rep. 12, 1975, pp 4966–70.

(*47*) Damadian, R. (1971). Tumor detection by nuclear magnetic resonance. *Science 171*, 1151–1153.

(*48*) Serber, Z., Ledwidge, R., Miller, S. M., and Dötsch, V. (2001). Evaluation of parameters critical to observing proteins inside living Escherichia coli by in-cell NMR spectroscopy. *Journal of the American Chemical Society 123*, 8895–8901.

(*49*) Theillet, F. X., Binolfi, A., Bekei, B., Martorana, A., Rose, H. M., Stuiver, M., Verzini, S., Lorenz, D., Van Rossum, M., Goldfarb, D., and Selenko, P. (2016). Structural disorder of monomeric α-synuclein persists in mammalian cells. *Nature 530*, 45–50.

(*50*) Luchinat, E., and Banci, L. (2018). In-Cell NMR in Human Cells: Direct Protein Expression Allows Structural Studies of Protein Folding and Maturation. *Accounts of Chemical Research 51*, 1550–1557.

(*51*) Lippens, G., Cahoreau, E., Millard, P., Charlier, C., Lopez, J., Hanoulle, X., and Portais, J. C. (2018). In-cell NMR: from metabolites to macromolecules. *The Analyst 143*, 620–629.

(*52*) Takeuchi, K., Arthanari, H., and Wagner, G. (2016). Perspective: revisiting the field dependence of TROSY sensitivity. *Journal of Biomolecular NMR 66*, 221–225.

(*53*) Hänsel, R., Luh, L. M., Corbeski, I., Trantirek, L., and Dötsch, V. (2014). In-cell NMR and EPR spectroscopy of biomacromolecules. *Angewandte Chemie (International ed. in English) 53*, 10300–10314.

(*54*) Bekei, B., Rose, H. M., Herzig, M., Dose, A., Schwarzer, D., and Selenko, P. In *Methods in molecular biology (Clifton, N.J.)* 2012; Vol. 895, pp 43–54.

(*55*) Van Zijl, P. C., Moonen, C. T., Faustino, P., Pekar, J., Kaplan, O., and Cohen, J. S. (1991). Complete separation of intracellular and extracellular information in NMR spectra of perfused cells by diffusion-weighted spectroscopy. *Proceedings of the National Academy of Sciences 88*, 3228–3232.

(*56*) Fieremans, E., Novikov, D. S., Jensen, J. H., and Helpern, J. A. (2010). Monte Carlo study of a two-compartment exchange model of diffusion. *NMR in Biomedicine 23*, 711–724.

(*57*) Meier, C., Dreher, W., and Leibfritz, D. (2003). Diffusion in compartmental systems. II. Diffusion-weighted measurements of rat brain tissue in vivo and postmortem at very large b-values. *Magnetic Resonance in Medicine 50*, 510–514.

(*58*) Szafer, A., Zhong, J., and Gore, J. C. (1995). Theoretical Model for Water Diffusion in Tissues. *Magnetic Resonance in Medicine 33*, 697–712.

(*59*) Le Bihan, D., and Johansen-Berg, H. (2012). Diffusion MRI at 25: Exploring brain tissue structure and function. *NeuroImage 61*, 324–341.

(*60*) Pfeuffer, J., Flögel, U., Dreher, W., and Leibfritz, D. (1998). Restricted diffusion and exchange of intracellular water: Theoretical modelling and diffusion time dependence of 1H NMR measurements on perfused glial cells. *NMR in Biomedicine 11*, 19–31.

(*61*) Andrasko, J. (1976). Measurement of membrane permeability to slowly penetrating molecules by a pulse gradient NMR method. *Journal of Magnetic Resonance 21*, 479–484.

(*62*) Kuchel, P. W., Coy, A., and Stilbs, P. (1997). NMR 'diffusion-diffraction' of water revealing alignment of erythrocytes in a magnetic field and their dimensions and membrane transport characteristics. *Magnetic Resonance in Medicine 37*, 637–643.

(*63*) Meier, C., Dreher, W., and Leibfritz, D. (2003). Diffusion in compartmental systems. I. A comparison of an analytical model with simulations. *Magnetic Resonance in Medicine 50*, 500–509.

(*64*) Stanisz, G. J., Li, J. G., Wright, G. A., and Henkelman, R. M. (1998). Water Dynamics in Human Blood via Combined Measurements. *Magnetic Resonance in Medicine 39*, 223–233.

(*65*) Stanisz, G. J., Szafer, A., Wright, G. A., and Henkelman, R. M. (1997). An analytical model of restricted diffusion in bovine optic nerve. *Magnetic Resonance in Medicine 37*, 103–111.

(*66*) Panagiotaki, E., Walker-Samuel, S., Siow, B., Johnson, S. P., Rajkumar, V., Pedley, R. B., Lythgoe, M. F., and Alexander, D. C. (2014). Noninvasive quantification of solid tumor microstructure using VERDICT MRI. *Cancer Research 74*, 1902–1912.

(*67*) Le Bihan, D. (2013). Apparent Diffusion Coefficient and Beyond: What Diffusion MR Imaging Can Tell Us about Tissue Structure. *Radiology 268*, 318–322.

(*68*) Kärger, J. (1969). Zur Bestimmung der Diffusion in einem Zweibereichsystem mit Hilfe von gepulsten Feldgradienten. *Annalen der Physik 479*, 1–4.

(*69*) Kärger, J. (1985). NMR self-diffusion studies in heterogeneous systems. *Advances in Colloid and Interface Science 23*, 129–148.

(*70*) Kärger, J., Pfeifer, H., and Heink, W. (1988). Principles and Application of Self-Diffusion Measurements by Nuclear Magnetic Resonance. *Advances in Magnetic and Optical Resonance 12*, 1–89.

(*71*) Neuman, C. H. (1974). Spin echo of spins diffusing in a bounded medium. *The Journal of Chemical Physics 60*, 4508–4511.

(*72*) Murday, J. S., and Cotts, R. M. (1968). Self-Diffusion Coefficient of Liquid Lithium. *The Journal of Chemical Physics 48*, 4938–4945.

(*73*) Price, W. S., Barzykin, A. V., Hayamizu, K., and Tachiya, M (1998). A Model for Diffusive Transport through a Spherical Interface Probed by. *Biophysical Journal 74*, 2259–2271.

(*74*) Balinov, B., Jönsson, B., Linse, P., and Söderman, O. The NMR self-diffusion method applied to restricted diffusion. simulation of echo attenuation from molecules in spheres and between planes., 1993.

(*75*) Tanner, J. E., and Stejskal, E. O. (1968). Restricted self-diffusion of protons in colloidal systems by the pulsed-gradient, spin-echo method. *The Journal of Chemical Physics 49*, 1768–1777.

(*76*) Li, H., Jiang, X., Xie, J., Gore, J. C., and Xu, J. (2017). Impact of transcytolemmal water exchange on estimates of tissue microstructural properties derived from diffusion MRI. *Magnetic Resonance in Medicine 77*, 2239–2249.

(*77*) Baldwin, A. J., Christodoulou, J., Barker, P. D., Dobson, C. M., and Lippens, G. (2007). Contribution of rotational diffusion to pulsed field gradient diffusion measurements. *The Journal of chemical physics 127*, 114505.

(*78*) Waudby, C. A., and Christodoulou, J. (2011). GPU accelerated Monte Carlo simulation of pulsed-field gradient NMR experiments. *Journal of Magnetic Resonance 211*, 67–73.

(*79*) Callaghan, P. T., Coy, A., MacGowan, D., Packer, K. J., and Zelaya, F. O. Diffraction-like effects in NMR diffusion studies of fluids in porous solids., 1991.

(*80*) Callaghan, P. T., Coy, A., Halpin, T. P., MacGowan, D., Packer, K. J., and Zelaya, F. O. (1992). Diffusion in porous systems and the influence of pore morphology in pulsed gradient spin-echo nuclear magnetic resonance studies. *The Journal of Chemical Physics 97*, 651–662.

(*81*) Grebenkov, D. S., *Diffusion NMR of Confined Systems*; Valiullin, R., Ed.; New Developments in NMR 9; Royal Society of Chemistry: Cambridge, 2016, pp 70–128.

(*82*) Connell, M. A., Bowyer, P. J., Adam Bone, P., Davis, A. L., Swanson, A. G., Nilsson, M., and Morris, G. A. (2009). Improving the accuracy of pulsed field gradient NMR diffusion experiments: Correction for gradient non-uniformity. *Journal of Magnetic Resonance 198*, 121–131.

(*83*) Zhao, L., Kroenke, C. D., Song, J., Piwnica-Worms, D., Ackerman, J. J. H., and Neil, J. J. (2008). Intracellular water-specific MR of microbead-adherent cells: the HeLa cell intracellular water exchange lifetime. *NMR in Biomedicine 21*, 159–164.

(*84*) Milo, R., Jorgensen, P., Moran, U., Weber, G., and Springer, M. (2010). BioNumbers—the database of key numbers in molecular and cell biology. *Nucleic Acids Research 38*, D750–D753.

(*85*) Grebenkov, D. S. (2007). NMR survey of reflected Brownian motion. *Reviews of Modern Physics 79*, 1077–1137.

(*86*) Pfeuffer, J., Fl??gel, U., Dreher, W., and Leibfritz, D. (1998). Restricted diffusion and exchange of intracellular water: Theoretical modelling and diffusion time dependence of 1H NMR measurements on perfused glial cells. *NMR in Biomedicine 11*, 19–31.

(*87*) Åslund, I., and Topgaard, D. (2009). Determination of the self-diffusion coefficient of intracellular water using PGSE NMR with variable gradient pulse length. *Journal of Magnetic Resonance 201*, 250–254.

(*88*) Tian, X., Li, H., Jiang, X., Xie, J., Gore, J. C., and Xu, J. (2017). Evaluation and comparison of diffusion MR methods for measuring apparent transcytolemmal water exchange rate constant. *Journal of Magnetic Resonance 275*, 29–37.

(*89*) Yang, D. M., Huettner, J. E., Bretthorst, G. L., Neil, J. J., Garbow, J. R., and Ackerman, J. J. (2018). Intracellular water preexchange lifetime in neurons and astrocytes. *Magnetic Resonance in Medicine 79*, 1616–1627.

(*90*) Eriksson, S., Elbing, K., Söderman, O., Lindkvist-Petersson, K., Topgaard, D., and Lasič, S. (2017). NMR quantification of diffusional exchange in cell suspensions with relaxation rate differences between intra and extracellular compartments. *PLoS ONE 12*, 1–18.

(*91*) Zakhartsev, M., and Reuss, M. (2018). Cell size and morphological properties of yeast Saccharomyces cerevisiae in relation to growth temperature. *FEMS yeast research 18*, 1–16.

(*92*) Srinorakutara, T. (1998). Determination of Yeast Cell Wall Thickness and Cell Diameter Using New Methods. *Fermentation and Bioengineering, Journal Of 86*, 253–260.

(*93*) Labadie, C., Lee, J. H., Vetek, G., and Springer, C. S. (1994). Relaxographic Imaging. *Journal of Magnetic Resonance, Series B 105*, 99–112.

(*94*) Åslund, I., Nowacka, A., Nilsson, M., and Topgaard, D. (2009). Filter-exchange PGSE NMR determination of cell membrane permeability. *Journal of Magnetic Resonance 200*, 291–295.

(*95*) Makarov, V. A., Andrews, B. K., Smith, P. E., and Pettitt, B. M. (2000). Residence Times of Water Molecules in the Hydration Sites of Myoglobin. *Biophysical Journal 79*, 2966–2974.

(*96*) Persson, E., and Halle, B. (2008). Cell water dynamics on multiple time scales. *Proceedings of the National Academy of Sciences of the United States of America 105*, 6266–6271.

(*97*) Waudby, C. A., Mantle, M. D., Cabrita, L. D., Gladden, L. F., Dobson, C. M., and Christodoulou, J. (2012). Rapid Distinction of Intracellular and Extracellular Proteins Using NMR Diffusion Measurements. *Journal of the American Chemical Society 134*, 11312–11315.

(*98*) Powles, J. G., Mallett, M. J. D., Rickayzen, G., and Evans, W. A. B. (1992). Exact Analytic Solutions for Diffusion Impeded by an Infinite Array of Partially Permeable Barriers. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 436*, 391–403.

(*99*) Xing, H., Lin, F., Wu, Q., and Gong, Q. (2013). Investigation of different boundary treatment methods in Monte-Carlo simulations of diffusion NMR. *Magnetic Resonance in Medicine 70*, 1167–1172.

(*100*) Holz, M., Heil, S. R., and Sacco, A. (2000). Temperature-dependent self-diffusion coefficients of water and six selected molecular liquids for calibration in accurate 1H NMR PFG measurements. *Physical Chemistry Chemical Physics 2*, 4740–4742.

(*101*) 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. *Journal of Biomolecular NMR 6*, 277–293.

(*102*) Helmus, J. J., and Jaroniec, C. P. (2013). Nmrglue: An open source Python package for the analysis of multidimensional NMR data. *Journal of Biomolecular NMR 55*, 355–367.

(*103*) Newville, M., Stensitzki, T., Allen, D. B., and Ingargiola, A. (2014). LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python., DOI: `10.5281/ZENODO.11813`.

(*104*) Forsén, S., and Hoffman, R. A. (1963). Study of Moderately Rapid Chemical Exchange Reactions by Means of Nuclear Magnetic Double Resonance. *The Journal of Chemical Physics 39*, 2892–2901.

(*105*) Vallurupalli, P., Bouvignies, G., and Kay, L. E. (2012). Studying "invisible" excited protein states in slow exchange with a major state conformation. *Journal of the American Chemical Society 134*, 8148–8161.

(*106*) Vallurupalli, P., Sekhar, A., Yuwen, T., and Kay, L. E. (2017). Probing conformational dynamics in biomolecules via chemical exchange saturation transfer: a primer. *Journal of Biomolecular NMR 67*, 243–271.

(*107*) Harston, G. W., Tee, Y. K., Blockley, N., Okell, T. W., Thandeswaran, S., Shaya, G., Sheerin, F., Cellerini, M., Payne, S., Jezzard, P., Chappell, M., and Kennedy, J. (2015). Identifying the ischaemic penumbra using pH-weighted magnetic resonance imaging. *Brain 138*, 36–42.

(*108*) Zhou, J., Payen, J.-F., Wilson, D. A., Traystman, R. J., and Zijl, P. C.M. V. (2003). Using the amide proton signals of intracellular proteins and peptides to detect pH effects in MRI. *Nature Medicine 9*, 1085–1090.

(*109*) Mayer, M., and Meyer, B. (2001). Group Epitope Mapping by Saturation Transfer Difference NMR To Identify Segments of a Ligand in Direct Contact with a Protein Receptor., DOI: 10.1021/ja0100120.

(*110*) Fawzi, N. L., Ying, J., Ghirlando, R., Torchia, D. a., and Clore, G. M. (2011). Atomic-resolution dynamics on the surface of amyloid-β protofibrils probed by solution NMR. *Nature 480*, 268–272.

(*111*) Vallurupalli, P., Bouvignies, G., and Kay, L. E. (2013). A computational study of the effects of 13c-13c scalar couplings on 13C CEST NMR spectra: Towards studies on a uniformly 13C-labeled protein. *ChemBioChem 14*, 1709–1713.

(*112*) Ward, K. M., and Balaban, R. S. (2000). Determination of pH using water protons and Chemical Exchange Dependent Saturation Transfer (CEST). *Magnetic Resonance in Medicine 44*, 799–802.

(*113*) Ward, K. M., Aletras, A. H., and Balaban, R. S. (2000). A New Class of Contrast Agents for MRI Based on Proton Chemical Exchange Dependent Saturation Transfer (CEST). *Journal of Magnetic Resonance 143*, 79–87.

(*114*) Di Gregorio, E., Iani, R., Ferrauto, G., Nuzzi, R., Aime, S., and Gianolio, E. (2018). Gd accumulation in tissues of healthy mice upon repeated administrations of Gadodiamide and Gadoteridol. *Journal of Trace Elements in Medicine and Biology 48*, 239–245.

(*115*) Zhou, J., Payen, J.-F., Wilson, D. A., Traystman, R. J., and van Zijl, P. C. M. (2003). Using the amide proton signals of intracellular proteins and peptides to detect pH effects in MRI. *Nature Medicine 9*, 1085–1090.

(*116*) van Zijl, P. C. M., and Yadav, N. N. (2011). Chemical exchange saturation transfer (CEST): What is in a name and what isn't? *Magnetic Resonance in Medicine 65*, 927–948.

(*117*) Helgstrand, M., Hard, T., and Allard, P. (2000). Simulations of NMR pulse sequences during equilibrium and non-equilibrium chemical exchange. *Journal of Biomolecular NMR 18*, 49–63.

(*118*) van Zijl, P. C. M., Jones, C. K., Ren, J., Malloy, C. R., and Sherry, A. D. (2007). MRI detection of glycogen in vivo by using chemical exchange saturation transfer imaging (glycoCEST). *Proceedings of the National Academy of Sciences 104*, 4359–4364.

(*119*) Neuhaus, D., and Williamson, M. P., *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, 2nd; Wiley-VCH: New York, 2000.

(*120*) van Zijl, P. C., Lam, W. W., Xu, J., Knutsson, L., and Stanisz, G. J. (2018). Magnetization Transfer Contrast and Chemical Exchange Saturation Transfer MRI. Features and analysis of the field-dependent saturation spectrum. *NeuroImage 168*, 222–241.

(*121*) Henkelman, R. M., Stanisz, G. J., and Graham, S. J. (2001). Magnetization transfer in MRI: A review. *NMR in Biomedicine 14*, 57–64.

(*122*) Liepinsh, E., and Otting, G. (1996). Proton exchange rates from amino acid side chains— implications for image contrast. *Magnetic Resonance in Medicine 35*, 30–42.

(*123*) Halle, B. (2003). Cross-relaxation between macromolecular and solvent spins : The role of long-range dipole couplings. *Journal of Chemical Physics 119*, 12372–12385.

(*124*) Modig, K., Liepinsh, E., Otting, G., and Halle, B. (2004). Dynamics of protein and peptide hydration. *Journal of the American Chemical Society 126*, 102–114.

(*125*) Otting, G., Liepinsh, E., and Wuthrich, K. (1991). Protein hydration in aqueous solution. *Science 254*, 974–980.

(*126*) Sun, P. Z. (2010). Simplified and scalable numerical solution for describing multi-pool chemical exchange saturation transfer (CEST) MRI contrast. *Journal of Magnetic Resonance 205*, 235–241.

(*127*) Jones, C. K., Huang, A., Xu, J., Edden, R. A., Schär, M., Hua, J., Oskolkov, N., Zacà, D., Zhou, J., McMahon, M. T., Pillai, J. J., and van Zijl, P. C. (2013). Nuclear Overhauser enhancement (NOE) imaging in the human brain at 7T. *NeuroImage 77*, 114 –124.

(*128*) Ray, K. J., Larkin, J. R., Tee, Y. K., Khrapitchev, A. A., Karunanithy, G., Barber, M., Baldwin, A. J., Chappell, M. A., and Sibson, N. R. (2016). Determination of an optimally sensitive and specific chemical exchange saturation transfer MRI quantification metric in relevant biological phantoms. *NMR in Biomedicine 29*, 1624–1633.

(*129*) Purushotham, A., Campbell, B. C. V., Straka, M., Mlynash, M., Olivot, J.-M., Bammer, R., Kemp, S. M., Albers, G. W., and Lansberg, M. G. (2015). Apparent Diffusion Coefficient Threshold for Delineation of Ischemic Core. *International Journal of Stroke 10*, 348–353.

(*130*) Harston, G. W. J., Minks, D., Sheerin, F., Payne, S. J., Chappell, M., Jezzard, P., Jenkinson, M., and Kennedy, J. (2017). Optimizing image registration and infarct definition in stroke research. *Annals of Clinical and Translational Neurology 4*, 166–174.

(*131*) Sipos, T., and Merkel, J. R. (1970). Effect of calcium ions on the activity, heat stability, and structure of trypsin. *Biochemistry 9*, 2766–2775.

(*132*) Aguilar, J. A., Nilsson, M., Bodenhausen, G., and Morris, G. A. (2012). Spin echo NMR spectra without J modulation. *Chemical Communications 48*, 811–813.

(*133*) Gillies, R. J., Raghunand, N., Garcia-Martin, M. L., and Gatenby, R. A. pH imaging. A review of pH measurement methods and applications in cancers. *IEEE engineering in medicine and biology magazine : the quarterly magazine of the Engineering in Medicine & Biology Society 23*, 57–64.

(*134*) Gillies, R. J., Raghunand, N., Karczmar, G. S., and Bhujwalla, Z. M. (2002). MRI of the tumor microenvironment. *Journal of Magnetic Resonance Imaging 16*, 430–450.

(*135*) Huang, Y., and McNamara, J. O. (2004). Ischemic stroke: 'Acidotoxicity' is a perpetrator. *Cell 118*, 665–6.

(*136*) Eriksson, M., Härd, T., and Nilsson, L. (1995). On the pH dependence of amide proton exchange rates in proteins. *Biophysical Journal 69*, 329–339.

(*137*) Charisiadis, P., Kontogianni, V., Tsiafoulis, C., Tzakos, A., Siskos, M., Gerothanassis, I., Charisiadis, P., Kontogianni, V. G., Tsiafoulis, C. G., Tzakos, A. G., Siskos, M., and Gerothanassis, I. P. (2014). 1H-NMR as a Structural and Analytical Tool of Intra- and Intermolecular Hydrogen Bonds of Phenol-Containing Natural Products and Model Compounds. *Molecules 19*, 13643–13682.

(*138*) van Zijl, P. C. M., Jones, C. K., Ren, J., Malloy, C. R., and Sherry, A. D. (2007). MRI detection of glycogen in vivo by using chemical exchange saturation transfer imaging (glycoCEST). *Proceedings of the National Academy of Sciences 104*, 4359–4364.

(*139*) Alderson, T. R., and Markley, J. L. (2013). Biophysical characterization of α-synuclein and its controversial structure. *Intrinsically Disordered Proteins 1*, e26255.

(*140*) Stefanis, L. (2012). α-Synuclein in Parkinson's disease. *Cold Spring Harbor perspectives in medicine 2*, a009399.

(*141*) Trojanowski, J. Q., and Lee, V. M. Y. (2002). Parkinson's disease and related synucleinopathies are a new class of nervous system amyloidoses. *Neurotoxicology 23*, 457–60.

(*142*) Zarranz, J. J., Alegre, J., Gómez-Esteban, J. C., Lezcano, E., Ros, R., Ampuero, I., Vidal, L., Hoenicka, J., Rodriguez, O., Atarés, B., Llorens, V., Tortosa, E. G., del Ser, T., Muñoz, D. G., and de Yebenes, J. G. (2004). The new mutation, E46K, of α-synuclein causes parkinson and Lewy body dementia. *Annals of Neurology 55*, 164–173.

(*143*) Bousset, L., Pieri, L., Ruiz-Arlandis, G., Gath, J., Jensen, P. H., Habenstein, B., Madiona, K., Olieric, V., Böckmann, A., Meier, B. H., and Melki, R. (2013). Structural and functional characterization of two alpha-synuclein strains. *Nature Communications 4*, 2575.

(*144*) Bodner, C. R., Dobson, C. M., and Bax, A. (2009). Multiple Tight Phospholipid-Binding Modes of α-Synuclein Revealed by Solution NMR Spectroscopy. *Journal of Molecular Biology 390*, 775–790.

(*145*) Heise, H., Hoyer, W., Becker, S., Andronesi, O. C., Riedel, D., and Baldus, M. (2005). Molecular-level secondary structure, polymorphism, and dynamics of full-length alpha-synuclein fibrils studied by solid-state NMR. *Proceedings of the National Academy of Sciences of the United States of America 102*, 15871–6.

(*146*) Tkachenko, O., Benesch, J. L. P., and Baldwin, A. J. (2018). αB-crystallin inhibits amyloidogenesis by disassembling aggregation nuclei. *bioRxiv*, 300541.

(*147*) Dixon, W. T., Ren, J., Lubag, A. J. M., Ratnakar, J., Vinogradov, E., Hancu, I., Lenkinski, R. E., and Sherry, A. D. (2010). A concentration-independent method to measure exchange rates in PARACEST agents. *Magnetic Resonance in Medicine 63*, 625–632.

(*148*) Aime, S., Delli Castelli, D., and Terreno, E. (2002). Novel pH-Reporter MRI Contrast Agents. *Angewandte Chemie International Edition 41*, 4334–4336.

(*149*) Zhang, S., Merritt, M., Woessner, D. E., Lenkinski, R. E., and Sherry, A. D. (2003). PARACEST Agents: Modulating MRI Contrast via Water Proton Exchange. *Accounts of Chemical Research 36*, 783–790.

(*150*) Wang, A. M., Leung, G. K., Kiang, K. M., Chan, D., Cao, P., and Wu, E. X. (2017). Separation and quantification of lactate and lipid at 1.3 ppm by diffusion-weighted magnetic resonance spectroscopy. *Magnetic Resonance in Medicine 77*, 480–489.

(*151*) Iglic, A., and Kulkarni, C., *Advances in planar lipid bilayers and liposomes. Volume 18*; Elsevier Academic: 2013.

(*152*) Lingwood, D., and Simons, K. (2010). Lipid rafts as a membrane-organizing principle. *Science (New York, N.Y.) 327*, 46–50.

(*153*) Nishimura, K., Matsuura, T., Sunami, T., Fujii, S., Nishimura, K., Suzuki, H., and Yomo, T. (2014). Identification of giant unilamellar vesicles with permeability to small charged molecules. *RSC Advances 4*, 35224.

(*154*) Pereno, V., Carugo, D., Bau, L., Sezgin, E., Bernardino De La Serna, J., Eggeling, C., and Stride, E. (2017). Electroformation of Giant Unilamellar Vesicles on Stainless Steel Electrodes. *ACS Omega 2*, 994–1002.

(*155*) Hyman, A. A., and Brangwynne, C. P. (2011). Developmental Cell Forum Beyond Stereospecificity: Liquids and Mesoscale Organization of Cytoplasm. *Developmental Cell 21*, 14–16.

(*156*) Boeynaems, S., Alberti, S., Fawzi, N. L., Mittag, T., Polymenidou, M., Rousseau, F., Schymkowitz, J., Shorter, J., Wolozin, B., Van Den Bosch, L., Tompa, P., and Fuxreiter, M. (2018). Protein Phase Separation: A New Phase in Cell Biology. *Trends in Cell Biology 28*, 420–435.

(*157*) Brangwynne, C. P., Eckmann, C. R., Courson, D. S., Rybarska, A., Hoege, C., Gharakhani, J., Jülicher, F., and Hyman, A. A. (2009). Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science (New York, N.Y.) 324*, 1729–32.

(*158*) Nott, T. J., Petsalaki, E., Farber, P., Jervis, D., Fussner, E., Plochowietz, A., Craggs, T. D., Bazett-Jones, D. P., Pawson, T., Forman-Kay, J. D., and Baldwin, A. J. (2015). Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Molecular cell 57*, 936–947.

(*159*) Weber, S. C., and Brangwynne, C. P. Getting RNA and protein in phase., 2012.

(*160*) Patel, A. et al. (2015). A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell 162*, 1066–77.

(*161*) Wheeler, R. J., Lee, H.-O. K., Pal, A., Murthy, A. C., Poser, I., Marrone, L., Doeleman, T., Sterneckert, J., Fawzi, N. L., A., M. B., Hermann, A., Alberti, S., and Hyman, A. A. Small molecules for modulating protein driven liquid-liquid phase separation in neurodegenerative disease. *Submitted*.

(*162*) Poser, I. et al. (2008). BAC TransgeneOmics: A high-throughput method for exploration of protein function in mammals. *Nature Methods 5*, 409–415.

(*163*) Karplus, M, and Kuriyan, J (2005). Molecular dynamics and protein function. *Proceedings of the National Academy of Sciences 102*, 6679–6685.

(*164*) Karplus, M., and Petsko, G. A. (1990). Molecular dynamics simulations in biology. *Nature 347*, 631–639.

(*165*) Boehr, D. D., Nussinov, R., and Wright, P. E. (2009). The role of dynamic conformational ensembles in biomolecular recognition. *Nature Chemical Biology 5*, 789–796.

(*166*) Onuchic, J. N., Luthey-Schulten, Z., and Wolynes, P. G. (1997). THEORY OF PROTEIN FOLDING: The Energy Landscape Perspective. *Annual Review of Physical Chemistry 48*, 545–600.

(*167*) Wolynes, P. G. (2005). Recent successes of the energy landscape theory of protein folding and function. *Quarterly Reviews of Biophysics 38*, 405.

(*168*) Dyson, H. J., and Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nature Reviews Molecular Cell Biology 6*, 197–208.

(*169*) Palmer, A. G. (2016). A dynamic look backward and forward. *Journal of Magnetic Resonance 266*, 73–80.

(*170*) Palmer, A. G. (1993). Dynamic properties of proteins from NMR spectroscopy. *Current Opinion in Biotechnology 4*, 385–391.

(*171*) Hansen, D. F., Vallurupalli, P., Lundström, P., Neudecker, P., and Kay, L. E. (2008). Probing chemical shifts of invisible states of proteins with relaxation dispersion NMR spectroscopy: How well can we do? *Journal of the American Chemical Society 130*, 2667–2675.

(*172*) Kleckner, I. R., and Foster, M. P. (2011). An introduction to NMR-based approaches for measuring protein dynamics. *Biochimica et biophysica acta 1814*, 942–68.

(*173*) Yuwen, T., Vallurupalli, P., Kay, L. E., Carr, A. A. H., and Meiboom, P. (2016). Enhancing the Sensitivity of CPMG Relaxation Dispersion to Conformational Exchange Processes by Multiple-Quantum Spectroscopy Angewandte. *Angewandte Chemie (International ed. in English) 55*, 11490–11494.

(*174*) Pervushin, K. (2000). Impact of transverse relaxation optimized spectroscopy (TROSY) on NMR as a technique in structural biology. *Quarterly Reviews of Biophysics 33*, 161–197.

(*175*) Tugarinov, V., and Kay, L. E. (2004). An isotope labeling strategy for methyl TROSY spectroscopy. *Journal of Biomolecular NMR 28*, 165–172.

(*176*) Laio, A., and Parrinello, M. (2002). Escaping free-energy minima. *Proceedings of the National Academy of Sciences of the United States of America 99*, 12562–6.

(*177*) Bernardi, R. C., Melo, M. C., and Schulten, K. (2015). Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochimica et Biophysica Acta (BBA) - General Subjects 1850*, 872–877.

(*178*) 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.

(*179*) Ahmed-Belkacem, A., Colliandre, L., Ahnou, N., Nevers, Q., Gelin, M., Bessin, Y., Brillet, R., Cala, O., Douguet, D., Bourguet, W., Krimm, I., Pawlotsky, J.-M., and Guichou, J.-F. (2016). Fragment-based discovery of a new family of non-peptidic small-molecule cyclophilin inhibitors with potent antiviral activities. *Nature Communications 7*, 12777.

(*180*) Schlegel, J., Armstrong, G., Redzic, J. S., Zhang, F., and Eisenmesser, E. Z. (2009). Characterizing and controlling the inherent dynamics of cyclophilin-A. *Protein Science 18*, NA–NA.

(*181*) Eisenmesser, E. Z., Bosco, D. A., Akke, M., and Kern, D. (2002). Enzyme dynamics during catalysis. *Science 295*, 1520–3.

(*182*) Sweeney, Z. K., Fu, J., and Wiedmann, B. (2014). From Chemical Tools to Clinical Medicines: Nonimmunosuppressive Cyclophilin Inhibitors Derived from the Cyclosporin and Sanglifehrin Scaffolds. *Journal of Medicinal Chemistry 57*, 7145–7159.

(*183*) Ni, S., Yuan, Y., Huang, J., Mao, X., Lv, M., Zhu, J., Shen, X., Pei, J., Lai, L., Jiang, H., and Li, J. (2009). Discovering Potent Small Molecule Inhibitors of Cyclophilin A Using de Novo Drug Design Approach. *Journal of Medicinal Chemistry 52*, 5295–5298.

(*184*) Juarez-Jimenez, J., Gupta, A. A., Karunanithy, G., Mey, A., Georgiou, C., Ioannidis, C., Simone, A. D., Barlow, P. N., Hulme, A. N., Walkinshaw, M. D., Baldwin, A. J., and Michel, J. (2018). Atomic Level Characterisation of Millisecond-Time Scale Protein Motions through a Combined Molecular Simulations and NMR Approach. *bioRxiv*, 490987.

(*185*) Pierce, L. C., Salomon-Ferrer, R., Augusto F. de Oliveira, C., McCammon, J. A., and Walker, R. C. (2012). Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *Journal of Chemical Theory and Computation 8*, 2997–3002.

(*186*) Pande, V. S., Beauchamp, K., and Bowman, G. R. (2010). Everything you wanted to know about Markov State Models but were afraid to ask. *Methods 52*, 99–105.

(*187*) Vögeli, B., Kazemi, S., Güntert, P., and Riek, R. (2012). Spatial elucidation of motion in proteins by ensemble-based structure calculation using exact NOEs. *Nature Structural & Molecular Biology 19*, 1053–1057.

(*188*) Chi, C. N., Vögeli, B., Bibow, S., Strotz, D., Orts, J., Güntert, P., and Riek, R. (2015). A Structural Ensemble for the Enzyme Cyclophilin Reveals an Orchestrated Mode of Action at Atomic Resolution. *Angewandte Chemie International Edition 54*, 11657–11661.

(*189*) Gamble, T. R., Vajdos, F. F., Yoo, S, Worthylake, D. K., Houseweart, M, Sundquist, W. I., and Hill, C. P. (1996). Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. *Cell 87*, 1285–94.

(*190*) Ottiger, M., Zerbe, O., Güntert, P., and Wüthrich, K. (1997). The NMR solution conformation of unligated human cyclophilin A 1 1Edited by P. E. Wright. *Journal of Molecular Biology 272*, 64–81.

(*191*) Kumar, A., Christy Rani Grace, R., and Madhu, P. K. (2000). Cross-correlations in NMR. *Progress in nuclear magnetic resonance spectroscopy 37*, 191–319.

(*192*) Vogeli, B., Segawa, T. F., Leitz, D., Sobol, A., Choutko, A., Trzesniak, D., van Gunsteren, W., and Riek, R. (2009). Exact Distances and Internal Dynamics of Perdeuterated Ubiquitin from NOE Buildups. *Journal of the American Chemical Society 131*, 17215–17225.

(*193*) Li, F., Lee, J. H., Grishaev, A., Ying, J., and Bax, A. (2015). High accuracy of Karplus equations for relating three-Bond J couplings to protein backbone torsion angles. *ChemPhysChem 16*, 572–578.

(*194*) Bax, A., Kontaxis, G., and Tjandra, N. (2001). Dipolar couplings in macromolecular structure determination. *Methods in Enzymology 339*, 127–174.

(*195*) Zweckstetter, M., and Bax, A. (2002). Evaluation of uncertainty in alignment tensors obtained from dipolar couplings. *Journal of Biomolecular NMR 23*, 127–137.

(*196*) Han, B., Liu, Y., Ginzinger, S. W., and Wishart, D. S. (2011). SHIFTX2: significantly improved protein chemical shift prediction. *Journal of Biomolecular NMR 50*, 43–57.

(*197*) Tamiola, K., Acar, B., and Mulder, F. A. A. (2010). Sequence-Specific Random Coil Chemical Shifts of Intrinsically Disordered Proteins. *Journal of the American Chemical Society 132*, 18000–18003.

(*198*) Millet, O., Loria, J. P., Kroenke, C. D., Pons, M., and Palmer, A. G. (2000). The static magnetic field dependence of chemical exchange linebroadening defines the NMR chemical shift time scale. *Journal of the American Chemical Society 122*, 2867–2877.

(*199*) Lipari, G., and Szabo, A. (1982). Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *Journal of the American Chemical Society 104*, 4559–4570.

(*200*) 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. *Journal of the American Chemical Society 104*, 4546–4559.

(*201*) Clore, G. M., Szabo, A., Bax, A., Kay, L. E., Driscoll, P. C., and Gronenborn, A. M. (1990). Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *Journal of the American Chemical Society 112*, 4989–4991.

(*202*) Mandel, A. M., Akke, M., and Palmer, A. G. (1995). Backbone dynamics of Escherichia coli ribonuclease HI: Correlations with structure and function in an active enzyme. *Journal of Molecular Biology 246*, 144–163.

(*203*) D'Auvergne, E. J., and Gooley, P. R. (2006). Model-free model elimination: A new step in the model-free dynamic analysis of NMR relaxation data. *Journal of Biomolecular NMR 35*, 117–135.

(*204*) Kay, L. E., Torchia, D. A., and Bax, A. (1989). Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry 28*, 8972–8979.

(*205*) Farrow, N. A., Zhang, O., Szabo, A., Torchia, D. A., and Kay, L. E. (1995). Spectral density function mapping using 15N relaxation data exclusively. *Journal of Biomolecular NMR 6*, 153–162.

(*206*) Bieri, M., D'Auvergne, E. J., and Gooley, P. R. (2011). RelaxGUI: A new software for fast and simple NMR relaxation data analysis and calculation of ps-ns and ls motion of proteins. *Journal of Biomolecular NMR 50*, 147–155.

(*207*) Vallurupalli, P., Bouvignies, G., and Kay, L. E. (2011). Increasing the Exchange Time-Scale That Can Be Probed by CPMG Relaxation Dispersion NMR. *The Journal of Physical Chemistry B 115*, 14891–14900.

(*208*) Reddy, J. G., Pratihar, S., Ban, D., Frischkorn, S., Becker, S., Griesinger, C., and Lee, D. (2018). Simultaneous determination of fast and slow dynamics in molecules using extreme CPMG relaxation dispersion experiments. *Journal of Biomolecular NMR 70*, 1–9.

(*209*) Hansen, D. F., Vallurupalli, P., and Kay, L. E. (2008). Using relaxation dispersion NMR spectroscopy to determine structures of excited, invisible protein states. *Journal of Biomolecular NMR 41*, 113–120.

(*210*) Carver, J., and Richards, R. (1972). A general two-site solution for the chemical exchange produced dependence of T2 upon the carr-Purcell pulse separation. *Journal of Magnetic Resonance (1969) 6*, 89–105.

(*211*) Baldwin, A. J. (2014). An exact solution for R2,eff in CPMG experiments in the case of two site chemical exchange. *Journal of Magnetic Resonance 244*, 114–124.

(*212*) Vallurupalli, P., Flemming Hansen, D, Stollar, E., Meirovitch, E., and Kay, L. E. (2007). Measurement of bond vector orientations in invisible excited states of proteins. *Proceedings of the National Academy of Sciences 104*, 18473–18477.

(*213*) Korzhnev, D. M., Orekhov, Y., Dahlquist, F. W., and Kay, L. E. Off-resonance R 1ρ relaxation outside of the fast exchange limit: An experimental study of a cavity mutant of T4 lysozyme., tech. rep., 2003, pp 39–48.

(*214*) Mulder, F. A., de Graaf, R. A., Kaptein, R., and Boelens, R. (1998). An Off-resonance Rotating Frame Relaxation Experiment for the Investigation of Macromolecular Dynamics Using Adiabatic Rotations. *Journal of Magnetic Resonance 131*, 351–357.

(*215*) Trott, O., and Palmer, A. G. (2002). R1ρ Relaxation outside of the Fast-Exchange Limit. *Journal of Magnetic Resonance 154*, 157–160.

(*216*) Baldwin, A. J., and Kay, L. E. (2013). An R 1ρ expression for a spin in chemical exchange between two sites with unequal transverse relaxation rates. *J Biomol NMR 55*, 211–218.

(*217*) Shaw, D. E. et al. (2008). Anton, a special-purpose machine for molecular dynamics simulation. *Communications of the ACM 51*, 91.

(*218*) Shaw, D. E., Maragakis, P., Lindorff-Larsen, K., Piana, S., Dror, R. O., Eastwood, M. P., Bank, J. A., Jumper, J. M., Salmon, J. K., Shan, Y., and Wriggers, W. (2010). Atomic-level characterization of the structural dynamics of proteins. *Science (New York, N.Y.) 330*, 341–6.

(*219*) De Simone, A., Georgiou, C., Ioannidis, H., Gupta, A. A., Juárez-Jiménez, J., Doughty-Shenton, D., Blackburn, E. A., Wear, M. A., Richards, J. P., Barlow, P. N., Carragher, N., Walkinshaw, M. D., Hulme, A. N., and Michel, J. (2019). A computationally designed binding mode flip leads to a novel class of potent tri-vector cyclophilin inhibitors. *Chemical Science 10*, 542–547.

(*220*) Sekhar, A., and Kay, L. E. (2013). NMR paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers. *Proceedings of the National Academy of Sciences of the United States of America 110*, 12867–74.

(*221*) Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., and Lindah, E. (2015). Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX 1-2*, 19–25.

(*222*) Scherer, M. K., Trendelkamp-Schroer, B., Paul, F., Pérez-Hernández, G., Hoffmann, M., Plattner, N., Wehmeyer, C., Prinz, J.-H., and Noé, F. (2015). PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *Journal of Chemical Theory and Computation 11*, 5525–5542.

(*223*) McGibbon, R., Beauchamp, K., Harrigan, M., Klein, C., Swails, J., Hernández, C., Schwantes, C., Wang, L.-P., Lane, T., and Pande, V. (2015). MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal 109*, 1528–1532.

(*224*) Zweckstetter, M. (2008). NMR: Prediction of molecular alignment from structure using the PALES software. *Nature Protocols 3*, 679–690.

(*225*) Robustelli, P., Trbovic, N., Friesner, R. A., and Palmer, A. G. (2013). Conformational Dynamics of the Partially Disordered Yeast Transcription Factor GCN4. *Journal of Chemical Theory and Computation 9*, 5190–5200.

(*226*) Wear, M. A., Nowicki, M. W., Blackburn, E. A., McNae, I. W., and Walkinshaw, M. D. (2017). Thermo-kinetic analysis space expansion for cyclophilin-ligand interactions - identification of a new nonpeptide inhibitor using Biacore™ T200. *FEBS Open Bio 7*, 533–549.

(*227*) Vallurupalli, P., Hansen, D. F., and Kay, L. E. (2008). Structures of invisible, excited protein states by relaxation dispersion NMR spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America 105*, 11766–71.

(*228*) Gupta, A. A., Reinartz, I., Karunanithy, G., Spilotros, A., Jonna, V. R., Hofer, A., Svergun, D. I., Baldwin, A. J., Schug, A., and Wolf-Watz, M. (2018). Formation of a Secretion-Competent Protein Complex by a Dynamic Wrap-around Binding Mechanism. *Journal of Molecular Biology 430*, 3157–3169.

(*229*) Cornelis, G. R. (2006). The type III secretion injectisome. *Nature Reviews Microbiology 4*, 811–825.

(*230*) Galán, J. E., and Wolf-Watz, H. (2006). Protein delivery into eukaryotic cells by type III secretion machines. *Nature 444*, 567–573.

(*231*) Loquet, A., Sgourakis, N. G., Gupta, R., Giller, K., Riedel, D., Goosmann, C., Griesinger, C., Kolbe, M., Baker, D., Becker, S., and Lange, A. (2012). Atomic model of the type III secretion system needle. *Nature 486*, 276–279.

(*232*) Worrall, L. J., Hong, C., Vuckovic, M., Deng, W., Bergeron, J. R., Majewski, D. D., Huang, R. K., Spreter, T., Finlay, B. B., Yu, Z., and Strynadka, N. C. (2016). Near-atomic-resolution cryo-EM analysis of the Salmonella T3S injectisome basal body. *Nature 540*, 597–601.

(*233*) Kuhlen, L., Abrusci, P., Johnson, S., Gault, J., Deme, J., Caesar, J., Dietsche, T., Mebrhatu, M. T., Ganief, T., Macek, B., Wagner, S., Robinson, C. V., and Lea, S. M. (2018). Structure of the core of the type III secretion system export apparatus. *Nature Structural & Molecular Biology 25*, 583–590.

(*234*) Page, A. L., and Parsot, C. (2002). Chaperones of the type III secretion pathway: Jacks of all trades. *Molecular Microbiology 46*, 1–11.

(*235*) Stebbins, C. E., and Galán, J. E. (2001). Maintenance of an unfolded polypeptide by a cognate chaperone in bacterial type III secretion. *Nature 414*, 77–81.

(*236*) Rodgers, L., Gamez, A., Riek, R., and Ghosh, P. (2008). The type III secretion chaperone SycE promotes a localized disorder-to-order transition in the natively unfolded effector YopE. *Journal of Biological Chemistry 283*, 20857–20863.

(*237*) Lilic, M., Vujanac, M., and Stebbins, C. E. (2006). A Common Structural Motif in the Binding of Virulence Factors to Bacterial Secretion Chaperones. *Molecular Cell 21*, 653–664.

(*238*) Lloyd, S. A., Sjostrom, M., Andersson, S., and Wolf-Watz, H. (2002). Molecular characterization of type III secretion signals via analysis of synthetic N-terminal amino acid sequences. *Molecular Microbiology 43*, 51–59.

(*239*) Guan, J, and Dixon, J. E. Protein Tyrosine Phosphatase Activity of an Essential Virulence Determinant in Yersinia., 1990.

(*240*) Evdokimov, A. G., Tropea, J. E., Routzahn, K. M., Copeland, T. D., and Waugh, D. S. (2001). Structure of the N-terminal domain of Yersinia pestis YopH at 2.0 Å resolution. *Acta Crystallographica Section D: Biological Crystallography 57*, 793–799.

(*241*) Lee, W., Tonelli, M., and Markley, J. L. (2015). NMRFAM-SPARKY: Enhanced Software for Biomolecular NMR Spectroscopy. *Bioinformatics 31*, 1325–1327.

(*242*) Hansen, D. F., Yang, D., Feng, H., Zhou, Z., Wiesner, S., Bai, Y., and Kay, L. E. (2007). An exchange-free measure of 15N transverse relaxation: An NMR spectroscopy application to the study of a folding intermediate with pervasive chemical exchange. *Journal of the American Chemical Society 129*, 11468–11479.