

**SCHOOL OF CHEMISTRY  
CARDIFF UNIVERSITY**

**Effect of organic co-solvents on the  
NMR chemical shifts and dynamics  
of dihydrofolate reductase from *E.*  
*coli***



A thesis submitted to Cardiff  
University for the degree of Master of  
Philosophy

By  
Stella Cosgrave

**2018**



## DECLARATION

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.

Signed ..... (candidate)      Date .....

## STATEMENT 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of  
MPhil

Signed ..... (candidate)      Date .....

## STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by explicit references. The views expressed are my own.

Signed ..... (candidate)      Date .....

## STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ..... (candidate)      Date .....

## **Acknowledgements**

First and foremost, my thanks go to my supervisor Rudolf Allemann and the extended Allemann group for the opportunity to work on this project and the ongoing support provided throughout the course of my work.

Many thanks to Dr Sara Whittaker (NMR facilities at School of Cancer Studies, University of Birmingham), Alain Oregioni and Geoff Kelly (MRI, Mill Hill), Rob Jenkins and Joel Loveridge (Cardiff University) and Matthew Crump (University of Bristol) for their expertise in NMR.

## Abstract

Enzymes are large biological molecules that can exist in many different states and each state may have multiple conformations. The probability of finding an enzyme in a particular state is governed by the free energy landscape and the energy barriers between the different states. The energy landscape of an enzyme can be perturbed through making changes to the external conditions such as the reaction medium, which can cause a shift in the delicate equilibrium between states. Experimental evidence suggests that the model for environmentally coupled tunnelling is inconsistent for catalysis by dihydrofolate reductase (DHFR) from *E coli*. It is proposed that catalysis is governed by electrostatics and characterised by populations of sub-states with distinct conformations and kinetics. Here, using solution NMR spectroscopy, the enzyme is probed at the point of the chemical reaction step by subjecting a mimic of the Michaelis complex (EcDHFR: NADP<sup>+</sup>: folate) to organic co-solvents in order to perturb the energy landscape of the protein by altering the viscosity and dielectric constant of the enzymatic reaction medium. The chemical shift assignment of the Michaelis complex mimic under standard buffer conditions, in the presence of 17% methanol and 17 % glycerol co-solvents (to effect almost isodielectric mediums that differ in viscosity) is reported. The majority of atoms in the protein complex in the presence of 17% glycerol and 17% methanol show small chemical shift perturbations (**Chapter 4.0**). The ps-ns relaxation dynamics of the Michaelis complex mimic were measured *via* solution NMR at 600 MHz and 900 MHz in the presence of co-solvents to determine any alterations in the timescale of dynamics on this timescale. The results indicate slightly increased mobility of some residues in loop regions in the presence of methanol and glycerol co-solvents and slightly decreased mobility of some residues in areas of defined secondary structure (**Chapter 5.0**).

## Abbreviations

<b>A</b>	Absorption
<b>A</b>	Arrhenius pre-exponential factor
<b>ADB</b>	adenosine-binding domain
<b>β-Me</b>	β-mercaptoethanol
<b>C</b>	Concentration
<b>CBD</b>	Cofactor binding domain
<b>CD</b>	Circular dichromism spectroscopy
<b>D</b>	Deuterium
<b>DAD</b>	Donor-acceptor distance
<b>dH<sub>2</sub>O</b>	Deionised water
<b>DHF</b>	7,8 -dihydrofolate
<b>DHFR</b>	dihydrofolate reductase
<b>E<sub>a</sub></b>	activation energy
<b>E</b>	Enzyme
<b>ES</b>	Enzyme-substrate complex
<b>FPLC</b>	Fast performance liquid chromatography
<b>H</b>	Planck's constant = $6.626176 \times 10^{-34} \text{ J s}$
<b>H</b>	Hydrogen
<b>IPTG</b>	isopropyl-β-D-1-thiogalactopyranoside
<b>KIE</b>	Kinetic isotope effect
<b>K</b>	Rate constant
<b>K<sub>B</sub></b>	Boltzmann constant = $1.3806503 \times 10^{-23} \text{ JK}^{-1}$
<b>K<sub>M</sub></b>	Michaelis constant
<b>K<sub>i</sub>PO<sub>4</sub></b>	potassium phosphate
<b>L</b>	Pathlength
<b>LB</b>	Luria-Bertani growth media
<b>LD</b>	Loop domain
<b>M</b>	atomic mass
<b>NADP<sup>+</sup></b>	Oxidised nicotine adenine dinucleotide phosphate
<b>NMR</b>	Nuclear Magnetic Resonance

<b>P</b>	product
<b>PAGE</b>	Polyacrylamide gel electrophoresis
<b>PDB</b>	Protein data bank
<b>R</b>	reactant
<b>R</b>	universal gas constant
<b>S</b>	substrate
<b>S<sup>2</sup></b>	The site-specific squared order parameter
<b>SBD</b>	substrate-binding domain
<b>SDS-PAGE</b>	Sodium dodecyl sulphate Polyacrylamide gel electrophoresis
<b>T</b>	Absolute temperature (K)
<b>T</b>	Tritium
<b>TS</b>	transition state
<b>TST</b>	Transition state theory
<b>TEMED</b>	N,N,N',N'-tetramethyldiamine
<b>UV/Vis</b>	ultraviolet/ visible spectroscopy
<b>ε</b>	extinction coefficient
<b>ZPE</b>	zero point energy
<b>ΔG<sup>0</sup></b>	reaction free energy
<b>ΔG<sup>‡</sup></b>	free activation energy of the reaction
<b>ΔH<sup>‡</sup></b>	activation enthalpy
<b>ΔS<sup>‡</sup></b>	activation entropy
<b>λ</b>	de Broglie wavelength
<b>σ</b>	standard deviation
<b>μ</b>	reduced mass
<b>‡</b>	reactive configuration

<b>Amino Acid</b>	<b>Letter code</b>	<b>Abbreviation</b>
<b>Alanine</b>	<b>A</b>	<b>Ala</b>
<b>Cysteine</b>	<b>C</b>	<b>Cys</b>
<b>Aspartic acid</b>	<b>D</b>	<b>Asp</b>
<b>Glutamic acid</b>	<b>E</b>	<b>Glu</b>
<b>Phenylalanine</b>	<b>F</b>	<b>Phe</b>
<b>Glycine</b>	<b>G</b>	<b>Gly</b>
<b>Histidine</b>	<b>H</b>	<b>His</b>
<b>Isoleucine</b>	<b>I</b>	<b>Ile</b>
<b>Lysine</b>	<b>K</b>	<b>Lys</b>
<b>Methionine</b>	<b>M</b>	<b>Met</b>
<b>Asparagine</b>	<b>N</b>	<b>Asn</b>
<b>Proline</b>	<b>P</b>	<b>Pro</b>
<b>Glutamine</b>	<b>Q</b>	<b>Gln</b>
<b>Arginine</b>	<b>R</b>	<b>Arg</b>
<b>Serine</b>	<b>S</b>	<b>Ser</b>
<b>Threonine</b>	<b>T</b>	<b>Thr</b>
<b>Valine</b>	<b>V</b>	<b>Val</b>
<b>Tryptophan</b>	<b>W</b>	<b>Trp</b>
<b>Tyrosine</b>	<b>Y</b>	<b>Tyr</b>

# Contents

1.0 Introduction .....	1
1.1 Enzymes .....	1
1.1.1 Enzyme – Substrate interactions .....	1
1.1.2 Enzyme-Substrate recognition and binding .....	2
1.1.3 Enzymatic Transition State Theory .....	4
1.1.4 Kinetic Isotope Effects for Characterisation of Transitions States .....	9
1.1.5 The Bell Adjusted Model of Tunnelling .....	11
1.1.6 Contemporary Theories of Enzyme Catalysis .....	12
1.2 Dihydrofolate Reductase .....	20
1.2.1 Dihydrofolate reductase as a model for catalysis .....	20
1.2.2 Dihydrofolate Reductase from <i>E. coli</i> .....	21
1.2.3 Gathering Insight into the Nature of Catalysis by Dihydrofolate Reductase .....	25
1.3 Introduction to Protein Nuclear Magnetic Resonance Spectroscopy .....	29
1.3.1 NMR Theory in brief .....	30
1.4 Relaxation NMR .....	33
1.4.1 Introduction to Relaxation NMR Spectroscopy .....	34
1.4.2 Longitudinal relaxation .....	34
1.4.3 Transverse Relaxation .....	35
1.4.4 The Nuclear Overhauser Effect .....	36
1.4.5 The Heteronuclear NOE .....	37
1.4.6 Measuring Protein Dynamics by NMR .....	38
2.0 Aims and Objectives .....	44
2.2 Analysis of chemical shift perturbation in the presence of co-solvents .....	48
2.3 ps-ns Relaxation Dynamics .....	48
3.0 Experimental .....	49
3.1 Buffers and Media .....	49

3.1.1 Purification and UV buffers.....	49
3.1.2 SDS-PAGE .....	49
3.2 Protein Expression Media .....	51
3.2.1 Preparation of <i>Luria-Bertani</i> (LB) medium .....	51
3.2.2 Preparation of M9 medium (1.0 L).....	51
3.3 Labelled co-solvents.....	51
3.4 Labelled protein expression methods .....	51
3.4.1 Preparation of M9 medium for perdeuteration (0.5 L).....	51
3.4.2 Expression of unlabelled EcDHFR.....	52
3.4.3 Expression of <sup>15</sup> N/ <sup>13</sup> C EcDHFR .....	52
3.4.4 Expression of perdeuterated EcDHFR.....	53
3.5 Protein purification.....	53
3.5.1 Purification of EcDHFR .....	53
3.6 Preparation of EcDHFR NMR samples .....	54
3.6.1 Buffer only sample preparation .....	54
3.6.2 Preparation of sample containing 17% glycerol.....	54
3.6.3 Preparation of sample containing 17% methanol .....	54
3.7 Analytical Methods .....	55
3.7.1 Ultraviolet photospectrometry .....	55
3.7.2 Nuclear magnetic resonance spectroscopy .....	56
4.0 Assignment and Analysis of the chemical shift perturbations EcDHFR: NADP <sup>+</sup> : Folate in the presence of co-solvents .....	60
4.1 Biosynthesis, purification and preparation of protein samples .....	63
4.2 Acquisition of NMR spectra.....	63
4.3 Assignment of EcDHFR: NADP <sup>+</sup> : folate in buffer .....	65
4.4 Assignment of EcDHFR in 17% methanol .....	72
4.5 Assignment of EcDHFR in 17% glycerol .....	73

4.6 Assignment Conclusions .....	74
4.7 Analysis of the chemical shift perturbations of EcDHFR: NADP <sup>+</sup> : folate in the presence of co-solvents .....	75
4.7.1 Backbone Amide <sup>1</sup> H <sup>15</sup> N Chemical Shift Perturbations.....	76
4.7.2 Weighted chemical shift Perturbations .....	79
4.7.3 Chemical shift perturbation mapping .....	81
4.8 C $\alpha$ and H $\alpha$ Chemical Shift perturbations .....	84
4.8.1 Chemical shift Perturbations.....	84
4.8.2 Chemical shift perturbation mapping .....	88
4.9 C $\beta$ and H $\beta$ Chemical Shifts.....	90
4.9.1 H $\beta$ and H $\beta$ Chemical Shift Perturbations .....	90
4.8.2 C $\beta$ and H $\beta$ Chemical shift mapping .....	91
4.9 Side chain Chemical Shift Perturbations.....	93
4.10 Chemical shift perturbations across the whole protein .....	94
4.11 Discussions .....	96
4.12 Conclusions .....	98
5.0 Relaxation dynamics of EcDHFR: NADP <sup>+</sup> : Folate in the presence of co-solvents..	99
5.1 Biosynthesis, purification and preparation of protein samples .....	100
5.2 Longitudinal relaxation data analysis .....	100
5.4 T1/T2 Analysis .....	113
5.5 The Heteronuclear NOE .....	118
5.6 Conclusions from phenomenological relaxation data analysis.....	121
6.0 Overall Conclusions and Further work .....	122
6.1 Overall Conclusions .....	122
6.2 Further Work .....	124
6.2.1 NMR Solution Structure Calculations .....	124
6.2.2 Lipari-Szabo .....	127

7.0 Published Work..... 128

8.0 References..... 129

## 1.0 Introduction

### 1.1 Enzymes

Enzymes are a class of proteins, which catalyse biochemical reactions, achieving rate enhancements of up to  $10^{19}$  when compared to un-catalysed reactions.<sup>1</sup> Enzymatic reactions account for the whole spectrum of cellular chemistry, yet reactions are often performed under challenging conditions at ambient temperature in aqueous solution and at physiological pH. Developing a full understanding of the way that enzymes achieve such efficiency is at the forefront of biochemistry.<sup>1</sup> With a better understanding of the way that enzymes work, on both atomic and global level, much needed insight will be available for the design of artificial enzymes and developing new drugs to target enzymatic based diseases.

#### 1.1.1 Enzyme – Substrate interactions

Enzymes function by binding substrates and co-factors in the active site through formation of energetically favorable interactions. Of these interactions non-covalent bonds are most typical, consisting of *electrostatic interactions* including ionic and van der Waals contacts, as well as *hydrogen bonds* and the *hydrophobic effect*. Electrostatic interactions form between opposite charges and are most obvious for substrates that have functional groups that are ionisable in aqueous solution and at physiological pH, such as carboxylate groups. In a vacuum, the strength of the electrostatic force between two charges follows Coulombs law and is dependent on the inverse squared distance between the charges. Typical electrostatic binding energies are between 25-50 kJ Mol<sup>-1</sup>. Electrostatic interactions are also present for partial charges, such as dipoles.<sup>2</sup> *Van der Waals interactions*, a subtype of electrostatic interactions, result from forces between permanent dipoles, permanent dipole induced dipoles and, instantaneously induced dipoles. Van der Waals interactions are short range (2-4 Å) as the field strength of the interactions is proportional to the distance of the interaction (r) ( $1/r^6$ ). The energy of these individual interactions are around 6-8 kJ mol<sup>-1</sup>, smaller than for discrete charge interactions, however still substantial for substrates that fit efficiently

into the enzyme active site, as significant binding energy can be achieved through maximising close surface contact between the enzyme and substrate.

*Hydrogen bonds* form between hydrogen bond donors and acceptors. Hydrogen bond donors are electronegative atoms possessing a lone pair of electrons; hydrogen bond acceptors possess electron deficient hydrogen atoms that are capable of accepting a lone pair of electrons. The strength of a hydrogen bond is dependent on whether the donors and acceptors are charged, which can result in bonds with energy of 2.0 -7.5 kJ mol<sup>-1</sup> and 12.5-25 kJ mol<sup>-1</sup> respectively, in solution.<sup>2</sup>

### **1.1.2 Enzyme-Substrate recognition and binding**

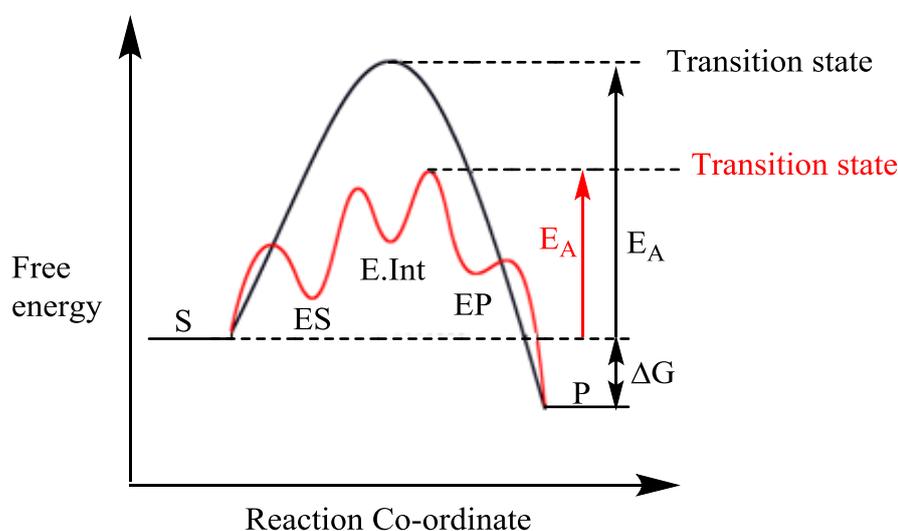
The “Lock and Key” approach to enzyme catalysis proposed by Emil Fischer provided a simplistic model of how a substrate (the key) binds to a rigid enzyme active site (the lock) prior to catalysis. In this model, the enzyme’s active site is complementary to the substrate, allowing for efficient binding (maximising surface interactions). Both the enzyme and substrate shapes were proposed to be fixed, and not to alter as a result of the enzyme and substrate coming in to close proximity with one another. Whilst this provided the first model of enzyme catalysis and described a role for specific binding interactions, dictating that the enzyme and the substrate must form complementary interactions, it did not provide a mechanism for how the catalytic reaction may be achieved once binding had occurred, nor provide a model for development of a transition state (**Section 1.1.3**).<sup>3</sup>

Koshland extended the lock and key theory through giving consideration to the role of water during catalysis. He postulated that as water in solution has a concentration of 55 molar, that this would fill the active site of the enzyme. Through considering a kinase with a glucose substrate, he recognised that the substrate (at a concentration of much less than water) would be in competition with water to occupy the enzyme active site. Koshland concluded that the process of glucose displacing water from the active site would require binding energy so consequentially, the enzyme must efficiently exclude water from the active site and maximise interactions to the substrate. The induced-fit theory was developed from these considerations, along with ideas regarding inhibition, whereby a substrate analogue binds to the enzyme at a higher affinity than the substrate but does not turnover to form a product. The induced fit theory describes a flexible

“hand in glove” model, where both the substrate and enzyme maximise interactions by altering their final shapes such that the catalytic groups are oriented correctly for enzyme action. A change in the three-dimensional relationship of the amino acids at the active site brings the catalytic groups in to the proper alignment for turnover to occur, whereas a non-substrate will not.<sup>3</sup>

### 1.1.3 Enzymatic Transition State Theory

Enzymes are biological catalysts that accelerate reaction rates without themselves having a net chemical change. Linus Pauling proposed that enzymes accelerate biochemical reaction rates by binding the transition state better than the substrate alone,<sup>4</sup> a theory which is supported by experimental work utilising synthetic approaches to mimic proposed enzyme transition states.<sup>5</sup> Wolfenden proposed chemically stable analogues resembling the transition state would bind more tightly than substrates by factors resembling the rate enhancement imposed by enzymes.<sup>6</sup>



**Figure 1** An illustration of the free energy change ( $\Delta G$ ) as enzyme catalysed (red) and un-catalysed (black) reactions progress through the reaction coordinate from substrate (S) to product (P). The activation energy ( $E_A$ ) for the enzyme catalysed reaction is lower than for the un-catalysed reaction but the overall free energy change for both reactions is the same. The transition state is a local energy maximum whereas any semi-stable intermediates formed in the process of the reaction such as **ES**, **E.Int** or **EP** are local energy minima.<sup>7</sup>

In a chemical reaction, the transition state (TS) describes the state with the highest free energy along the reaction coordinate. Through binding to and lowering the energy of the TS of the substrate(s) or product(s), enzymes stabilise the TS, reducing the activation barrier to the reaction and enhancing the reaction rate (**Figure 1**). Enzymes must therefore adopt electrostatic and geometric interactions with the reactants that are complementary to the TS in order to lower the energy of the TS. Maximisation of the favourable interactions with the TS results in a reduction of the activation energy, the rate limiting barrier to the chemical reaction.<sup>8</sup>

Transition state theory (TST) is a statistical mechanical approach to understanding complex chemical reactions, proposed by Henry Eyring. This branch of theoretical physics uses probability theory to study the average behaviour of a mechanical system. TST utilises statistical mechanics to calculate the equilibrium constant for a chemical system from the partition functions of the species involved which in turn, allows the chemical potential or the Gibbs free energy of a species to be calculated.<sup>9</sup> The rate of enzyme catalysed reactions and the relationship to the free energy ( $\Delta G^\ddagger$ ) of the TS is described by the Arrhenius equation (**Equation 1**).<sup>10</sup> In TST, the enzyme lowers the barrier to the reaction by stabilisation of the transition state such that the ratio of the rate of the catalysed reaction to the un-catalysed reaction =  $k_{\text{cat}}/k_{\text{uncat}} \exp^{-(E_{\text{uncat}} - E_{\text{cat}})/RT}$  is typically many orders of magnitude greater, for example, in the case of phosphoryl transfer reactions catalysed by kinases can be up to  $10^{20}$ .<sup>11, 12</sup>

$$k = A(T) \exp^{-\Delta G^\ddagger / (RT)}$$

**Equation 1** Arrhenius equation where T is the absolute temperature (Kelvin), R is the universal gas constant and A is the pre-exponential function which includes factors such as the transmission coefficient ( $\kappa$ ), friction factors and re-crossing events.

TST can be applied to an enzymatic chemical reaction provided that the timescale for re-equilibration of the pool of reactants is faster than the transformation of product to reactant.<sup>13</sup> This assumption holds provided that the activated complexes are in a state of quasi-equilibrium with the reactants and that the motion of the system along the reaction coordinate can be treated as a free translational motion and expressed by using kinetic theory.<sup>14</sup> The limitations of the theory were widely recognised at the time, particularly as the TST model fails for short-lived complexes (quasi-equilibrium assumption) and at high temperatures because it assumes the TS is the lowest energy saddle point on the potential energy surface (PES - *see below*). The TST model further assumes that molecules must collide with enough energy in order to react, which is contrary to the quantum mechanical view allowing hydrogen tunnelling.<sup>14</sup>

Chemical reactions involve bonds between atoms being broken, which requires energy. A quantum mechanical description of a static diatomic molecule can be modeled as a simple harmonic oscillator. In a simple model, this can be envisioned as two balls connected by a spring. The potential energy stored in the molecule or spring changes as

the length of the spring is distorted from the equilibrium position. For small strains, Hooke's law (**Equation 2**) provides the basis for the model of the change in potential energy ( $E_{P.E.}$ ), where the change in potential energy is dependent on the spring or force constant and the displacement of the bond from the equilibrium position.<sup>15</sup>

$$E_{P.E.} = \frac{1}{2} k x_d^2$$

**Equation 2** Hooke's law provides the basis for changes in the potential energy for small strains. The change in potential energy is proportional to the force constant ( $k$ ) and the displacement from the equilibrium bond length ( $x_d$ ) squared ( $x_d = x - x_{Eq}$ ).

In reality, molecules are not static. Vibrations are an indication of the kinetic energy possessed in the system. The solution to the Schrodinger equation allows derivation of the permitted vibrational energy levels in a molecule (**Equation 3**). The zero point energy corresponds to the lowest irremovable vibrational energy, for example, when the lowest or zeroeth vibrational energy level is populated ( $v=0$ ) (**Equation 4**).<sup>16</sup>

$$E_v = \left( v + \frac{1}{2} \right) \frac{h}{2\pi} \sqrt{\frac{k}{\mu}}$$

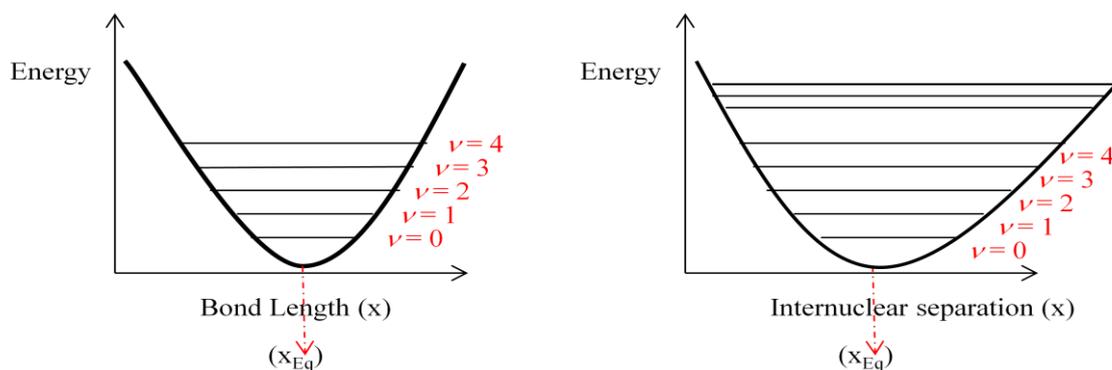
**Equation 3** The energy levels of a harmonic oscillator are evenly spaced and calculated from the equation above (where permitted energy levels of  $v = 0, 1, 2, 3, \dots$ ). In the lowest energy state (where  $v = 0$ ), the term  $(v + 1/2)$  simplifies to  $1/2$ , the oscillator has a vibrational energy greater than zero.  $h$  = Planck's constant,  $k$  is a force constant and  $\mu$  is the reduced mass of the two atoms involved in the bond breaking reaction,  $m_1$  and  $m_2$  ( $\mu = m_1 m_2 / m_1 + m_2$ ).

$$E_{z.p.} = \frac{h}{2\pi} \sqrt{\frac{k}{\mu}}$$

**Equation 4** For calculation of the zero point energy of a harmonic oscillator where  $h$  = Planck's constant,  $k$  is a force constant and  $\mu$  is the reduced mass of the two atoms involved in the bond breaking reaction,  $m_1$  and  $m_2$  ( $\mu = m_1 m_2 / m_1 + m_2$ ).

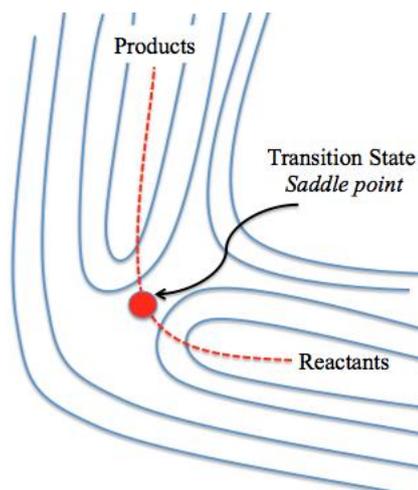
The relationship between the potential energy and bond length is linear for small bond strains. Larger bond strains are poorly described by the parabola (**Figure 2, left**) and the potential energy profile of the molecule is better described by (**Figure 2, right**) at the bond dissociation limit at which the bond vibration is anharmonic and the vibrational energy levels converge. In diatomic molecules, there is need for only one geometric

parameter (bond length) to be described in the energy profile. The energy profile can be extended for molecule with more than one geometric parameter, such a triatomic molecule with 2 bonds and a bond angle. Consideration of a 3 dimensional energy profile, along with molecular symmetry and group theory gives rise to potential energy surfaces.<sup>17</sup>



**Figure 2** Potential energy forms of a harmonic oscillator. The left hand side diagram shows the potential energy profile for the vibrational energy levels in an ideal harmonic oscillator. The right hand diagram represents the convergence of vibrational energy levels in a diatomic molecule at large inter-nuclear separation.

The potential energy surface (PES) of a reaction is a mathematical function providing an alternative method to describing the enzymatic reaction coordinate in terms of molecular energy and geometry. This expansion on transition state theory allows the enzyme catalysed reaction to be viewed as a 3D energetic surface. The PES can be constructed from experimental data and results from quantum chemical calculations, displaying the potential energy as a function of the relative positions of all of the atoms taking part in the reaction. It is noted above, that the PES for a diatomic molecule is two dimensional; this is because a diatomic molecule has just one degree of freedom, the chemical bond and no bond angles to consider. For each additional degree of freedom possessed by the molecule, there will be one more dimension to the PES (Degrees of freedom =  $3N-6$  for a linear molecule and  $3N-5$  for non-linear, where  $N$  = number of atoms in molecule). These surfaces provide a visual display of the paths available to the system that correspond to the least potential energy. Invoking the Born-Oppenheimer approximation simplifies the calculations by assuming that electrons move much faster than any change in electronic configuration would occur. Each possible electronic state can be defined by a unique PES, however in general the ground state PES is used.<sup>18</sup>



**Figure 3** The potential energy surface displays the reactant(s) and product(s) at local energy minima. The transition state is a *saddle point* or local energy maximum through which the reactant(s) must pass in order to react. The reaction trajectory is shown by the dotted red line and denotes the lowest energy pathway from substrate(s) to product(s).

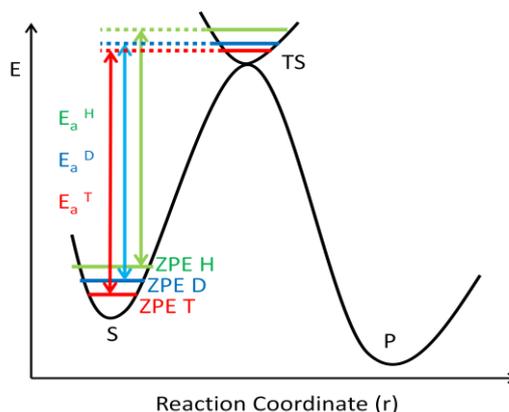
PES demonstrates that TST describes an ensemble of different states and that the transition state represents a critical geometry or set of configurations that every trajectory must go through in order to progress from reactant to product. On the PES, the TS is represented by a “saddle point” or region of the trajectory with maximum potential energy (**Figure 3**). Enzymes can therefore accelerate reactions by either lowering the energy of the TS or by progressing through alternative TS of lower energy when compared to that of the un-catalysed reaction.<sup>8</sup>

### 1.1.4 Kinetic Isotope Effects for Characterisation of Transitions States

Kinetic isotope measurements are a very popular method to characterise the transition state of enzyme catalysed reactions when the chemical step involves a bond breaking or formation.<sup>19</sup> The kinetic rate of the reaction can be investigated by systematically replacing atoms present in the substrate with their heavier isotope counterparts. If one of the atoms in the bond being broken is a heavier isotope, any change in reaction rate that is observed is called a primary kinetic isotope effect (pKIE). If the heavier isotope is present anywhere else in the substrate molecule other than being one of the atoms involved in the bond breakage or formation, any change observed in the reaction rate is called a secondary isotope effect (sKIE).<sup>20</sup>

According to the semi-classical theory, primary kinetic isotope effects (KIEs) arise due to the difference in reduced mass ( $\mu$ ) of the bonds containing the heavier and lighter isotopes, which leads to a difference in zero point energy ( $E_{z.p.}$ ) of two bonded isotopes (1.1.3. Enzymatic Transition State Theory), e.g.  $X-^1H$  and  $X-^2H$ .

**Equation 4** simplifies analysis such that for the  $X-^1H$  and  $X-^2H$  bonds, the only difference is assumed to be between the reduced mass of the atoms forming the bond. In this example, the reduced mass for the bond containing deuterium is greater than hydrogen so that the energy barrier for the breakage of the bond to the deuterium atom will be greatest. Secondary KIEs occur as a result of changes in the vibrational frequencies of motions orthogonal to the reaction coordinate.<sup>21</sup>



**Figure 4** According to the semi-classical theory, primary kinetic isotope effects (KIEs) arise due to the difference in reduced mass ( $\mu$ ) of the bonds containing the heavier and lighter isotopes, which leads to a difference in zero point energy ( $E_{z.p.}$ ) of two bonded isotopes, e.g.  $X-^1H$  and  $X-^2H$  in the substrate and transition state.

The difference between the zero point vibrational energies for  $^{12}\text{C}-^1\text{H}$  and  $^{12}\text{C}-^2\text{H}$  bonds is assumed to be the difference in energy required to break the two bonds. Heavier isotopes possess lower ZPEs therefore the energy required to break the bond is greater (**Figure 4**). This difference leads to a slower rate of reaction in comparison to the lighter isotope reaction. The ratio of the reaction rates for the light and heavy isotope reactions ( $\text{X}-^1\text{H}$  and  $\text{X}-^2\text{H}$ ) is derived from the ratio of the Arrhenius equations. The semi-classical model assumes that isotope effect on the pre-exponential factor  $A$  is close to unity, so a simplified ratio of ( $k_{\text{H}}/k_{\text{D}}$ ) predicts the primary KIE (**Equation 5**).<sup>22</sup> The ratio is based on the Arrhenius equation, so is accordingly temperature-dependent and tends to zero at infinite temperatures. The maximum ratio of  $k_{\text{H}}/k_{\text{D}}$  predicted from this classical model at room temperature (298 K) is 7.

KIEs values that are less than this maximum value of 7 may occur if the hydride transfer step of the reaction is only partially rate limiting, or if vibrational frequencies are present in the transition state that cancel the ground state frequencies. Meanwhile, experimental values much greater than the theoretical maximum have also been observed. Whilst it is recognised that KIEs greater than the theoretical maximum are not reliable indicators of quantum mechanical effects, inflated KIEs have been used as an indicator that the classical model is limited and that quantum mechanical tunnelling may play an important role in the reaction mechanism. These experimental results underline that the TST fails to account for important mechanisms in the reaction.<sup>20</sup>

$$\frac{k_{\text{H}}}{k_{\text{D}}} = \frac{\exp^{-\Delta H^{\ddagger}/RT}}{\exp^{-\Delta H^{\ddagger} + 4.8\text{kJmol}^{-1}/RT}} = \exp^{4.8\text{kJmol}^{-1}/RT}$$

**Equation 5** Primary kinetic isotope effect where  $k_{\text{H}}$  and  $k_{\text{D}}$  are the rates of reaction in which the bond broken is to hydrogen and deuterium, respectively;  $R$  is the universal gas constant;  $T$  is the temperature in Kelvin

The Swain-Schaad exponential was proposed to explain the kinetic relationship between the KIEs for hydrogen, deuterium and tritium (**Equation 6**). The exponents are calculated from the reduced masses of the different isotopes. A comprehensive derivation of these ratios is discussed in a recent review.<sup>21</sup> In enzymatic reactions, provided that the reaction step giving rise to the KIE is fully rate limiting, the Swain-Schaad relationship can be used as an indicator of tunnelling.

$$\frac{k_H}{k_T} = \left(\frac{k_H}{k_D}\right)^{1.4}$$

**Equation 6** The Swain-Schaad exponential relationship between the KIEs of the different isotopes of hydrogen. In this example, hydrogen is used as a reference.

### 1.1.5 The Bell Adjusted Model of Tunnelling

C-H bond cleavage occurs in around 50% of biological reactions, many of these enzymatic reactions employ the mechanism of hydrogen tunnelling.<sup>23</sup> The Bell model was introduced as a correction to TST due to the classical model failing to account for KIEs outside the expected range as a result of quantum mechanical tunnelling. Tunnelling allows small and light particles such as H<sup>-</sup>, H<sup>+</sup> and H<sup>•</sup> to pass through the activation barrier when the width of the barrier is comparable to the de Broglie wavelength of the particle (wave-particle duality theory). As a consequence, the activation energy at the transition state energy does not need to be realised.<sup>24, 25</sup>

The Bell correction has successfully accounted for the observed elevated KIEs that occur due to quantum tunnelling. This model makes the assumption that KIEs are temperature independent so fails to account for experimentally measured temperature-dependent KIEs, breakpoints in the temperature dependence of KIEs, KIE values up to 100 and values of  $\Delta E$  greater than expected which have also been observed. The semi-classical model also assumes no change in the pre-exponential factor in contrast to experimental Arrhenius pre-factors outside the semi-classical limits being observed (**Equation 1**).<sup>26</sup>

## 1.1.6 Contemporary Theories of Enzyme Catalysis

Within a biochemical framework, enzyme catalysis results from the binding of substrate, followed by activation of the substrate by the enzyme. The limitations of TST theory and the failure of the Bell adjusted model to explain inflated KIEs, temperature dependent KIEs and unusual Arrhenius pre-factor ratios in hydrogen tunnelling reactions led to the proposition of extended theories, which build upon transition state theory and new theories. This section aims to give a brief overview of the current status of ideas in enzyme catalysis.

### 1.1.6.1 Dynamical Treatment

The importance of taking a more dynamic treatment of the H-transfer process and considering barrier width, in addition to barrier height was realised with the limitations of TST when determining the rate of hydride transfer reactions. These ideas provided a platform for theories such as environmentally coupled tunnelling,<sup>27, 26, 28</sup> promoting vibrations,<sup>29</sup> vibrationally enhanced ground state tunnelling and,<sup>30</sup> multidimensional tunnelling.

Enzyme motions range from the fast fluctuations of bond vibrations and electrostatic changes to domain and loop movements. This range of timescales stretches from seconds to femto-seconds and gives an indication of the complexity of protein motions. The role of large protein motions, such as changes in conformation during the catalytic cycle that affect the active site are well known to be critical to enzyme activity however the understanding of how the enzyme achieves this cooperation between chemical and conformational change is not well understood. Controversially, it has also been proposed that protein motions coupled to the chemical step contribute to the acceleration of the chemical step of the enzyme catalysed reaction. This idea that enzymes can accelerate the chemical reaction through selecting one specific mode of motions is one that is hotly contested.<sup>31</sup>

Marcus theory was developed by Rudolph Marcus in 1956 to explain rates of electron transfer reactions and has since been applied to develop further models for hydrogen transfer reactions in enzymes termed promoting vibrations, environmentally coupled tunnelling and multidimensional tunnelling.<sup>32, 33</sup> These models include a role for dynamics in driving hydrogen tunnelling and are able to account for both temperature dependent and temperature independent KIEs.<sup>34,35,36</sup> The environmentally coupled

model of hydrogen tunnelling is just one of the models that implements Marcus theory to explain electron transfer reactions in solution and to justify why, conversely to the general understanding of kinetics and thermodynamics, thermodynamics may influence kinetics.<sup>37</sup> Within the environmentally coupled tunnelling model, an explanation for the occurrence of temperature dependent and temperature independent KIEs are provided. The environmentally coupled model of hydrogen tunnelling separates motions in to two classes, each of which are important for catalysis:

- Short-range, fast gating motions have been attributed to increasing the efficiency of hydrogen tunnelling by altering the width of the potential energy barrier,<sup>38</sup> although it has been argued that similar motions occur in free solution.<sup>39,40</sup>
- Long-range motions, which pre-organise the active site and lead to a tunnelling ready conformation, have been inferred from analysis of sequence conservation, mutant kinetics and molecular dynamics simulations of hydride transfer.<sup>41,42</sup>

According to the environmentally coupled model, unaccounted for temperature dependent KIEs imply that the enzyme requires conformational changes to reach the active conformation, whereas the temperature independent KIEs do not as the enzyme adopts the tunnelling ready conformation.<sup>43</sup>

The light mass of the H nucleus is such that QM tunnelling of the particle can occur once the wavefunction of the particle in the reactants state overlaps with that of the product state. The point at which tunnelling occurs is barrier width and height sensitive and, sensitive to the particle mass. The region along the reaction coordinate in which the particle is capable of tunnelling is called the tunnelling-ready-state (TRS) and is effectively a QM delocalised TS. The Arrhenius pre-factor (A) (**Equation 1**) is not assumed to change with isotopic substitution in TST but in the environmentally coupled model, the Arrhenius pre-factor ( $A_T$ ) is separated into individual components (**Equation 7**).<sup>44</sup>

$$k = C_{(T)} \exp^{-\Delta G^\ddagger/RT} \int_0^\infty P_{(m,DAD)} \exp^{-(E_{(DAD)})/(k_B T)} dDAD$$

**Equation 7** Transition state equation for strong electronic coupling between reactant and product in which the Arrhenius pre-factor  $A_T$  is divided in to the heavy atom motions ( $C_T$ ) and an integral summing the tunnelling ready states by their tunnelling probability ( $P$ ) with respect to mass and donor acceptor distance (DAD) and the Boltzmann distribution of the range of donor acceptor distances. Further models have been developed for non-adiabatic reactions.<sup>44</sup>

The first part of the Arrhenius pre-factor  $A_{(T)}$  contains a temperature dependent term for heavy atom motions and is mostly insensitive to isotopic substitutions. The second part of the pre-factor (the Franck-Condon term) is an integral summing the tunnelling ready states (TRS) by their tunnelling probability ( $P$ ). This term is isotopic mass sensitive and measures the probability of hydrogen transfer taking place once the system reaches TRS. Strong coupling between the reactant and product states is assumed in this case of transition state theory however, there are non adiabatic models which consider the case of weak coupling between the reactant and product potentials. In this situation, the first part of the Arrhenius pre-factor (the Marcus term) describes the rate of reaching a TRS based on the electronic coupling between the reactant and product, the reorganisation energy ( $\gamma$ ) and the driving force for the reaction ( $\Delta G^0$ ) (**Equation 8**). In either case of strong or weak coupling or reactant and product, the Marcus terms are mostly insensitive to the mass of the tunnelling particle and so the KIE can be described as a ratio of the Franck-Condon terms for the light and heavy isotopes.

$$k = C_{(T)} \exp^{(-\Delta G^0 + \gamma)/(4\gamma k_B T)} \int_0^{\infty} P_{(m, DAD)} \exp^{-E_{(DAD)}/(k_B T)} dDAD$$

**Equation 8** Transition state equation for weakly coupled reactant and product in which the Arrhenius pre-factor  $A_{(T)}$  is divided in to the heavy atom motions ( $C_T$ ) and an integral summing the tunnelling ready states by their tunnelling probability ( $P$ ) with respect to mass and donor acceptor distance (DAD) and the Boltzmann distribution of the range of donor acceptor distances.<sup>44</sup>

Theoretical and experimental studies have established an important role for protein dynamics in driving catalysis,<sup>45,46</sup> whilst other studies have concluded that protein dynamics do not contribute to catalysis.<sup>47, 48</sup> The evidence for the existence protein motions that drive catalysis is discussed with reference to the enzyme dihydrofolate reductase (**1.2.3 Gathering Insight into the Nature of Catalysis by Dihydrofolate Reductase**).

### 1.1.6.2 The Electrostatic Basis for Catalysis

Electrostatic interactions are formed between ions that carry opposite charges, polar functional groups which carry permanent dipoles with opposite polarities and atoms containing relatively large numbers of electrons enabling dipole induced dipole or van der Waals interactions. In solution, the application of Coulombs law reveals that the

strength of electrostatic interactions is determined by the distance between charges or dipoles and the dielectric constant of the solvent. The catalytic theory of pre-organised electrostatic interactions surmises that the reaction is accelerated through the enzyme using pre-organised dipoles which stabilise the transition state of the enzyme catalysed reaction relative to the ground state. Warshel argues electrostatic effects give the largest contributions to enzyme catalysis and the enzyme accomplishes binding of the transition state tighter than the substrate through electrostatic stabilisation.<sup>47, 49</sup> The main difficulty with the theory is in quantifying the relative importance of the different energy contributions. The argument for electrostatic pre-organisation is in direct contrast to the theory that catalysis is driven by long range or promoting motions.<sup>50</sup> The argument for electrostatics being the main driver for catalysis is discussed with reference to the enzyme dihydrofolate reductase (**1.2.3 Gathering Insight into the Nature of Catalysis by Dihydrofolate Reductase**).

#### **1.1.6.3 Ground state destabilisation of the substrate**

The concept of ground state destabilisation focuses on the idea that enzymes function by reducing the transition state energy relative to that of the substrate, and therefore by binding the substrate to the enzyme active site and destabilising it, it is possible to reduce the activation energy and accelerate the reaction.<sup>51</sup> Hence the barrier to the enzyme catalysed reaction can be reduced by either lowering the free energy of the transition state (transition state stabilisation) or by alternatively by raising the energy of the enzyme-substrate and enzyme-product complexes (ground state destabilisation) (**Figure 1**). Ground state destabilisation is quantified as the free energy difference between the ligand in its bound state and the ligand in its aqueous solvation state. The methods of ground state destabilisation are outlined below:<sup>52</sup>

*Electronic strain* has been identified and quantified by vibrational spectroscopy in the case of retinol bound to rhodopsin<sup>53</sup> and later, in the red shifting of the carbonyl bond of dihydroxyacetone phosphate by  $19\text{ cm}^{-1}$  upon binding to the active site of triose phosphate isomerase.<sup>54</sup>

The idea of *bond distortion* was proposed by Haldane in which he proposed that two sugars flanking a glycosidic bond could create strain in the glycosidic bond “the rack” bond length and bond angle distortion have been detected in crystal structures.<sup>55</sup>

*Conformation restriction*, a form of conformational selection, has been determined from stereochemical studies and from crystal structures with ligands bound in active sites. As substrates bind to enzymes in few conformations and this conformational restriction has been proposed to be a source of catalytic power in enzymes.<sup>56</sup>

#### **1.1.6.4 Desolvation**

Investigation of solvent effects on the rate of reactions have shown that for some reactions, desolvation is known to enhance rates of reaction in which charges are delocalised or neutralised in the transition state.<sup>57</sup> It has been reported that through the process of binding their substrates, enzymes effectively strip substrates of water and replace water-substrate hydrogen bonds with enzyme-substrate hydrogen bonds. The resulting enzyme-substrate complex is mostly free of water molecules that are present in the free enzyme and ligands. This void of water in the enzyme-substrate complex has led to the suggestion that the active site resembles a gas phase, more than a liquid phase. The point is made that reactions in the gas and liquid phases have very different properties;  $S_N2$  reactions in the gas phase have small activation barriers compared to the corresponding liquid reaction.<sup>58</sup> As a consequence, it has therefore been suggested that simple elimination of the solvent may account for the specificity and rate enhancements of enzyme catalysis.<sup>59</sup> Nevertheless, it is apparent that this view does not take into account that the enzyme is replacing the displaced solvent and effectively solvating substrates with interactions directly from the protein's environment. This is important as for reactions that have large charge separations the rate is much reduced in the gas phase, as the charges are not solvated. Enzymes still effectively accelerate these reactions and as such this hypothesis that enzymes create a gas phase like environment does not stand. Instead, it has been demonstrated that enzymes act by largely removing the solvent from the substrates in order to replace the environment for a polar active site that does not resembling the gas phase but is an environment designed for electrostatic stabilisation of ionic transition states, capable of solvating these states more specifically than water does.<sup>60</sup>

### 1.1.6.5 Low-barrier hydrogen bond

Hydrogen bonding plays an important role in biochemical processes, low-barrier hydrogen bonds are proposed to play a role in the efficiency of catalysis.<sup>61</sup> The low-barrier hydrogen bond (LBHB) theory of catalysis centers on the idea that catalysis is driven by the formation of a partial covalent bond between the enzyme and the ionic transition state through a LBHB that is stabilized through quantum resonance interactions. A low-barrier hydrogen bond (LBHB) encompasses hydrogen on an electronegative atom, which is involved in strong hydrogen bonds to two different hydrogen bond acceptors simultaneously. The strength of the covalent interaction is at its strongest when the  $pK_a$  values of the hydrogen donor and acceptor are equal.<sup>62</sup>

It is argued that LBHBs are rarely formed in water and that any extra covalent stabilization in the LBHBs is not large enough to win over the largely electrostatic solvation effects. As a result it was concluded that LBHBs would only exist in weakly polar environments with low solvation.<sup>63</sup>

A group of enzymes that are considered to exhibit examples of the low-barrier hydrogen bond hypothesis are the serine proteases. The hydrogen bonding network of serine protease trypsin contains a His57-Asp-102-Ser-195 catalytic triad. The mechanism of the serine protease is one which has been contested due to the apparent high reactivity of Ser 195, despite the hydroxyl group of Ser 195 being a poor nucleophile. The observation of a low-field proton signal at ~19 ppm of  $H_{\delta 1}$  of His 57 from NMR spectroscopy has been attributed to a proton participating in a LBHB between His 57 and Asp 102,<sup>64</sup> although it should be noted that  $^{15}N$  NMR data has yielded conflicting results for  $\alpha$ -lytic protease.<sup>65</sup> It was proposed this potential LBHB increases the basic reactivity of His 57 through a charge-relay mechanism.<sup>66</sup>

In order to determine any catalytic advantage of LBHBs, Ishida took the NMR data of the His 57 imidazole and correlated this with the structural rearrangement of the active site in the acylation.<sup>67</sup> In order to take into account the surrounding electrostatic environment of the protein, ab initio QM/MM framework was combined with the NMR data. Additionally, free energy profile of proton transfer from His 57 to Asp 102 in the tetrahedral intermediate obtained by ab initio QM/MM calculations alongside molecular dynamics free energy perturbation simulations. It was determined that downfield

chemical shifts were only observed when the tetrahedral intermediate formed in the acylation process. The free energy profile that was calculated showed an asymmetrical single well, rather than the symmetrical double well that is expected for a delocalized proton. The analyses led the authors to conclude that the unusually downfield proton chemical shift can be rationalized in terms of electrostatic stabilisation and does not provide unequivocal evidence for LBHB driving catalysis.<sup>67</sup>

It has been more recently argued, through the measuring ionic hydrogen bond strengths, that the debate over the possible role of LBHBs has not taken in to account the extent of the strength of hydrogen bonds that can be achieved in environments where there is no competition to form alternative hydrogen bonds, such as an enzyme active site. As hydrogen bonds exhibit their maximal strengths in isolation, it is argued the contributing strengths of these interactions have been underestimated.<sup>68</sup>

#### 1.1.6.6 Near-attack conformation

Near attack conformers (NACs) are conformations ready for catalysis and can be viewed as precursors to the conformation of the substrate in the transition state. Near attack conformers are defined as substrate conformations in which the reacting atoms are within van der Waals contact and at an angle resembling the bond to be formed in the transition state. NACs are characterised as having reacting atoms within 3.2 Å and an approach angle of  $\pm 15^\circ$  of the bond angle in the transition state.<sup>69</sup> The free energy for formation of a NAC relative to the substrate ( $\Delta G_{N^\circ}$ ) can be calculated from the mole fraction of NAC relative to substrate or by free energy calculation methods by consideration of the activation energy from going from NAC to transition state ( $\Delta G_{TS^\ddagger}$ ) and the overall activation energies ( $\Delta G^\ddagger$ ), in the form ( $\Delta G^\ddagger = \Delta G_{TS^\ddagger} + \Delta G_{N^\circ}$ ).  $\Delta G_{N^\circ}$  is comprised of the kinetic advantage of the enzyme compared with water in formation of the NAC ( $\Delta\Delta G_{N^\circ}$ ) and the advantage of conversion of NAC to TS ( $\Delta\Delta G_{TS^\ddagger}$ ), such that  $\Delta G_{N^\circ} = \Delta\Delta G_{N^\circ} + \Delta\Delta G_{TS^\ddagger}$ .<sup>70</sup>

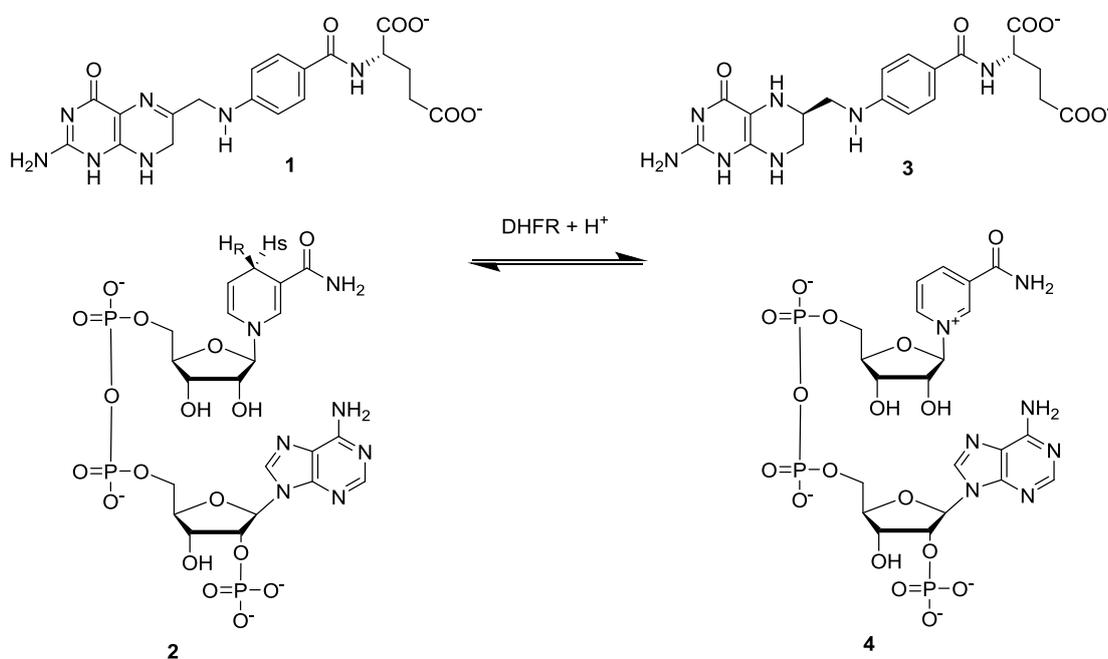
Chorismate mutase (CM), a well established enzyme system, catalyses the unimolecular Claisen rearrangement of chorismate to prephenate; both are important precursors in the biosynthesis of tyrosine and phenylalanine amino acids. The substrate exists in an extended pseudoequatorial conformation in free solution, however it must form a

pseudoaxial conformer to form the TS. The reactions in water and the enzyme are both kinetically first order and comparable. Computational studies have found that the chorismate dianion NAC is present at about 10% of the population of ground state conformers in water but at 30 % of the E.NAC in the enzyme Michaelis complex.  $\Delta G_N^\circ$  in water and enzyme were determined as 8.4 and 0.6 kcal/mol, respectively, indicating that the enzymatic reaction had the advantage of efficient formation of NACs. In addition to chorismate mutase, determination of  $\Delta G_N^\circ$  for two wild type enzymes, mutants and a catalytic antibody indicated that the common mechanism did not include TS stabilisation. It was concluded that the free energy of conversion of NAC to TS ( $\Delta G_{TS}^\ddagger$ ) is independent of the overall free energy of the reaction ( $\Delta G^\ddagger$ ), and the stability of NAC ( $\Delta G_N$ ) determines the difference in  $\Delta G^\ddagger$ .<sup>70</sup> In the case of chorismate mutase it was concluded that the enzyme proficiency arises from entropic advantage of enzyme in forming the reactive ground state conformations compared with water.<sup>70</sup>

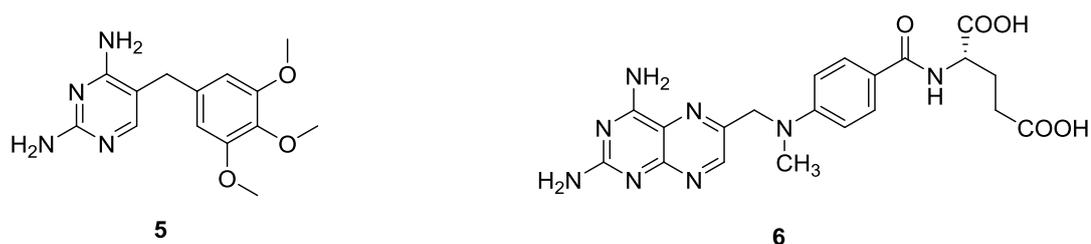
## 1.2 Dihydrofolate Reductase

### 1.2.1 Dihydrofolate reductase as a model for catalysis

Dihydrofolate reductase (DHFR), an abundant enzyme in nature, catalyses the reduction of 7,8-dihydrofolate (DHF) to 5,6,7,8-tetrahydrofolate (THF). Consequentially, DHFR is important for maintaining intracellular levels of THF. THF is a carrier of one-carbon units in biosynthetic pathways including syntheses of amino acids, purines and thymidylate.<sup>71</sup> The chemical step is a reduction which involves stereo-selective transfer of the pro-*R* hydrogen of the NADPH cofactor to C6 of DHF with concomitant protonation of the nitrogen at position 5 (N5) (**Scheme 1**).<sup>72</sup> The pK<sub>a</sub> of N5 is elevated in the active site complex to 6.5 through specific interactions, this is in contrast to the solution pK<sub>a</sub> of 2.6.<sup>72</sup> The well-elucidated structure and catalytic mechanism of dihydrofolate reductase has made the enzyme a paradigm for enzyme catalysis.<sup>73</sup> Due to the important role of dihydrofolate reductase in the synthesis of DNA, it is a target for anti-cancer drug methotrexate (**5**),<sup>74</sup> and trimethoprim (**6**), an antibacterial agent which functions through preferentially inhibiting bacterial DHFRs over mammalian orthologues (**Figure 5**).<sup>75</sup>



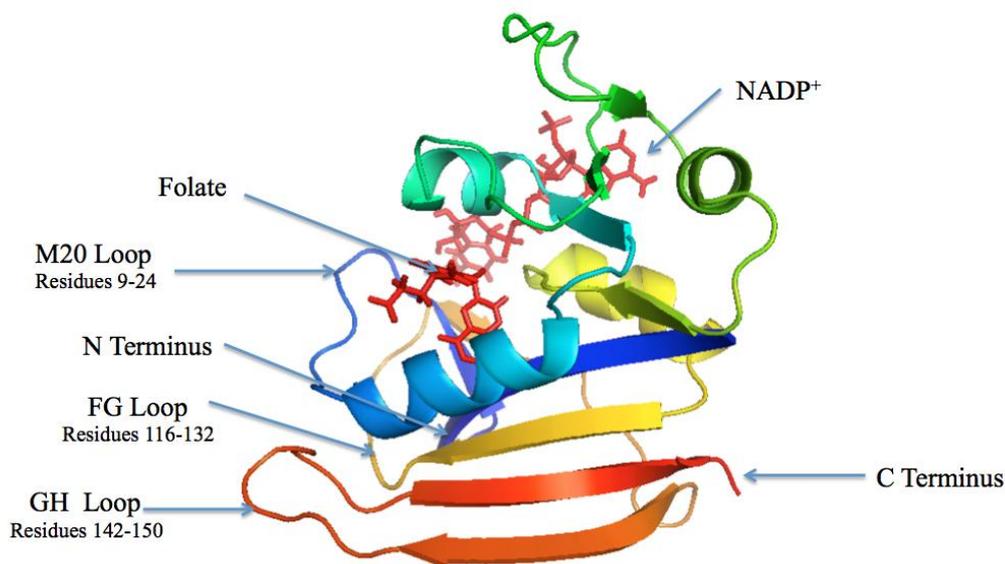
**Scheme 1** The chemical step of the DHFR catalysed reduction of dihydrofolate **1** to tetrahydrofolate **3** through oxidation of NADPH **2** to NADP<sup>+</sup> **4**.



**Figure 5** Chemical structures of methotrexate **5** and trimethoprim **6**

### 1.2.2 Dihydrofolate Reductase from *E. coli*

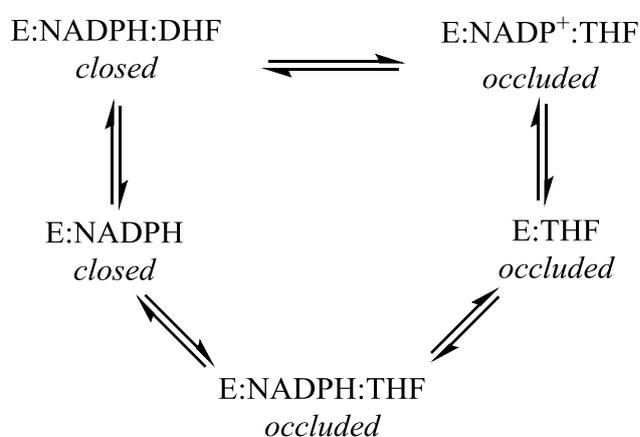
DHFR from *E. coli* (EcDHFR) is a monomeric, 159 amino acid enzyme with a molecular mass of 18 kDa. X-Ray crystallography revealed that the DHFR structure is formed from an 8 stranded  $\beta$ -sheet with 4  $\alpha$ -helices, connected by flexible loop regions (**Figure 6**).<sup>76</sup> The secondary structure content of EcDHFR consists of 29.0 %  $\beta$ -sheets and 24.6 %  $\alpha$ -helices.<sup>77</sup> DHFRs from many species have been characterised and it should be noted that these proteins show a low degree of sequence homology but a high degree of structural homology.<sup>78</sup>



**Figure 6** X-Ray Crystallographic structure of Dihydrofolate Reductase from *E. coli* (1RX2.pdb)

The full kinetic pathway for dihydrofolate reductase from *E. coli* has been elucidated under steady state and pre-steady state turnover conditions<sup>79</sup> and five major intermediates identified in the catalytic cycle (**Scheme 2**).<sup>78</sup> The structures of all of the

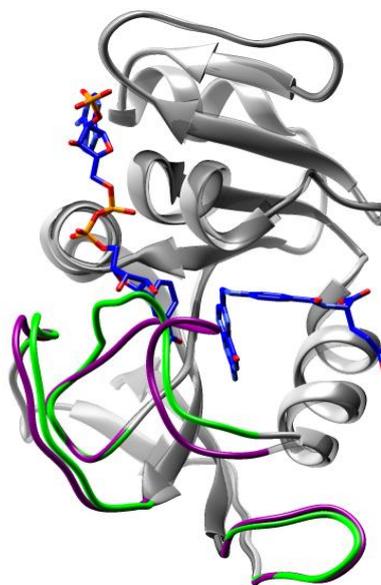
models of the intermediates have been determined by X-ray crystallography and the associated conformational changes outlined. Three of the loop regions in EcDHFR: M20, FG and GH have been identified as being important in catalysis. X-ray crystallographic studies have revealed open, closed, occluded and disordered conformations of the M20 loop section of the enzyme.<sup>80</sup> The closed and occluded conformations have been observed in solution NMR experiments.<sup>81</sup> The closed conformation has been observed in certain crystal forms and has been attributed to stabilisation by crystal lattice contacts. The disordered M20 loop conformation has been observed in X-ray structures of the apo-enzyme and the methotrexate bound complex.<sup>82</sup>



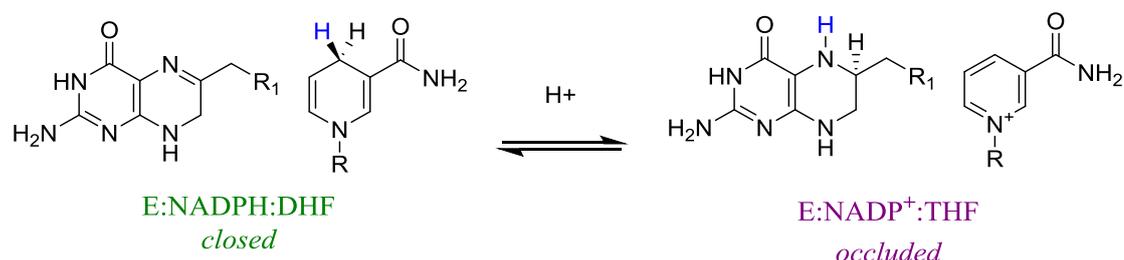
**Scheme 2** Catalytic cycle of DHFR from *E. coli* showing the intermediate states on the pathway and the conformation of the M20 Loop

The NADPH cofactor binds to the apo-enzyme, E, to form the holoenzyme (E:NADPH). The concentration of NADPH in the cell is saturating such that the free enzyme is not detected *in vivo*. The substrate, DHF, binds to the holoenzyme forming the ternary (Michaelis) complex (E:NADPH:DHF). In the Michaelis complex, the M20 loop adopts a closed conformation, packing closely to the nicotinamide ring of the NADPH cofactor and positioning the substrate for the hydride transfer. The M20 loop closes the active site to solvent in the closed conformation while forming advantageous hydrogen bonds to the nicotinamide ring, which assists positioning of the substrate of the chemical step. The closed conformation is stabilised by formation of hydrogen bonds between the backbone NH and O $\gamma$  of D122 in the FG loop and the backbone CO and NH of G15 and E17 in the M20 loop. The formation of these hydrogen bonds positions M16 and E17 outside the active site, allowing the bonding of the nicotinamide

and allowing N18 and M20 side chains to cover the bound substrate and cofactor (**Figure 7**).<sup>82</sup>



**Figure 7** X-Ray crystallographic cartoon structures of EcDHFR displaying M20 loop with closed (green) and occluded (purple) conformations.<sup>83</sup>



**Scheme 3** Chemical step of the catalytic cycle of EcDHFR: Hydride transfer is accompanied by a change in conformation of the M20 loop from closed to occluded

Hydride transfer from NADPH to DHF occurs rapidly compared to the overall turnover rate of the enzymatic reaction, upon which the M20 loop alters to an occluded conformation through breaking the hydrogen bonds that stabilised the closed conformation and replacing with new hydrogen bonds, which stabilise the occluded conformation (**Scheme 3**). In the occluded conformation, the M20 loop is positioned in the active site to prevent the nicotinamide ring of the cofactor from accessing the active site. This is achieved by the formation of hydrogen bonds between D23 in the M20 loop and S148 in the GH loop which stabilises the loop conformation. The oxidised cofactor is released, followed by binding of a reduced cofactor and then product release. It has been suggested that the control over the cofactor binding may be an adaptation of

bacterial cells to their NADP<sup>+</sup> rich cellular environment as otherwise the cellular levels of NADP<sup>+</sup> would favour product binding.<sup>84</sup> NMR and kinetic studies have determined that the rate limiting step at pH 7 is release of the THF product, which requires the enzyme to switch from the occluded (**purple**) to closed (**green**) conformation (**Figure 7**).<sup>85</sup>

This thesis focuses on the structure of the enzyme complex right at the point of the chemical reaction, principally the Michaelis complex. A good model of the Michaelis complex is the ternary E: NADP<sup>+</sup>: folate complex in which the M20 loop is in the closed position.<sup>86</sup> This complex is inert to the chemical step of the reaction and provides a structure similar to the Michaelis complex with the M20 loop in a closed conformation. The E: NADP<sup>+</sup>: folate complex is widely utilised as a Michaelis complex mimic, but that said it must be remembered that this is a model state and will inevitably lack some of the fluidity demonstrated by the actual enzymatic complex due to not being able to turn over the substrate to product and may give rise to artefacts.<sup>87, 82</sup>

### 1.2.3 Gathering Insight into the Nature of Catalysis by Dihydrofolate Reductase

As an established paradigm for catalysis (**Section 1.2**), dihydrofolate reductase from *E. coli* is a well studied enzyme for which there is a large amount of data available. It has been proposed that hydrogen tunnelling contributes significantly to the chemical step in catalysis by dihydrofolate reductase (DHFR), for which the hydride transfer step has been characterised by extensive KIE measurements.<sup>88, 48, 43, 105, 106</sup> Two principal views are held regarding how EcDHFR catalyses hydride transfer. In one view, a network of coupled promoting motions spanning the whole protein is proposed to drive the catalysis and accelerate the reaction.<sup>89</sup> In an alternative and contradictory view, DHFR is thought to adopt a specific conformation at the point of the reaction. The electrostatic environment of the active site in this conformational state stabilises the transition state, lowering the activation energy and accelerating the reaction. This section aims to give an overview of the research completed to date and provide the justification for the work completed in this thesis.

The environmentally coupled tunnelling model was proposed for catalysis by EcDHFR to explain a correlation between protein dynamics and bond cleavage and formation.<sup>90</sup> The combination of results from techniques including genomic analysis, kinetic measurements on multiple mutation versions of EcDHFR and hybrid quantum-classical molecular dynamics gave rise to the proposition that a network of coupled promoting motions plays an important role in catalysis by dihydrofolate reductase.<sup>89</sup> Catalysis in EcDHFR was proposed to be driven by a global dynamic network of residues, both near and far from the active site, coupled to each other and to the reaction co-ordinate.<sup>91</sup>

In the catalytic cycle of EcDHFR, the M20 loop switches between closed and occluded conformations (**Scheme 2**). X-ray crystallography and NMR data analysis have confirmed that these large-scale motions occur in association with binding and release of the substrate, cofactor and product.<sup>82, 92, 93, 94</sup> These observations gave rise to the suggestion that the motion of the M20 loop (**Figure 7**) may have an important role in limiting the rate of turnover of the enzyme by limiting the rates of product dissociation. NMR studies of EcDHFR in complex with a variety of ligands have led to suggestions that changes in the dynamics of the protein may be correlated with kinetic events along

the catalytic cycle.<sup>86</sup> The dynamic behaviour of residues remote from the active site led to the proposal that these residues might influence events at the active site.<sup>95, 96, 97</sup>

Site directed mutagenesis allows one or more specific amino acid residues in a wild-type protein to be substituted with different amino acids in mutant protein. The replacement amino acids can be more or less bulky, have the opposite charge or polarity or prevent formation of hydrogen bonds, resulting in altered properties in the mutant enzyme compared to the wild-type. Combinations of single and double mutants were conducted on EcDHFR at residue positions close to and further away from the active site.<sup>89</sup> It was determined that the effect on the turnover rate was less for some double mutants than the sum of the corresponding single mutations. These results were proposed to serve as indication that long-range protein motions have a key role in coupling to the reaction coordinate and driving the chemical reaction during catalysis. It was proposed that the motions across the whole protein provide the environmental reorganisation required to maintain optimal donor-acceptor distance for efficient tunnelling **(1.1.6.1 Dynamical Treatment)**.<sup>98</sup>

A range of kinetic isotope effect experiments have been conducted with the aim of clarifying a role for dynamics in catalysis by EcDHFR. It was earlier described how KIEs can be used to characterise transition states **(1.1.4 Kinetic Isotope Effects for Characterisation of Transitions States)** and that KIEs can be interpreted in terms of Marcus-like theories **(1.1.6.1 Dynamical Treatment)**. The intrinsic primary and secondary KIEs that have been measured in wild type DHFR enzymes have been temperature independent, which is interpreted in Marcus theory that the enzyme adopts a tunnelling ready configuration with optimal DAD distances.<sup>99</sup>

Temperature dependence and DAD have been investigated with regards to mutation of Ile14 in EcDHFR. Ile14 was mutated to aliphatic valine, alanine and glycine amino acids to determine the extent to which the side chain length plays a role in holding the nicotinamide ring of the NADPH cofactor close the dihydrofolate hydride acceptor and helps to maintain optimal DAD.<sup>100</sup> Gly, Ala and Val amino acids were chosen to avoid significant perturbation of the electrostatics in the active site. The distributions of DADs were assessed by molecular dynamics simulations. The results indicated an increased and broader range of DADs with shorter AA chain lengths, indicating larger and

broader DADs result in steeper temperature dependence of the intrinsic KIE.<sup>101</sup> A more recent QM/MM study has revealed that the mutations cause an increase in the free energy barrier height of the hydride transfer reaction relative to the wild type enzyme. Extended dynamics simulations of the reactive Michaelis complex exposed enhanced flexibility of the mutants and fluctuations in DADs compared with the wild type enzyme, implying the mutations partially impair the active site environment.<sup>102</sup>

A ‘dynamic knockout’ mutant of EcDHFR (N23PP/S148A) was reported in several publications. The mutations were designed to prevent formation of the occluded conformation through loss of hydrogen bonding between the M20 and GH loops and eliminate millisecond to timescale motions of the M20 loop in the Michaelis complex. This set of mutations was proposed to severely impair hydride transfer in one report.<sup>46</sup> A computational study concluded the reduced rate constant was a result of changes to the reorganisation free energy of the reaction as a consequence of modifications to the electrostatic pre-organization within the active site.<sup>47</sup> Kinetic measurements determined the nature of the chemistry in the ‘dynamic knockout’ of EcDHFR was unaffected by the mutations and effectively the same as the wild type enzyme. This concluded that no link between conformational motions and chemical motions could be determined for EcDHFR.<sup>103</sup> A recent study revealed that the link between closed and occluded forms and progression through the catalytic cycle seems to be specific to EcDHFR and is not characteristic of other bacterial DHFRs, such as DHFR from *Moritella Profunda*, which does not form the occluded conformation.<sup>104</sup>

Heavy or “Born-Oppenheimer” enzyme equivalents (<sup>15</sup>N, <sup>13</sup>C and <sup>2</sup>H for all non-exchangeable protons) are a popular method of perturbing bond vibrations without significantly perturbing the electrostatics of the protein. Heavy enzymes have been used as a tool to quantify the relationship between fast protein vibrations (ps-fs) on the chemical step of the enzymatic reaction. Light and heavy EcDHFRs have been analysed through a combination of experimental results, QM/MM simulation and theoretical analyses. The heavy EcDHFR was found to have a slightly slower reaction rate, which was attributed to the difference in the environmental coupling to the hydride transfer step. The lower activity measured for the heavy enzyme led to the proposal that protein dynamics have a small but measurable effect on the chemical reaction rate. Modelling experiments found that the barrier and contribution to quantum tunnelling was not

affected, concluding that it is unlikely that promoting motions have a role in driving tunnelling or modulating the barrier to reaction.<sup>105</sup>

The heavy enzyme variant of the N23PP/S148A dynamic knockout EcDHFR was analysed in a similar manner to that described above.<sup>105</sup> A reduced rate constant for the heavy enzyme in this case was attributed to an increase in the quasi-classical free energy barrier and to a minor extent, an increase in the number of re-crossing trajectories on the transition state dividing surface. A higher degree of active site reorganisation is needed to reach the transition state. It was concluded that for this variant, motions on a millisecond time scale are blocked but motions on the femtosecond timescale are increased, which leads to an increase in dynamic coupling to the chemical coordinate that is detrimental to catalysis.<sup>106</sup> TmDHFR, a dimeric hyperthermophilic enzyme, was subjected to same treatment as the light and heavy EcDHFRs. It was found that the hydride transfer step was not affected by isotopic substitution, leading to the conclusion that TmDHFR, in contrast to EcDHFR, lacks coupling of protein motions to the reaction coordinate.<sup>107</sup>

It has been argued heavy enzymes electrostatics are altered by the isotopic substitution. Isotopic substitution causes slight change in geometry and dipole moment in C-H which in turn causes aliphatic groups to become more hydrophobic and aromatic groups become more hydrophilic. This effect could multiply over the whole enzyme and may lead to a change in the enzyme dynamics, particularly on the slower timescales involved in the physical steps of catalysis.<sup>108</sup>

The effect of altering the dielectric constant and viscosity of the medium of the enzymatic reaction have been quantified through measuring rate constants for hydride transfer in the presence of organic co-solvents. It was determined that the rate constant but not the KIE is proportional to the dielectric constant, but no trend for viscosity was concluded from the results. This outcome provides evidence against long range motions being involved in the chemical step and implies a significant role for electrostatics in controlling catalysis.<sup>109</sup>

### 1.3 Introduction to Protein Nuclear Magnetic Resonance Spectroscopy

A large number of X-ray crystal structures are available for both the apo-enzyme and various complexes of *Ec*DHFR.<sup>110, 78,79, 80</sup> X-ray crystallography is a useful tool for measuring inter-atomic distances, however is not possible to determine dynamics from these structures due to crystallisation effects which prevent determination of the energy of different states from their simple populations. X-ray structures provide a snapshot view of molecules but cannot provide dynamical information. Solution nuclear magnetic resonance (NMR) spectroscopy can provide a range of possible structure solutions close to the global minimum representing the structure of the enzyme in solution.<sup>111</sup> In particular, NMR spectroscopy measures the distinct chemical shift of each atom in the protein allowing structural and dynamical information on an atomic scale to be determined. Moreover, NMR has expanded to provide details regarding low populated states and dynamics on a range of timescales from seconds to nanoseconds.<sup>116, 112</sup> Combined NMR spectroscopy and molecular modelling (MM) methods can provide information about the individual atoms in the protein and the dynamic correlations between atoms. Semi-empirical NMR chemical shift prediction from molecular dynamics (MD) simulations of proteins can be used to characterise protein motions in atomic detail.<sup>116</sup>

NMR spectroscopy is a powerful tool for studying the structure, function and dynamics of biological macromolecules. To date, there have been 13 NMR structures of dihydrofolate reductases deposited in the protein data bank (PDB) compared to 333 X-Ray crystallographic structures. This section aims to give a brief overview of the NMR theory relevant for structure elucidation, the important points for the structural elucidation process and the relevant NMR structures that have been elucidated to date. The aim of this section is to give an overview of the NMR theory and methods relevant to the proposed work (**Chapter 2.0**).

### 1.3.1 NMR Theory in brief

The intrinsic spin angular momentum of elementary particles gives rise to magnetic properties. Quantum mechanical theory has derived that fermions such as protons and neutrons have a nuclear spin quantum number of  $\frac{1}{2}$ . Consequently, nuclei with particular combinations of unpaired protons and/or neutrons are spin-active. The nucleus of a hydrogen atom consists of one proton, so can be described as having a nuclear spin ( $I$ ) of  $\frac{1}{2}$ . Four important nuclei in protein NMR ( $^1\text{H}$ ,  $^2\text{H}$ ,  $^{15}\text{N}$  and  $^{13}\text{C}$ ) are displayed below (**Table 1**).<sup>141</sup>

Nucleus	$I$	$\gamma$ (T s) <sup>-1</sup>	Natural abundance (%)
$^1\text{H}$	$\frac{1}{2}$	$2.6752 \times 10^8$	99.99
$^2\text{H}$	1	$4.107 \times 10^7$	0.012
$^{13}\text{C}$	$\frac{1}{2}$	$6.728 \times 10^7$	1.07
$^{15}\text{N}$	$\frac{1}{2}$	$-2.713 \times 10^7$	0.37

**Table 1** Nuclear spin angular momentum quantum number ( $I$ ), gyromagnetic ratio ( $\gamma$ ) and the natural isotopic abundance for nuclei in biological NMR spectroscopy.

Spin angular momentum is a vector quantity; the magnitude of the spin angular momentum is represented below (**Equation 9**). A condition of quantum mechanical theory is that only one of the three Cartesian components of  $I$  can be specified and by default, this is  $I_z$ . The magnetic quantum number  $m$  (where  $m = I, I-1, \dots, -I$ ) is restricted to  $2I+1$  values. Relative to the  $z$ -axis, the spin and therefore magnetic moment of the hydrogen nucleus may lie in two ( $2I+1 = 2$ ) different orientations with  $m = +\frac{1}{2}$  and  $m = -\frac{1}{2}$  (**Equation 10**).<sup>8</sup>

$$|I| = [\mathbf{I} \cdot \mathbf{I}] = \{I(I+1)\}^{1/2}\hbar$$

**Equation 9** The magnitude of the spin angular momentum

$$I_z = \hbar m$$

**Equation 10** Magnetic moment of the hydrogen nucleus relative to the  $z$  axis

The gyromagnetic ratio ( $\gamma$ ) quantifies the magnitude of the nuclear magnetic moment ( $\mu_z$ ) relative to the spin angular momentum ( $I_z$ ) for a particular nucleus. When the spin angular momentum of a nucleus is not equal to zero, a nuclear magnetic moment will be present which will give rise to a small magnetic field (**Equation 11**).<sup>8</sup>

$$\mu_s = \gamma \mathbf{I}_z = \gamma \hbar m$$

**Equation 11** The gyromagnetic ratio  $\gamma$  of a nucleus is the ratio of the nuclear magnetic moment ( $\mu_s$ ) and the spin angular momentum

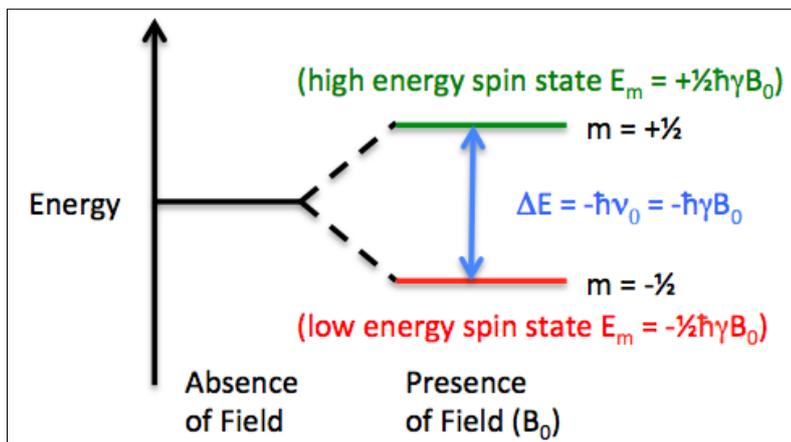
A typical one molar NMR sample contains about  $10^{26}$  molecules. Each nucleus in the molecule possesses nuclear spin causing it to behave as a small bar magnet would. NMR is concerned with the bulk properties of the sample of nuclei. In the absence of an applied magnetic field, the individual magnetic moments and angular spins are randomly orientated. When an external magnetic field is applied ( $B_0$ ), the spin angular momentum causes the magnetic moment to precess about the magnetic field at a rate equal to the Larmor frequency ( $\nu_0$ ).<sup>113</sup>

$$E_m = -\gamma \mathbf{I}_z \mathbf{B}_0 = -\gamma \hbar m B_0$$

**Equation 12** Energy of the spin states, which are quantised and proportional to their projection on to the magnetic field vector along the z-axis. Gives rise to the well known  $E = \pm \frac{1}{2} \gamma \hbar B_0$ , where the maximum energy difference between the two states is  $\gamma \hbar B_0$

The orientation of the nuclear magnetic moment ( $\mu_z$ ) to the applied magnetic field vector determines the energy of the spin state; the lowest energy is achieved when the  $\mu_z$  is aligned with the applied magnetic field and the highest energy is achieved when the  $\mu_z$  is aligned against the applied magnetic field (**Figure 8**). It is important to note that there is no requirement for the spins to align with or against the magnetic field and they may adopt any alignment and therefore any energy. (**Equation 12**) represents the energy of the spin states, which are quantised and proportional to their projection on to the magnetic field vector along the z-axis. At equilibrium, the energy level population distribution can be described by the

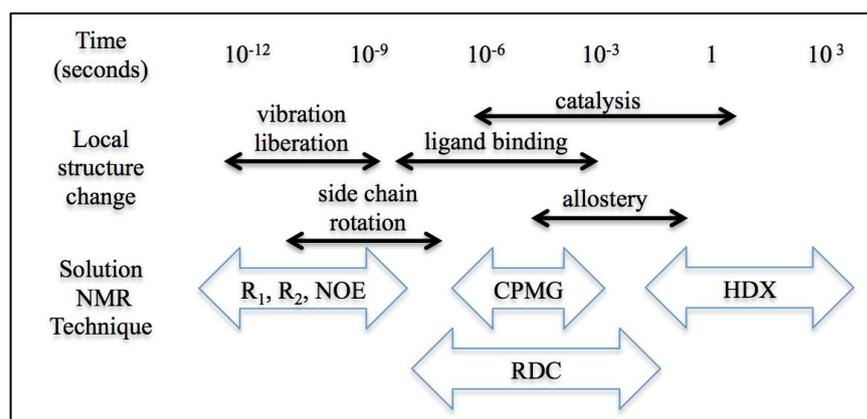
Boltzmann ratio:  $N^-/N^+ = e^{-E/kT}$ . At equilibrium, the bulk magnetisation (average) is aligned with the external magnetic field as this is the state with the lowest energy. NMR perturbs the bulk magnetisation and it is the relaxation back to equilibrium is observed.



**Figure 8** Representation of energy of spin systems in the absence and presence of an applied magnetic field. In the presence of a magnetic field, the size of the net magnetisation is given by the energy difference between the Zeeman levels

## 1.4 Relaxation NMR

Protein structure and function are traditionally understood and viewed in terms of the time-averaged structure of the molecule. The time averaged structure is an ensemble of time dependent structures with distinct conformations in equilibrium with one another. These distinct conformations are constantly sampled over a range of timescales and distances. It has been established that the dynamic processes involved with these motions are important for biological processes including catalysis and protein folding and that different NMR techniques provide methods to directly measure these processes (Figure 9).<sup>114</sup> Gaining further understanding of the structural bases for these dynamical processes is important for understanding the full spectrum of macromolecular function, particularly when predicting and engineering protein behaviour such as artificial enzymes and for understanding how enzymes function.<sup>115</sup>



**Figure 9** Solution NMR techniques for measuring protein conformational changes (CPMG: Carr-Purcell-Meiboom-Gill, RDC: Residual Dipolar Coupling, HDX: Hydrogen-Deuterium Exchange, R<sub>1</sub>, R<sub>2</sub>, NOE: Nuclear Spin Relaxation).<sup>116</sup>

NMR is a useful tool for probing motions over the various timescales, often achieved through measuring the effects of motions on the relaxation of nuclei. Relaxation is the process of the nucleus returning to a state of equilibrium, which corresponds to loss of coherence of the excited state prior to excitation. Relaxation mainly occurs due to interaction of the excited nucleus with nearby NMR active nuclei. The strength of the interaction between the excited state nucleus and nearby NMR active nuclei determines the time taken for relaxation. The rate of relaxation is affected by atomic motion, therefore measuring and interpreting relaxation data can yield important information on the dynamics and motions of molecules.

### 1.4.1 Introduction to Relaxation NMR Spectroscopy

Nuclear spin relaxation (NSR) is a method of indirectly probe ps-ns dynamics in proteins by measuring the site-specific observables  $R_1$ ,  $R_2$  and the heteronuclear NOE (*het*NOE). Fast motions of the backbone and side chain fluctuations that occur in a protein occur on a pico-second to nanosecond timescale and can be studied by measuring the relaxation rates of  $^{13}\text{C}$  or  $^{15}\text{N}$  hetero-nuclei in the protein. Molecular motions on this timescale that have been highlighted as potentially being implicated in catalysis as these are on the timescale of the chemical step and include dynamics such as bond vibrations, bond liberation, side-chain rotamer inter-conversion, random coil and loop motions and backbone torsion angle rotation. The time taken for relaxation is used to infer information on the local dynamics of the protein. Furthermore, the relaxation rates can be measured at different field strengths to yield more information regarding the proportion of dynamics that contribute to the relaxation.

### 1.4.2 Longitudinal relaxation

Longitudinal or spin-lattice relaxation is the recovery of net magnetisation aligned with the applied  $B_0$  field and corresponds to the process of re-establishing the Boltzmann distribution of spin states in the magnetic field. In this relaxation process, energy is lost to the surroundings or the ‘lattice’ due to magnetic field fluctuations in the x-y plane. The time constant for spin-lattice relaxation,  $T_1$ , is field dependent because fluctuations in the x-y plane are the most effective when they occur at the Larmor precession frequency ( $\nu_0$ ). The longitudinal relaxation rate constant is a measurable value (**Equation 13**).

$$I_{(t)} = I_{(0)} (1 - \exp^{-R_1 t})$$

**Equation 13** Longitudinal relaxation rate constant is determined by measurement of the time dependent exponential restoration of intensity of the net magnetisation aligned with the applied  $B_0$  field (where  $R_1 = 1/T_1$ )

### 1.4.3 Transverse Relaxation

Transverse or spin-spin relaxation is the loss of coherence of x-y magnetisation among the ensemble of spins. The loss of coherence amongst the spins results in greater disorder in the system and therefore an increase in entropy. The time constant for spin-spin relaxation,  $T_2$  is always less than or equal to  $T_1$  as return of the magnetisation vector in the z-direction causes loss of magnetisation in the x-y plane (**Equation 14**).<sup>113</sup> There are two main mechanisms by which relaxation occurs and a third in certain circumstances which are explained below:

$$I_{(t)} = I_{(0)} (1 - \exp^{-R_1 t})$$

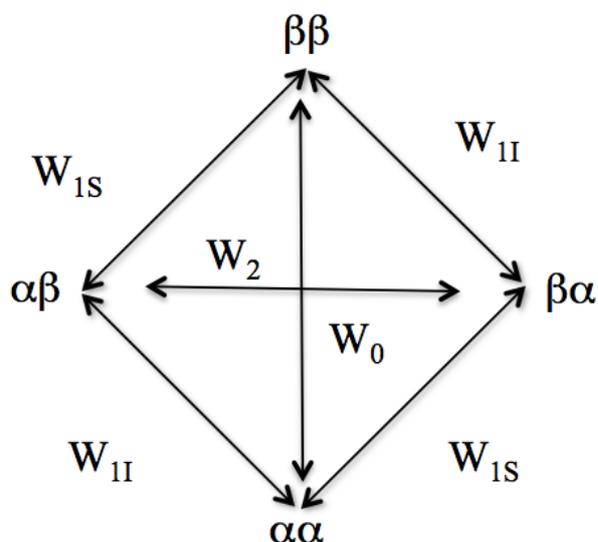
**Equation 14** Longitudinal relaxation rate constant is determined by measurement of the time dependent exponential restoration of intensity of the net magnetisation aligned with the applied  $B_0$  field (where  $R_1 = 1/T_1$ )

*Chemical shift anisotropy* describes the dependency of the chemical shift on the orientation of a bond vector relative to a static magnetic field  $B_0$ . The orientation of the bond vector relative to the static magnetic field ( $B_0$ ) can change due to rotation of the protein or, rotation of the bond vector. These motions occur on a ps-ns timescale. Changes to the orientation of the bond vector alters the local magnetic field experienced by a nucleus resulting in production of an oscillating magnetic field of MHz frequency, which can induce nuclear relaxation.

*Dipolar coupling* causes a change in the magnetic field experienced by a nucleus due to through space interaction with other nuclei. Interaction between spins is dependent on their through space distance and orientation relative the static magnetic field ( $B_0$ ). Through space distance and orientation can change with rotation of the protein or rotation of the bond vector on a ps-ns timescale, which as with the chemical shift anisotropy, can result in an oscillating local magnetic field that induces relaxation.

*Quadrupolar interactions* occur between nuclei with a non-zero quadrupolar moment ( $^2\text{H}$ ,  $^{14}\text{N}$ ,  $^{17}\text{O}$ ) and an electric field gradient due to locally bonded electrons or ions. Relaxation due to quadrupolar interactions exceeds chemical shift anisotropy and dipolar coupling effects by up to 100 fold. In protein NMR, samples are often labelled with  $^{15}\text{N}$  so this method of relaxation is uncommon.

### 1.4.4 The Nuclear Overhauser Effect



**Figure 10** The energy transitions in a two-spin system, where  $W_{1S}$  and  $W_{1I}$  are single quantum transition probability rates and  $W_0$  and  $W_2$  are the probability rates for double and zero quantum transitions, respectively.

The nuclear Overhauser effect (NOE) occurs when a spin that has been saturated with a radio frequency pulse (spin S) perturbs an unsaturated spin through space (spin I) *via* a dipolar interaction (**Figure 10**). For a two-spin system, single quantum transitions cause spin-lattice (longitudinal) relaxation. The magnitude of the NOE is strongly dependent on competing double and zero quantum transitions (**Equation 15**) and the correlation time ( $\tau_c$ ) of the molecule, as spin transitions are induced when the molecule rotates with a frequency similar to the Larmor frequency.

$$\epsilon_{NOE} = 1 + \frac{\gamma_I}{\gamma_S} \left( \frac{W_2 - W_0}{W_0 + 2W_{1S} + W_2} \right)$$

**Equation 15** Magnitude of the NOE where  $\gamma_I$  and  $\gamma_S$  are the gyromagnetic ratios of saturated spin (S) and unsaturated spin (I), respectively.  $W_{1S}$  and  $W_{1I}$  are single quantum transition probability rates and  $W_0$  and  $W_2$  are the probability rates for double and zero quantum transitions, respectively.

### 1.4.5 The Heteronuclear NOE

Dipolar coupling between different types of nuclei induces through-space magnetisation transfer which gives rise to the heteronuclear NOE (*hetNOE*). Measurement of the *hetNOE* is widely used for quantification of the thermal fluctuations on the picosecond to nanosecond timescale that occur in a protein on a per residue basis. *hetNOE* values are measured by comparing the intensity of signal transferred from  $^{15}\text{N}$  to  $^1\text{H}$  in the absence and presence of proton saturation (**Equation 16**).<sup>117</sup> Magnetisation exchange between amide protons and water protons, either via NOE or hydrogen exchange can cause amide protons to relax to their thermal equilibrium value with a time constant that can be much longer than their inherent  $T_1$ .

$$\text{hetNOE} = \left( \frac{I_{\text{saturated}}}{I_{\text{unsaturated}}} \right) - 1$$

**Equation 16** The steady-state NOE values are established from the ratio of peak intensities ( $I$ ) in the presence and absence of dipolar coupling.

$^{15}\text{N}$ -[ $^1\text{H}$ ] *hetNOE* experiments are widely used in protein NMR, providing a good basis to start studying protein structure as the 2 dimensional experiment can be used to determine the extent to which protein residues are unfolded in a protein structure and to validate secondary structure assignments. Partially flexible parts of the protein, such as exposed surface loops, are readily distinguished from the folded core. Unstructured regions tend to be more mobile than structured regions of the protein, this motion causes the NOE signal to decay quickly. Unstructured parts or unfolded proteins frequently yield signals of zero or opposite sign and are readily recognisable from very structured parts of the protein which tend to yield a *hetNOE* value close to the theoretical maximum.<sup>118</sup> The steady-state NOE values are established from the ratio of peak intensities ( $I$ ) (**Equation 16**).

## 1.4.6 Measuring Protein Dynamics by NMR

The dynamics exhibited by proteins occur over a range of timescales. For NMR methods specific experimental setups provide access to study protein dynamics within particular time windows. Two timescales of motion are of particular interest for NMR measurements; microsecond-millisecond and picosecond-nanosecond. The distinction between these timescales is emphasised in terms of long-range motions that require large-scale protein rearrangements, and therefore occur on a longer timescale and short-range motions that occur on a shorter timescales. As discussed previously, short range motions have been connected with hydrogen tunnelling as they occur on a similar timescale to hydride transfer.

In NMR, the experimental methods to measure fast and slow dynamics are separated by the rotational correlation time. The *correlation time* describes the time taken for a molecule to rotate through an angle of one radian in solution (**Equation 17**). The correlation function decays to zero when the timescale for the motion is equal to the correlation time. Molecular motions that occur on a timescale faster than the rotational correlation time can be measured using nuclear spin relaxation methods whereas motions occurring slower than the rotational correlation time are measured through relaxation-dispersion methods.<sup>115</sup>

$$\tau_c = 4\pi\eta a^3/3kT$$

**Equation 17** The upper limit of the correlation time ( $\tau_c$ ) for a spherical particle is approximated by the Stokes-Einstein equation:  $a$  is the particle radius,  $\eta$  is the solvent viscosity,  $k$  is Boltzmann's constant,  $T$  is the absolute temperature (K).

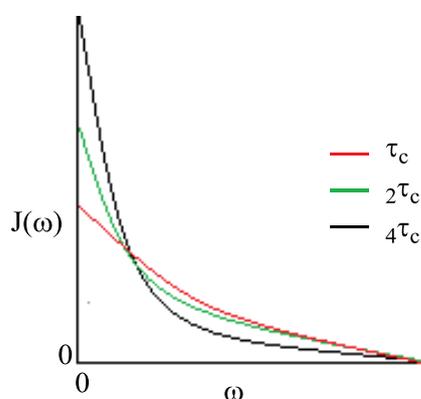
### Nuclear Spin Relaxation (NSR)

Picosecond -nanosecond motions in proteins cause oscillating fields at MHz frequency through the following mechanisms, which were discussed earlier: Chemical shift anisotropy (CSA), dipolar coupling (DC) and quadrupolar interactions. The orientation correlation function  $C_{(t)}$  describes the time dependence of the bond reorientation. Given that in the excited state, all the vectors are coherent and during the relaxation process they gradually become reoriented,  $C_{(t)}$  quantifies the rate at which the bond vectors becomes randomised. The time taken for the reorientation of the bond vectors is described by an exponential relationship (**Equation 18**). Fourier transformation of the

correlation function gives the spectral density function  $J(\omega)$  for each bond vector, which describes the proportion that each frequency of motion leads to the relaxation.<sup>115</sup>

$$C_{(t)} = A \exp(-|\tau|/\tau_c) \xrightarrow{FT} J(\omega) = \frac{\tau_c}{1+(\omega\tau_c)^2}$$

**Equation 18** The *correlation function*  $C_{(t)}$  is the dot product of the bond orientation vector at time zero and a later time  $|\tau|$  and is represented by exponential decay equation, where  $\tau_c$  is the *correlation time*, describing the timescale of tumbling in solution (time taken for the molecule to rotate through one radian). The *spectral density function*,  $J(\omega)$  is the Fourier transformation of the correlation function (**Figure 11**).



**Figure 11** The spectral density describes the frequency profile of the fluctuation. Larger proteins (larger values of  $\tau_c$ ) undergo more low frequency motions.

The spectral density function and the correlation function include contributions from molecular rotation and motion due to internal flexibility. The spectral density function and correlation function are linked to the NMR observables  $R_1$  (**Equation 19**),  $R_2$  (**Equation 20**) and the *hetNOE* (**Equation 21**). In the equations below,  $d_{HN} = \left( \frac{\mu_0 h \gamma_H \gamma_N}{8\pi^2 r_{HN}^3} \right)$  represents the dipolar interaction between the nitrogen nucleus and the attached  $^1\text{H}$  nucleus,  $\mu_0$  is the permeability of free space,  $h$  is Planck's constant,  $\gamma$  is the gyromagnetic ratio,  $r_{HN}$  represents the length of the  $^1\text{H}$ - $^{15}\text{N}$  bond,  $c_N = \frac{(\gamma_N B_0 \Delta\sigma)}{\sqrt{3}}$  is the CSA interaction for  $^{15}\text{N}$  and  $N$  is the number of  $^1\text{H}$  atoms attached to the nucleus.

$$R_1(X) = \frac{d_{HX}^2}{4} [J(\omega_N - \omega_H) + 3J(\omega_N) + 6J(\omega_N + \omega_H)] + c_N^2 J(\omega_N)$$

**Equation 19** The relationship between the longitudinal rate relaxation constant, the spectral density function and the correlation function.

$$R_2(X) = \frac{d_{NH}^2}{8} [4J(0) + 3J(\omega_N) + J(\omega_N - \omega_H) + 6J(\omega_H) + 6J(\omega_N + \omega_H)] \\ + \frac{c_N^2}{6} [4J(0) + 3J(\omega_N)]$$

**Equation 20** The relationship between the transverse rate relaxation constant, the spectral density function and the correlation function.

$$hetNOE = 1 + \left( \frac{\gamma_H}{\gamma_N} \frac{N}{R_1} \right) \frac{d_{HN}^2}{4} [6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)]$$

**Equation 21** The relationship between the heteronuclear NOE, the spectral density function and the correlation function.

NMR relaxation experiments measure  $R_1$ ,  $R_2$  and the *hetNOE*. Using the equations above, interpretation of frequencies of motions can be taken. Each frequency describes a particular point of the spectral density. It is standard practice for the data to be recorded at two field strengths to obtain more points along the spectral density curve in order to ensure better-restrained fitting parameters. In theory, the data could be obtained at more than two field strengths but this can lead to difficulty fitting the data. There are a variety of methods for interpretation of the nuclear relaxation rates:<sup>115</sup>

### Phenomenological interpretation of $R_1$ , $R_2$ and *hetNOE*

In this approach to analysing the data, it is not processed beyond calculating the site-specific  $R_1$ ,  $R_2$  and *hetNOE*. The values can be mapped along the protein sequence to show regions of dynamic behaviour. Site-specific  $R_1$ ,  $R_2$  and *hetNOE* contain different combinations of internal motions, which are interpreted as follows:

- $R_1$  is decreased by ps-ns flexibility and faster tumbling
- The value of  $R_2$  displays increased values with  $\mu$ s-ms chemical exchange and in-homogeneity in the magnetic field. As with  $R_1$  values,  $R_2$  values also

decreases with ps-ns flexibility and faster tumbling. Additionally,  $R_2$  is affected by anisotropic rotation.

- $R_1/R_2$  is the site-specific ratio which is constant for spherical molecules. Differences in the ratio infer anisotropic molecular rotation.
- The  $^1\text{H}^{15}\text{N}$ -*het*NOE values are reduced in the presence of internal flexibility. Typically, values fall between 1 and -4 (due to the negative gyromagnetic ratio ( $\gamma$ ) of  $^{15}\text{N}$ ).

### **Spectral density mapping**

The site-specific values  $R_1$ ,  $R_2$  and *het*NOE are used to reconstruct each spectral density function  $J(\omega)$  to quantify motions over a range of frequencies. This method makes no underlying assumptions made about the nature of molecular motion.<sup>119</sup>

### **A specific model of internal motions**

Site specific models of internal motion make assumptions about the structure and interactions of the protein at the atomic level. Motions of the bond vector are described in terms of the cone angle ( $\theta$ ) defined by the order parameter ( $S^2$ ) and the diffusion time ( $\tau_e$ ).

## Model-free Analysis

In the two parameter approach to the model free formalism of Lipari and Szabo, the spectral density function for a molecule undergoing isotropic tumbling is given by **(Equation 22)** which describes fast internal motions in terms of two independent quantities: the generalised order parameter ( $S$ ) and an effective correlation time ( $\tau_e$ ).<sup>120</sup>

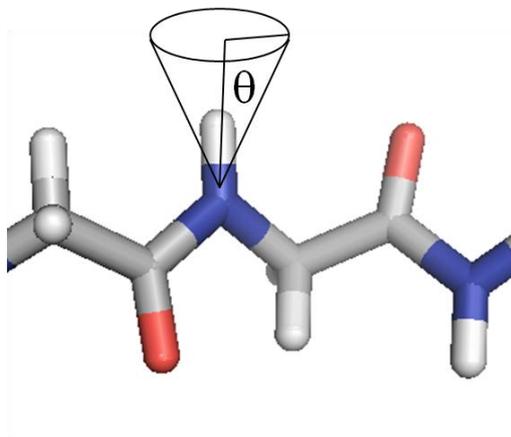
$$J(\omega) = S^2 \frac{\tau_R}{(1 + \omega^2 \tau_R^2)} + (1 - S)^2 \frac{\tau}{(1 + \omega^2 \tau^2)}$$

**Equation 22** Spectral density function for a molecule undergoing isotropic tumbling (corresponding to the correlation function:  $C_{1(t)} = S^2 + (1 - S^2)e^{-t/\tau}$ ) where  $S$  is the generalised order parameter,  $\tau_R$  is the overall isotropic rotational correlation time of the molecule  $\tau = \frac{\tau_R \tau_e}{(\tau_R + \tau_e)}$  (where  $\tau_e$  is an effective correlation time describing the internal motion).<sup>120</sup>

Extended model free analysis is the most popular approach to interpretation of  $R_1$ ,  $R_2$  and *het*NOE. The analysis is described as model free because no structural model is used to describe the nature of the motion. The model treats internal motions and molecular tumbling independently by making the assumption that internal motions are much faster than molecular motion. Four parameters are used to model the dynamics of each bond vector:<sup>115</sup>

- The *site-specific correlation time* ( $\tau_e$ ) quantifies the timescale of bond vector reorientation and can be partitioned into fast and slow components ( $\tau_{e, \text{fast}}$  and  $\tau_{e, \text{slow}}$ )
- The *site specific squared order parameter* ( $S^2$ ) quantifies the amplitude of motions that reorient the bond vector and can be further partitioned in to fast and slow components ( $S^2_{\text{fast}}$  and  $S^2_{\text{slow}}$ )
- The *site-specific exchange broadening* ( $R_{\text{ex}}$ ) quantifies the contribution to  $R_2$  from  $\mu\text{s}$ - $\text{ms}$  chemical exchange
- The *rotational diffusion tensor* ( $D$ ) quantifies the rate of molecular rotation about each of the three axis, which can also be parameterised by a single nanosecond correlation time  $\tau_c$

$S^2$  describes the amplitude of motion within a cone formed by the bond vector; as the bond becomes less rigid, the cone angle ( $\theta$ ) increases.  $S^2$  takes a value of zero for a completely disordered bond vector and 1 for a completely rigid bond vector (**Figure 12**).<sup>121</sup>

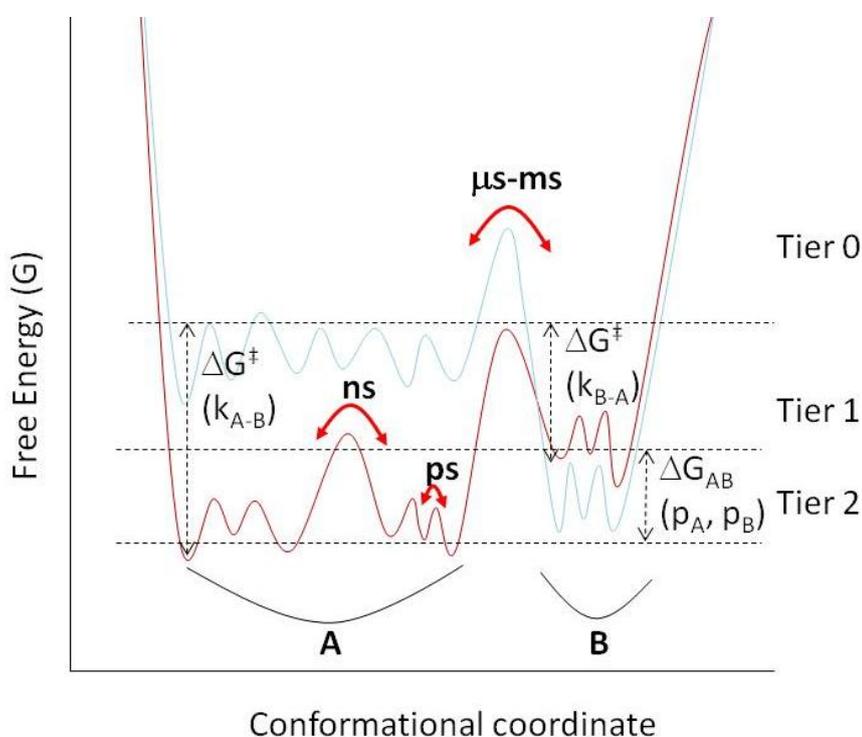


**Figure 12**  $S^2$  describes the amplitude of motion of the bond vector (N-H) within a cone of angle ( $\theta$ ).  $S^2$  takes a value of zero for a completely disordered bond vector and one for a completely rigid bond vector.

The site-specific squared order parameter ( $S^2$ ) is the most commonly quoted parameter from model free because it is the most accurately fitted parameter compared to the internal correlation time, which is often not fitted very strictly. The  $S^2$  values are often mapped on to the protein sequence to demonstrate areas of local ps-ns rigidity. Typically,  $S^2$  values for  $^{15}\text{N}$  backbone sites are found to be  $>0.8$  in the secondary structure and between 0.5 and 0.8 for loops, turns and termini.<sup>166</sup>  $S^2$  values can be used to quantify changes in conformational entropy and rigidity upon ligand binding and estimate heat capacity.<sup>115</sup>

## 2.0 Aims and Objectives

The hierarchy of enzyme motions, probability of states and the energy barriers between states can be understood in terms of a multidimensional landscape, a cross section of which is shown below (**Figure 13**).<sup>97</sup> A protein molecule in a particular state can assume a very large number of conformational states. Different sub-states have the same overall structure but differ in details- performing the same function but with different rates.<sup>122</sup> The tiered model in Figure 13 was introduced to provide order to the complexity of the motions and states displayed by proteins.<sup>122</sup>



**Figure 13** The dynamics in proteins and the associated energy barriers are represented by a one dimensional cross-section through the high-dimensional energy landscape of a protein. States are defined as minima on the energy surface and transition states are represented by maxima between wells, as with the PES. Populations of tier 0 states A and B ( $p_A$ ,  $p_B$ ) are defined as Boltzmann distributions based on their difference in free energy  $\Delta G_{AB}$ . The rate of inter-conversion ( $k$ ) is determined by the barrier between states A and B. Lower tiers describe faster fluctuations between a large number of closely related sub-states within each tier 0 state.<sup>97</sup>

The energy landscape can be altered by changes to the system such as: ligand binding, protein mutation and changes in external conditions which shift the equilibrium between states.<sup>97</sup> Changes to the external conditions in which the enzyme resides can be made through changes to the solvent medium. The viscosity or dielectric constant of the solvent medium can be modulated through addition of organic solvents. This approach was taken to investigate catalysis by dihydrofolate reductases.<sup>43, 129</sup>

The reaction kinetics of the catalysis by EcDHFR in the presence of organic co-solvents was initially investigated and the results interpreted in line with the environmentally coupled tunneling model of catalysis.<sup>125</sup> To determine the effect of viscosity on the steady state turnover kinetics ( $k_{\text{cat}}/K_m$ ) and rapid hydride transfer ( $k_H$ ) for the EcDHFR catalysed reaction, these parameters were measured in the presence of increasing concentrations of iso-viscous and iso-dielectric co-solvents. Iso-viscous media have comparable viscosities but the dielectric constant of the different mediums may vary, whereas iso-dielectric media have comparable dielectric constants but may vary in viscosity. Methanol, ethanol, isopropanol, ethylene glycol, glycerol, sucrose and tetrahydrofuran co-solvents were added to the enzyme reaction medium in final volume concentrations of 17 %, 33 % and 50 %. The KIE (kinetic isotope effect),  $k_H$  (rate constant for hydride transfer) and  $k_{\text{cat}}$  (steady state rate constant) were measured at both pH 7 (product release is rate limiting) and pH 9.5 (hydride transfer is rate limiting).

The rate constants  $k_{\text{cat}}$  and  $k_H$  at pH 7 were found to decrease in a manner proportional to decreasing dielectric constant but not proportional to viscosity. The pH 7 KIE on  $k_H$  was found to be independent of solvent composition. The rate constants  $k_{\text{cat}}$  and  $k_H$  at pH 9.5 showed a slight increase with decrease of dielectric constant between 80 and 70 and then a decrease proportional to decreasing dielectric constant. The pH 9.5 KIE on  $k_{\text{cat}}$  was found to decrease as the dielectric constant decreased which implied a change in the rate limiting step and demonstrated that the change in solvent composition has a greater effect on the physical steps of the catalysis, rather than the chemistry.<sup>123</sup>

Interpreted within the environmentally coupled tunneling framework, the absence of a dependency of the reaction rate on viscosity led to a conclusion that major contributions to tunneling from long-range motions do not drive hydride transfer in EcDHFR. The results did not rule out the possibility that short-range motions could

contribute to catalysis in EcDHFR or that conformationally and kinetically distinct sub-states are present at the point of catalysis.<sup>123</sup> It has been argued that the rate constant for the reaction is most likely determined by the conformational state of the enzyme at the point of the chemical reaction because this specific conformation will determine the electrostatics in and around the active site and will therefore control the reaction free energy.<sup>103</sup>

The previous work conducted by Loveridge *et al.* did not rule out short-range motions, un-slaved to the solvent, that could drive hydride transfer or conformationally and kinetically distinct sub-states are present at the point of catalysis. Localised motions are suggested to be important for hydride transfer in lactate dehydrogenase,<sup>124</sup> morphinone reductase,<sup>125</sup> aromatic amine dehydrogenase<sup>126</sup> and EcDHFR.<sup>127</sup> Previous studies of EcDHFR implied that hydride transfer in DHFR is not described well by the environmentally coupled tunnelling model of catalysis but better described by a model in which temperature- and pH-dependent changes to the conformational ensemble of the enzyme are important for catalysis.<sup>128</sup>

The effect of co-solvents on the EcDHFR (EcDHFR: NADP<sup>+</sup>: folate) model Michaelis complex was investigated by NMR spectroscopy. Large chemical shift perturbations in the <sup>1</sup>H-<sup>15</sup>N HSQC were observed for certain amide groups upon addition of methanol and glycerol, while other amide groups exhibited little or no change. The chemical shift changes were not uniform and different in the presence of glycerol and methanol. The residue specific amide chemical shift perturbation magnitudes and directions were such that no general solvent effect could be described. The chemical shift perturbations could not be correlated with the surface exposure of the residue, ligand binding, or whether the residue shows chemical exchange (to the excited reactive state) on a microsecond-millisecond time scale in the absence of co-solvent.<sup>129</sup>

From the non-linearity of some of the chemical shift perturbations with increasing solvent concentrations, it was inferred that the presence of organic co-solvent may induce slow conformational changes, which give rise to long-lived intermediates within the Michaelis complex of EcDHFR at the point of reaction. Conformational transitions would occur on a slower timescale than the chemical step thus giving rise to conformations of the enzyme complex with altered electrostatics in the active site, which may give rise to the observed, altered rate constants for hydride transfer.<sup>129</sup>

In other studies with organic co-solvents, molecular dynamics simulations have predicted that protein mobility should be decreased in more viscous solvent and that the influence of the increased viscosity would be felt by the protein both internally and externally. This study also found that the average structure, fluctuations and secondary structure of the protein did not appear to show a dependence on viscosity of the medium.<sup>130</sup> Vibrational echo experiments measuring the dephasing of CO bound to the active site of four haeme proteins have revealed weak viscosity dependence from ps-ns timescale motions. Slower but larger amplitude fluctuations requiring protein topology changes were found to be affected by viscosity. It was proposed that increasing the viscosity reduces the presence of larger amplitude fluctuations and lengthens the characteristic timescale of fluctuations in the time window investigated.<sup>131</sup> Alterations to the electrostatics of the solvent medium should have consequences for dynamics, if the protein structure and motions are governed in full or partially by electrostatics. Molecular dynamics simulations have found that motions in chymotrypsin become more rapid with increasing dielectric constant and it is proposed that protein motions are faster in more hydrophilic solvents.<sup>132</sup>

**To address the question of whether catalysis by DHFR from *E. coli* in the presence of co-solvents characterised by sub-populations of the Michaelis complex mimic (EcDHFR: NADP<sup>+</sup>: folate) with different conformations and kinetics,** sufficient NMR data will be acquired to achieve complete assignment of the enzyme complex under standard buffer conditions, in the presence of 17% glycerol and, in the presence of 17% methanol.

Initial changes to the chemical shifts of amide groups in the Michaelis complex mimic in the presence of co-solvents which affect the electrostatic or viscosity of the environment could be a direct consequence of the solvent interacting with the enzyme complex and its surrounding solvent sphere, which has very few regions that are inaccessible to solvent, or could be a sensitive indicator of changes to the structure of the protein, both outer and inner-core. Analysis of the chemical shifts of the whole protein will provide further clarity on the nature of any changes, on an atomic scale. The backbone amide and carbon chemical shifts of the Michaelis complex mimic in buffer have previously been published, which will provide a starting basis for assignment.<sup>133</sup>

## **2.2 Analysis of chemical shift perturbation in the presence of co-solvents**

Chemical shift values of nuclei are sensitive indicators to the electronic environment in which they reside so can be used as probes to determine any changes to protein residues in solution. Sets of chemical shift assignments for the Michaelis complex mimic under standard buffer conditions, in the presence of 17% glycerol and, in the presence of 17% methanol allow for direct comparisons of chemical shift values of all the measureable carbon, nitrogen and proton atoms in the protein. Chemical shift is a manifestation of the electronic environment of the nucleus. Chemical shift perturbations can be exhibited in proteins due to changes in molecular structure, ligand binding or a change in the protein environment, such as a change to the solvent composition. The direct comparison of the chemical shifts of the protein under different co-solvent conditions will report on any co-solvent induced, atom specific changes to the protein environment or structure.

## **2.3 ps-ns Relaxation Dynamics**

Dynamical information from ps-ns T1, T2 and NOE experiments in the presence of 17% glycerol, 17% methanol and a buffer only control will initially be interpreted using the phenomenological method and *via* the Lipari-Szabo “model free” formalism if time permits. The comparison of the phenomenological data will report on any co-solvent induced changes to the ps-ns dynamics of the protein at the point of reaction.

## 3.0 Experimental

### 3.1 Buffers and Media

#### 3.1.1 Purification and UV buffers

The purification buffers are outlined below (**Table 2**). All buffers were prepared freshly with distilled water, the pH adjusted and then filtered and degassed.

<b>Buffer A</b>	50 mM $\text{KiPO}_4$ , 10 mM $\beta$ -ME, pH 7.0
<b>Buffer B</b>	50mM $\text{KiPO}_4$ , 10 mM $\beta$ -ME, 1 M NaCl, pH 7.0
<b>Buffer C</b>	50 mM $\text{KiPO}_4$ , 10 mM $\beta$ -ME, 1 mM NaCl, pH 7.0
<b>Buffer D</b>	75.8 mM $\text{KiPO}_4$ , 15 mM $\beta$ -ME, 1.5 mM NaCl, pH 7.0
<b>UV Kinetics Buffer</b>	50mM $\text{KiPO}_4$ , 10 mM $\beta$ -ME, 1 M NaCl at pH 7.0
<b>UV Concentration Buffer</b>	50 mM $\text{KiPO}_4$ at pH 7.0

**Table 2** Purification and UV buffer constituents

#### 3.1.2 SDS-PAGE

##### **Gel buffer**

To deionised water (20 mL) was added Tris base (36.42 g) and 10% SDS (3 mL). The pH was adjusted to 8.45 and a final volume of 50 mL.

##### **Anode running buffer** (0.2 M tris base)

To deionised water (900 mL) was added Tris base (24.28 g). The pH was adjusted to 8.90 and the final volume adjusted to 1 L.

##### **Cathode running buffer** (0.1 M Tris, 0.1 M Tricine and 0.1 % SDS)

To deionised water (900 mL) was added Tris base (12.14 g), tricine (17.92 g) and 10% SDS (10 mL). The pH was adjusted to 8.3 and the final volume adjusted to 1 L by addition of deionised water.

##### **Running gel** (10 % acrylamide/bis-acrylamide)

To deionised water (0.62 mL) was added 30 % acrylamide/bis-acrylamide (3.33 mL), gel buffer (3.33 mL) and 50 % glycerol (2.6 mL). Immediately prior to pouring, freshly made 10% ammonium persulfate (APS, 100  $\mu$ L) and N,N,N',N'-tetramethylethylenediamine (TEMED, 20  $\mu$ L) were added and the solution gently mixed to initiate polymerisation.

#### **Stacking gel (4 % acrylamide/bis-acrylamide)**

To deionised water (3 mL) was added 30 % acrylamide/bis-acrylamide (0.67 mL) and gel buffer (1.25 mL). Immediately prior to pouring, freshly made 10 % APS (50  $\mu$ L) and TEMED (20  $\mu$ L) were added and the solution gently mixed to initiate polymerisation.

#### **Sample buffer (4x)**

To 0.5 mL of deionised water was added 1 M Tris-HCl pH 6.8 (2.5 mL), SDS (1.0 g), 0.1 % bromophenol blue (0.8 mL), 100 % glycerol (4 mL) and  $\beta$ -mercaptoethanol (2 mL of 14.3 M (stock))

#### **Acrylamide gel stain**

#### **Acrylamide gel destain**

To distilled water (500 mL) was added glacial acetic acid (100 mL) and isopropanol (400 mL).

#### **SDS-PAGE**

The running gel was poured and once set the stacking gel poured directly on top. A comb was inserted to create loading wells and the gel allowed to set. Fraction samples (15  $\mu$ L) were incubated with sample buffer (5  $\mu$ L) for 10 mins at 85 °C and loaded into the gel wells. Anode and cathode running buffers were added and a 150 V potential applied for 45 minutes or until the sample fractions had run the length of the gel. Acrylamide gels were stained with Coomassie blue stain and destained (10% Acetic acid, 40% ethanol, 50% water).

## **3.2 Protein Expression Media**

### **3.2.1 Preparation of *Luria-Bertani* (LB) medium**

Tryptone (10 g), Yeast extract (5 g) and NaCl (10 g) were added to 1 L of distilled water and the pH adjusted to 7.0. LB was autoclaved and appropriate antibiotic added once the LB had cooled to ambient temperature.

### **3.2.2 Preparation of M9 medium (1.0 L)**

Deionised water (1 L) was autoclaved and cooled to room temperature. To sterile deionised water (200 mL) was added Na<sub>2</sub>HPO<sub>4</sub> (6.7 g), KH<sub>2</sub>PO<sub>4</sub> (3.0 g), NaCl (0.5 g) and [<sup>15</sup>N]-Ammonium Chloride (1 g). [<sup>13</sup>C<sub>6</sub>]-D-glucose (1 g)\* was dissolved in sterile deionised water (100 mL) and added to the salt solution whilst mixing. Trace elements (1.8 mL), CaCl<sub>2</sub> (1.0 M, 0.3 mL) and MgSO<sub>4</sub> (1.0 M, 6.0 mL) were added to the solution and sterile deionised water added to a final volume of 1L. The pH of the solution was adjusted to pH 7 and the solution sterile filtered and transferred to sterile Erlenmeyer flasks.

\* Depending on sample labelling: for <sup>15</sup>N labelling only, use [<sup>12</sup>C<sub>6</sub>]-α-D-glucose (3 g)

## **3.3 Labelled co-solvents**

Deuterated glycerol (D-4) and deuterated methanol (D-4) were sourced from Sigma-Aldrich.

## **3.4 Labelled protein expression methods**

### **3.4.1 Preparation of M9 medium for perdeuteration (0.5 L)**

To sterile deuterated water (100 mL) was added Na<sub>2</sub>HPO<sub>4</sub> (3.35 g), KH<sub>2</sub>PO<sub>4</sub> (1.5 g), NaCl (0.25 g) and [<sup>15</sup>N]-Ammonium Chloride (0.5 g). D-glucose\* (1.5 g) was dissolved in sterile deuterated water (100 mL) and added to the salt solution whilst mixing. Trace elements (300 µL), CaCl<sub>2</sub> (1.0 M, 50 µL) and MgSO<sub>4</sub> (1.0 M, 1.0 mL) were added to the solution and sterile deionised water added to a final volume of 1L. The pH of the

solution was adjusted to pH 7 and the solution sterile filtered and transferred to sterile Erlenmeyer flasks.

\* Depending on sample labelling: for up to 70 % deuteration  $^{13}\text{C}$  ( $[^{13}\text{C}_6]$ - $\alpha$ -D-glucose or  $[^{12}\text{C}_6]$ - $\alpha$ -D-glucose was used. For levels of deuteration higher than 70 %, deuterated  $^{13}\text{C}$  ( $[^{13}\text{C}_6]$ - $\alpha$ -D-glucose was used.

### 3.4.2 Expression of unlabelled EcDHFR

A scraping from a glycerol stock of *E. coli* BL21 (DE3) competent cells that had been transformed with the EcDHFR expression plasmid was used to inoculate LB medium (100 mL) containing ampicillin (100 mg/L) and incubated overnight (37 °C, 180 rpm). Each large-scale flask containing 500 mL of LB and ampicillin (100 mg/L) was then inoculated with 5 mL of the sub-culture. The cells were grown (37 °C, 180 rpm, 5 h) to an  $\text{OD}_{600}$  between 0.6 and 0.8. Isopropyl  $\beta$ -D-1 thiogalactopyranoside (IPTG) (120 mg/L) was added to induce protein expression and incubation continued (37 °C, 180 rpm, 18 h). The cells were harvested by centrifuge (6000 rpm, 20 min.) and the pellet stored at -20 °C.

### 3.4.3 Expression of $^{15}\text{N}/^{13}\text{C}$ EcDHFR

Sterile minimal media M9 salt solution for the desired labelling was prepared, as above. As previously described an overnight culture was prepared from a scraping from a glycerol stock of *E. coli* BL21 (DE3) containing the EcDHFR plasmid. In contrast to unlabelled expression, the sub-culture cells were harvested by centrifuge (6000 rpm, 10 mins). Harvested cells were then washed to remove traces of LB by re-suspending in M9 salt solution (5 mL) and centrifuging twice. The washed cells were re-suspended in M9 medium (5 ml) and 1 mL used to inoculate each M9 medium flask (4x500 mL) containing ampicillin (100 mg/L). The cells were grown (37 °C, 180 rpm, 9 h) to an  $\text{OD}_{600}$  of 0.6. Isopropyl  $\beta$ -D-1 thiogalactopyranoside (IPTG) (120 mg/L) was added and incubation continued for protein expression (37 °C, 180 rpm, 18 h). The cells were harvested by centrifuge (6000 rpm, 20 min.) and the pellet stored at -20 °C.

### **3.4.4 Expression of perdeuterated EcdHFR**

Deuterated minimal media M9 salt solution (0.5 L) for the desired labelling was prepared, as above. 100 mL of sterile water M9 media (unlabelled) was also prepared, as above. In addition, a further 70% deuterated M9 media (100 mL) containing unlabelled glucose and ammonium chloride was prepared. A scraping from a glycerol stock of *E. coli* BL21 (DE3) that had been transformed was used to inoculate LB medium (100 mL) containing ampicillin (100 mg/L). The media was incubated overnight (37 °C, 180 rpm). The cells were harvested by centrifuge (6000 rpm, 10 mins). Harvested cells were washed to remove traces of LB by re-suspending in M9 salt solution (5 mL) and centrifuging twice. The washed cells were re-suspended in M9 medium (5 mL) and 1 mL of the pellet re-suspension used to inoculate the sterile water M9 media (100 mL) containing ampicillin (100 mg/L). The cells were incubated in the M9 subculture for 10 hours (37 °C, 180 rpm). 1 mL of the M9 sub-culture was used to inoculate the 100 mL solution of 70% deuterated M9 media. The cells were grown overnight (37 °C, 180 rpm). 4 mL of the 70 % deuterated subculture was used to inoculate 400 mL of deuterated M9 media. The media was incubated (37 °C, 180 rpm) until OD<sub>600</sub> of 0.6. Expression was induced by addition of IPTG (120 mg/L) and further incubation (37 °C, 200 rpm) for 24 hours. The pellet stored at -20 °C.

## **3.5 Protein purification**

### **3.5.1 Purification of EcdHFR**

The cell pellet was defrosted and re-suspended in a minimal volume of buffer (50 mM KiPO<sub>4</sub>, 10 mM βME (β-mercaptoethanol), pH 7.0). The cell suspension was sonicated for 5 min on ice (35% amplitude, 5 seconds on, 10 seconds off) and the lysed cell mixture centrifuged (16000 rpm, 30 min.). The supernatant was collected and syringe filtered to remove any remaining cell debris. The supernatant was applied to a Q-Sepharose column and the bound protein was eluted with a gradient from 0 to 1 M NaCl buffer A and B (**3.1.1 Purification and UV buffers**). Fractions were tested for activity, concentrated to 10 mL using a stirred cell concentrator (Amicon) with a 10 kDa membrane and the solution loaded onto a size exclusion column (HiLoad 26/60 Superdex G75) previously equilibrated in buffer C\*. The pure protein was eluted over 2-3 fractions (10 mL) and confirmed by SDS-PAGE.

## **3.6 Preparation of EcDHFR NMR samples**

### **3.6.1 Buffer only sample preparation**

EcDHFR: NADP<sup>+</sup>: Folate Complex in 100 % KiPO<sub>4</sub> buffer: The volume of the protein solution was reduced to 0.5 mL (1-2 mM concentration) using a spin column (10 kDa molecular weight cut off). An equimolar mixture of Folate and NADP<sup>+</sup> (100 μL, 100 mM) in buffer C was added to the protein solution to give a ligand concentration in 8-fold excess of the protein and much greater than the ligand dissociation constants. Deuterated water (10 % by volume) was added to the sample. A sample volume of 500 μL was added to a standard 600 MHz rated NMR tube, unless otherwise indicated.

\*An equal molar mixture of Folate and NADP<sup>+</sup> (100 mM) was prepared in KiPO<sub>4</sub> buffer C, the pH adjusted to 7.0 and 100 μL aliquots stored in the freezer, wrapped in tin foil to protect from light exposure.

### **3.6.2 Preparation of sample containing 17% glycerol**

The volume of the protein solution in buffer C was reduced to 5 mL. Buffer D (5 mL) was added and the volume further reduced to 0.5 mL and around 1 mM concentration (3.6.1). To deuterated glycerol (85 μL, 113.1 mg) was added D<sub>2</sub>O (50 μL), an equal molar mixture of Folate and NADP<sup>+</sup> (100 mM, 30 μL) and the volume made up to 500 μL by addition of the concentrated protein to give a final volume concentration of glycerol of 17%.

### **3.6.3 Preparation of sample containing 17% methanol**

The volume of the protein solution in buffer C was reduced to 5 mL. Buffer D (5 mL) was added and the volume further reduced to 0.5 mL. To deuterated methanol (85 μL) was added D<sub>2</sub>O (50 μL), an equal molar mixture of Folate and NADP<sup>+</sup> (100 mM, 30 μL) and the volume made up to 500 μL by addition of the concentrated protein to give a final volume concentration of methanol of 17%.

## 3.7 Analytical Methods

### 3.7.1 Ultraviolet photospectrometry

Steady state kinetic measurements were taken on a spectrometer (Jasco V-660) with both the cofactor and substrate in excess over the enzyme in a cuvette with path length of 10 mm and volume of 1 mL. The reaction temperature was controlled using a water bath (ref) and the kinetic buffer was incubated prior to the reaction at the set temperature.

Steady state turnover of the enzymatic reaction was monitored at 340 nm using a combined extinction coefficient for the substrate and cofactor ( $11,800 \text{ M}^{-1} \text{ cm}^{-1}$ ). In a standard  $k_{\text{cat}}$  measurement, enzyme of  $0.05 \text{ }\mu\text{M}$  concentration was incubated with cofactor ( $100 \text{ }\mu\text{M}$ ) in kinetic buffer (**Table 2**) for 1 minute. Substrate ( $100 \text{ }\mu\text{M}$ ) was rapidly mixed into the reaction mixture. In a standard  $k_{\text{M}}$  measurement set, the concentrations of substrate and cofactor ranged from  $0.5$  to  $200 \text{ }\mu\text{M}$ , whilst the saturating concentration was set to  $200 \text{ }\mu\text{M}$ . Initial velocities were recorded for 40 seconds, the change in absorbance per second measured and reaction rates calculated using the Beer-Lambert law.

The absorption of the concentrated protein eluent solution was determined *via* UV spectroscopy at four wavelengths (**Table 3**). The concentration was calculated from the Beer-Lambert law:  $A = \epsilon cl$  ( $A$  = absorption,  $c$  = concentration,  $l$  = path length (1 cm)), using the associated molecular extinction coefficients ( $\epsilon$ ). Typical protein concentrations ranged from 1-2 mM in 0.5 mL of concentrated eluent.

Extinction coefficient $\epsilon$ ( $\text{mL mg}^{-1} \text{ cm}^{-1}$ )	Wavelength (nm)
31	205
20	210
15	215
11	220

**Table 3** EcDHFR absorption maxima and extinction co-efficients for concentration calculations.

### 3.7.2 Nuclear magnetic resonance spectroscopy

NMR spectra were acquired at 298 K and the spectrometer specifics are presented with the acquisition logs below. Sample labelling and respective experiments run on the sample are given (**Table 4**, **Table 5**, **Table 6**, **Table 7**, **Table 8** and **Table 9**). Spectra were processed using NMRPipe<sup>134</sup> or Bruker Topspin software and analysed using CcpNmr Analysis 2.1.5 for Mac.<sup>135,136</sup> TOCSY mixing times were set to 20 ms and NOESY mixing times set to 120 ms for uniformity. Backbone assignment was completed with a standard set of triple resonance spectra including <sup>1</sup>H-<sup>15</sup>N-HSQC, HNCA, HN(CO)CA, HNCACB, C(CO)NH-TOCSY, HNCO, HN(CA)CO. Side chains were assigned utilising C(CO)HN-TOCSY, CCH-TOCSY, HCCH-TOCSY and <sup>13</sup>C-HSQC. Through space inter-proton protein-protein distance restraints were derived from 3D <sup>1</sup>H-<sup>15</sup>N NOESY-HSQC and 3D <sup>1</sup>H-<sup>13</sup>C NOESY-HSQC, inter-proton protein-ligand distance restraints were derived from 13C/15N F1 Filtered 13C Edited NOESY and 13C/15N F1 Filtered 15N Edited NOESY HSQC. Dihedral angle restraints were calculated using DANGLE. Hydrogen bonding restraints were inferred from a set of hydrogen-deuterium exchange data. Backbone amides with long half-lives were considered as H-bond acceptors and donors were found manually by inspection of the crystal structure 1RX2.pdb.

Reference <sup>1</sup>H-<sup>15</sup>N HSQC were taken at the start and end of each set of acquisition to determine and changes over time to the sample. <sup>1</sup>H referencing to sodium 2,2-dimethyl-2-silapentane-5-sulphonate (DSS) at 0.000 ppm was undertaken in separate samples and the referencing then automatically propagated to <sup>13</sup>C and <sup>15</sup>N in Bruker Topspin through consideration of the ratios of gyromagnetic ratios.

I acquired the data with the exception of spectra denoted with **EJL** (in these cases, the data was acquired by E. Joel Loveridge) at the following locations:

- Department of Chemistry, Cardiff University (**CU**) at **600 MHz** (Bruker Avance 600 MHz (1H) spectrometer equipped with cryogenically cooled 5 mm z-gradient HCN probe).
- Henry Wellcome Building, national facility for biomedical research, University of Birmingham (**HWB-NMR**) at either **900 MHz** (5mm cold probe equipped) or **600 MHz** (Bruker, 1.7 mm cryoprobe).
- MRC Biomedical NMR Centre (now known as The Francis Crick Institute) (**MRC-NMR**) at either **800 MHz** (Bruker Avance III HD 800 equipped with

cryogenically cooled 5 mm z-gradient HCN probe) or **700 MHz** (Bruker Avance III HD 700 with a 5mm  $^1\text{H}/^{31}\text{P}/^{13}\text{C}/^{15}\text{N}$  quadruple-resonance PFG cryoprobe with cooled C13 and P31 preamplifiers)

- Department of Chemistry, University of Bristol (**UOB-NMR**) at **600 MHz**

Backbone	Data Acquisition Record		
	0% Co-solvent	17% Glycerol	17% Methanol
15N-HSQC	MRC-NMR 800 MHz	MRC-NMR 800 MHz	MRC-NMR 800 MHz
13C-HSQC	MRC-NMR 800 MHz	HWB-NMR 900 MHz	MRC-NMR 800 MHz
HNCA	MRC-NMR 800 MHz	CU-NMR 600 MHz	NA
HNCACB	CU-NMR 600 MHz	CU-NMR 600 MHz	MRC-NMR 800 MHz
HNCACO	MRC-NMR 700 MHz	CU-NMR 600 MHz	NA
HNCO	MRC-NMR 700 MHz	CU-NMR 600 MHz	MRC-NMR 800 MHz
HNCOCA	MRC-NMR 700 MHz	CU-NMR 600 MHz	NA
CBCACONH	BRI+CF 600 MHz	NA	NA

**Table 4** Log of NMR data acquired for protein backbone assignment. All data was acquired on double-labelled  $^{13}\text{C}/^{15}\text{N}$  protein, with the exception of  $^{15}\text{N}$ -HSQC, which was solely  $^{15}\text{N}$  labelled. Final protein concentrations were ~ 1 mM.

Data for Side-chain Assignment			
Spectrum	0% Co-solvent	17% Glycerol	17% Methanol
CcoNH-TOCSY	MRC-NMR 700 MHz	CU-NMR 600 MHz	NA
$^{15}\text{N}$ -TOCSY-HSQC	EJL HWB-NMR *	NA	NA
CCH-TOCSY	EJL HWB-NMR *	HWB-NMR 900MHz	NA
HCCH-TOCSY	EJL HWB-NMR *	CU-NMR 600 MHz	MRC-NMR 800 MHz
CBHD	HWB-NMR 900 MHz	MRC-NMR 800 MHz	MRC-NMR 800 MHz
CBHE	HWB-NMR 900 MHz	MRC-NMR 800 MHz	MRC-NMR 800 MHz

**Table 5** Log of NMR data acquired for protein side chain assignment. All data was acquired on double-labelled  $^{13}\text{C}/^{15}\text{N}$  protein, with the exception of  $^{15}\text{N}$ -TOCSY-HSQC, which was solely  $^{15}\text{N}$  labelled. Final protein concentrations were ~ 1 mM.

Data for Through-Space Distance Restraints			
Spectrum	0% Co-solvent	17% Glycerol	17% Methanol
<sup>15</sup> N-NOESY-HSQC	EJL UOB-NMR *	HWB-NMR 900 MHz	MRC-NMR 800 MHz
13C-NOESY AROMATIC	EJL HWB-NMR*	NA	NA
13C-NOESY ALIPHATIC	MRC-NMR 800 MHz	HWB-NMR 900MHz	MRC-NMR 800 MHz
HNHA	MRC-NMR 800 MHz	NA	MRC-NMR 800 MHz
13C/15N F1 Filtered 15N Edited NOESY HSQC	HWB-NMR 900 MHz	EJL HWB-NMR	MRC-NMR 800 MHz
13C/15N F1 Filtered 13C Edited NOESY	HWB-NMR 900 MHz	MRC-NMR 800 MHz	MRC-NMR 800 MHz
Long range HNC0	HWB-NMR 900 MHz	NA	NA

**Table 6** Acquisition log of through space NMR data. All data Acquired on double labelled <sup>13</sup>C/<sup>15</sup>N, with the exception of <sup>15</sup>N-NOESY-HSQC, which was solely <sup>15</sup>N labelled. Final protein concentrations were ~ 1 mM.

Acquired T2 data						
Delay (ms)	Buffer		17% Glycerol		17% Methanol	
	HWB-NMR 900MHz	HWB-NMR 600 MHz	HWB-NMR 900MHz	CU-NMR 600 MHz	HWB-NMR 900 MHz	HWB-NMR 600 MHz
10	X	X	X	X	X	X
30	X	X	X	X	X	X
50	X	X	X	X	X	X
70	X	X	X	X	X	X
90	X	X	X	X	X	X
110	X	X	X	X	X	X
130	X	X	X	X	X	X
150	X	X	NA	X	X	X
170	X	X	NA	X	X	X
190	X	X	NA	X	X	X
210	X	X	NA	X	X	X
250	X	X	NA	X	X	X

**Table 7** T2 Relaxation NMR data acquisition log. All data was acquired from perdeuterated <sup>15</sup>N labelled protein samples with final protein concentrations ~ 500 μM.

Acquired T1 Data						
	Buffer		17% Glycerol		17% Methanol	
Delay (ms)	HWB-NMR 900MHz	HWB-NMR 600 MHz	HWB-NMR 900MHz	CU-NMR 600 MHz	HWB-NMR 900 MHz	HWB-NMR 600 MHz
10	X	X	X	X	X	X
30	X	X	X	X	X	X
50	X	X	X	X	X	X
70	X	X	X	X	X	X
100	X	X	X	X	X	X
200	X	X	X	X	X	X
300	X	X	X	X	X	X
400	X	X	X	X	X	X
600	X	X	X	X	X	X
1000	X	X	X	X	X	X
1500	X	X	X	X	X	X
2000	X	X	X	X	X	X

**Table 8** T1 Relaxation NMR data acquisition log. All data was acquired from perdeuterated  $^{15}\text{N}$  labelled protein samples with final protein concentrations  $\sim 500 \mu\text{M}$ .

Acquired hetNOE (NOE/noNOE) data						
	Buffer		17% Glycerol		17% Methanol	
Recycle delay (s)	HWB-NMR 900MHz	HWB-NMR 600 MHz	HWB-NMR 900MHz	CU-NMR 600 MHz	HWB-NMR 900 MHz	HWB-NMR 600 MHz
10	X	X	X	X	X	X

**Table 9** hetNOE Relaxation NMR data acquisition log. All data was acquired from perdeuterated  $^{15}\text{N}$  labelled protein samples with final protein concentrations  $\sim 500 \mu\text{M}$ .

## 4.0 Assignment and Analysis of the chemical shift perturbations EcDHFR: NADP<sup>+</sup>: Folate in the presence of co-solvents

The results from work previously completed on the effects of organic co-solvents (methanol, ethanol, isopropanol, ethylene glycol, glycerol, sucrose and tetrahydrofuran) on the catalytic reaction rates of Dihydrofolate reductase (DHFR) from *Escherichia coli* and DHFRs from *Thermotoga maritime* and *Mortellia profunda* were deemed inconsistent with the environmentally coupled model of hydride transfer proposed for catalysis by EcDHFR due to the absence of a correlation between solvent viscosity and the rate constants ( $k_{\text{cat}}$  and pH 7 and 9 and  $k_{\text{H}}$  at pH 7).<sup>43, 137</sup> The environmentally coupled tunnelling model dictates that an increase in viscosity would be expected to dampen any long range promoting motions coupled to the hydride transfer step and consequentially, should show a dependence of the rate constant on viscosity.<sup>138</sup>

The rate constants were shown to decrease proportionally to the dielectric constant of the solvent medium, implying an overriding role for electrostatics in driving the rate constant. A low dielectric constant of the solvent strengthens hydrogen bonding networks and makes the protein more stable and less flexible and is therefore expected to favour catalysis by reducing the shielding of electrostatic effects within the active site. However, the results of this work revealed that catalysis was favoured by high dielectric constant, which was in contrast to the expected result.<sup>157</sup> It was proposed that reducing the dielectric constant “*reduces the gradient between the dielectric constant within the active site and that of the surrounding medium, thereby reducing protection of the perfectly aligned dipoles within the active site from more remote dipoles that may be deleterious to transition state stabilisation*”.<sup>47</sup>

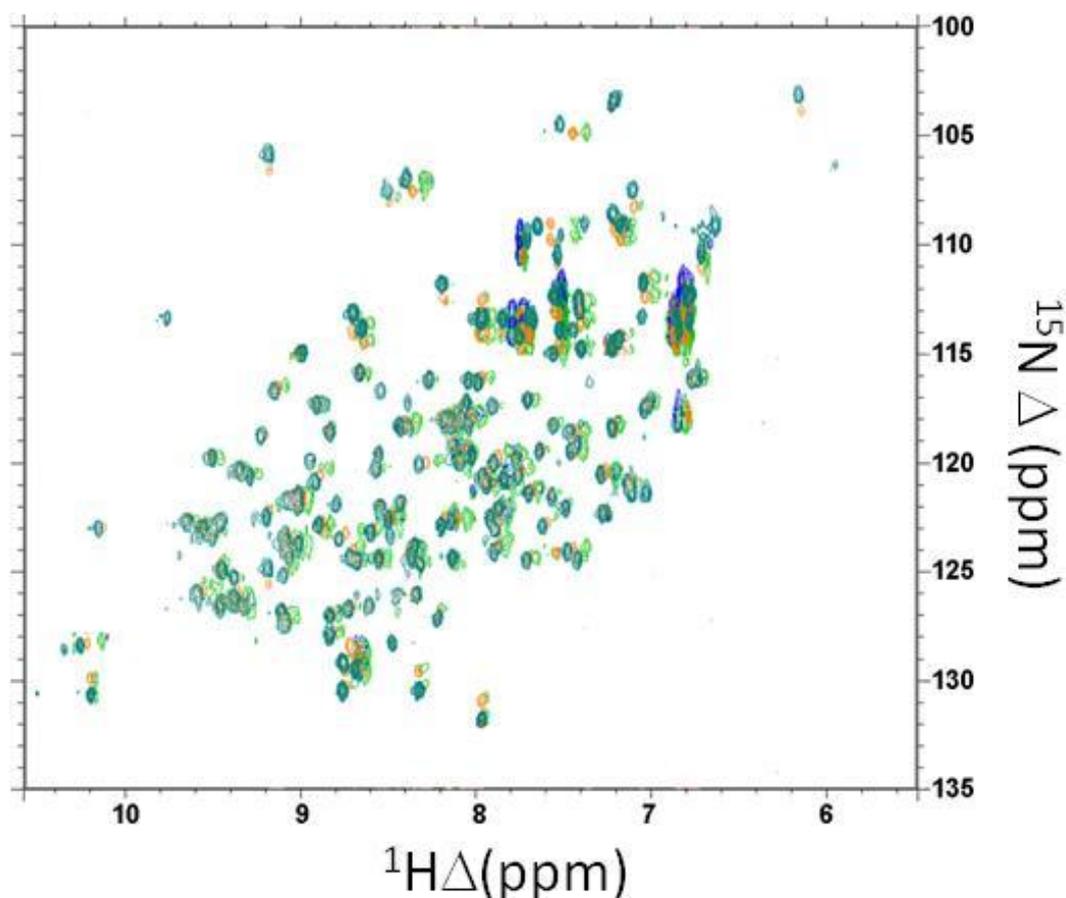
Of the organics co-solvents, glycerol and methanol have similar dielectric constants, making them almost iso-dielectric and allowing for any viscosity effects to be extracted through comparison of the data, assuming similar behaviours in the bulk solvent mixtures. Changes to the chemical shifts of atoms due to the changes in electrostatic or

viscous environment could be a direct consequence of the solvent interacting with certain areas of the protein or report on global, indirect changes to the structure of the protein. The location of atoms that display a chemical shift perturbation can be used as an indicator to the likelihood of the change in the chemical shift being due to an interaction with the co-solvent, which may lead to a change in the protein structure or a structural change in a non-solvent accessible region of the protein. The structure of EcDHFR in complex with NADP<sup>+</sup> and folate has already been solved by x-ray crystallography (1RX2.pdb) and will be initially used to infer the location of atoms and residues in the protein complex.

Organic co-solvents have been implemented to study the effects of perturbing protein motions by altering the viscosity and dielectric constants of the solvents in which enzymatic reactions take place. Molecular dynamics (MD) simulations have predicted that alterations to the electrostatics of the solvent medium should have consequences for dynamics, if the protein structure and motions are governed in full or partially by electrostatics. MD simulations have found that motions in chymotrypsin become more rapid with increasing dielectric constant and have proposed that protein motions are faster in more hydrophilic solvents.<sup>132, 139</sup>

Analysis of chemical shift perturbations are often carried out on enzyme complexes to study ligand binding and can provide wealth of information on structural changes caused by ligand binding close to and far from the actual ligand binding site.<sup>140, 145</sup>

**The objective of this results chapter is to report on any direct changes to the chemical shifts of atoms in the EcDHFR: NADP<sup>+</sup>: Folate complex that are induced in the presence of methanol and glycerol co-solvents.**

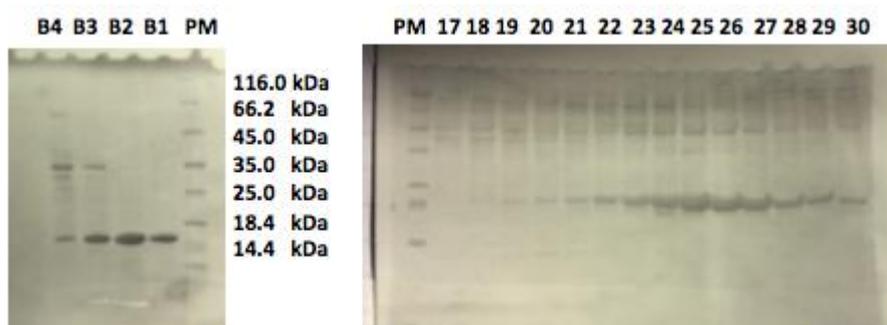


**Figure 14**  $^1\text{H}$   $^{15}\text{N}$ -HSQC backbone amide shifts of EcDHFR:Folate: $\text{NADP}^+$  complex in 0% (teal), 17% (red/orange), 33% (light green) and 50% (navy blue) glycerol. Spectra acquired on 1.1mM protein in 50 mM potassium phosphate buffer (pH 7.0) containing 1 mM NaCl, 10mM  $\beta$  - mercaptoethanol and 6.7 mM each  $\text{NADP}^+$  and folate with 10 %  $\text{D}_2\text{O}$  at 298 K.

Initial HSQC's showing backbone nitrogen and proton shifts indicate a non-uniform chemical shift of amide residues that are not in contact with the solvent with increasing co-solvent concentration (0%, 17%, 33%, 50% glycerol) (**Figure 14**). The chemical shift perturbations measured were residue specific and non-linear. Notably, it was observed that the chemical shift perturbations were uncorrelated to surface exposure of the residue, ligand binding or whether the residue shows chemical exchange on a microsecond-millisecond timescale, in the absence of any co-solvent. The peaks observed in the spectra did not display broadening which led to the conclusion that presence of glycerol as a co-solvent gives rise to slow conformational transitions within the EcDHFR:  $\text{NADP}^+$ : folate complex and that structures in different percentages of co-solvent represent long lived intermediate conformations of the enzyme, at the point of the chemical reaction.<sup>129</sup>

## 4.1 Biosynthesis, purification and preparation of protein samples

Dihydrofolate reductase from *E. coli* was overexpressed with regard to the required labelling strategy (Section 3.4), purified (Section 3.5.1) and prepared accordingly (Section 3.6). The protein purity was judged to be pure after two-step purification by ion exchange chromatography and size exclusion chromatography by the presence of a single molecular weight band at 18 kDa (SDS-PAGE) (Figure 15).



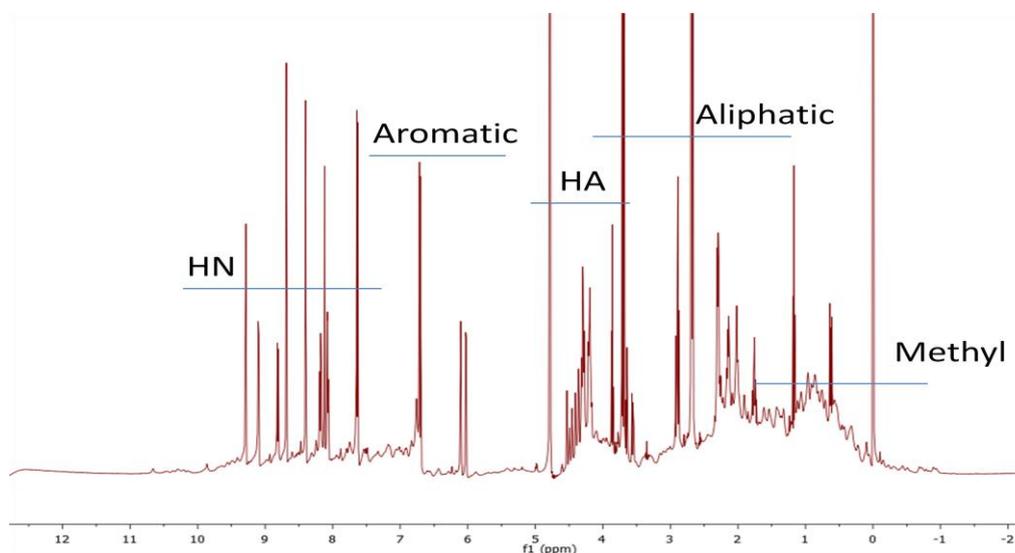
**Figure 15** SDS-PAGE gels from purification of EcDHFR. Fractions 17-30 eluted from Q-sepharose ion exchange column and fractions B4-B1 eluted from Superdex size exclusion column. Concentration of fractions B2-B1 gave > 95% purity EcDHFR.

## 4.2 Acquisition of NMR spectra

In order to characterise the effect of solvent on EcDHFR chemical shifts it was necessary to assign peaks to their corresponding residue within the primary sequence separately for all the conditions being studied. This was important as solvent effects result in large chemical shift differences within the  $^1\text{H}$   $^{15}\text{N}$ -HSQC making previous assignment of EcDHFR difficult to use. Spectra acquired for assignment of the backbone chemical shifts included the traditional suite of triple resonance experiments HNCA, HN(CO)CA, HNCACB, HN(CO)CACB, HNCACO and HNCO.<sup>141</sup> Assignment of side chain resonances was facilitated by acquisition of C(CO)NH-TOCSY, HCCH-TOCSY, CCH-TOCSY,  $^{15}\text{N}$ -TOCSY-HSQC and  $^{13}\text{C}$  HSQC spectra. In more challenging side chain assignment cases, through space  $^{13}\text{C}/^{15}\text{N}$  3D separated NOESY spectra were used as additional reference points. Through space NOE signals were identified with the help of  $^{13}\text{C}/^{15}\text{N}$  3D separated NOESY.  $^{13}\text{C}/^{15}\text{N}$  F1 Filtered  $^{13}\text{C}$  Edited NOESY and  $^{13}\text{C}/^{15}\text{N}$  F1 Filtered  $^{15}\text{N}$  Edited NOESY HSQC were acquired to

determine protein-ligand distance restraints. The comprehensive list of acquired spectra can be seen in (Chapter 3.7.2).

The protein purity and folding was confirmed by 1D and 2D NMR spectroscopy. 1D NMR was used to screen the protein complex in order to gauge the signal-to-noise ratio to determine the feasibility of basic correlation experiments. The folding of the protein was judged from the dispersion of resonances in the spectrum. Folded proteins exhibit a range of chemical shifts due to the anisotropic chemical shifts of near-by aromatic or carbonyl groups, whereas unfolded proteins exhibit little dispersion.<sup>34</sup> Sharp peaks observed in the spectrum are due to the folate and NADP<sup>+</sup> ligands (Figure 16).



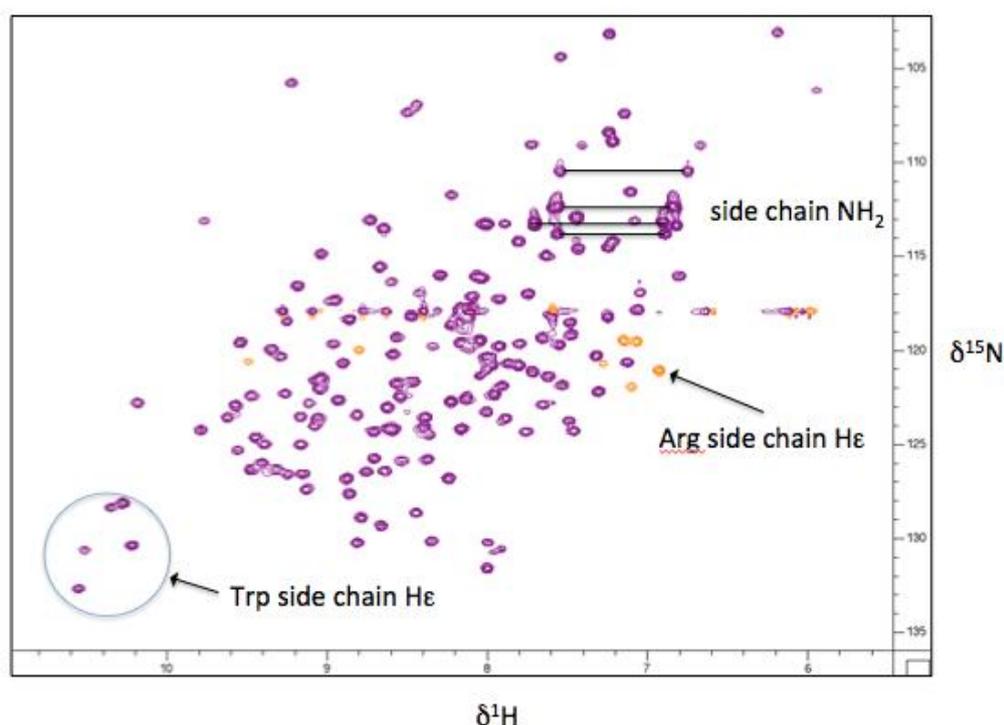
**Figure 16** <sup>1</sup>H NMR spectrum of EcdHFR:folate:NADP<sup>+</sup> under standard phosphate buffer conditions with DSS. The <sup>1</sup>H spectrum of proteins exhibits fingerprint regions. Figure labelling adapted from reference <sup>141</sup>

HSQC experiments were run prior to and post 3D experiments to ensure the integrity of the protein sample throughout the experimental time period. The spectra were over-laid to observe any changes in the protein and to confirm the sample's stability.

### 4.3 Assignment of EcDHFR: NADP<sup>+</sup>: folate in buffer

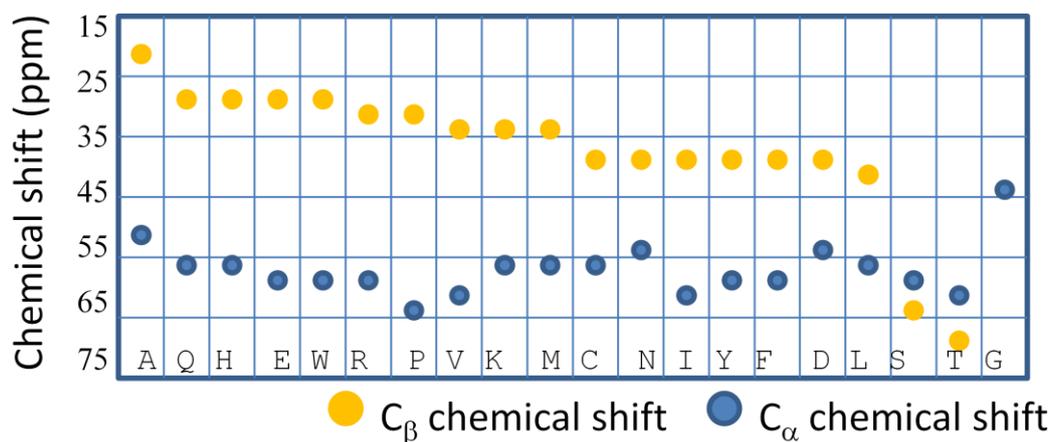
NMR spectra were assigned in CCPNMR Analysis *via* a triple resonance sequential backbone assignment method, which involves the use of a HSQC spectrum and the described suite of three dimensional triple resonance backbone spectra to assign the backbone resonances to protein primary sequence.

The <sup>1</sup>H - <sup>15</sup>N HSQC spectrum shows one peak for each backbone amide N-H bond in a uniformly <sup>15</sup>N labelled protein. As such, all amino acid residues in a protein, with the exception of proline, which does not have a backbone N-H group when polymerised, give rise to one peak in the HSQC spectrum. Side chain N-H bonds also give rise to one peak in the HSQC and side chain NH<sub>2</sub> groups give rise to two peaks with equal nitrogen chemical shifts but differing hydrogen chemical shifts. The NH side chains of tryptophan and arginine appear in characteristic regions of the spectrum (**Figure 17**).



**Figure 17** <sup>1</sup>H-<sup>15</sup>N-HSQC of EcDHFR: NADP<sup>+</sup>: folate complex in phosphate buffer only displaying characteristic amino acid side chain regions. Arginine side chains appear orange as they are typically recorded folded into the spectrum to reduce the spectral width and improve resolution; the average <sup>15</sup>N chemical shift is around 80 ppm.

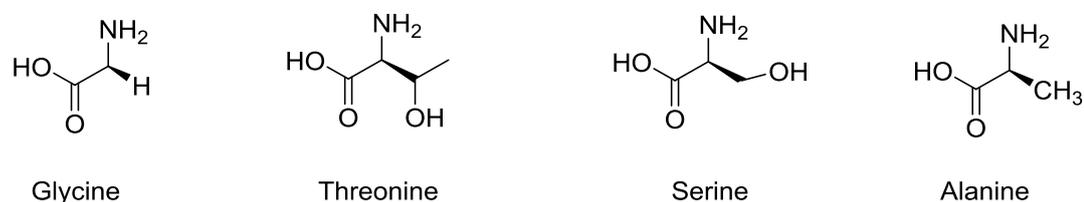
Each peak in the  $^{15}\text{N}$  –H HSQC was assigned an ambiguous spin system (with the exception of side chain peaks, which are identified in a straightforward manner), which was then disambiguated through analysis of the triple resonance 3D spectra. The 3D spectra allow each ambiguous residue to be identified as an amino acid type through analysis of the spectra. A spin system is a group of spins that are connected to each other by scalar spin-spin couplings. The amide bond in a protein prevents  $^1\text{H}$ - $^1\text{H}$  J-coupling between different amino acids, meaning that each amino acid in a protein has one or more separate spin system. Sometimes amino acids are comprised of two or more spin systems, in the case of amino acids with the aromatic side chains of tyrosine, tryptophan and histidine and the amide protons of glutamine and asparagine because the protons in these regions of the side chain are not scalar coupled to the protons in rest of the side chain.



**Figure 18**  $\text{C}_\alpha$  and  $\text{C}_\beta$  chemical shifts of all amino acids. The chemical shifts of some amino acids are very characteristic which allows facile identification.

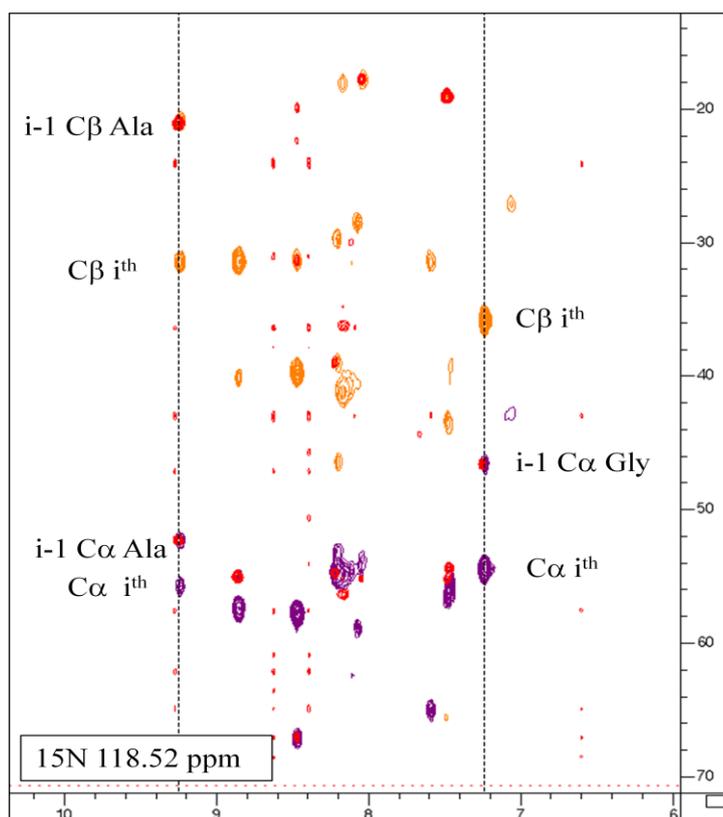
The different nuclei in the backbone of amino acids have very characteristic chemical shifts, which aids the significant challenge of assigning the chemical shifts of a whole protein. It is therefore possible to distinguish between the  $\text{C}_\alpha$ ,  $\text{C}_\beta$  and CO chemical shifts for a particular amino acid in a facile manner (**Figure 18**). The challenge is further alleviated because most amino acid spins systems exhibit distinctive peak patterns in the triple resonance and TOCSY (Total Correlation Spectroscopy) spectra. In the case of glycine, the side chain consists solely of a proton therefore in the HNCACB spectrum, the absence of a  $\text{C}_\beta$  is indicative of the amino acid type. The amino acid type is further confirmed because  $\text{C}_\alpha$  in glycine has a distinctive random coil chemical shift of  $\sim 45\text{ppm}$ . The side chain of alanine consists of one methyl group such that  $\text{C}_\beta$  in the

alanine side chain has a random coil chemical shift of ~ 20ppm which is distinctive from other amino acids. The C $\beta$  resonance in most amino acid side chains appears more upfield than C $\alpha$ . Serine and threonine are immediately identifiable from other amino acids due to C $\beta$  having a more downfield chemical shift than C $\alpha$  (**Figure 19**).



**Figure 19** Amino acids with characteristic C $\alpha$  and C $\beta$  chemical shifts. During the initial backbone assignment process, these amino acids can often be identified outright.

The spectra acquired for the triple resonance backbone approach to the protein NMR assignment are shown below (**Table 10**) These spectra include either carbon resonances for the  $i^{\text{th}}$  amino acid and the  $i^{\text{th}}-1$  amino acid (HNCA, HNCACB) residues or only resonances for the  $i-1$  amino acid (HNCOCA, HNCOCACB and CCONH-TOCSY). Combined use of the five experiments in this case enabled the intra ( $i^{\text{th}}$ ) amino acid resonances to be distinguished from those of the preceding inter ( $i-1$ ) residue. The CCONH-TOCSY shows the amide HN correlated to all aliphatic carbons in the immediate spin system on  $i-1$  amino acid residue only (**Figure 20**). Although the CCONH TOCSY that was acquired was of relatively low signal to noise, it significantly simplified the task of determining the amino acid type of spin systems on the preceding residue containing C $\gamma$ , C $\delta$  and C $\epsilon$  in the cases of lysine, arginine, proline, leucine, isoleucine and valine and assisted later on in the process of assigning the side chain resonances. HNCO and HNCACO spectra were acquired for assignment of the carbonyl carbons.

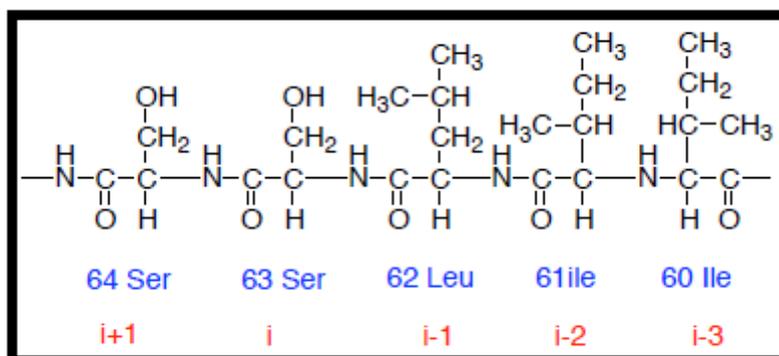


**Figure 20** Combination of CCONH-TOCSY (red) and HNCACB ( $C_{\alpha}$  shown in purple and  $C_{\beta}$  shown in orange). Complete side chain resonances on the preceding residue appear in the CCONH-TOCSY allowing identification of the amino acid type.

Spectrum	Spectrum description
HNCO	Shows amide HN correlated to the backbone carbonyl carbon on i-1 amino acid residues
HNCACO	Shows amide HN correlated to the backbone carbonyl carbon on the i and i-1 amino acid residues
HNCA	Shows amide HN correlated to $C_{\alpha}$ on i and i-1 amino acid residues
HNCOCA	Shows amide HN correlated to $C_{\alpha}$ on i-1 amino acid residue only
HNCACB	Shows amide HN correlated to $C_{\alpha}$ and $C_{\beta}$ on i and i-1 amino acid residues
HNCOCACB	Shows amide HN correlated to $C_{\alpha}$ and $C_{\beta}$ on i-1 amino acid residue only
CCONH-TOCSY	Shows amide HN correlated to all aliphatic carbons in the immediate spin system on i-1 amino acid residue only

**Table 10** Description of spectra acquired for backbone assignment

Assignment of the side chain resonance were achieved using several total correlation spectroscopy (TOCSY) experiments (**Table 11**), which allowed determination of the amino acid type of both inter ( $i^{\text{th}}$ ) and intra ( $i-1$ ) residue and hence the position of the residue in the polypeptide sequence (**Figure 21**).

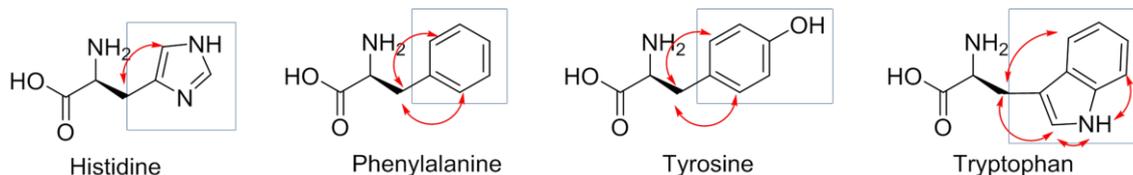


**Figure 21** The amino residues in the protein sequence were initially labelled ambiguously (red) until the true residue number could be determined (blue).

CCH-TOCSY	Shows the side-chain hydrogen nuclei and directly attached $^{13}\text{C}$ nuclei correlated to all other aliphatic carbon nuclei in the spin system
HCCH-TOCSY	Shows the side-chain hydrogen nuclei and directly attached $^{13}\text{C}$ nuclei correlated to all other hydrogen nuclei in the spin system
HHN-TOCSY	Shows amide HN correlated to all other hydrogen nuclei in the spin system

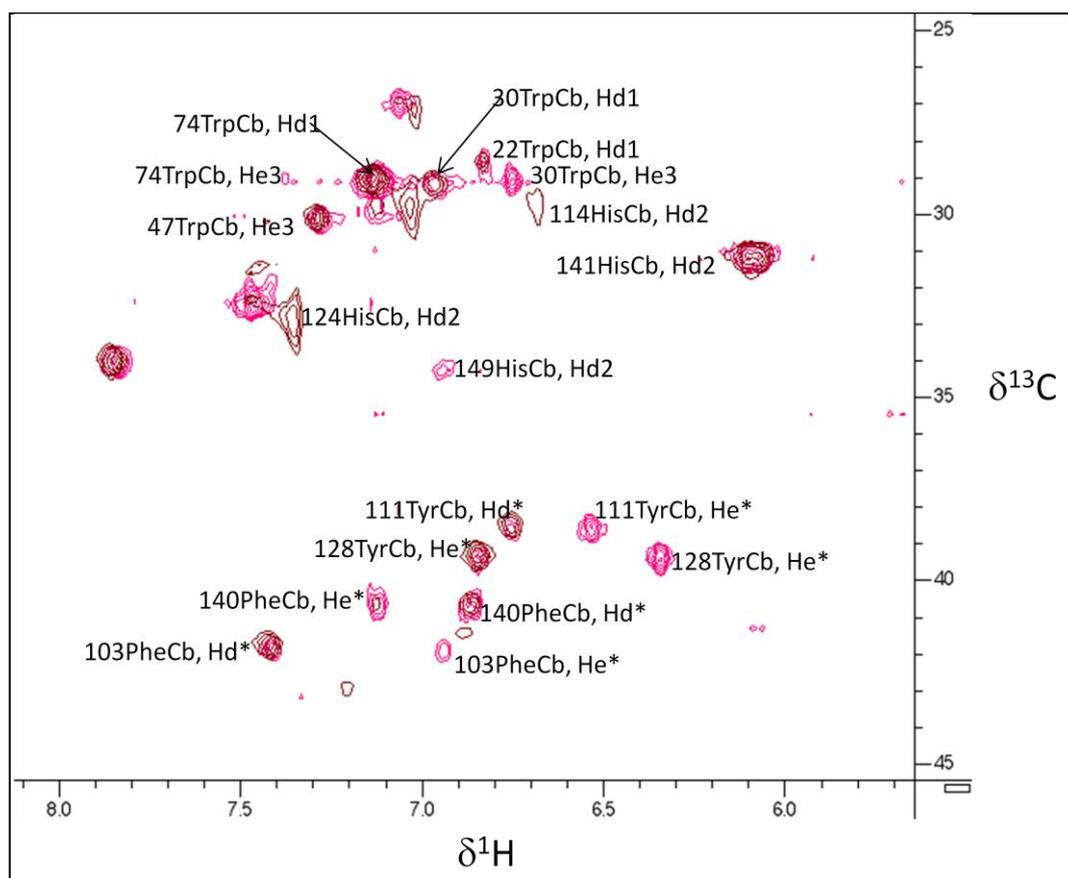
**Table 11** Description of the spectra acquired for assignment of the side chain resonances.

The backbone and side chain amide resonances were assigned to a high completion in the HSQC. Some backbone resonances were not assigned (1MET, 7ALA, 14ILE, 17GLU, 46THR, 47TRP, 48GLU, 56GLY, 97GLY, 113THR, 123THR, and 124HIS) due to peaks not being visible in the HSQC most likely due to unfavourable relaxation properties although side chain resonances for some of these residues were assigned.



**Figure 22** The aromatic rings of the amino acids above are not coupled to the backbone and side-chain and it was necessary to use specific pulse sequences which propagate the magnetisation in order to acquire the data.

Assignment of aromatic residues was achieved through acquisition of CBHD and CBHE. Of the aromatic residues: tryptophan, phenylalanine, tyrosine and histidine, it is possible to assign the  $\delta$  and  $\epsilon$  protons provided the chemical shift of  $C\beta$  is already known and there is good dispersion of peaks in the spectrum (**Figure 22**). The assignments were confirmed by establishing the presence of the aromatic proton resonance in the NOESY spectra. Only HD is visible in histidine as the HE is isolated by the nitrogen bonded either side of CE.<sup>142</sup> There is breakthrough on the CBHE spectrum from HD due to the propagation of the pulse sequence causing excitation of CD. Experiments are insensitive because of the distance of propagation of magnetisation, so not all of the peaks appear in the spectrum (**Figure 23**).



**Figure 23** Combined CBHD and CBHE spectra for assignment of aromatic side chain protons. The spectrum shows the CB chemical shift on one axis and the HD chemical shift (brown) and the HE chemical shift (pink).

The completeness of the chemical shift assignments for the EcDHFR complex is shown below (**Table 12**). The comprehensive list of chemical shift data for this complex under buffer only conditions is presented in **Appendix 8.1.3**.

Atom Type	Percentage assigned
C	74.3
H	84.5
N	67.4

**Table 12** Completeness of assignment of EcDHFR: NADP<sup>+</sup>: Folate in phosphate buffer. Figures were taken directly from ccpNMR Analysis project and not corrected for un-assignable atoms.

#### 4.4 Assignment of EcDHFR in 17% methanol

The spectra acquired for the assignment of EcDHFR: NADP<sup>+</sup>: folate in the presence of 17% methanol are listed comprehensively (**Chapter 3.7.2**) The assignment of EcDHFR: NADP<sup>+</sup>: folate in the presence of 17% methanol was completed by comparison of the already assigned protein in buffer only alongside the newly acquired protein spectra in the presence of 17% methanol. The data set that was acquired was large enough to confirm assignments but for example only a HNCO was recorded in this case whereas for the buffer only assignment, both a HNCO and HNCACO were acquired for assignment of carbonyl carbons.

Addition of methanol to the protein sample decreased the stability but not significantly enough to impact the acquisition of the longest experiment, which was around 2 days.

The completeness of the chemical shift assignments for the EcDHFR complex is shown below (**Table 13**). The comprehensive list of chemical shift data for this complex under buffer only conditions is presented in **Appendix 8.1.2**.

<b>Atom Type</b>	<b>Percentage assigned</b>
C	72.7
H	86.1
N	69.1

**Table 13** Completeness of assignment of EcDHFR: NADP<sup>+</sup>: Folate in phosphate buffer with 17% methanol. Figures were taken directly from ccmNMR Analysis project and uncorrected for non-assignable atoms.

## 4.5 Assignment of EcDHFR in 17% glycerol

Assignment of the protein complex [EcDHFR:folate: NADP<sup>+</sup>] in the presence of 17% glycerol was completed in the same manner as the protein in 17% methanol. The spectra acquired for the assignment of EcDHFR: NADP<sup>+</sup>: Folate in the presence of 17% glycerol are listed comprehensively (**Chapter 3.7.2**) The increased viscosity of the solution caused by the presence of glycerol led to some line broadening and a loss of peaks in the spectra due to slower tumbling. Any increased ambiguity was alleviated through comparison with the data for the protein in buffer only.

The completeness of the chemical shift assignments for the EcDHFR complex is shown below (**Table 14**). The comprehensive list of chemical shift data for this complex under buffer only conditions is presented in **Appendix 8.1.1**.

Atom Type	Percentage assigned
C	73.3
H	84.8
N	68.7

**Table 14** Completeness of assignment of EcDHFR: NADP<sup>+</sup>: Folate in phosphate buffer with 17% glycerol. Figures were taken directly from ccmNMR Analysis project and uncorrected for non-assignable atoms.

## 4.6 Assignment Conclusions

The proton, carbon and nitrogen chemical shifts of the Michaelis complex mimic (EcDHFR: NADP<sup>+</sup>: folate) in buffer have been assigned to a good completion, which significantly builds on the previously published backbone results.<sup>143</sup> The assignments of the protein in the presence of 17% methanol and 17% glycerol have been completed to an equally high level of completion, which will facilitate the aim of analysing the chemical shift perturbations caused by the solvent. The comprehensive chemical shift data is presented in **Appendix 8.1**. For the protein in the presence of 17% glycerol (**Appendix 8.1.1**), 17% methanol (**Appendix 8.1.2**) and the protein in buffer (**Appendix 8.1.3**).

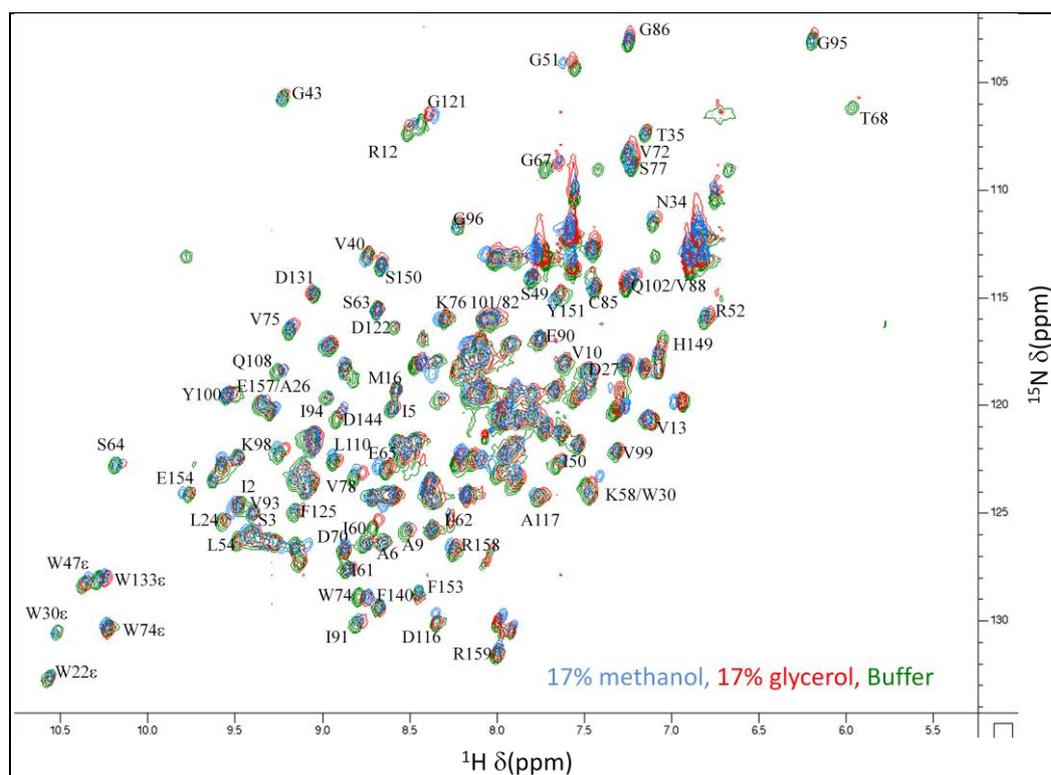
#### **4.7 Analysis of the chemical shift perturbations of EcDHFR: NADP<sup>+</sup>: folate in the presence of co-solvents**

Chemical shift perturbation (CSP) is a commonly implemented analysis technique for studying ligands or small molecules binding to proteins or RNAs. As the chemical shift is very sensitive to the local electronic environments of surrounding nuclei, areas of structural changes and can be indicated quickly and accurately (with respect to the residue) by consideration of two dimensional HSQCs.<sup>144</sup> Whilst this type of analysis is typically used to study binding interactions to a protein, it will also serve as a useful tool to investigate any changes induced in the structure of the protein as a consequence of the addition of the methanol and glycerol co-solvents at 17% by final volume concentrations.<sup>145</sup>

The average chemical shifts for each atom of the protein under the three different solution conditions were exported from ccpNMR Analysis along with the standard deviation for each assigned atom. The chemical shifts of the protein in buffer only were considered to be the reference point from which any perturbation induced by the presence of 17% glycerol or 17% methanol was measured. The chemical shifts of the protein in 17% glycerol and 17% methanol were deducted from the chemical shifts of the equivalent atoms of the protein in buffer only to yield the chemical shift perturbation for each assigned atom. The chemical shift differences (ppm) between the protein in 17% methanol and buffer and, 17% glycerol and buffer were plotted for the various assigned atoms in each residue.

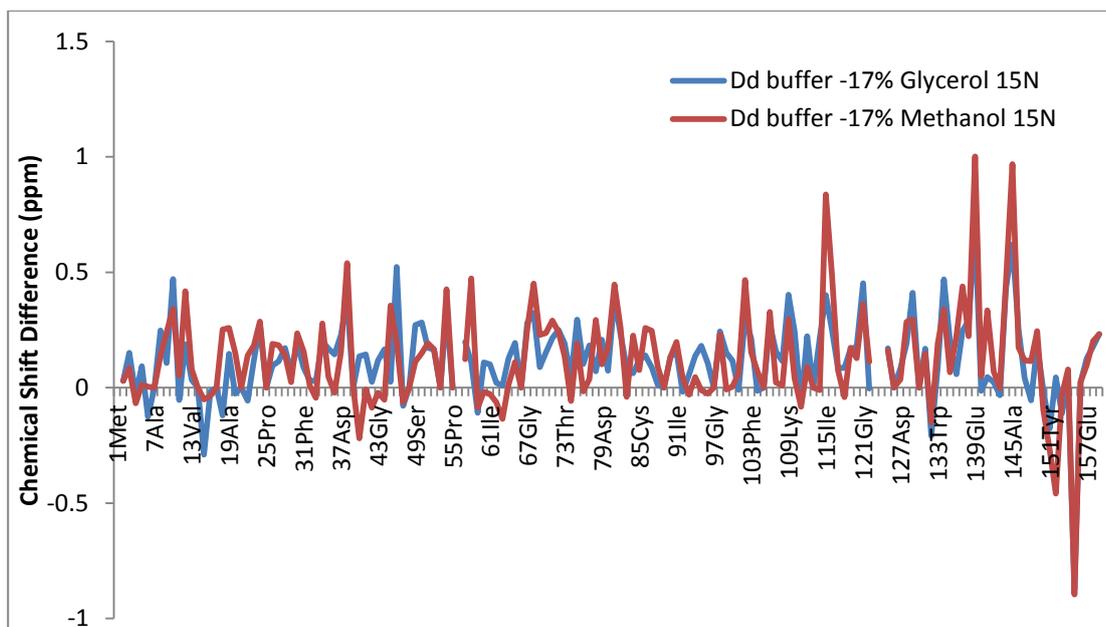
#### 4.7.1 Backbone Amide $^1\text{H}$ $^{15}\text{N}$ Chemical Shift Perturbations

The  $^1\text{H}$ - $^{15}\text{N}$ -HSQC spectra of the EcDHFR complex under the buffer and co-solvent conditions were combined to reveal any obvious movement in chemical shifts due to the presence of the co-solvents (**Figure 24**). All three  $^1\text{H}$  - $^{15}\text{N}$ HSQC spectra were referenced with DSS ( $^1\text{H} = 0.00$  ppm) and the referencing was propagated to  $^{15}\text{N}$  and  $^{13}\text{C}$  in Bruker TopSpin. Residues clearly showing the largest movements of amide peaks are labelled in the spectrum below and the comprehensive list of chemical shift assignments can be found in (**Appendix 8.1**). Perturbation of both  $^1\text{H}$  and  $^{15}\text{N}$  resonances are observed in the presence of the co-solvents and it can be seen that for residues that do show chemical shift perturbations, that the presence of 17% methanol (blue) induces larger shifts in ppm than 17% glycerol (red). These observations are concurrent with the previous observations of EcDHFR in the presence of glycerol and methanol.<sup>129</sup> The chemical shift perturbations were calculated in order to quantify the effect of the co-solvents on the backbone amide resonances.



**Figure 24**  $^1\text{H}$ - $^{15}\text{N}$ -HSQC spectrum showing amide chemical shifts for the EcDHFR:folate:NADP<sup>+</sup> complex in the presence of 17% methanol (blue), 17% glycerol (red) and buffer only (green), all samples references to DSS at 0.00 ppm.

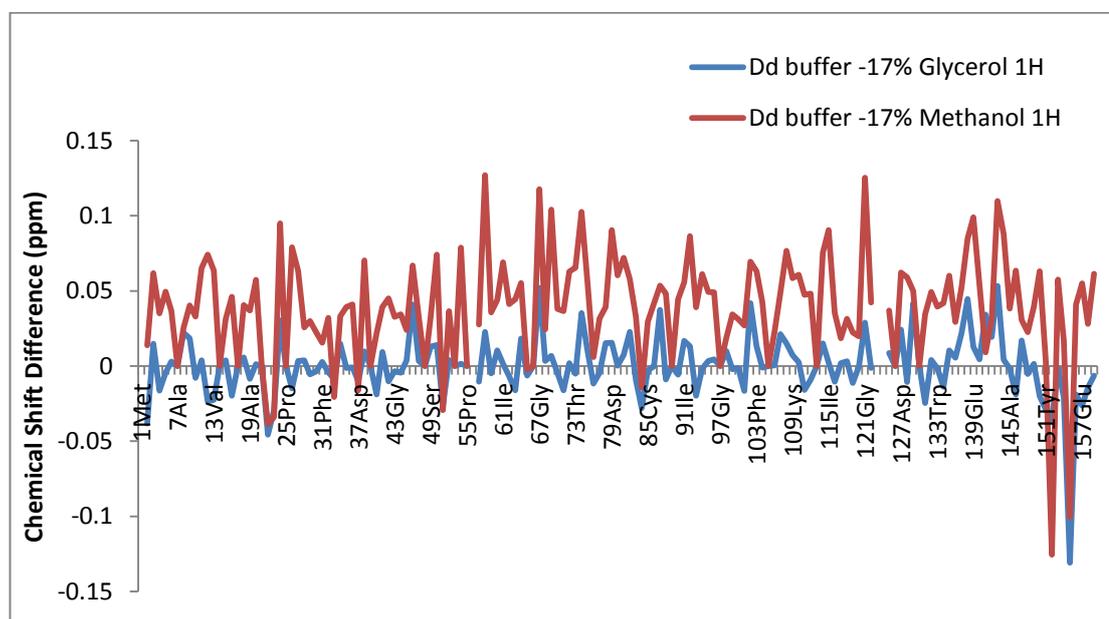
It is known that the chemical shifts of backbone  $^{15}\text{N}$  atoms in proteins are affected by a variety of factors including hydrogen bonding, backbone conformational effects, the amino acid type and the side chain conformation of the residue itself.<sup>145</sup> Additionally, the chemical shift is affected by the backbone conformation of the amino acid and the  $\psi$  angle of the preceding amino acid residue.<sup>146</sup> The exact individual contributions to the  $^{15}\text{N}$  chemical shifts of these factors is relatively unknown, it has been suggested that an additive model may explain contributions to the chemical shift and application of density functional theory (DFT) is ongoing to aid understanding.<sup>147, 146</sup> These combined effects can cause secondary chemical shifts of up to 22ppm in  $^{15}\text{N}$  and therefore the chemical shift perturbations observed in the presence of the co-solvents here are small in comparison (**Figure 25**).<sup>145</sup> The analysis of  $^{15}\text{N}$  amide chemical shift perturbations in the presence of 17% glycerol revealed chemical shift differences ( $\Delta\delta$ ) of up to 0.87ppm. The analysis of  $^{15}\text{N}$  amide chemical shift perturbations in the presence of 17% methanol revealed chemical shift differences ( $\Delta\delta$ ) of up to 1.00 ppm. Overall, the chemical shift perturbations in the presence of 17% methanol are greater than for the changes in the presence of 17% glycerol. The backbone amide  $^{15}\text{N}$  atoms that show the greatest perturbations are in the same regions in both 17% methanol and 17% glycerol co-solvent conditions. A DFT study has suggested that  $^{15}\text{N}$  chemical shifts are most sensitive to hydrogen bonding and that slight changes in hydrogen bonding length can result in chemical shift differences less than 1 ppm.<sup>146</sup> The largest perturbations are seen at the C terminus which may be a reflection that the C terminus is more flexible which leads to broader peaks in the NMR spectra and therefore larger error or may indicate that the change to the solvent medium has the greatest impact on solvent exposed, disordered parts of the protein.<sup>148</sup> The absence of any particularly large perturbations in loop regions however, implies this may not be the case.



**Figure 25** Amide  $^{15}\text{N}$  Chemical Shift Differences between the EcDHFR: Folate:  $\text{NADP}^+$  protein complex in the buffer only solvent conditions and the 17% Methanol and 17% Glycerol solvent conditions.

The origins of chemical shift in backbone  $^1\text{H}_\text{N}$  atoms are different to  $^{15}\text{N}$ . In folded proteins, backbone amide protons range from 6.5 – 9.5 ppm whereas in unfolded proteins, the dispersion is much lower. Random coil chemical shifts describe the characteristic chemical shifts for amino acid types in the absence of any secondary or tertiary structural preferences.<sup>136</sup>  $^{136}$   $^1\text{H}_\text{N}$  chemical shifts are highly dependent through bond effects, secondary structure and whether the proton is involved in hydrogen bonding.<sup>149</sup> The effect that the external environment has on the proton including effects from paramagnetic effects, magnetic anisotropy, aromatic currents and electric fields from charged atoms is often referred to as the secondary chemical shift. Chemical shifts are summative and so these secondary chemical shift contributions “add on” to the random coil chemical shift and are representative of the structural environment of atom in the amino acid. The random coil chemical shifts all 20 amino acids have been determined.<sup>150</sup> It has been determined that the largest  $\text{H}_\text{N}$  chemical shifts arise from hydrogen bonding interactions but it is difficult to quantify the effects due to little being known regarding hydrogen bonding interactions with the solvent.<sup>151</sup> The chemical shift perturbations caused by the presence of 17% glycerol are generally much smaller than those caused by the presence of 17% methanol and are not much greater than 0.02 ppm, with a few exceptions (**Figure 26**). The backbone amide protons of certain residues in 17% methanol show much greater perturbations on average of ppm. Overall, residues

which show perturbations in the chemical shift of the amide proton show similar trends in the relative values in the presence of both co-solvents.



**Figure 26** Amide  $^1\text{H}$  Chemical Shift perturbations in the EcdHFR: folate:  $\text{NADP}^+$  complex caused by 17% Methanol and 17% Glycerol solvent conditions.

#### 4.7.2 Weighted chemical shift Perturbations

The chemical shifts of the amide group can be analysed individually or in combination. It is common to calculate the average Elucidate distance moved by the amide peak in the  $^{15}\text{N}$ - $^1\text{H}$ -HSQC by considering the weighted chemical shift change (**d**). The weighted chemical shift change is the square root of the sum of the individual chemical shift change of  $^1\text{H}$  and the scaled chemical shift change of  $^{15}\text{N}$  in ppm (**Equation 23**).<sup>145</sup>

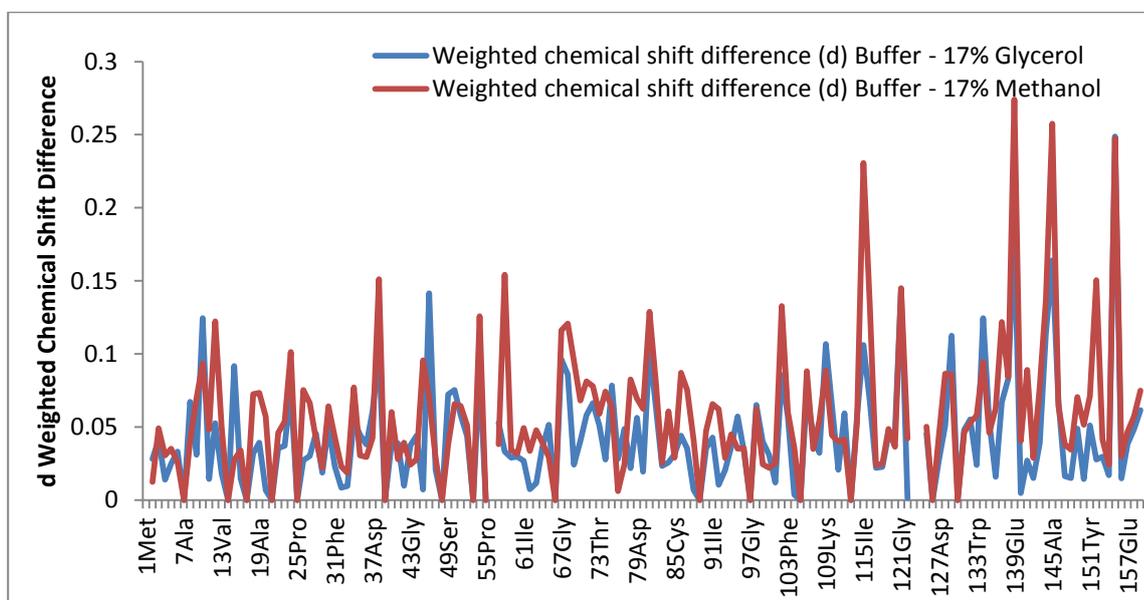
$$d = \sqrt{1/2[(\delta_H^2 + (\alpha \cdot \delta_N^2)]}$$

**Equation 23** Calculation of weighted chemical shift values for amino acid amide groups.<sup>145</sup>

The optimal value of the weighting factor ( $\alpha$ ) for the  $^{15}\text{N}$  chemical shift change has been greatly discussed particularly because the origins of  $^1\text{H}$  and  $^{15}\text{N}$  chemical shifts are

different, so there is no clear theory to provide a weighting. It is commonly accepted that  $\alpha = 0.14$  for all amino acids, with the exception of glycine where  $\alpha = 0.2$ . Glycine is treated differently because the spread of chemical shifts is different than for other amino acids.<sup>152</sup>

The  $^{15}\text{N}$ - $^1\text{H}$ -amide weighted chemical shift differences for EcDHFR: folate:  $\text{NADP}^+$  protein complex in the buffer only solvent conditions and the 17% methanol and 17% glycerol solvent condition were calculated and plotted (**Figure 27**). The individual chemical shift differences for each of the nuclei were analysed to determine the individual contributions of the  $^1\text{H}$  and  $^{15}\text{N}$  chemical shift differences to the weighted chemical shift (**Figure 25**) and (**Figure 26**).



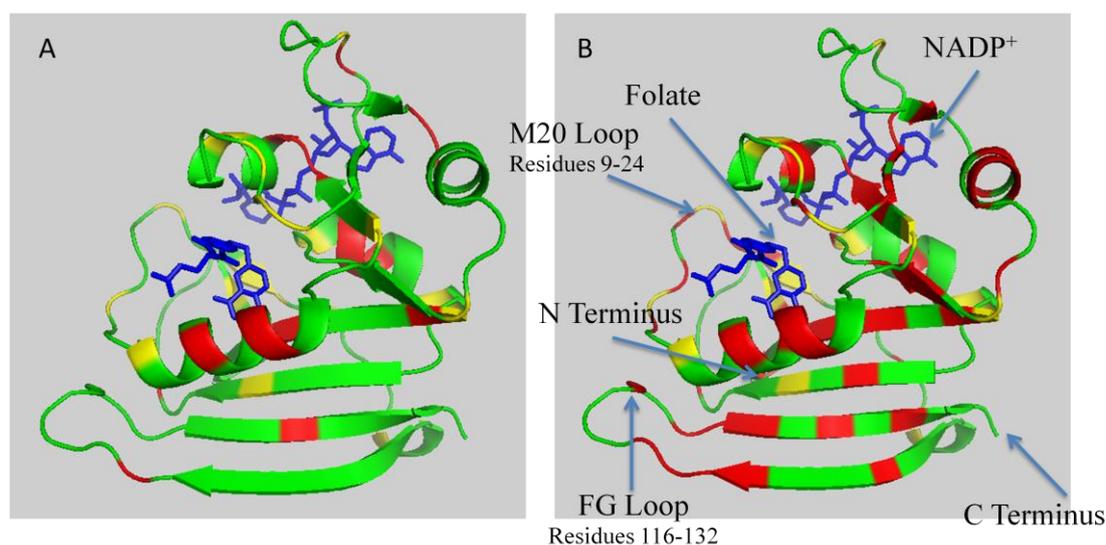
**Figure 27**  $^{15}\text{N}$ - $^1\text{H}$ -Amide Weighted Chemical Shift Differences (d) between the EcDHFR: folate:  $\text{NADP}^+$  protein complex in the buffer only solvent conditions and the 17% Methanol and 17% Glycerol solvent conditions.

The weighted chemical shift perturbations for the protein in 17% glycerol and 17% methanol were plotted to reveal similar trends for both solvent conditions. Certain residues do show changes in  $^1\text{H}$  and  $^{15}\text{N}$  chemical shift and weighted chemical shift (d) of differing magnitude at different locations in the polypeptide chain. Calculation of the weighted chemical shifts partially conceals that the root of the larger perturbations caused by the addition of 17% methanol is in the larger perturbations caused to the amide proton. Particularly large perturbations are seen at C terminus residues in

addition to Ile 115 which is located in the same  $\beta$  sheet as the C- terminus. It is again likely that these large perturbations are caused by the more flexible residues becoming more solvent exposed.

### 4.7.3 Chemical shift perturbation mapping

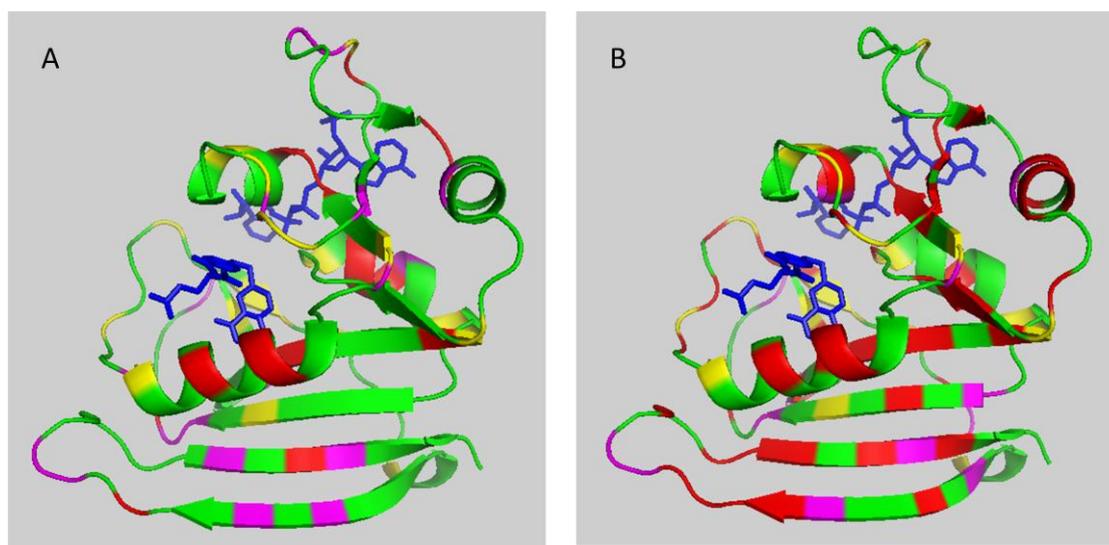
In order to determine whether there is any spatial correlation between the residues which show the greatest amide group perturbations, the location of the residues most perturbed were highlighted on the crystal structure of the protein (1RX2.pdb). The combined standard deviation for each of for the weighted amide chemical shift differences was calculated. Residues displaying weighted amide chemical shift perturbations greater than the combined standard deviation were mapped on to the crystal structure in green; residues for which the weighted chemical shifts were less than the standard deviation were depicted red and yellow represents residues for which there was no data available (**Figure 28**). The comprehensive list of all atoms in the protein displaying chemical shift perturbations greater than the standard deviation is included (**Appendix 8.2**).



**Figure 28** Residues in the EcDHFR: Folate: NADP<sup>+</sup> protein complex that show weighted <sup>15</sup>N-<sup>1</sup>H chemical shift perturbations greater than the standard deviation on the addition of co-solvent 17% methanol (A) and 17% glycerol (B).

The addition of 17% methanol to the solution in which the protein resides results in  $^{15}\text{N}$ - $^1\text{H}$  weighted chemical shift perturbations greater than the standard deviation in 124 of the 159 residues. No backbone amide  $^{15}\text{N}$ - $^1\text{H}$  chemical shift perturbations were calculated for 17 of the 159 protein residues, of these 10 are the proline residues and one is the N terminal methionine residue. The remaining residues for which there was no data (7 Ala, 14 Ile, 17 Gly, 48 Glu, 97 Gly and 113 Thr) did not show peaks in one or both of the  $^{15}\text{N}$ - $^1\text{H}$ -HSQC spectra.

The addition of 17% glycerol to the solution in which the protein resides gave rise to perturbations greater than the standard deviation in 85 residues. No data was available for 20 of the 159 residues including the 10 prolines and 1 N terminal methionine. The remaining residues for which no data was found (7 Ala, 14 Ile, 17 Glu, 48 Glu, 56 Gly, 97 Gly, 113 Thr, 123 Thr, 124 His) did not show peaks in one or more of the HSQC spectra. Residues for which the weighted chemical shift perturbations (d) were greater than 0.1 were determined and mapped on to the crystal structure in magenta to indicate their spatial location for both 17% glycerol and 17% methanol (**Figure 29**).



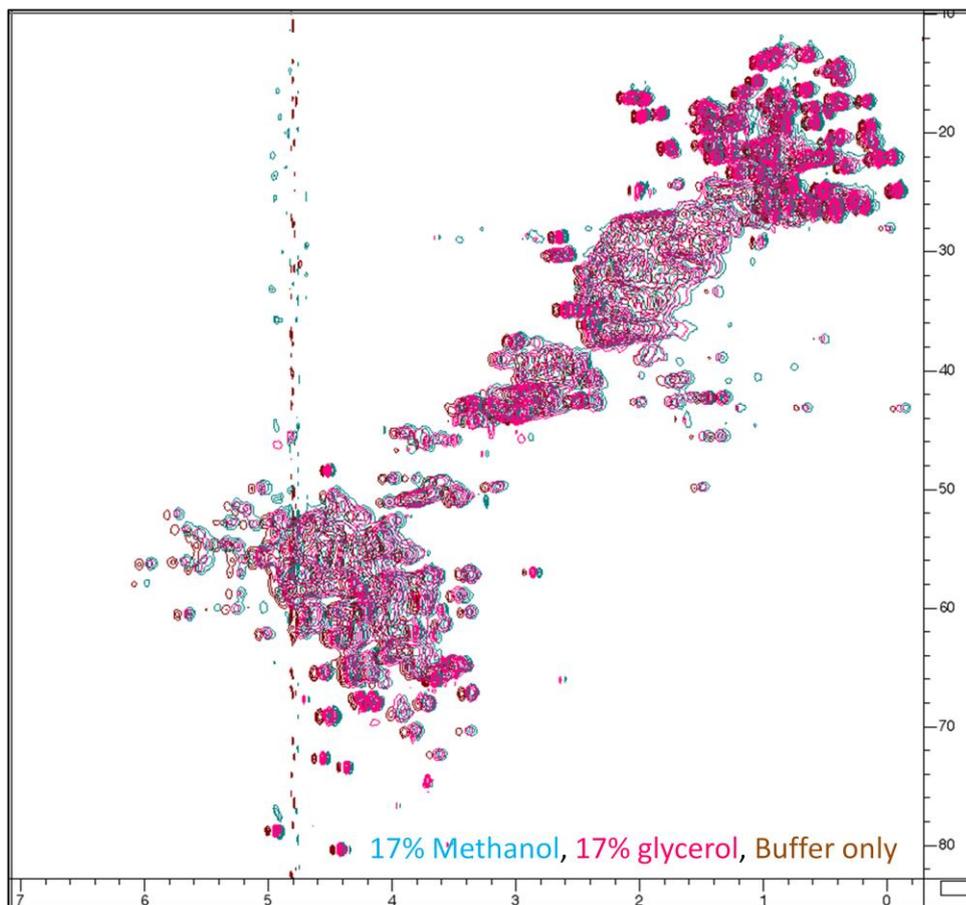
**Figure 29** Residues in the EcdHFR: Folate:  $\text{NADP}^+$  protein complex that show weighted  $^{15}\text{N}$ - $^1\text{H}$  chemical shift perturbations greater than 0.1 ppm on the addition of co-solvent 17% Methanol (A) and 17% Glycerol (B) are coloured magenta.

Figure 29 (A), the Michaels complex mimic in the presence of 17% methanol shows slightly more amide chemical shift perturbations greater than the standard deviation than the complex in the presence of 17 % glycerol. Amide chemical shift perturbations

of this size are likely due to slight changes in the hydrogen bonding of the protein, which due to the high dispersion of perturbations across the protein could be a result of non-specific bulk solvent effects. Significant chemical shift perturbations occur mainly in solvent exposed regions, within loops or the bottom beta sheet strands for both methanol and glycerol. Significantly the chemical shifts are relatively small and likely to occur due to perturbations to hydrogen bonds.

## 4.8 C $\alpha$ and H $\alpha$ Chemical Shift perturbations

### 4.8.1 Chemical shift Perturbations



**Figure 30**  $^1\text{H}$ - $^{13}\text{C}$ -HSQC of EcdHFR: folate:  $\text{NADP}^+$  protein complex in the presence of 17% methanol co-solvent (blue), 17% glycerol co-solvent (pink) and buffer only (brown). All spectra were referenced to DSS at 0.00 ppm.

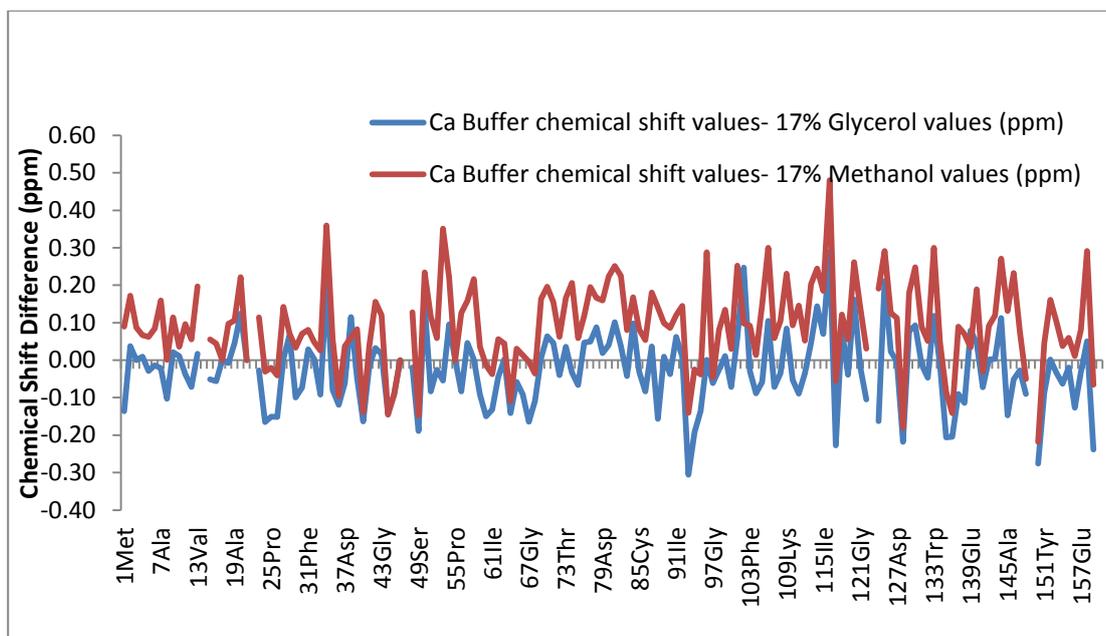
The  $^1\text{H}$ - $^{13}\text{C}$ -HSQC of the EcdHFR: folate:  $\text{NADP}^+$  complex in the presence of the co-solvents shows some clear perturbations for proton chemical shifts, particularly in the methyl (~2.0-0.0 ppm) and aromatic (5.5-6.5 ppm) regions which as discussed above, depend on a range of external factors but predominantly ring current effects which can, for example, considerably extend the H $\alpha$  chemical shift range well outside the normal 3.5 -5.5 ppm (**Figure 30**).

At first glance, there does not appear to be much deviation in the carbon chemical shifts in the presence of the co-solvents. It should however, be noted that carbon chemical

shifts in proteins cover a large range from between ~75 – 10 ppm for backbone and side chain resonances and 140-110 ppm for carbonyl side chain and backbone resonances.<sup>141</sup> The range of carbon chemical shifts is very large in comparison to proton chemical shifts and therefore the secondary effects that contribute significantly to proton chemical shifts do not alter carbon chemical shifts to the same degree. External effects contribute less to carbon chemical shifts than in protons as carbon atoms tend to be less surface exposed.

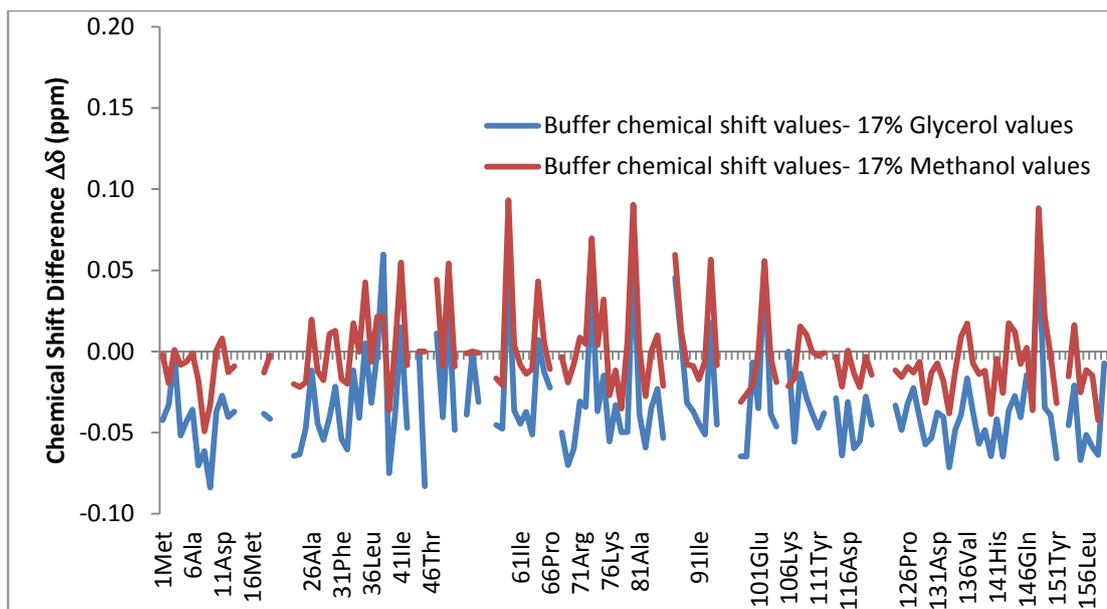
C $\alpha$  and C $\beta$  chemical shifts are also affected little by neighbouring residue types due to their position in the centre of the amino acid. Unlike proton chemical shifts, carbon secondary chemical shifts are mostly dependent on the intraresidue dihedral angle, which can be attributed to the presence of electrons in hybridised orbitals.<sup>145</sup> Backbone conformation can affect the secondary chemical shift by around 4 ppm, side chain conformation by 0.6 ppm and hydrogen bonding by 0.9 ppm. With these factors in mind, it is unlikely that any significant perturbations would be observed for carbon chemical shifts particularly unless significant structural rearrangement occurs.<sup>145</sup> In order to quantify any backbone and side chain chemical shift perturbations caused by the presence of co-solvents, the chemical shift differences were plotted.

The C $\alpha$  chemical shift values of the protein in 17% methanol and 17% glycerol were subtracted from the C $\alpha$  chemical shifts of the protein in buffer only. Therefore, values of chemical shift difference ( $\delta\Delta\text{ppm}$ ) (**Figure 31**) less than zero represent that the chemical shift perturbation caused by the co-solvent is downfield and values of the chemical shift difference greater than zero represent that chemical shift perturbations caused by the co-solvent are upfield. The presence of 17% glycerol causes chemical shift perturbations that are overall more downfield than those of methanol.



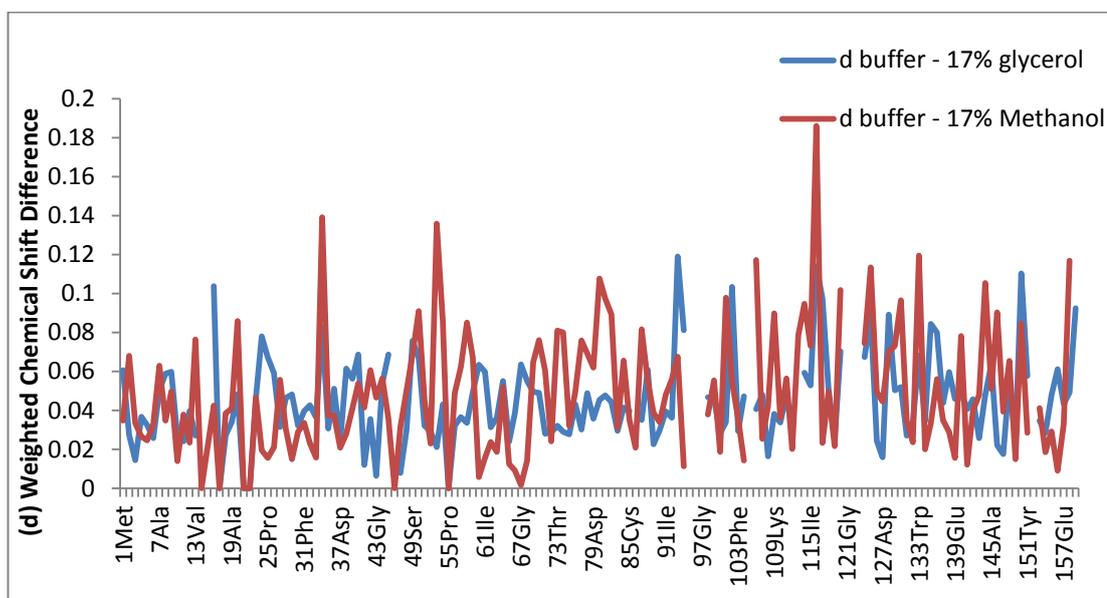
**Figure 31**  $\alpha$  Chemical Shift Differences between the EcdHFR: Folate:  $\text{NADP}^+$  protein complex in the buffer only solvent conditions and the 17% methanol and 17% glycerol solvent conditions.

The chemical shift perturbations caused by the presence of 17% glycerol are small for the majority of residues and slightly larger and more upfield in the presence of 17% methanol. The absence of any significant changes to the chemical shift implies little or no change to the secondary structure of the protein is induced by addition of the co-solvents however in the case of residues showing larger perturbations  $\sim 0.4$  ppm, changes to the conformation of side-chains cannot be ruled out and will be discussed further in the chapter (**Figure 31**). Carbon is relatively insensitive to changes in hydrogen bonding, displaying changes in chemical shift below 1 ppm, small changes in hydrogen bonding cannot be ruled out.<sup>146</sup> The chemical shift differences for  $\text{H}\alpha$  were plotted in the same manner as  $\text{C}\alpha$  below (**Figure 32**) to reveal similar trends in perturbations caused by 17% glycerol and 17% methanol but overall large perturbations are caused by the presence of the glycerol co-solvent, compared to the methanol co-solvent. The maximum chemical shift difference is around 0.08 ppm, which is again quite small and implies that no major changes in structure are occurring, instead these changes are likely a result of secondary chemical shift effects, possible due to slight changes in the charge density of the protons caused by the changes in dielectric constant of the solvent medium due to the presence of the co-solvents .



**Figure 32**  $H\alpha$  Chemical Shift Differences between the EcDHFR: folate:  $NADP^+$  protein complex in the buffer only solvent conditions and the 17% methanol and 17% glycerol solvent conditions.

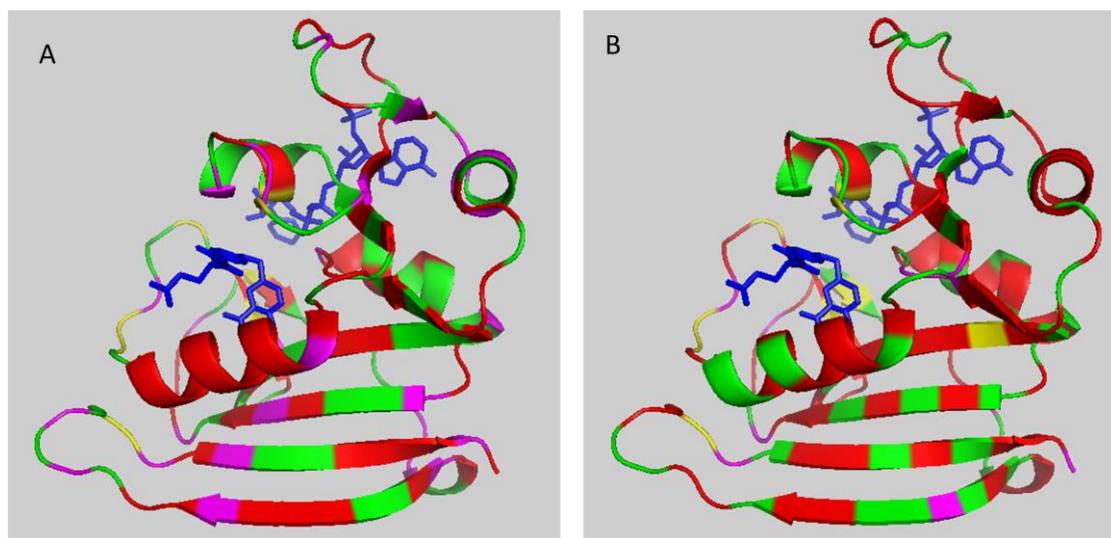
The  $^{13}C\alpha$ - $^1H\alpha$  weighted chemical shift perturbations (d) were calculated in a similar manner to the amide perturbations (**4.7.2 Weighted chemical shift Perturbations**). The optimal value of the weighting factor ( $\alpha$ ) for the  $^{13}C\alpha$  chemical shift is 0.3 based on the chemical shift range.<sup>145</sup> The weighted chemical shift perturbations are show below (**Figure 33**).



**Figure 33**  $^{13}C\alpha$ - $^1H\alpha$  weighted chemical shift differences between the EcDHFR: folate:  $NADP^+$  protein complex in the buffer only solvent conditions and the 17% methanol and 17% glycerol solvent conditions.

## 4.8.2 Chemical shift perturbation mapping

Residues in the EcDHFR: folate: NADP<sup>+</sup> protein complex that show weighted <sup>13</sup>Cα-<sup>1</sup>Hα chemical shift perturbations greater than the standard deviation upon the addition of co-solvents were highlighted on the crystal structure to give a spatial representation of the perturbations (**Figure 34**). Residues displaying perturbations in weighted <sup>13</sup>Cα-<sup>1</sup>Hα chemical shifts greater than the standard deviation in the presence of co-solvent are coloured green. Residues for which the weighted <sup>13</sup>Cα-<sup>1</sup>Hα chemical shift perturbation were calculated to be less than or equal to the standard deviation are coloured red. Residues for which no data was available are depicted yellow and residues showing weighted chemical shift perturbations greater than 0.1 (magenta). The comprehensive list of all atoms in the protein displaying chemical shift perturbations greater than the standard deviation is included (**Appendix 8.2**).



**Figure 34** Residues in the EcDHFR: folate: NADP<sup>+</sup> protein complex that show chemical shift perturbations (weighted <sup>13</sup>Cα-<sup>1</sup>Hα) upon the addition of 17% methanol co-solvent (A) and 17% glycerol co-solvent (B).

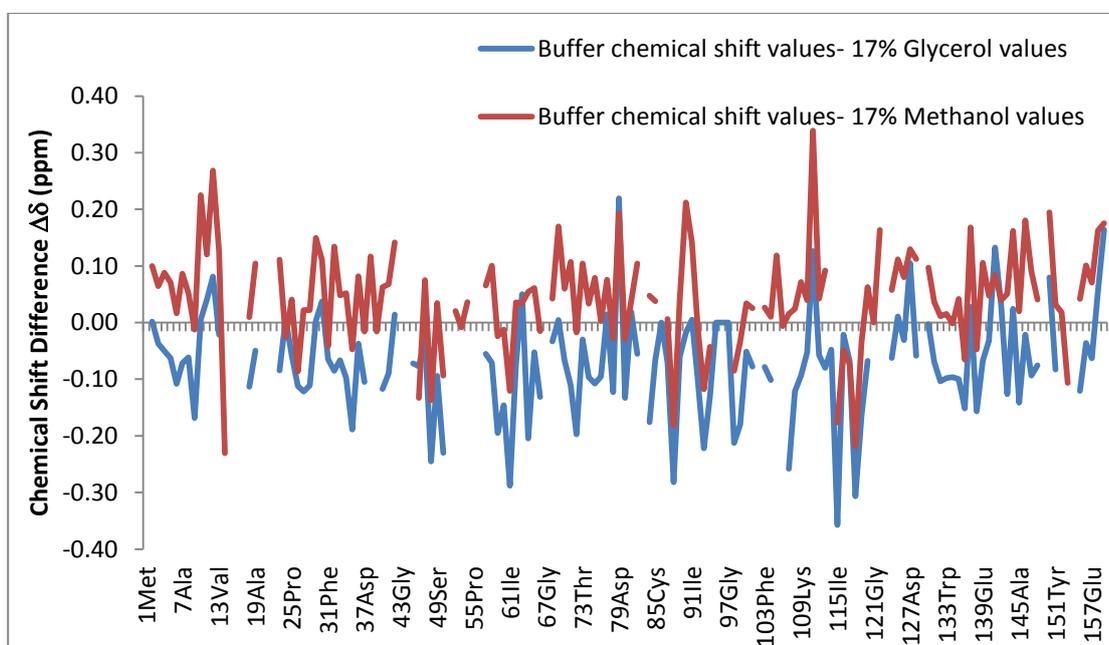
The regions of the protein affected by the two different solvent conditions are broadly similar however it can be seen that the methanol co-solvent induces more, larger perturbations, particularly in the central β-sheet region of the protein complex and the β-turn region. In general, all regions of the protein show small changes in the presence

of both co-solvents. From viewing Figure 34, ligand binding areas show the least perturbations under both co-solvent conditions. It is likely that the ligand binding regions are preserved to a high degree, which is unsurprising as solvent is excluded from the active site of the protein and because these regions are involved in electrostatic interactions between the substrate and co-factor. It is equally plausible that these regions also show very small chemical shift perturbations that are not visible due to the error or that they show no chemical shift changes at all, given that the value of perturbations are only as accurate as the error.

## 4.9 C $\beta$ and H $\beta$ Chemical Shifts

### 4.9.1 H $\beta$ and H $\beta$ Chemical Shift Perturbations

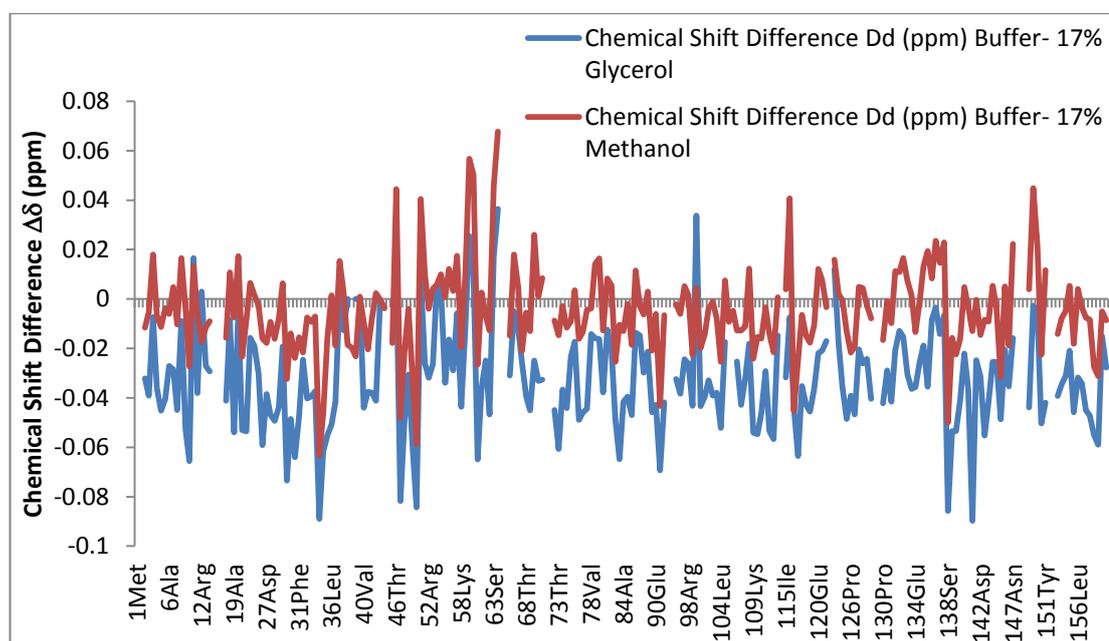
In proteins, C $\beta$ , like C $\alpha$  chemical shifts are mostly affected by changes to protein side chain conformation, backbone conformations and hydrogen bonding.<sup>145</sup> The chemical shift perturbations of the C $\beta$  atoms in the Michaelis complex mimic of EcdHFR were plotted (**Figure 35**) revealing similar differences in chemical shift caused by the addition of 17% methanol and 17% glycerol. The magnitude chemical shift changes caused by the presence of glycerol are greater than those caused by methanol, in general.



**Figure 35** C $\beta$  Chemical Shift Differences between the EcdHFR: folate: NADP<sup>+</sup> protein complex in the buffer only solvent conditions and the 17% methanol and 17% glycerol solvent conditions.

A similar trend is seen for the H $\beta$  atoms, for which most atoms in the presence of 17% methanol show very small chemical shift perturbations with the exception of H $\beta$ 's in around 10 residues, the majority of perturbations are less than 0.02 ppm, which are within the general error expected (**Figure 36**). A significant number of larger perturbations (0.04- 0.08 ppm) are observed in the presence of glycerol, which are

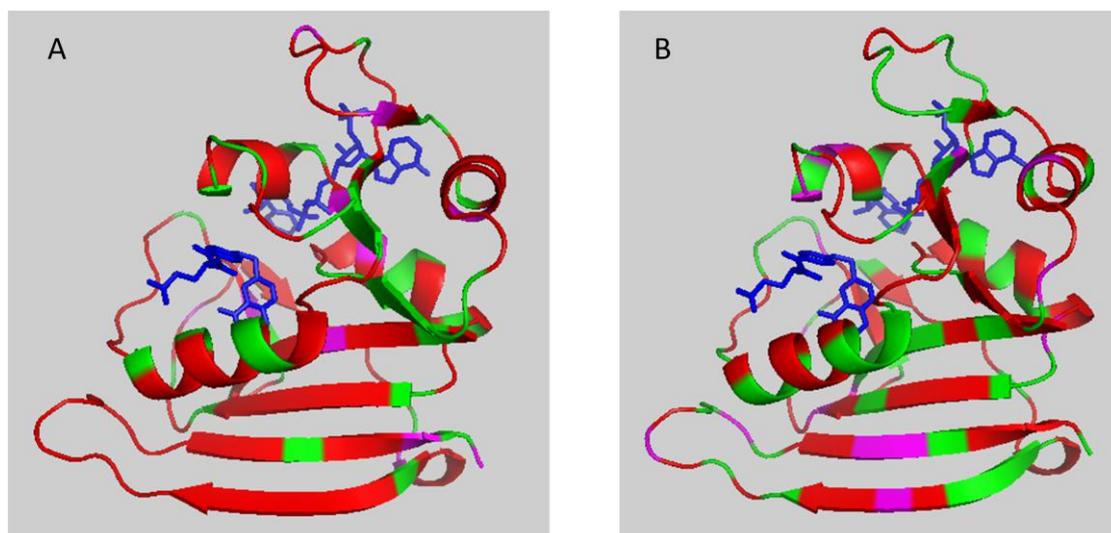
outside the expected error margin and strongly indicate these side chain atoms are affected by the addition of co-solvents.



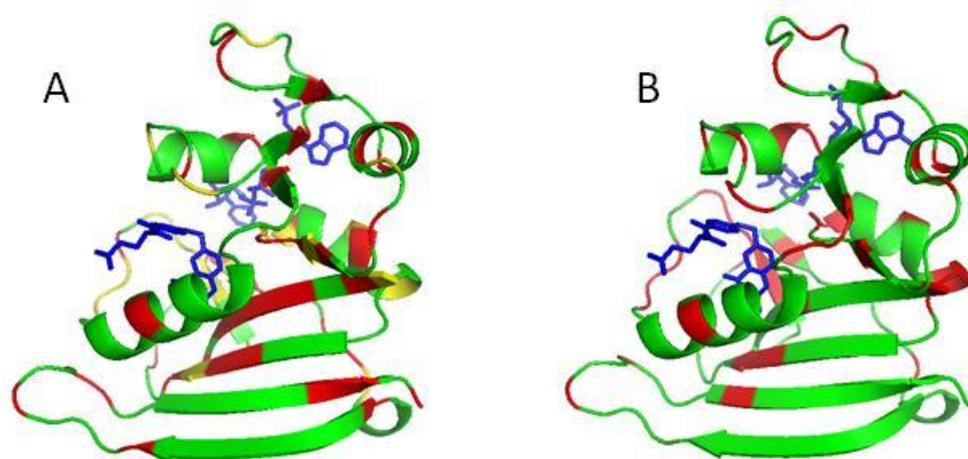
**Figure 36** H $\beta$  Chemical Shift Differences between the EcDHFR: folate: NADP<sup>+</sup> protein complex in the buffer only solvent conditions and the 17% methanol and 17% glycerol solvent conditions.

#### 4.8.2 C $\beta$ and H $\beta$ Chemical shift mapping

Residues in the EcDHFR: folate: NADP<sup>+</sup> protein complex that show <sup>13</sup>C $\beta$  chemical shift perturbations greater than the standard deviation upon the addition of co-solvents were highlighted on the crystal structure to give a spatial representation of the perturbations (**Figure 37**). Residues displaying perturbations in <sup>13</sup>C $\beta$  chemical shifts greater than the standard deviation in the presence of co-solvent are coloured green. Residues for which the <sup>13</sup>C $\beta$  chemical shift perturbation were calculated to be less than or equal to the standard deviation are coloured red. Residues for which no data was available are depicted yellow and residues showing chemical shift perturbations greater than 0.1 (magenta). The comprehensive list of all atoms in the protein displaying chemical shift perturbations greater than the standard deviation is included (**Appendix 8.2**).

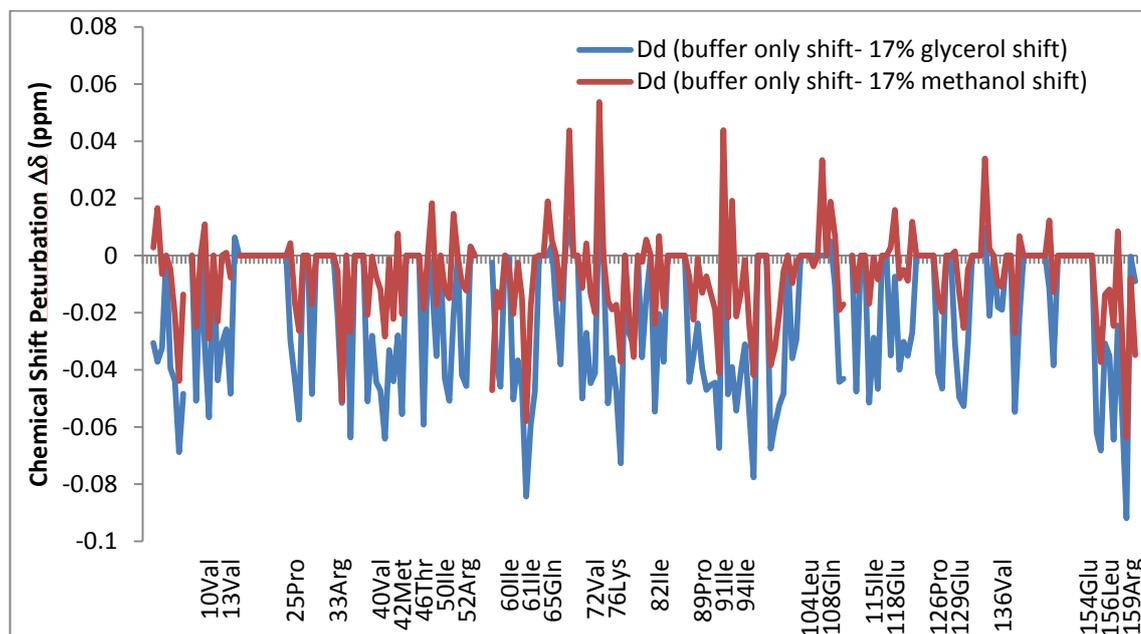


**Figure 37** Residues in the EcDHFR: folate: NADP<sup>+</sup> protein complex that show chemical shift perturbations (<sup>13</sup>Cβ) upon the addition of 17% methanol co-solvent (A) and 17% glycerol co-solvent (B).

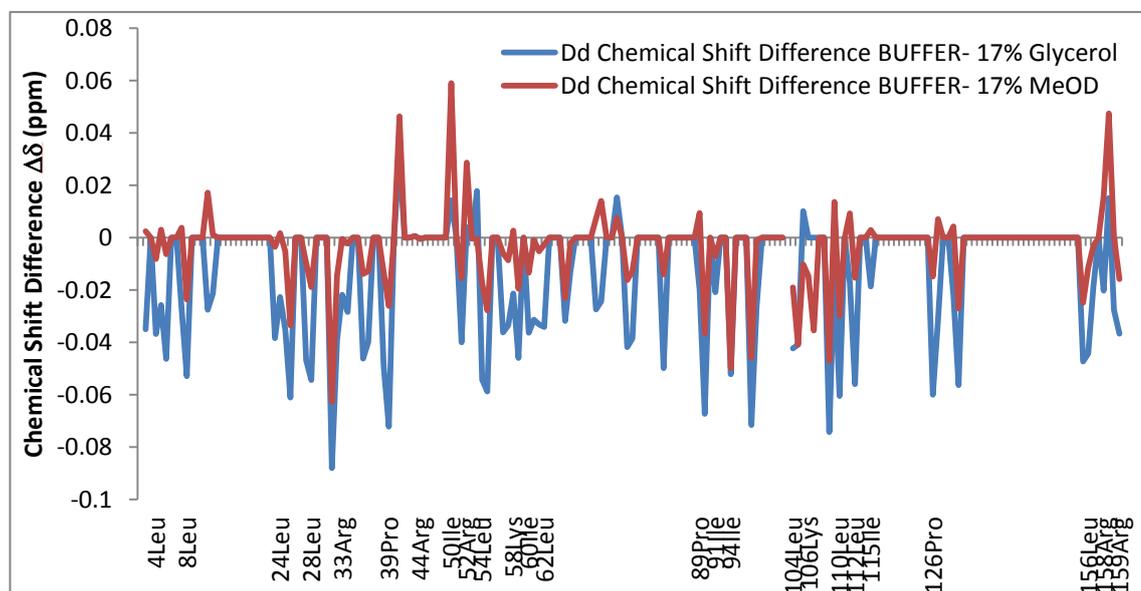


**Figure 38** Residues in the EcDHFR: folate: NADP<sup>+</sup> protein complex that show chemical shift perturbations (<sup>1</sup>Hβ) upon the addition of 17% methanol co-solvent (A) and 17% glycerol co-solvent (B).

## 4.9 Side chain Chemical Shift Perturbations



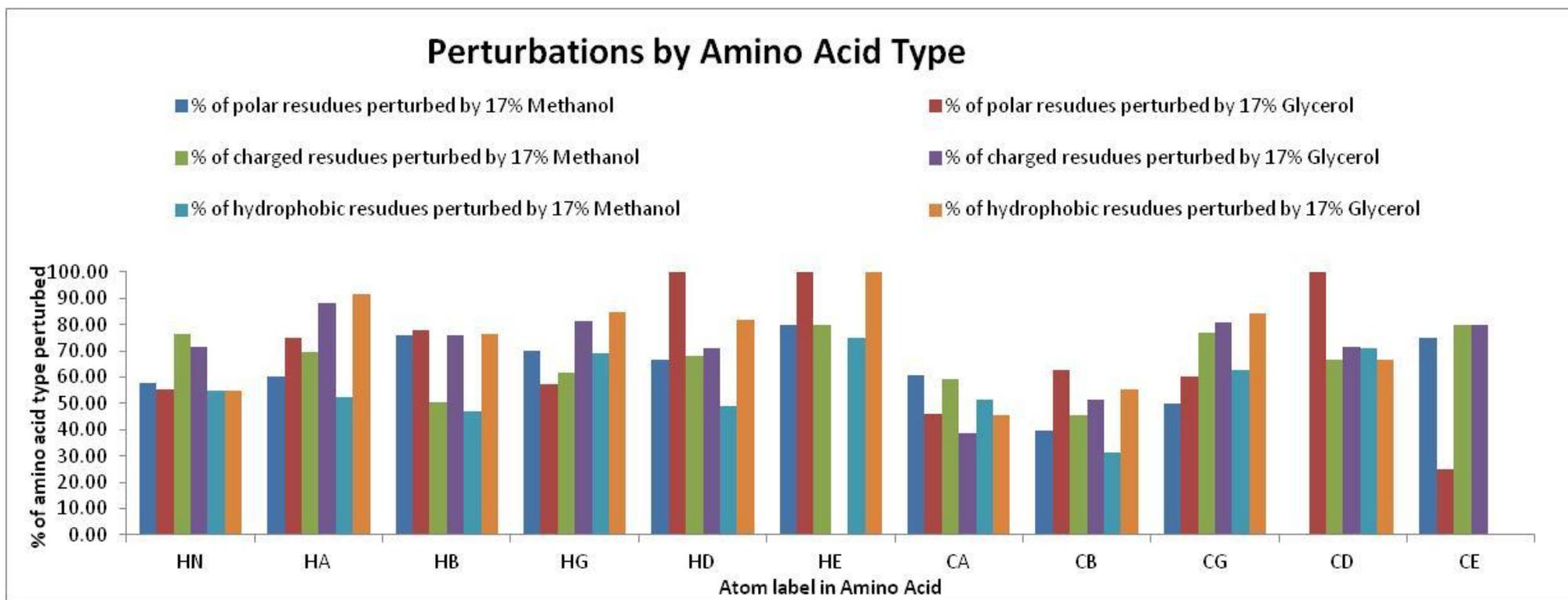
**Figure 39** H $\gamma$  Chemical Shift Differences between the EcdHFR: folate: NADP<sup>+</sup> protein complex in the buffer only solvent conditions and the 17% methanol and 17% glycerol solvent conditions.



**Figure 40** H $\gamma$  Chemical Shift Differences between the EcdHFR: folate: NADP<sup>+</sup> protein complex in the buffer only solvent conditions and the 17% methanol and 17% glycerol solvent conditions.

## 4.10 Chemical shift perturbations across the whole protein

NMR measurable atoms in the EcDHFR: folate: NADP<sup>+</sup> protein complex that showed weighted chemical shift perturbations greater than the standard deviation upon the addition of co-solvents are presented in a series of tables (**Appendix 8.2**) which are derived from a spreadsheet allowing the viewer to assess any trends in chemical shift perturbation caused by the addition of co-solvent with respect to the hydrogen bonding status of the residue, amino acid properties (polar, charged or hydrophobic), amino acid type and secondary structure type. Atoms displaying perturbations in weighted <sup>13</sup>C $\alpha$ -<sup>1</sup>H $\alpha$  chemical shifts greater than the standard deviation in the presence of co-solvent are coloured green. Atoms for which the weighted chemical shift perturbation were calculated to be less than or equal to the standard deviation are coloured red. Atoms for which no data was available are depicted yellow. It should be noted well that the values in each cell (**Appendix 8.2**) represent the root squared magnitude of the chemical shift perturbation in ppm minus the combined standard deviation of the chemical shift values hence, positive values imply a chemical shift perturbation greater than the standard deviation and negative values imply the chemical shift perturbation was less than the standard deviation.



**Figure 41** Chemical shift perturbations by amino acid type. The amino acid residues in EcdHFR: NADP<sup>+</sup>: folate complex were categorised according to whether they are hydrophobic, charged or polar. The percentage of atoms perturbed by the addition of 17% methanol and 17% glycerol in each residue type is reported above.

## 4.11 Discussions

The native cell environment of *E. coli* consists of large amounts of protein, RNA and DNA at 200-320 mg/ mL, 75-120 mg/mL and 11-18 mg/ mL respectively.<sup>153</sup> The intrinsic viscosity of the cell in which enzymes normally functions is therefore, far different from the standard buffer conditions, frequently used to studying enzyme catalysis in vitro. Protein solutions used to measure NMR chemical shift data and dynamics often only include standard buffer components necessary to maintain the pH and the integrity of the sample for acquisition, and therefore it might be anticipated that these do not truly mimic the effects of the cellular environment on enzyme catalysis. As such, the chemical shifts of the protein under solely buffer conditions can only represent a reference point from which viscosity and electrostatic attempts can be measured. Within this chapter the effect of different solutes (glycerol and methanol) on the chemical shifts of EcDHFR was probed for a number of nuclei so to have a better understanding of the effects on protein structure and on the local electrostatic environments.

To accurately consider the effect that the co-solvents may have on the structure of the protein, it is important to consider the behaviour of the water-glycerol and water-methanol solvent mixtures. Alterations to the medium in which the protein resides are known to have various effects on the protein structure. Protein molecules in solution are surrounded by a hydration shell of water molecules, some of which can form hydrogen bonds with the surface of the protein. The presence of organic solvents can cause structural loss to the protein or cause it to denature by disturbing the hydration shell or disrupting the hydrophobic core of the protein.<sup>154</sup>

The dielectric constant ( $\kappa$ ) or relative permittivity ( $\epsilon_r$ ), is the ratio of the absolute permittivity and permittivity in a vacuum and is typically used as a measure of solvent polarity.<sup>155</sup> In physical terms, solvents with a high dielectric constant encourage complete dissociation of electrolytes and solvents with low dielectric constant encourage ion pairing.<sup>156</sup> In terms of the effect that this can have on proteins in solution, reducing the dielectric constant of the solvent strengthens H-bonding networks, making the protein more stable and less flexible.<sup>157</sup>

The molar ratio of methanol to glycerol at 17% by final volume in a sample of a given volume is 1.74:, consequentially methanol is present in a concentration 74% higher than glycerol. In a simple water-glycerol or water-methanol mixture, this corresponds to molar fractions of 0.05 and 0.08, respectively. This difference in concentration of the co-solvents is not insignificant on the microscopic level as it will surely have consequences for the hydrogen bonding network of the bulk solvent. Glycerol is a larger molecule than water or methanol. Research is available on the hydrogen-bonding patterns of glycerol/water mixtures. In mixtures of glycerol and water, at low glycerol concentrations (mole fraction less than 0.15) there are areas of bulk-like water. Upon increasing the concentration of glycerol, the areas of bulk-water are depleted. Water in the first hydration shell becomes concentrated around the polar groups of glycerol and the alkyl groups of glycerol self associate.<sup>158</sup> It has been observed that the spectral characteristics of proteins alter upon addition of a critical concentration of organic solvent; this provides evidence that conformational rearrangements in the protein occur, such as destruction of hydrophobic contacts and formation of new hydrogen bonds. Hydrophilic solvents have a high affinity for water and as such, can strip water from the surface of the protein unless the water-protein interactions are strong enough to prevent this. The hydrophilicity of a solvent is dependent on hydrogen bonding donor-acceptor abilities and the dielectric constant.<sup>159</sup>

It is clear from the data that the presence of co-solvents is having small effects on the chemical shifts of the protein. The small chemical shift perturbations is most likely a result of the change in electrostatic environment due to the presence of the almost isodielectric glycerol and methanol co-solvents. It is unclear whether the difference in the size of the chemical shift perturbations in some areas of the protein can be attributed to the difference in viscosity or whether the relative smaller size of the methanol co-solvent compared to the glycerol co-solvent molecule allows penetration of hydration layers closer to the protein and therefore greater perturbation of chemical shifts through hydrogen bonding interactions. It is even possible, that because these perturbation values are so small, that there is no change in chemical shift for the majority of residues.

## 4.12 Conclusions

The native structure of the protein has the lowest free energy in conformational space and according to the free energy theory, the energy landscape of the protein folding is funnel shaped.<sup>112</sup> The balance of the electrostatics in proteins is so delicate, that small changes can have huge consequences for catalysis. Most of the measurable nuclei show some perturbation in chemical shift, albeit very small in some cases, which could be an indication of larger overall errors or an indication of a global co-solvent effect on the protein. The magnitude of changes are not uniform or localized to particular areas of the protein. For the amide proton and nitrogen, the effects of methanol and glycerol co-solvents are similar but of differing magnitudes. Methanol causes perturbations of larger magnitude for  $^1\text{H}$  and  $^{15}\text{N}$ , which could be due to the smaller size of the methanol molecule compared to the glycerol molecule or the larger mole fraction of methanol.

Methanol co-solvent causes more and larger chemical shift changes to amide groups and  $\text{C}\alpha\text{H}\alpha$  whereas glycerol causes more and larger chemical shift perturbations to side chains, which may agree with the proposal that the effects of viscosity changes are proposed to reach deep inside protein interior, however this is dubious to conclude due to the magnitude of many perturbations being so small. Differences in chemical shift values in both the protein interior and exterior are unlikely caused through direct interaction with co-solvent molecules in all areas but may be due to a conduction effect.

It is not possible to state whether the altered chemical shift values are indicating changes in the overall structure of EcDHFR when co-solvents are present however, NMR solution structures may offer insight in to any specific changes in protein structure in the presence of co-solvents, however it is possible that the changes, if any, are too small to quantify through this technique. If the motions and structure of proteins are governed fully or partly by electrostatics, than the altering the electrostatic medium and consequently, the electrostatics of the protein would be expected to have consequential changes on the dynamics.<sup>160</sup>

## 5.0 Relaxation dynamics of EcDHFR: NADP<sup>+</sup>: Folate in the presence of co-solvents

It has previously been discussed that the energy landscape of an enzyme can be altered by making changes to the external conditions which may shift the equilibrium between kinetically and conformationally distinct states.<sup>97</sup> Addition of co-solvents to the protein solution is one approach that has been implemented in order to change the properties of the protein environment. It has been suggested that varying the solution viscosity can change the rotational correlation time without affecting the internal motions.<sup>161</sup> Obtaining relaxation data of the EcDHFR: NADP<sup>+</sup>: folate complex in the presence of a greater or less viscous organic co-solvent could reveal the presence of dynamical processes which cannot be elucidated under normal solution conditions. The objective of this results chapter is to report on any differences in the dynamical processes of the EcDHFR: NADP<sup>+</sup>: folate complex that are induced by the presence of methanol and glycerol co-solvents at 17% by final volume concentrations.

The solution NMR backbone dynamics of the EcDHFR: NADP<sup>+</sup>: folate complex and another two complexes, EcDHFR: folate: DHNADPH and EcDHFR: folate, which mimic the occluded conformation of the M20 loop before and after the NADP<sup>+</sup> cofactor release, respectively have been previously measured under standard buffer conditions. The *het*NOE, R1 and R2 relaxation values for each of these complexes have been reported, as have the model free relaxation parameters, all of which can be used as reference points.<sup>87</sup>

Studies on proteins using molecular dynamics simulations predict that protein mobility should be decreased in more viscous solvent and that the influence of the increased viscosity would be felt by the protein both internally and externally. This study also found that the average structure, fluctuations and secondary structure of the protein did not appear to show a dependence on viscosity of the medium.<sup>162</sup> Vibrational echo experiments measuring the dephasing of CO bound to the active site of four haeme proteins have revealed weak viscosity dependence from ps-ns timescale motions. Slower but larger amplitude fluctuations requiring protein topology changes were found to be affected by viscosity. It was proposed that increasing the viscosity reduces the

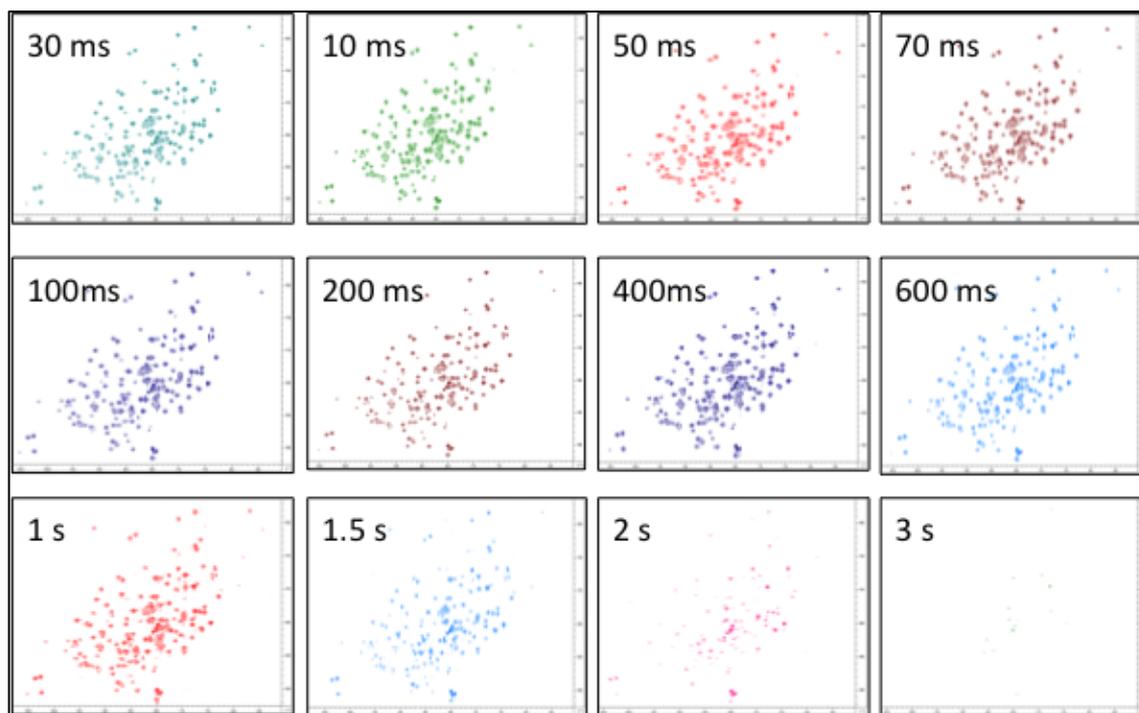
presence of larger amplitude fluctuations and lengthens the characteristic timescale of fluctuations in the time window investigated.<sup>163</sup> Alterations to the electrostatics of the solvent medium should have consequences for dynamics, if the protein structure and motions are governed in full or partially by electrostatics. Molecular dynamics simulations have found that motions in chymotrypsin become more rapid with increasing dielectric constant and it is proposed that protein motions are faster in more hydrophilic solvents.<sup>164</sup>

## **5.1 Biosynthesis, purification and preparation of protein samples**

Dihydrofolate reductase from *E. coli* was overexpressed with regard to the required labelling strategy, which in this case was perdeuterated <sup>15</sup>N (**Section 3.4.4**), purified (**Section 3.5.1**) and prepared accordingly (**Section 3.6**). Relaxation measurement values were also determined for solely <sup>15</sup>N labelled-EcDHFR in the presence of 17% glycerol at 600 Mhz as two attempts at acquiring data at 600 MHz for the perdeuterated 17% glycerol sample yielded acceptable hetNOE data but very poor T1 and T2 data.

## **5.2 Longitudinal relaxation data analysis**

T<sub>1</sub> relaxation data were measured by acquiring a series of HSQCs using the Bruker pulse sequence HSQCT1ETF3GPSI (Phase-sensitive ge-2D <sup>1</sup>H-<sup>15</sup>N HSQC using PEP and 1H-decoupling) pulse sequence, which inverts the spins and then leaves variable delays for inversion-recovery over a range of 10 ms to 2000 ms. 128 dummy scans were completed to ensure sample heating due to the large number of pulses had reached equilibrium prior to acquisition. The experiments were incremented randomly to ensure that any errors accumulated due to changes in the sample over the course of the experiment were randomly and not systematically introduced to the dataset.



**Figure 42**  $^{15}\text{N}$ - $^1\text{H}$ -HSQC T1 relaxation spectra with differing relaxation delays. Data presented represents T1 relaxation data of  $^{15}\text{N}$  only-labelled-EcDHFR: NADP<sup>+</sup>: folate in 17% glycerol at 600 MHz

The data were processed in nmrPipe, imported into ccpNMR and the assignments propagated from the previously assigned data.  $T_2$  for each heteronucleus were extracted from the data by measuring the peak volume as it appears in each HSQC, which is a measure of the degree to which the signal has returned to its equilibrium value during the inversion-recovery time. The relative peak volume was calculated with respect to the peak volume at 10 ms inversion-recovery and plotting as a first order rate process (**Equation 24**). A normal distribution of the measurement errors was assumed and the “covariance” error method used in plotting the data. This method estimates the error (standard deviation) as the square root of “the chi squared value times the covariance matrix diagonal term for that parameter”.<sup>165</sup>

$$I_{(t)} = I_{(0)} (1 - \exp^{-R_1 t})$$

**Equation 24** Longitudinal relaxation rate constant is determined by measurement of the time dependent exponential restoration of intensity of the net magnetisation aligned with the applied  $B_0$  field (where  $R_1 = 1/T_1$ ).

The T1 relaxation values were calculated for each residue of the EcDHFR protein in the buffer conditions and also in the presence of 17% glycerol and 17% methanol at both 600 MHz and 900 MHz using the measurements function in ccpNMR Analysis. The data were fitted to curves of the function in Equation 24 through which the relaxation time constant T1 was extracted. During the analysis, data sets for any residues that showed large chemical shift variations or significant overlap with other residues were excluded from the final analysis, as were data sets for residues, which showed significant doubling or splitting of the peaks in the HSQC experiments. The 10-proline residues, and the N-terminal methionine do not give rise to signals in the relaxation HSQC experiments, meaning that relaxation time constant values are possible for a total of 148 residues. The data including the fit errors are presented for the enzyme complex in buffer only (**Appendix 8.3**), 17% glycerol co-solvent (**Appendix 8.4**), and 17% methanol co-solvent (**Appendix 8.3**). The T1 values were plotted as a function of residue number (**Figure 42**).

#### **T1 for EcDHFR: folate: NADP+ in Buffer**

T1 values of the EcDHFR complex in buffer at 600 MHz are reported for 109 residues (**Appendix 8.3.1**). The maximum T1 values were found for 112 Leu (1499.33 ms,  $s = 1519.49$  ms), 54 Leu (1156.84 ms,  $error = 526.90$  ms), 136 Val (1122.80 ms,  $error = 506.41$  ms) and 102 Gln (1001.23 ms,  $error = 14.51$  ms) however, the standard deviation in T1 values for 112 Leu, 54 Leu and 136 Val were all large implying a poor fit of the data for these residues. To further validate the data in this analysis, 10 residues for which the standard deviation of the T1 values was greater than 10% of T1 were discounted (**Appendix 8.3.1**). This resulted in yielding a maximum value of T1 for 102 Glu (1001.23 ms,  $error = 14.51$ ) and a minimum T1 value at residue 8 Leu (742.02 ms,  $error = 24.50$ ). The average T1 value of the 99 values included in the analysis was calculated with the standard deviation (834.93 ms,  $s = 48.82$  ms,  $n = 99$ ).

A similar analysis was completed for the EcDHFR complex in buffer T1 data at 900 MHz; T1 values for 113 residues are reported (**Appendix 8.3.2**) a total of 12 of the T1 values were discounted from the analysis because the SD error was greater than 10% of the T1 value, those values are highlighted in the reported data (**Appendix 8.3.2**). This resulted in yielding a maximum value of T1 for 60 Ile (1216.25 ms,  $error = 56.86$ ) and

a minimum T1 value at residue 68 Thr (746.36 ms, *error* = 45.15). The average T1 value of the 101 values included in the analysis was calculated with the standard deviation (997.01 ms, *s* = 75.78 ms, *n* = 101).

#### **T1 for EcDHFR: folate: NADP+ in 17% Methanol**

The T1 data for the EcDHFR complex in 17% methanol at 600 MHz was analysed in the same manner as above. T1 values for 120 residues are reported (**Appendix 8.5.3**) a total of 13 of the T1 values were discounted from the analysis because the SD error was greater than 10% of the T1 value, those values are highlighted in the reported data (**Appendix 8.5.3**). This resulted in yielding a maximum value of T1 for 80 Glu (1327.28 ms, *error* = 55.00) and a minimum T1 value at residue 68 Thr (699.07 ms, *error* = 55.40). The average T1 value of the 101 values included in the analysis was calculated with the standard deviation (999.94 ms, *s* = 79.46 ms, *n* = 107).

T1 data for the EcDHFR complex in 17% methanol at 900 MHz are reported for 116 residues (**Appendix 8.5.1**). A total of 17 of the T1 values were discounted from the analysis because the SD error was greater than 10% of the T1 value, those values are highlighted in the reported data (**Appendix 8.5.1**). This resulted in yielding a maximum value of T1 for 134 Glu (1739.65 ms, *error* = 166.57) and a minimum T1 value at residue 68 Thr (826.52 ms, *error* = 79.97). The average T1 value of the 101 values included in the analysis was calculated with the standard deviation (1401.82 ms, *s* = 140.99 ms, *n* = 99).

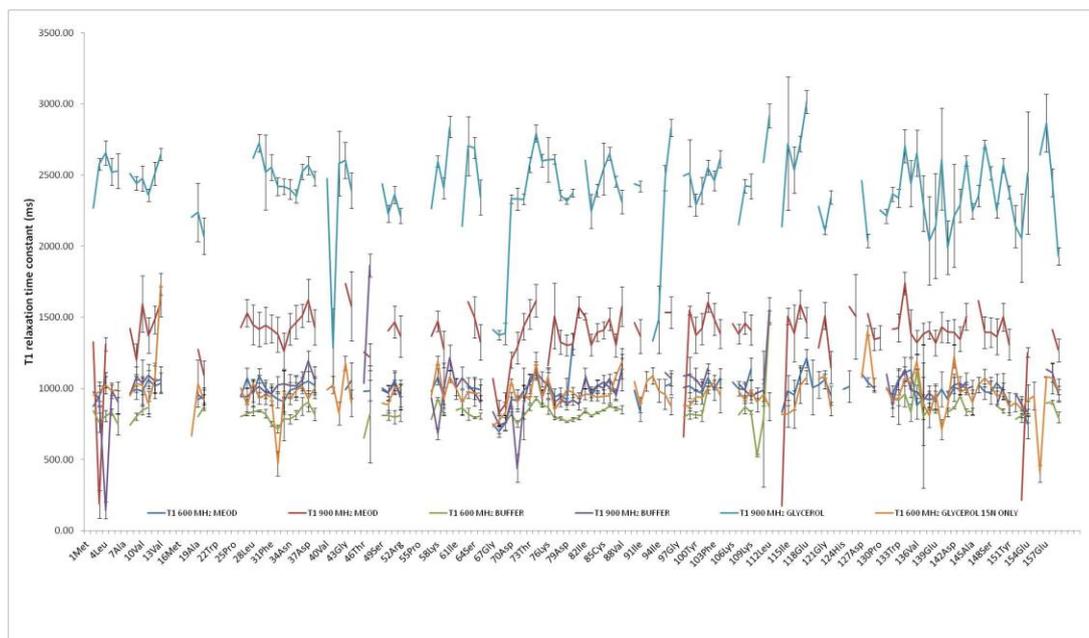
#### **T1 for EcDHFR: folate: NADP+ in 17% Glycerol**

T1 data for the EcDHFR complex in 17% glycerol at 900 MHz are reported for 129 residues (**Appendix 8.4.2**). A total of 20 of the T1 values were discounted from the analysis because the SD error was greater than 10% of the T1 value, those values are highlighted in the reported in the attached data (**Appendix 8.4.2**). This resulted in yielding a maximum value of T1 for 157 Glu (2865.65 ms, *error* = 98.24) and a minimum T1 value at residue 68 Thr (1375.63 ms, *error* = 74.30). The average T1 value of the 109 values included in the analysis was calculated with the standard deviation (2416.10 ms, *s* = 263.81 ms, *n* = 109). It should be noted that the curve fit errors on this

data set were large over the majority, implying that the data set did not give a good overall fit to the curve function.

Two attempts to acquire the T1 data set for the EcDHFR complex at 600 MHz yielded extremely poor fitting data with considerable change of the sample over the course of the incremented T1 experiments, which take a considerably larger amount of time than the T2 experiments to acquire. The data sets were judged to be of too poor fit to be of analytical value to the overall data set. Instead, a set of data T1 data that was acquired at 600 MHz on a  $^{15}\text{N}$ -labelled EcDHFR complex is included as a point of reference.

T1 data for the  $^{15}\text{N}$  labelled EcDHFR complex in 17% glycerol at 600 MHz are reported for 131 residues (**Appendix 8.4.4**). A total of 15 of the T1 values were discounted from the analysis because the SD error was greater than 10% of the T1 value, those values are highlighted in the reported in the attached data (**Appendix 8.4.4**). This resulted in yielding a maximum value of T1 for 128 Tyr (1406.15 ms, *error* = 49.60) and a minimum T1 value at residue 140 Phe (713.50 ms, *error* = 71.05). The average T1 value of the 116 values included in the analysis was calculated with the standard deviation (972.44 ms, *s* = 97.08 ms, *n* = 116).



**Figure 43** Longitudinal relaxation time constant (T1) values for the EcDHFR: NADP<sup>+</sup>: folate complex in the presence of 17% methanol, 17% glycerol and buffer at 600 MHz and 900 MHz.

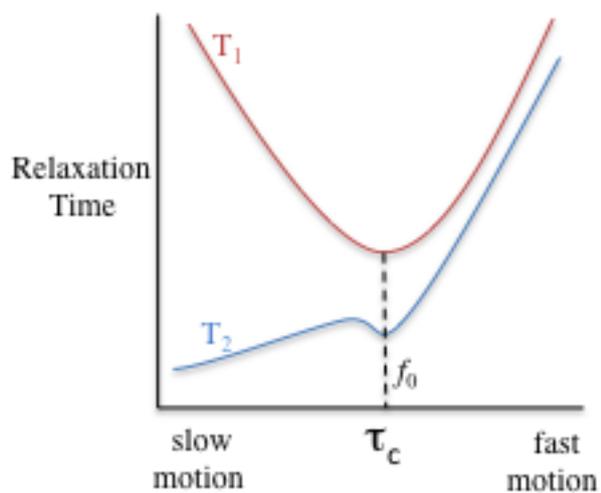
The longitudinal relaxation time (T1) gives a measure of the rate of energy transfer from the nuclear spin system to the surrounding lattice in order to re-establish a Boltzmann distribution of the energies in the system. T1 is dependent on factors such as the magnetic field strength, molecular size, overall correlation time of the molecule, macroscopic viscosity of the solvent and, temperature of the sample.

The chart above (**Figure 43**) shows that the overall magnitude of the T1 value for the EcDHFR: NADP<sup>+</sup>: Folate complex is affected by the magnetic field strength and the solution properties of the protein however, the variation of the T1 values with regards to protein sequence under each set of conditions appear to be similar. Meaning that, for residues with relatively short T1 values, T1 is relatively short regardless of the field strength and solution properties. The dependence of T1 on the magnetic field strength is a direct result of the frequency of the magnetic field fluctuations, which cause the relaxation. The closer the magnetic field fluctuations are to the Larmor frequency, the more efficient the relaxation is, *i.e.* T1 is shortest when the correlation time  $\tau_c$  is closest to the Larmor frequency. The Larmor frequency is dependent on the applied magnetic field strength (T) and the gyromagnetic ratio for a particular isotope. It follows that at stronger magnetic field strengths; the Larmor frequency will be greater. The 17 % glycerol sample at 900 MHz did not yield particularly well fitting data which could be due to the reduced concentration of this particular sample, relative to samples with no-solvent and methanol. The lower concentration of the protein would alter the viscosity and in turn the correlation time. The results set would benefit from a repeat run of these particular experiments in the presence of 17 % glycerol co-solvent.

Large proteins tend to have T1 values of the order of seconds, which increases for measurements conducted at higher fields (**Figure 44**). The increased viscosity of the protein medium in the presence of 17% glycerol causes an increase in the overall correlation time of the protein complex. Additionally, the increase in magnetic field from 600 MHz to 900 MHz causes a change in the frequency of the magnetic field fluctuations, making the relaxation less efficient and increasing the relaxation time.

Addition of methanol reduces the viscosity of the protein solution, relative to the protein in buffer only, which might be expected to cause an increase in the overall correlation time  $\tau_c$  of the protein complex. This increase in the correlation time may act to offset the

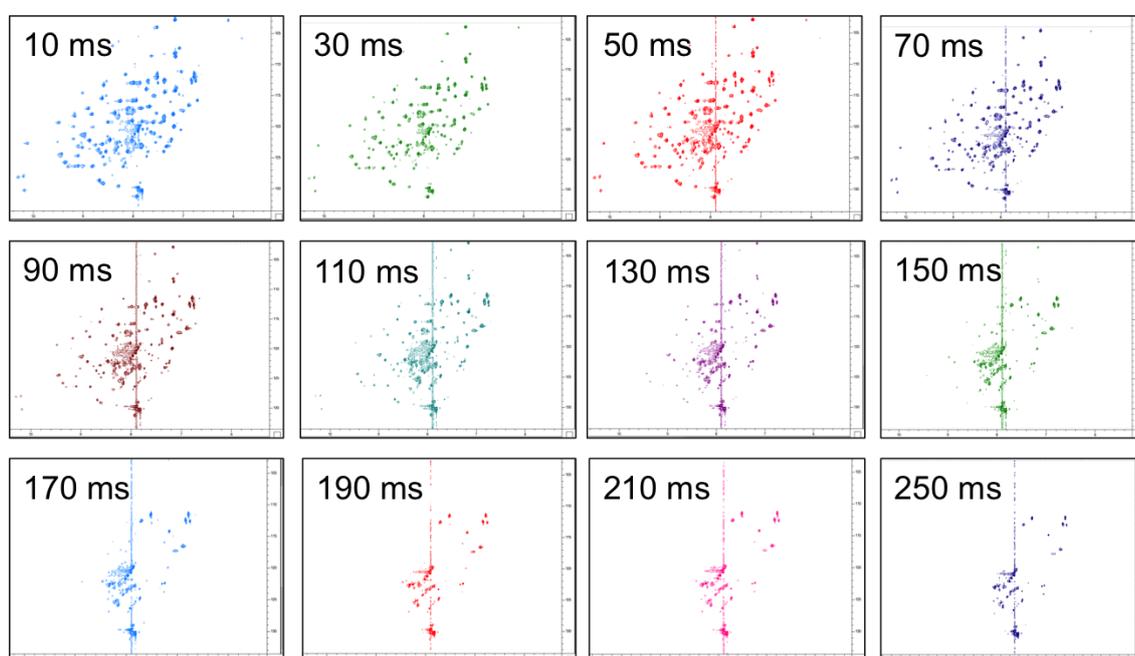
effect on T1 that the increase in magnetic field from 600 MHz to 900 MHz would be expected to have, as in the case of 17% glycerol co-solvent.



**Figure 44** The variation of T1 and T2 relaxation time with molecular correlation time ( $\tau_c$ ). Relaxation is most efficient when the correlation is close to the Larmor frequency ( $f_0$ )

### 5.3 Transverse relaxation data analysis

T<sub>2</sub> relaxation data were measured by acquiring a series of HSQC's using the pulse sequence HSQCT2ETF3GPSI (Phase-sensitive ge-2D <sup>1</sup>H-<sup>15</sup>N HSQC using PEP) with variable diffusion delays between 10ms and 250ms. The maximum number of loops for compensation of heating effects was set to 15 for T<sub>2</sub> with delay value of 250 ms and the minimum number of loops set to 1 at 10 ms. 128 dummy scans were completed to combat sample heating. The D1 recycle delay between scans was set to 4 seconds and the experiments were incremented randomly.



**Figure 45** <sup>15</sup>N-<sup>1</sup>H-HSQC T<sub>2</sub> relaxation spectra with differing relaxation delays. Data presented represents T<sub>2</sub> relaxation data of <sup>15</sup>N <sup>2</sup>H-labelled-EcdHFR: NADP<sup>+</sup>: folate in 17% methanol at 900 MHz

The data were processed in nmrPipe, imported into ccpNMR and the assignments propagated from the previously assigned data. T<sub>2</sub> for each heteronucleus was extracted from the data by measuring the peak volume as it appears in each HSQC, which is a measure of the degree to which the signal has returned to its equilibrium value during the inversion-recovery time (**Figure 45**). The relative peak volume was calculated with respect to the peak volume at 10ms inversion-recovery and plotting as a first order rate process (**Equation 25**). The covariance error method was again used, assuming a normal distribution of the error methods (**Chapter 6.2**).

$$I_{(t)} = I_{(0)} (1 - \exp^{-R_1 t})$$

**Equation 25** Longitudinal relaxation rate constant is determined by measurement of the time dependent exponential restoration of intensity of the net magnetisation aligned with the applied  $B_0$  field (where  $R_2 = 1/T_2$ ).

### **T2 for EcDHFR: folate: NADP+ in Buffer**

T2 values of the EcDHFR complex in buffer at 600 MHz are reported for 105 residues (**Appendix 8.3.3**). To further validate the data in this analysis, a total of 9 residues for which the standard deviation of the T2 values was greater than 10% of T2 were discounted (**Appendix 8.3.3**). This resulted in yielding a maximum value of T2 for 93 Val (165.82 ms, *error* =12.40), and a minimum T2 value at residue 5 Ile (60.09 ms, *error* = 3.12). The average T2 value of the 96 values included in the analysis was calculated with the standard deviation (78.59 ms, *s* =12.64, *n* = 96)

The analysis was completed for the EcDHFR complex in buffer T2 data at 900 MHz; T2 values for 103 residues are reported (**Appendix 8.3.4**) a total of 10 of the T2 values were discounted from the analysis because the SD error was greater than 10% of the T2 value, those values are highlighted in the reported data (**Appendix 8.3.4**). This resulted in yielding a maximum value of T2 for 102 Gln (96.62 ms, *error* = 3.19) and a minimum T2 value at residue 149 His (43.34 ms, *error* = 1.24). The average T2 value of the 93 values included in the analysis was calculated with the standard deviation (63.62 ms, *s* =9.06, *n* = 93).

### **T2 for EcDHFR: folate: NADP+ in 17% Methanol**

The T2 values for 116 residues are reported for the EcDHFR complex in buffer at 600 MHz (**Appendix 8.5.4**). A total of 5 of the T2 values were discounted from the analysis because the SD error was greater than 10% of the T2 value, those values are highlighted in the reported data (**Appendix 8.5.4**). This resulted in yielding a maximum value of T2 for 107 Ala (113.59 ms, *error* = 9.41) and a minimum T2 value at residue 95 Gly (44.05 ms, *error* = 1.74). The average T2 value of the 93 values included in the analysis was calculated with the standard deviation (66.39 ms, *s* = 1.40 ms, *n* = 111).

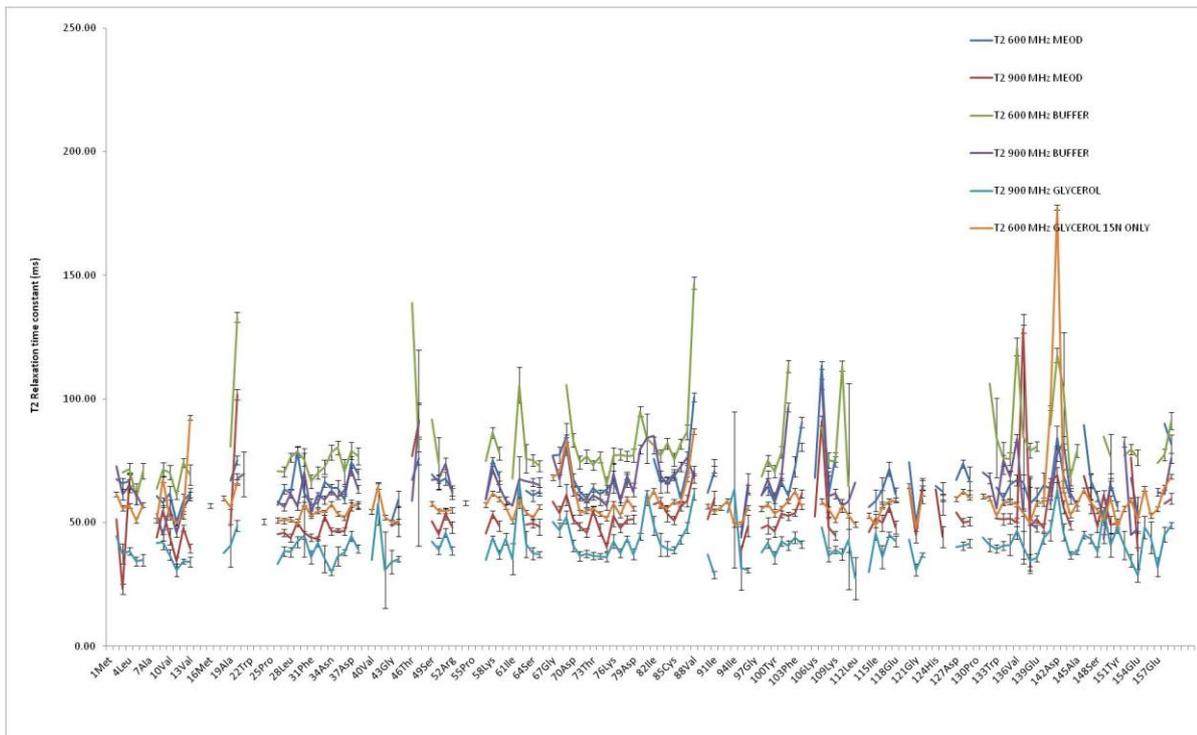
The analysis was completed for the EcDHFR complex in 17% methanol T2 data at 900 MHz; T2 values for 116 residues are reported (**Appendix 8.5.2**) a total of 4 of the T2 values were discounted from the analysis because the SD error was greater than 10% of the T2 value, those values are highlighted in the reported data (**Appendix 8.5.2**). This resulted in yielding a maximum value of T2 for 20 Met (101.73 ms, *error* = 9.13) and a minimum T2 value at residue 3 Ser (23.14 ms, *error* = 2.55). The average T2 value of the 93 values included in the analysis was calculated with the standard deviation (52.90 ms, *s* = 10.46 ms, *n* = 112).

### **T2 for EcDHFR: folate: NADP+ in 17% Glycerol**

T2 900MHz of the EcDHFR: NADP+: folate complex in the presence of 17% glycerol were recorded (**Appendix 8.4.1**). Values for which the SD of the error is greater than 10% of the T2 value are highlighted red and were excluded from the T2 analysis. T2 values for a total of 130 residues were calculated, 24 of these are discounted due to large errors. The average T2 value of the resulting data was calculated (40.96 ms, *s* = 6.47 ms, *n* = 106). The maximum T2 value determined was for 62 Leu (66.39 ms, *error* = 5.41) and the minimum T2 values was derived for 114 His (29.97 ms, *error* = 2.48). It should be noted that the curve fit errors on this data set were large over the majority, implying that the data set did not give a good overall fit to the curve function.

The T2 relaxation data at 600MHz of the EcDHFR: NADP<sup>+</sup>: folate complex in the presence of 17% glycerol was recorded (**Appendix 8.4.3**). Values for which the SD of the error is greater than 10% of the T2 value are highlighted red and were excluded from the T2 analysis. T2 values for a total of 131 residues were calculated, 5 of these are discounted due to large errors. The average T2 value of the resulting data was calculated (57.59 ms,  $s = 7.57$  ms,  $n = 126$ ). The maximum T2 value was derived for 142 Asp (177.58 ms,  $error = 10.83$ ), however the fit error for this value was one order of magnitude larger than the other fit errors so this value was excluded from the analysis, along with the 4 other values due to large standard deviations. The maximum T2 value included in the analysis was found for residue 13 Val (92.61 ms,  $error = 9.17$ ) and the minimum T2 value was found for residue 121 Gly (47.70, ms  $error = 1.52$ ).

The chart below (**Figure 46**) shows that the overall magnitude of the T2 value for the EcDHFR: NADP<sup>+</sup>: folate complex varies only slightly with magnetic field strength and co-solvent conditions. The variation in the T2 values of the protein under each set of conditions appears to be similar. Meaning that, for residues with relatively short T2 values, T2 is relatively short regardless of the field strength and solution properties, although the relative magnitude may be slightly altered. This is in agreement that the overall effect of changes in solvent is to alter only the overall correlation time and not the internal motions. Transverse relaxation (T2) is caused by loss of coherence amongst nuclei and always occurs where there is T1 relaxation, however, when molecular motion is below the Larmor frequency, T2 relaxation occurs in the absence of T1 relaxation and without energy exchange, such that T2 decreases for large molecules as the correlation time decreases (**Figure 44**).



**Figure 46** Transverse relaxation time constant (T2) values for the EcdHFR: NADP<sup>+</sup>: folate complex in the presence of 17% methanol, 17 % glycerol and buffer at 600 MHz and 900 MHz

## 5.4 T1/T2 Analysis

The internal motions of the backbone amide groups of a protein can be described in terms of generalised order parameters and effective correlation times in accordance with the model-free approach (**Section 1.5.6.**)<sup>166</sup> The model-free approach assumes that the global diffusion of a protein and the internal motions of the protein are uncorrelated, allowing the two types of motion to be separated with individual correlation functions (**Equation 26.**)<sup>167</sup>

$$C(t) = C_o(t) C_i(t)$$

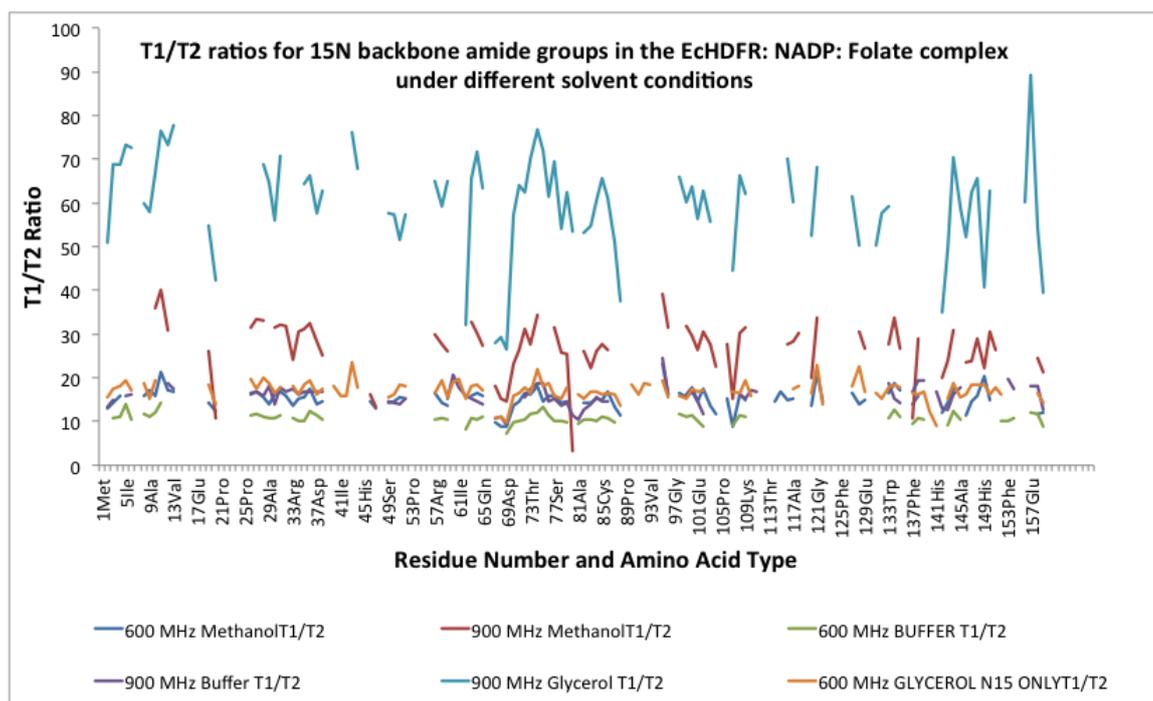
**Equation 26** Autocorrelation function  $C(t)$  is the product of the global correlation function  $C_o(t)$  and the internal correlation function  $C_i(t)$

It is possible to determine the rotational correlation time for a molecule from the ratio of T1/T2 for the <sup>15</sup>N protein backbone relaxation data provided that the intramolecular motions of most of the protein <sup>15</sup>N nuclei are fast (<100 ps).<sup>168</sup> If the internal and rotational motions of the protein are coupled, then model-free formalism may lead to incorrect estimation of some parameters. The rotational correlation time ( $\tau_R$ ) will be underestimated if the internal motions of most of the protein are of a similar timescale to the overall correlation time of the molecule. Acquiring the data at multiple fields cannot combat the underestimation of  $\tau_R$  and as a consequence, the underestimation of  $\tau_R$  can lead to an overestimation of the model-free order parameters and underestimation of correlation times of internal motions.<sup>169</sup> As there are a total of nine model-free models that the parameters can be fit to, these incorrect values may lead to the data being fit to the wrong model.

The ratio of T1/T2 was calculated for all residues for which the T1 and T2 errors were less than 10% of the total values and summarised. From those T1/T2 ratios, the mean T1/T2 ratio was calculated for the overall protein molecule (**Table 15**). The T1/T2 ratio values were plotted on a residue-by-residue basis for the protein under the three solution conditions at both 600 MHz and 900 MHz (**Figure 47**). The average T1/T2 ratios vary considerably between samples.

<b>EcDHFR: NADP<sup>+</sup>: Folate complex solvent conditions and NMR Field</b>	<b>Number of Residues</b>	<b>Mean T1/T2</b>	<b>SD</b>
EcDHFR complex in 17% MeOD 600 MHz	103	15	±2.3
EcDHFR complex in 17% MeOD 900 MHz	90	27	±5.9
EcDHFR complex under standard buffer conditions at 600 MHz	88	10	±1.2
EcDHFR complex under standard buffer conditions at 900 MHz	92	15	±2.4
EcDHFR complex in the presence of 17% glycerol at 900 MHz	95	60	±11.3
<sup>15</sup> N labelled EcDHFR complex in the presence of 17% glycerol at 600 MHz	113	17	±2.3

**Table 15** The mean ratio of T1/T2 for all residues in the EcDHFR: NADP<sup>+</sup>: folate complex for which the T1 and T2 errors were less than 10% of the total values, along with the standard deviation.



**Figure 47** T1/T2 for all residues in the EcDHFR: NADP<sup>+</sup>: folate complex for which the T1 and T2 errors were less than 10% of the total values.

The T1 / T2 ratio provides an approximation of the rate for which N-H vectors within the protein re-orientate due to overall anisotropic rotational diffusion.<sup>170</sup> Anisotropic diffusion causes residues to have different relaxation depending on the orientation of the N-H vector in relation to the protein structure. As proteins tend not to be uniformly spherical then they diffuse anisotropically, meaning that they favour one axis over another and rotate at different rates about these axes. Consequently, the relaxation of N-H vectors depends on the orientation with relation to the overall protein structure as the frequency of fluctuations in local magnetic fields due to the overall protein tumbling is different depending on which axis of the protein they are orientated. However, by observing the T1 / T2 ratios for the majority of residues within a protein an approximate overall correlation time can be established. Moreover, by comparison to X-ray crystal structures it is possible to identify residues which significantly differ from the normal and distinguish between whether this is because they are orientated differently or because of an alternative effect such as fast internal motions. Chemical shift is dependent on the orientation of molecules relative to the applied magnetic field. Chemical shift anisotropy exists if there are significant variations in the chemical shift in different directions. In the case of 17% glycerol samples, chemical shift anisotropy (CSA) effects may be important due to increased density of the solvent and may explain the shorted T2 values, compared to the buffer only and 17% methanol cases. CSA relaxation is proportional to the square of the magnetic field so becomes more important at higher magnetic fields and higher viscosities.

The ratio of T1/T2 reports on ratios that may be involved in chemical exchange or those residues that may exhibit internal motion. The mean T1/T2 ratio was calculated for all residues falling within one standard deviation of the mean T1/T2 (**Table 16**) The residues falling outside one standard deviation of the mean T1/T2 ratio were analysed separately.

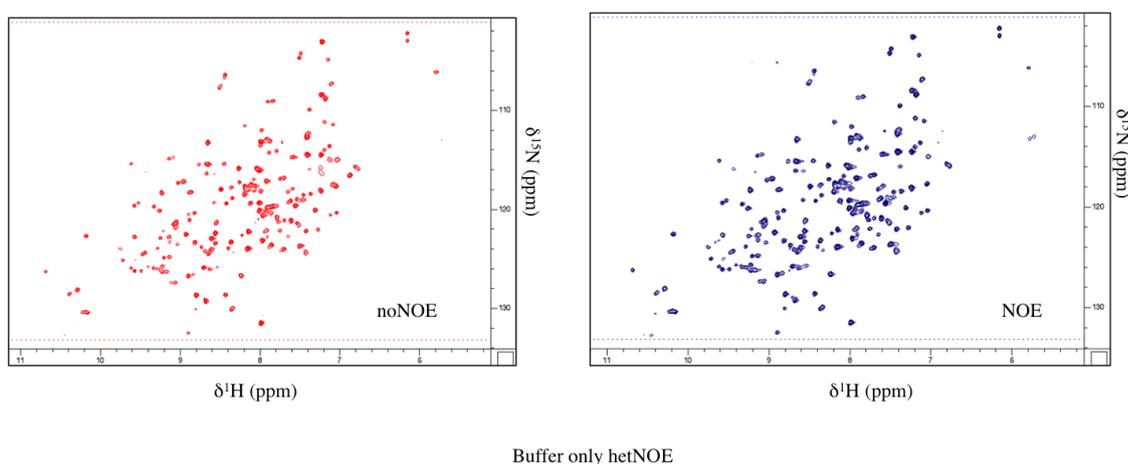
<b>EcDHFR: NADP<sup>+</sup>: Folate complex solvent conditions and NMR Field</b>	<b>Number of Residues</b>	<b>Mean T1/T2</b>	<b>SD</b>
EcDHFR complex in 17% MeOD 600 MHz	84	15	±1.1
EcDHFR complex in 17% MeOD 900 MHz	71	28	±3.1
EcDHFR complex under standard buffer conditions at 600 MHz	67	11	±0.5
EcDHFR complex under standard buffer conditions at 900 MHz	69	16	±1.2
EcDHFR complex in the presence of 17% glycerol at 900 MHz	74	61	±5.8
<sup>15</sup> N labelled EcDHFR complex in the presence of 17% glycerol at 600 MHz	88	17	±1.1

**Table 16** The mean T1/T2 ratio for all residues falling within one standard deviation of the mean T1/T2 with standard deviation of the residues included.

## 5.5 The Heteronuclear NOE

The sensitivity of the *het*NOE experiment, as with the case of  $T_1$  and  $T_2$  experiments, is naturally low due to the pulse sequence beginning with equilibrium of  $^{15}\text{N}$  rather than  $^1\text{H}$  magnetisation. The gyromagnetic ratio of  $^{15}\text{N}$  is one tenth of that of  $^1\text{H}$ , which directly leads to a 10-fold reduction in the magnetisation strength.<sup>171</sup> In everyday NMR acquisition, it is standard to increase the scan repetition rate (or decrease the relaxation delay) in order to reduce the experimental time. The  $T_1$  of water protons is around  $\sim 4.5$  s whereas the  $T_1$  value of protein protons  $\sim 1.4$  s which can lead water to remain in a semi saturated state when short relaxation times are implemented.

Short relaxation delays in the acquisition parameters of the *het*NOE lead to systematic errors in the *het*NOE values including artificially increase *het*NOE ratios due to chemical exchange of amide protons with saturated water protons, obliteration of the signals of rapidly exchanging protons and attenuation of the entire  $^1\text{H}$  spectrum due to spin diffusion of direct NOE with water.<sup>172</sup> A relaxation delay of 10 seconds is sufficient for all degrees of flexibility observed in structures parts of proteins and this value was implemented in the following experiments. For completely unstructured residues, a relaxation delay of 20 seconds is recommended to allow full relaxation of the water magnetisation.<sup>173</sup>



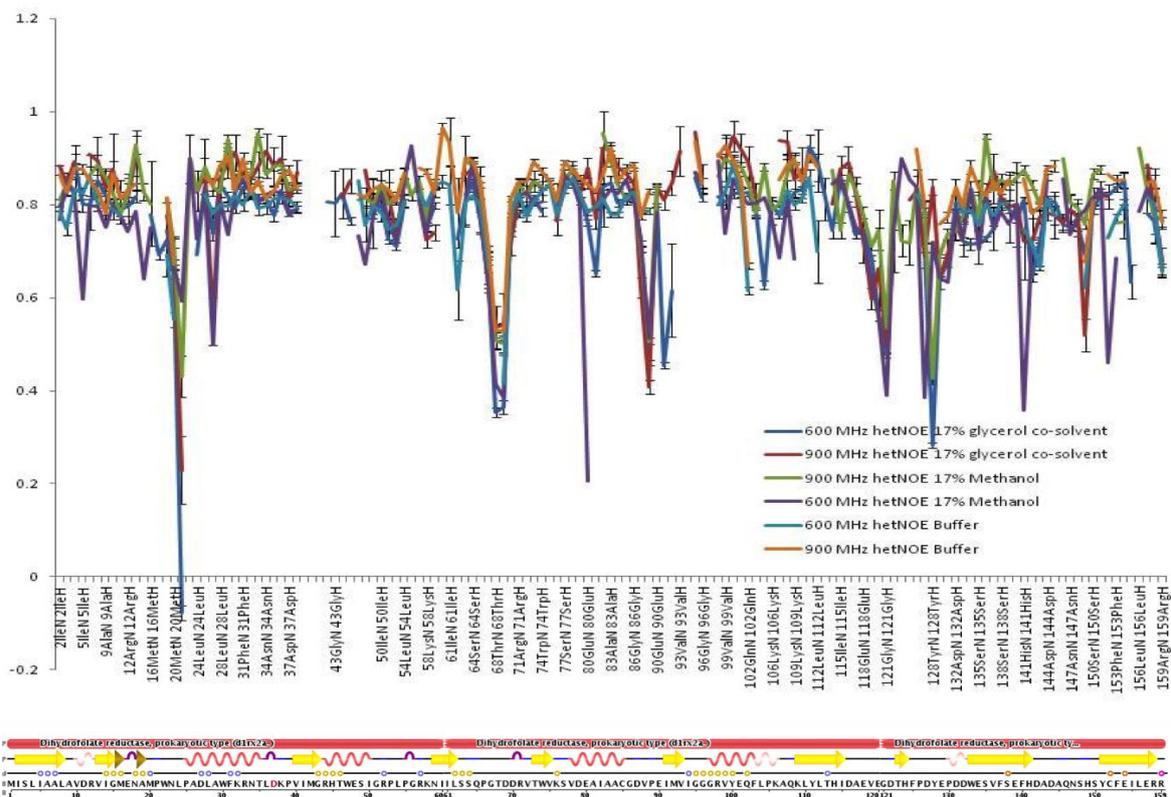
**Figure 48** The steady-state *het*NOE values were established from the ratio of peak intensities (I) in the presence and absence of *het*NOE.

The heteronuclear NOE data were measured using the Bruker pulse sequence HSQCNOEF3GPSI (Phase sensitive ge-2D  $^1\text{H}$ - $^{15}\text{N}$  HSQC to measure heteronuclear  $^{15}\text{N}$ - $^1\text{H}$  NOEs using PEP), which records both sets of data, interleaved. The spectra were recorded in the presence (NOE) and absence (NONOE) of  $^1\text{H}$  saturation at 600 MHz 900 MH (**Figure 48**). The data were processed in nmrPipe, imported into ccpNMR and the assignments propagated from the previously assigned data. The steady-state NOE values were established from the ratio of peak intensities ( $I$ ) (**Equation 27**). The hetNOE values were plotted in combination with the standard deviation (**Figure 49**)

$$\text{hetNOE} = \left( \frac{I_{\text{saturated}}}{I_{\text{unsaturated}}} \right) - 1$$

**Equation 27** The steady-state NOE values are established from the ratio of peak intensities ( $I$ ) in the presence and absence of dipolar coupling.

The  $^{15}\text{N}$  heteronuclear NOE (hetNOE) value gives an indication of the degree of mobility for a particular residue in the protein. Significantly decreased HetNOE values are found for mobile residues such as those occurring in loops, the N-terminus and the C-terminus. This is found in the case of the EcDHFR: NADP<sup>+</sup>: folate complex (**Figure 49**) the regions of the protein with a decreased hetNOE value correspond to disordered regions of the protein, as shown in the PDB sequence. There are slight differences in the magnitude of the hetNOE values. The heteronuclear NOE is five times more sensitive to internal dynamics than T1 and thirty times more sensitive than T2, so any changes to the internal dynamics of the protein complex caused by the presence of co-solvents should be most noticeable here.<sup>174</sup> In comparison with the hetNOE intensity ratio values reported for the Michaelis complex mimic at 600 MHz in previous work in buffer, the results acquired here are in relatively good agreement.<sup>87</sup> The addition of both methanol and glycerol co-solvents at 17 % (bfv) appear to cause some increased flexibility, particularly in the loop region containing Ser 64. The C terminus shows more flexibility in the presence of co-solvents but methanol and glycerol do not appear to have particularly different effects, implying that altering the dielectric constant of the protein medium may have the effect of increasing protein motions on ps-ns timescale but that the protein is largely unaffected by changes in viscosity.



**Figure 49** heteronuclear NOE intensity ratio values at 600 MHz and 900 MHz for all backbone amide residues of EcDHFR: NADP<sup>+</sup>: Folate complex in the presence of 17% glycerol and 17% methanol co-solvents and in the absence of co-solvent (buffer only). The sequence at the bottom of the diagram is taken from 1RX2.pdb. This figure is also included in (**Appendix 8.6.4.**)

## 5.6 Conclusions from phenomenological relaxation data analysis

Solute molecules commonly used NMR buffers are known to cause changes to protein chemical shifts, relaxation-dispersion parameters and solute-dependent changes in conformational exchange motions.<sup>175</sup> These findings highlight the changes that buffer alone can make to the structure and function of an enzyme. Further knowledge of how the solvent medium affects enzyme dynamics and function can offer insight into the understanding of how enzymes work.

The T1, T2 and T1/T2 values show similar trends for all amide backbone residues. The heteronuclear NOE is much more sensitive to any changes in internal dynamics. The hetNOE data reveals that the addition of both methanol and glycerol co-solvents at 17 % (bfv) appear to cause some increased flexibility, particularly in the loop region containing Ser 64 and to a lesser extent in other loop regions. Methanol and glycerol also appear to reduce internal motions (cause an increase in the hetNOE values) in regions of defined secondary structure which may be as a result of strengthened hydrogen bonding networks and electrostatic stabilisation. The C terminus shows more flexibility in the presence of co-solvents but methanol and glycerol do not appear to have particularly different effects, implying that altering the dielectric constant of the protein medium may have the effect of increasing protein motions on ps-ns timescale but that the protein is largely unaffected by changes in viscosity.

## 6.0 Overall Conclusions and Further work

### 6.1 Overall Conclusions

The chemical shift assignment of the Michaelis complex mimic under standard buffer conditions, in the presence of 17% methanol and 17 % glycerol co-solvents (to effect isodielectric mediums that differ in viscosity) is reported to a good level of completeness and any perturbations in the chemical shifts of the protein caused by the presence of co-solvents were analysed (**Chapter 4.0**). It was determined that the majority of atoms in the protein complex under 17% glycerol and 17% methanol show very small chemical shift perturbations from the buffer only values, these perturbations in most cases are not much greater than the error. In general, perturbations follow the same trend for both co-solvent conditions, allowing the presumption that, as 17 % glycerol and 17 % methanol are almost iso-dielectric but differ in viscosity, that electrostatics most likely have some effect the protein structure but that no viscosity effect can be detected under these conditions. In some cases, the magnitude of perturbations is greater in 17% methanol than 17% glycerol although these differences in magnitude could be a reflection of the mole fractions of the co-solvents (0.05 and 0.08 for glycerol and methanol, respectively) and bulk solvent behaviour, rather than a viscosity effect. The multiple variables in this data set might make it a candidate for principal component analysis, which may reveal further trends.

Phenomenological interpretation of relaxation data has yielded interesting insights in to the effect that co-solvents have on the ps–ns dynamics of EcDHFR: NADP<sup>+</sup>: folate.

The T1, T2 and T1/T2 values show similar trends for all amide backbone residues. The heteronuclear NOE is much more sensitive to any changes in internal dynamics. The het NOE data reveals that the addition of both methanol and glycerol co-solvents at 17 % (bfv) appear to cause some increased flexibility, particularly in the loop regions and that methanol and glycerol also appear to reduce internal motions (cause an increase in the hetNOE values) in regions of defined secondary structure which may be as a result of strengthened hydrogen bonding networks and electrostatic stabilisation. The C terminus shows more flexibility in the presence of co-solvents but methanol and glycerol do not appear to have particularly different effects, implying that altering the dielectric constant of the protein medium may have the effect of increasing protein motions on ps-

ns timescale but that the protein is largely unaffected by changes in viscosity. It has been postulated that protein properties can be isolated by keeping macroscopic solvent viscosity constant which may be possible, particularly at 0.05 mole fraction of glycerol co-solvent.<sup>176</sup>

These findings are consistent with the view that changes to the solvent composition affect the free energy surface of the protein and therefore the reorganisation energy. Although small, the large number of atoms showing chemical shift perturbations implies changes to the electrostatics of the protein and the hetNOE results display small changes to the flexibility of residues on a ps-ns timescale that are dependent on the dielectric constant of the medium. The lack of a clear viscosity effect on the structure and dynamics of the protein does not mean that there is no effect; however it is likely that any electrostatic effects far outweigh viscosity effects.

## 6.2 Further Work

### 6.2.1 NMR Solution Structure Calculations

The time-averaged ensemble that results from the restraints generated in the presence of the different co-solvents may reveal whether there are differences in the conformation of the ensemble at the point of reaction that are not normally visible under standard conditions. The chemical shift comparisons and perturbation plots confirm that there are small differences in the chemical shifts of the Michaelis complex under the different co-solvent conditions. The question arises as to whether these small differences in chemical shift translate to larger differences in the protein structure or whether the differences are too small to see using established techniques. Elucidation solution structures using NMR techniques may reveal whether these small changes in chemical shift translate into differences in protein structure in the presence of co-solvents. NMR solution structure calculations require assignment of the protein in question to a high level of completion, in addition to restraints such as a protein sequence, distance, dihedral angle, hydrogen bonding and any ligand-protein distance restraints (below). As it currently stands, these restraints have been acquired for the Michaelis complex under the three difference conditions, acquisition and assignment of this data is the most time consuming part of the NMR solution structure process. Collaboration with a computational specialist would allow the restraints to be computed using suitable protocols to generate structures which can be refined in water, a brief review of which is given below.

#### Overview of proposed computational approach

Ambiguous Restraints for Iterative Assignment (ARIA) uses an unassigned NOESY spectrum, a defined molecular system and a list of chemical shift assignments to oversee a complete structure determination process.<sup>177</sup> The data is converted into eXchange Markup Language (XML) by CCPNMR, which includes mapping between nuclei and frequency dimensions, NOESY mixing times and restraints file destinations. Distances (D) between atoms are derived from cross-peak volumes (V) in accordance with the relation  $D = (\alpha^{-1}V)^{-1/6}$  using ARIA core protocols. The ARIA generated restraints are then used to define molecular constraints that are consistent with the

observed cross peaks. Crystallography and NMR System (CNS) is a program used to calculate structures in an iterative format. The simulated annealing process gradually reduces the temperature over the 8 iterations of the calculation and generates a user-defined number of structures per iteration; the 7 structures that are lowest in energy are taken forward to the next iteration. At the end of the run, the generated structures, restraints and violations and NOESY seed assignments are reported back into CCPNMR. Restraints can be accepted or rejected based on the percentage of the generated structures that are violated and the calculated distance (Å). Alterations to any assignments can be made at this point. The final iteration, water refinement, is run after several rounds of refinement.

### **Distance Restraints**

$^{13}\text{C}$ -NOESY and  $^{15}\text{N}$ -NOESY spectra have been acquired and cross-peaks have been picked in ccpNmr Analysis for the EcDHFR: NADP<sup>+</sup>: folate complex under 17% glycerol, 17 % methanol and standard buffer protein samples. The inter-proton NOEs have been manually assigned when possible and the uncertain ones left for ARIA. Assignment tolerances have been set to the typical values of 0.400 ppm, 0.05 ppm and 0.20 ppm for indirect  $^{15}\text{N}$ ,  $^1\text{H}$  and  $^{13}\text{C}$  chemical shifts, respectively. Many of the chemical shift differences discussed earlier (**4.7 Analysis of the chemical shift perturbations of EcDHFR: NADP<sup>+</sup>: folate in the presence of co-solvents**) are smaller than these chemical shift tolerances but as the structure calculations calculate distances based on the calibrated volume of the peaks in the NOESY spectra, this does not infer that solution structure elucidation would be fruitless.

### **Dihedral angle restraints**

The dihedral restraints were calculated using Dihedral Angles from Global Likelihood Estimates (DANGLE), which is embedded in CCPNMR. DANGLE estimates protein Phi and Psi angles from backbone chemical shifts secondary structure assignments based on sequence information and comparison to database of known protein structures and chemical shifts. The maximum number of islands was set to 2 and those residues showing defined regions on the Ramachandran plot were assigned secondary structure.

## Assignment of Bound Ligands

The Michaelis complex consists of the EcDHFR enzyme, along with folate and NADP<sup>+</sup>. Chemical shift assignments of the ligands are required in order to determine distance restraints between the enzyme and the bound ligands. Without distance restraints between the enzyme and the bound ligands, it would not be possible to generate a correctly folded structure of the Michaelis complex of EcDHFR.

Chemical shifts of unbound folate and NADP<sup>+</sup> are available in the literature, but there are no reported values of the bound ligands. Spectra of labelled ligands in unlabelled protein would need to be acquired, however because the ligands are required in high excess for a completely bound protein, this would be expensive. Also, as the ligands are present at an 8 fold over protein concentration, filtering (to select for signals from protons attached to non isotopically labelled bodies and reject signals from protons attached to isotopically labelled bodies or *vice versa*) for ligands attached to <sup>12</sup>C would result in an overwhelming dominance in the spectrum of unbound ligand. A solution was to acquire isotope edited/filtered experiments at 900 MHz: 3D <sup>13</sup>C/<sup>15</sup>N f1 filtered, <sup>15</sup>N-edited NOESY-HSQC, yielding NOEs between the protons of ligand/ cofactor and amide protons and side-chain Asn/Gln protons of protein. Additionally, a 3D <sup>13</sup>C/<sup>15</sup>N f1 filtered, <sup>13</sup>C-edited NOESY-HSQC, was acquired yielding NOEs between <sup>1</sup>Hs of ligand and <sup>13</sup>C-attached protons of protein.<sup>178</sup>

As the experiments measure the NOE from the protein to the ligand, diagonal peaks correspond to the resonances of the protein amide and side chains. The cross peak resonances should correspond to the ligand proton resonances in the indirect proton dimension. The 3D <sup>13</sup>C/<sup>15</sup>N F1 filtered, <sup>15</sup>N-edited NOESY-HSQC was largely unsuccessful. The spectrum showed a large number of aliphatic protein-protein interactions breaking through between 0 and 3 ppm. There were not many resonances in the indirect proton dimension in regions that the resonances of NADP<sup>+</sup> and folate would be expected. Some restraints have been extracted from the acquired spectra but the confidence level for these restraints is low. Overall, the more effective approach may be to purchase ligand and cofactor with some degree of labelling and record spectra with the ligand/ cofactor in slight excess over unlabelled protein which would allow for less ambiguous data.

## Hydrogen Bonding Restraints

The long range HNCO experiment uses the trans-hydrogen scalar-coupling constant to allow unambiguous identification of the acceptor group involved in the hydrogen bond. The magnitude of the coupling constant can also be used to determine the geometry of the hydrogen bond.<sup>179</sup> Initial observation of the long range HNCO: The majority of peaks are due to the intra-molecular correlation between  $H_iN_i$  and  $CO_{i-1}$ . Very few (perhaps 1 or 2), if any, long-range correlations are shown in the spectrum. The explanation for this lies with the poor sensitivity of the experiment coupled with slow tumbling of the molecule, which results in the short  $^{15}\text{N}$  T2 relaxation rate. Possible solutions include deuteration of the protein to reduce the number of relaxation pathways, thus increasing the T2 relaxation rate or increasing the temperature. An alternative method of determining amides involved in hydrogen bonding can be employed by analysis of the hydrogen – deuterium exchange spectra. As a set of hydrogen-deuterium exchange data has previously been acquired on the Michaelis complex,<sup>180</sup> this was used in combination with the x-ray crystal structure (1RX2) to determine residues involved in hydrogen bonding. Those residues are listed in **Appendix 2**.

## 6.2.2 Lipari-Szabo

### Model Free Analysis

Application of model free to analyse the relaxation data will provide residue specific quantified values of the amplitude of internal motions. The order parameter ( $S^2$ ) can be used in combination with structures can help to provide an insight in to the importance of fast motions for catalysis. The internal motions of the backbone amide groups of a protein can be described in terms of generalised order parameters and effective correlation times in accordance with the model-free approach.<sup>181</sup>

## 7.0 Published Work

The following published work was contributed to during the course of this work being undertaken:

*Role of the occluded conformation in bacterial dihydrofolate reductase*, Enas M. Behiry, Louis Y. P. Luk, **Stella M. Matthews**, E. Joel Loveridge and Rudolf K. Allemann, *Biochemistry*, 59, **2014**, 4761.

*Aliphatic <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N chemical shift assignments of dihydrofolate reductase from psychropiezophile Moritella profunda in complex with NADP<sup>+</sup> and folate*, E. Joel Loveridge, **Stella M. Matthews**, Christopher Williams, Sara B.-M. Whittaker, Ulrich L. Günther, Rhiannon M. Evans, William M. Dawson, Matthew P. Crump and Rudolf K. Allemann, *Biomolecular NMR assignments*, 17, **2013**, 61.

## 8.0 References

- 
- [1] Wolfenden, R., Snyder, M. J., *Acc. Chem. Res.*, **2001**, *34*, 938
- [2] Bugg, T., *An introduction to enzyme and co-enzyme chemistry*, Blackwell Science Ltd., **1997**, 17
- [3] Koshland, D. E., *Proc. Natl. Acad. Sci.*, **1995**, *33* (23-24), 2375
- [4] Pauling, L., *Chem. Eng. News.*, **1946**, *24*, 1375
- [5] Schramm, V. L., *J. Bio. Chem.*, **2007**, *282* (39), 28297
- [6] Wolfenden, R., *Nature*, **1969**, *223*, 704
- [7] Bugg, T., *An introduction to enzyme and co-enzyme chemistry*, Blackwell Science Ltd., **1997**, 28
- [8] Atkins, P.; de Paula, J. *Physical Chemistry*, 8th ed. Oxford university press, **2006**, 876
- [9] Truhlar, D.G., Garrett, B. C., Klippenstein, S.J., *J. Phys. Chem.*, **1996**, *100* (31), 12771
- [10] Laidler, K., *J. Chem. Educ.*, **1984**, *61* (6), 494
- [11] Kohen, A., *Acc. Chem. Res.*, **2015**, *48*, 466
- [12] Wolfenden, R., *Ann. Rev. Biochem.*, **2011**, *20*, 645
- [13] Glowacki, D.R., Harvey, J.N., Mullholland, A.J., *Biochem. Soc. Trans*, **2012**, *40*, 515
- [14] Laidler, K. J., King, M. C., *J. Phys. Chem.*, **1983**, *87* (15), 2657
- [15] Atkins, P., de Paula, J. *Physical Chemistry*, 8th ed. Oxford university press, **2006**, 673
- [16] Atkins, P., de Paula, J. *Physical Chemistry*, 8th ed. Oxford university press, **2006**, 280
- [17] Lewars, E. G., *Computational chemistry; Introduction to the theory and applications of molecular and quantum mechanics*, **2011**, 664
- [18] Munstroph, F., *ChemPhysChem*, **2016**, *17*, 2616
- [19] Francis, K., Kohen, A., *Perspectives in Science*, **2014**, *1*, 110
- [20] Atkins, P., de Paula, J. *Physical Chemistry*, 8th ed. Oxford university press, **2006**, 816
- [21] Roston, D., Islam, Z., Kohen, A., *Molecules*, **2013**, *18*, 5543
- [22] Carpenter, B. A., *Quantum tunnelling in enzyme-catalysed reactions*, Royal Society of Chemistry, **2008**, 1
- [23] Sutcliffe, M.J., Masgrau, L., Roujeinikova, A., Johannissen, L.O., Hothi, P., Basran, J., Ranaghan, K.E., Mullholland, A.J., Leys, D., Scrutton, N.S., *Phil. Trans. R. Soc. B.*, **2006**, *361*, 1375
- [24] Loveridge, E. J., Allemann, R. K., *Direct Methods for the Analysis of Quantum-Mechanical Tunnelling: Dihydrofolate Reductase*, RSC publishing, **2009**
- [25] Kohen, S., Sen, A., *J. Phys. Org. Chem.*, **2010**, *23*, 613
- [26] Rickert, K.W., Klinman, J.P., *Biochemistry*, **1999**, *38*, 12218
- [27] Nagel, Z.D., Klinman, J.P., *Chem. Rev.*, **2006**, *106*, 3095
- [28] Nesheim, J.C., Lipscomb, J.D., *Biochemistry*, **1996**, 10240
- [29] Antoniou, D., Caratzoulas, S., Kalyanaraman, C., Mincer, J.S., Schwartz, S.D., *Eur. J. Biochem*, **2002**, *269*, 3103
- [30] Scrutton, N.S., Basran, J., Sutcliffe, M.J., *Eur. J. Biochem.*, **1999**, *264*, 666
- [31] McGeagh, J. D., Ranaghan, K. E., Mullholland, A. J., *Biochimica et Biophysica Acta*, **2011**, *1814*, 1077

- 
- [32] Marcus, R. A., *J. Chem. Phys.*, **1956**, 24 (5), 966
- [33] Marcus, R. A., Sutin, N., *Biochim Biophys Acta*, **1985**, 811, 265
- [34] Sutcliffe, M. J., Scrutton, N. S., *Eur. J. Biochem.* **2002**, 269, 3096
- [35] Villà, J., Warshel, A., *J. Phys. Chem. B*, **2001**, 105, 7887
- [36] Marcus, R. A., *J. Phys. Chem. B*, **2007**, 111, 6643
- [37] Silverstein, T. P., *J. Chem. Edu.*, **2012**, 89 (9), 1159
- [38] Knapp, M.J., Klinman, J.P., *Eur. J. Biochem.*, **2002**, 269, 3113
- [39] Masgrau, L., Roujeinikova, A., Johannissen, L.O., Hothi, P., Basran, J., Ranaghan, K.E., Mullholland, A.J., Sutcliffe, M.J., Scrutton, N.S., Leys, D., *Science*, **2006**, 312, 237
- [40] Olsson, M.H.M., Siegbahn, P.E.M., Warshel, A., *J. Am. Chem. Soc.*, **2004**, 126, 2820
- [41] Benkovic, S.J., Hammes-Schiffer, S., *Science*, **2003**, 301, 1196
- [42] Hammes-Schiffer, S.; Benkovic, S.J., *Annu. Rev. Biochem.* **2006**, 75, 519
- [43] Loveridge, E.J., Tey, L.-H., Allemann, R. K., *J. Am. Chem. Soc.*, **2010**, 132, 1137
- [44] Kohen, A., *Acc. Chem. Res.*, **2015**, 48, 466
- [45] Antoniou, D., Basner, J., Nunez, S., Schwartz, S. D., *Chem. Rev.*, **2006**, 106, 3170
- [46] Bhabha, G., Lee, J., Ekiert, D. C., Gam, J., Wilson, I. A., Dyson, H. J., Benkovic, S. J., Wright, P. E., *Science*, **2011**, 332, 234
- [47] Adamczyk, A. J., Cao, J., Kamerlin, S. C. L., Warshel, A., *Proc. Natl. Acad. Sci.*, **2011**, 108 (34), 14115
- [48] Loveridge, E. J., Behiry, E. M., Guo, J., Allemann, R. K., *Nature Chemistry* **2012**, 4, 292
- [49] Warshel, A., Sharma, P. K., Kato, M., Xiang, Y., Liu, H., Olsson, H. M., *Chem. Rev.*, **2006**, 106 (8), 3210
- [50] Brown, K. L., *Chem. Rev.*, **2005**, 105 (6), 2075
- [51] Bayliss, WM, *The Nature of Enzyme Action*, 5<sup>th</sup> edn., **1925**, London; Longmans, Green & Co.
- [52] Anderson, V. E., *Ground State Destabilisation, Encyclopaedia of life sciences*, **2001**, John Wiley & Sons Ltd.
- [53] Warshel, A., Russell, S. T., *Quarterly Review of Biophysics*, **1984**, 17, 283
- [54] Belasco, J. G., Knowles, J. R., *Biochemistry*, **1980**, 19, 472
- [55] Haldane, J. B. S., *Enzymes*, **1930**, London: Longmans, Green & Co.
- [56] Menger, F. M., *Accounts of Chemical Research*, **1985**, 18, 128
- [57] Hughes, E. D. I., Ingold, C. K., *J. Am. Chem. Soc.*, **1935**, 244
- [58] Warshel, A., Åqvist, J., Creighton, S., *Proc. Natl. Acad. Sci.*, **1989**, 86, 5820
- [59] Dewar, M. J., Storch, D. M., *Proc. Natl. Acad. Sci.*, **1985**, 82, 2225
- [60] Warshel, A., Åqvist, J., Creighton, S., *Proc. Natl. Acad. Sci.*, **1989**, 86, 5820
- [61] Jeffrey, G. A. , Saenger, *Hydrogen Bonding in Biological Structures*, **1991**, W., Springer-Verlag, Berlin
- [62] Shan, S., Loh, S., Herschlag, D., *Science*, **1996**, 272, 97
- [63] Warshel, A., Papazyan, A., *Proc. Natl. Acad. Sci.*, **1996**, 93, 13665
- [64] Cassidy, C. S., Lin, J., and Frey, P. A., *Biochemistry*, **1997**, 36, 4576

- 
- [65] Ash, E. L., Sudmeier, J. L., De Fabo, E. C., Bachovchin, W. W., *Science*, **1997**, 278, 1128
- [66] Blow, D. M., Birktoft, J. J., and Hartley, B. S., *Nature*, **1969**, 221, 337
- [67] Ishida, T., *Biochemistry*, **2006**, 45, 5413
- [68] Graham, J. D., Buytendyk, A. M., Wang, D., Bowen, K. H., Collins, K. D., *Biochemistry*, **2013**, 53 (2), 344
- [69] Hur, S., Bruice, T. C., *J. Am. Chem. Soc.*, **2003**, 125 (19), 5964
- [70] Hur, S., Bruice, T. C., *Proc. Natl. Acad. Sci.*, **2003**, 100 (21), 12015
- [71] Voet, D., Voet, J. G., *Biochemistry*, 2<sup>nd</sup> Edition, Wiley, Hoboken, NJ, USA, **1995**
- [72] Chen, Y. Q., Kraut, J., Blakeley, R.L., Callendar, R., *Biochemistry*, **1994**, 33, 702
- [73] Fan, Y., Cembrarr, A., Ma, S., Gao, J., *Biochemistry*, **2013**, 52, 2036
- [74] Schweitzer, B., Dicker, A. P., Bertino, J. R., *FASEB J.*, **1990**, 4 (8), 2441
- [75] Dauber-Osguthorpe, P., Roberts, V. A., Osguthorpe, D. J., Wolff, J., Genest, M., Hagler, A. T., *Proteins: Structure, Function and Genetics*, **1988**, 4, 31
- [76] Dams, T., Auerbach, G., Bader, G., Jacob, U., Ploom, T., Huber, R., Jaenicke, R., *J. Mol. Biol.*, **2000**, 297, 659
- [77] Hay, S., Evans, R. M., Levy, C., Loveridge, E. J., Wang, X., Leys, D., Allemann, R. K., Scrutton, N.S., *ChemBioChem*, **2009**, 10, 2348
- [78] Schnell, J.R., Dyson, H.J., Wright, P.E., *Annu. Rev. Biophys. Biomol. Struct.*, **2004**, 33, 119
- [79] Fierke, C.A., Johnson, K.A., Benkovic, S.J., *Biochemistry*, **1987**, 26 (13), 4085
- [80] Venkitakrishnan, R.P.; Zaborowski, E; McElheny, D; Benkovic, S.J.; Dyson, H.J.; Wright, P.E., *Biochemistry*, **2004**, 43, 16046
- [81] Epstein, D. M., Benkovic, S. J., Wright, P. E., *Biochemistry*, **1995**, 34, 11037
- [82] Sawaya, M.R.; Kraut, J., *Biochemistry*, **1997**, 36, 586
- [83] Provided by Dr. Louis Luk, Cardiff University.
- [84] Appleman, J.R., Beard, W.A., Delcamp, T.J., Prendergast, N.J., Freisheim, J.H., Blakley, R.L., *J. Biol. Chem.*, **1990**, 265, 2740
- [85] Fierke, C.A., Johnson, K.A., Benkovic, S.J., *Biochemistry*, **1987**, 26 (13), 4085
- [86] Boehr, D. D., McElheny, D., Dyson, J. H., Wright, P.E., *Science*, **2006**, 313, 1638
- [87] Osborne, M.J., Schnell, J.R., Benkovic, S.J., Dyson, H.J., Wright, P.E., *Biochemistry*, **2001**, 40, 9846
- [88] Sen, A., Kohen, A., *J. Phys. Org. Chem.*, **2010**, 23, 613
- [89] Agarwal, P.K., Billeter, S.R., Rajagopalan, P.T.R., Benkovic, S.J., Hammes-Schiffer, S., *Proc. Nat. Am. Soc.*, **2002**, 99, 2794
- [90] Nagel, Z.D., Klinman, J.P., *Chem. Rev.*, **2006**, 106, 3095
- [91] Maglia, G., Allemann, R. K., *J. Am. Chem. Soc.*, **2003**, 125, 13372
- [92] Mc Elheny, D., Schnell, J. R., Lansing, J. C., Dyson, H. J., Wright, P. E., *Proc. Natl. Acad. Sci. USA*, **2005**, 102, 5032
- [93] Falzone, C. J., Wright, P. E., Benkovic, S. J., *Biochemistry*, **1994**, 33, 439
- [94] Falzone, C. J., Wright, P. E., Benkovic, S. J., *Biochemistry*, **1992**, 31, 7826
- [95] Fraser, J. S., Clarkson, M. W., Degnon, S. C., Eiron, R., Kern, D., Alber, T., *Nature*, **2009**, 462, 669

- 
- [96] Loria, J. P., Barlow, R. B., Watt, E. D., *Acc. Chem. Res.*, **2008**, *41*, 214
- [97] Henzler-Wildman, K., Kern, D., *Nature*, **2007**, *450*, 964
- [98] Singh, P., Sen, A., Fancis, K., Kohen, A., *J. Am. Chem. Soc.*, **2014**, *136*, 2575
- [99] Sikovski, R. S., Wang, L., Markham, K. A., Rajagopalan, P. T., Benkovic, S. J., Kohen, A., *J. Am. Chem. Soc.*, **2004**, *126*, 4778
- [100] Stojkovic, V., Perissinotti, L. L., Lee, J., Benkovic, S. J., Kohen, A., *Chem. Commun.*, **2010**, *46*, 8974
- [101] Stojkovic, V., Perissinotti, L. L., Willmer, D., Benkovic, S. J., Kohen, A., *J. Am. Chem. Soc.*, **2012**, *134*, 1738
- [102] Doron, D., Stojkovic, V., Gakhar, L., Vardi-Kilshtain, A., Kohen, A., Major, D., *J. Phys. Chem. B*, **2014**, *119* (3), 906
- [103] Loveridge, E. J., Behiry, E. M., Guo, J., Allemann, R. K., *Nature Chemistry*, **2012**, *4*, 292
- [104] Behiry, E. M., Luk, L. Y. P., Matthews, S. M., Loveridge, E. J., Allemann, R. K., *Biochemistry*, **2014**, *53*, 4761
- [105] Luk, L. Y. P., Ruiz-Pernía, J. J., Dawson, W. M., Roca, M., Loveridge, E. J., Glowacki, D. R., Harvey, J. N., Mulholland, A. J., Tuñón, I., Moliner, V., Allemann, R. K., *Proc. Natl. Acad. Sci.*, **2013**, *110*, 16344
- [106] Ruiz-Pernia, J. J., Luk, L. Y. P., García-Meseguer, R., Martí, S., Loveridge, E. J., Tuñón, I., Moliner, V., Allemann, R. K., *J. Am. Chem. Soc.*, **2013**, *135*, 18689
- [107] Luk, L. Y. P., Loveridge, E. J., Allemann, R. K., *J. Am. Chem. Soc.*, **2014**, *136*, 6862
- [108] Singh, P., Abeysinghe, T., Kohen, A., *Molecules*, **2015**, *20*, 1192
- [109] Luk, L. Y. P., Loveridge, E. J., Allemann, R. K., *Phys. Chem. Chem. Phys.*, **2015**, *17*, 30817
- [110] Bystroff, C., Oatley, S. J., Kraut, J., *Biochemistry*, **1990**, *29* (13), 3263
- [111] Kwan, A. H., Mobli, M., Gooley, P. R., King, G. F., Mackay, J. P., *FEBS*, **2011**, *278*, 687
- [112] Boehr, D. D., McElheny, D., Dyson, H. J., Wright, P. E., *Proc. Natl. Acad. Sci.*, **2010**, *107*(4), 1373
- [113] J. Keeler, *Understanding NMR spectroscopy*. The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, England: Wiley Subscription Services, Inc., A Wiley Company, 1st ed., **2005**.
- [114] Chen, J., Brooks, C.L., Wright, P.E., *J. Bio. Mol. NMR*, **2004**, *29*, 243
- [115] Kleckner, I. R., Foster, M. P., *Biochemica et Biophysic Acta*, **2011**, *1814*, 942
- [116] Boehr, D. D., Dyson, H. J., Wright, P. E., *Chem. Rev.*, **2006**, *106*, 3055
- [117] Gong, Q., Ishima, R., *J. Biomol. NMR*, **2007**, *37*, 147
- [118] Renner, C., Schleicher, M., Moroder, L., Holak, T.A., *J. Biomol. NMR*, **2002**, *23*, 23
- [119] Farrow, N.A., Zhang, O., Szabo, A., Torchia, D.A., Kay, L.E., *J. Biomol. NMR*, **1995**, 153.
- [120] Clore, G.M., Szabo, A., Bax, A., Kay, L.E., Driscoll, P.C., Gronenborn, A.M., *J. Am. Chem. Soc.*, **1990**, 4989
- [121] Bronowska, A.K., *Thermodynamics of Ligand-Protein Interactions: Implications for Molecular Design, Thermodynamics - Interaction Studies - Solids, Liquids and Gases*, Dr. Juan Carlos Moreno,

---

**2011**, DOI: 10.5772/19447. (Available from: <http://www.intechopen.com/books/thermodynamics-interaction-studies-solids-liquids-and-gases/thermodynamics-of-ligand-protein-interactions-implications-for-molecular-design>)

- [122] Ansari, A., Berendzen, J., Bowne, S. F., Frauenfelder, H., Iben, I. E. T., Sauke, T. B., Shyamsunder, E., Young, R. D., *Proc. Natl. Acad. Sci.*, **1985**, *82*, 5000
- [123] Loveridge, E.J., Tey, L.-H., Allemann, R.K., *J. Am. Chem. Soc.*, **2010**, *132* (3), 2010
- [124] Basner, J.E., Schwartz, S.D., *J. Am. Chem. Soc.*, **2005**, *127*, (40), 13822
- [125] Hay, S., Pudney, C.R., Sutcliffe, M.J., Scrutton, N.S., *Angew. Chem, Int. Ed.*, **2008**, *47*, 537
- [126] Masgrau, L., Roujeinikova, A., Johannissen, L.O., Hothi, P., Basran, J., Ranaghan, K. E., Mulholland, A. J., Sutcliffe, M.J., Scrutton, N.S., Leys, D., *Science*, **2006**, *312* (5771), 237
- [127] Arora, K., Brooks, C. L., *J. Am. Chem. Soc.*, **2009**, *131* (15), 5642
- [128] Loveridge, E.J., Allemann, R.K., *ChemBioChem*, **2011**, *12* (8), 1258
- [129] Loveridge, E.J., Tey, L.-H., Behiry, E.M., Dawson, W.M., Evans, R.M., Whittaker, S.B.-M., Günther, U.L., Williams, C., Crump, M.P., Allemann, R. K., *J. Am. Chem. Soc.*, **2011**, *133*, 20561
- [130] Walser, R., van Gunsteren, W. F., *Proteins*, **2001**, *42* (3), 414
- [131] Finkelstein, I. J., Massari, A. M., Fayer, M. D., *Biophysical Journal*, **2007**, *92*, 3652
- [132] Affleck, R., Haynes, C. A., Clark, D. S., *Proc. Natl. Acad. Sci.*, **1992**, *89*, 5167
- [133] Osborne, M. J., Venkitakrishnan, R. P., Dyson, H. J., Wright, P. E., *Protein Science*, **2003**, *12*, 2230
- [134] Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G.; Pfeifer, J., Bax, A., *J. Biomol. NMR*, **1995**, *6*, 277.
- [135] Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., Laue, E.D., *Proteins*, **2005**, *59*, 687.
- [136] Higman, V.A., [www.protein-nmr.org.uk](http://www.protein-nmr.org.uk).
- [137] Loveridge, E. J., Evans, R. M., Allemann, R.K., *Eur. J.*, **2008**, *14*, 10782
- [138] Frauenfelder, H., Chen, G., Berendzen, J., Fenimore, P. W., Jansson, H., McMahon, B. H., Strope, I. R., Swenson, J., Young, R. D., *Proc. Natl. Acad. Sci.*, **2009**, *106* (13), 5129
- [139] Affleck, R., Haynes, C. A., Clark, D. S., *Proc. Natl. Acad. Sci.*, **1992**, *89*, 5167
- [140] Song, J., Markley, J. L., *J. Mol. Recognition*, **2001**, *14* (3), 166
- [141] Cavanagh, J., Fairbrother, W. J., Palmer (III), A. G., Rance, M., Skelton, N. J., *Protein NMR Principles and Practice*, 84 Theobald's Road, London, WC1X 8RR, UK, Elsevier Academic Press, 2<sup>nd</sup> edition, **2007**, 783
- [142] Yamazaki, T., Forman-Kay, J. D., and Kay, L. E., *J. Am. Chem. Soc.*, *115*, **1993**, 11054.
- [143] Osborne, M. J., Venkitakrishnan, R. P., Dyson, H. J., Wright, P. E., *Protein Science*, **2003**, *12*, 2230

- 
- [144] Schlundt, A., Heinz, G. A., Janowski, R., Geerlof, A., Stehle, R., Heissmeyer, V., Niessing, D., Sattler, M., *Nature Structural & Molecular Biology*, **2014**, *21*, 671
- [145] Williamson, M. P., *Progress in Nuclear Magnetic Resonance spectroscopy*, **2013**, *73*, 1
- [146] Xu, P-X., Case, D. A., *Biopolymers*, **2002**, *65*, 408
- [147] Luman, N. R., King, M. P., Augspurger, J. D. J., *Comput. Chem.*, **2002**, *22*, 366
- [148] Etai, J., Unger, R., *Bioinformatics*, **2007**, *23* (2), 225
- [149] Yao, J., Dyson, H. J., Wright, P. E., *FEBS Letters*, **1997**, *419*, 285
- [150] Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., Sykes, B. D., *J. Biomol. NMR*, **1995**, *5*, 67
- [151] Han, B., Liu, Y., Ginzinger, S. W., Wishart, D. S., *J. Biomol. NMR*, **2011**, *50*, 43
- [152] Urbaniak, M. D., Muskett, F. W., Finucane, M. D., Caddick, S., Woolfson, D. N., *Biochemistry* **2002**, *41*, 11731
- [153] Elowitz, M. B., Surette, M. G., wolf, P. E., stock, J. B., Leibler, S., *Journal of Bacteriology*, **1999**, *181* (1), 197
- [154] Rupley, J. A., Gratton, E., Careri, G., *Trends. Biochem. Sci.*, **1983**, *8*, 18
- 155 Dagenet, C., Dyson, P. J., Krossing, I., Oleinikova, A., Slattery, J., Wakai, C., Weingärtner, H., *J. Phys. Chem. B*, **2006**, *110*, 1268
- [156] Jouyban, A., Soltanpour, S., Chan, H-K., *International Journal of Pharmaceutics*, **2004**, *269*, 353
- [157] Hartsough, D.S., Merz, K.M., *J. Am. Chem. Soc.*, **1993**, *115*, 6529
- [158] Dashnau, J. L., Nucci, N. V., Sharp, K. A., Vanderkooi, J. M., *J. Phys. Chem. B*, **2006**, *110*, 13670
- [159] Natarajan, K. R., *J. Chem. Educ.*, **1991**, *68*, 13-16
- [160] Affleck, R., Haynes, C. A., Clark, D. S., *Proc. Natl. Acad. Sci.*, **1992**, *89*, 5167
- [161] Korchuganov, S. S., Gagnidze, I. E., Tkach, E. N., Schulga, A. A., Kirpichnikov, M. P., Arseniev, A. S., *J. Biomol. NMR*, **2004**, *30*, 431
- [162] Walser, R., van Gunsteren, W. F., *Proteins*, **2001**, *42* (3), 414
- [163] Finkelstein, I. J., Massari, A. M., Fayer, M. D., *Biophysical Journal*, **2007**, *92*, 3652
- [164] Affleck, R., Haynes, C. A., Clark, D. S., *Proc. Natl. Acad. Sci.*, **1992**, *89*, 5167
- [165] [www2.ccpn.ac.uk/documentation/analysis/popups/calcratespopup.html](http://www2.ccpn.ac.uk/documentation/analysis/popups/calcratespopup.html)
- [166] Lipari, G., Szabo, A., *J. Am. Chem. Soc.*, **1982**, *104* (17), 4546
- [167] Tugarinov, V., Liang, Z., Shapiro Y.E., Freed, J.F., Meirovitch, E., *J. Am. Chem. Soc.*, **2001**, *123*, 3055
- [168] Tjandra, N., Feller, S.E., Pastor R.W. and Bax, A., *J. Am. Chem. Soc.*, **1995**, *117*, 12562
- [169] Korzhnev, D.M., Orekhov, V.Yu. and Arseniev, A.S., *J. Magn. Reson.*, 1997, *127*, 184
- [170] Tjandra, N., Feller, S.E., Pastor R.W. and Bax, A., *J. Am. Chem. Soc.*, 1995, *117*, 12562
- [171] Grzesiek, S., Bax, A., *J. Am. Chem. Soc.*, **1993**, *115*, 12593
- [172] Grzesiek, S., Bax, A., *J. Am. Chem. Soc.*, **1993**, *115*, 12593
- [173] Renner, C., Schleicher, M., Moroder, L., Holak, T.A., *J. Biomol. NMR*, **2002**, *23*, 23
- [174] Kay, L. E., Torchia, D. A., Bax, A., *Biochemistry*, **1989**, *28*, 8972

- 
- [175] Wong, M., Khirich, G., Loria, J. P., *Biochemistry*, **2013**, 52 (37), 6548
- [176] Beece, D., Eisenstein, L., Frauenfelder, H., Good, D., Marden, M. C., Reinisch, L., Reynolds, A. H., Sorensen, L. B., Yue, K. T., *Biochem.*, **1980**, 19 (23), 5147
- [177] Habeck, M.; Rieping, W.; Linge, J.P.; Nilges, M., *Methods in Molecular Biology*, Protein NMR Techniques, 278, 18
- [178] Zwahlen, C., Vincent, S.J.F., Kay, L.E., *J. Magn. Reson.*, **1998**, 130, 169
- [179] Cordier, F., Nisus, L., Dingly, A.J., Grzesiek, S., *Nature protocols*, **2008**, 3, 235
- [180] Acquired by E. J. Loveridge, Cardiff University.
- [181] Lipari, G., Szabo, A., *J. Am. Chem. Soc.*, **1982**, 104, 4546