The Role of Enzyme Dynamics in Catalysis by β-Phosphoglucomutase

Hugh R. W. Dannatt

Department of Molecular Biology & Biotechnology
The University of Sheffield

Thesis submitted for the degree
of Doctor of Philosophy

August 2012

Abstract

This thesis primarily concerns the use of nuclear magnetic resonance (NMR) spectroscopy to study enzyme dynamics. Recent improvements in NMR and other biophysical methods have allowed the detailed study of protein dynamics, and have led to much speculation as to their involvement in the catalytic prowess of enzymes. β-Phosphoglucomutase (βPGM) is a phosphoryl transfer enzyme, and must therefore bring one of the slowest chemical reactions in nature onto a biologically-relevant timescale. As well as being exceptionally proficient catalysts, phosphate transfer enzymes are useful objects of study through the use of metal fluorides and ground-state and transition-state analogues which allow for the capture of conformations which relate to various stages along the reaction coordinate. Together then, βPGM and metal fluorides allow for a critical analysis of the role of enzyme dynamics in catalysis.

Formation of the transition state analogue (TSA) complex of βPGM captured by using MgF_3^- ions to mimic the transferring phosphate allows the study of both protein & TSA NMR signals. It is demonstrated that there are two conformational exchange processes occurring on the catalytic timescale: one which is coincident in rate to catalysis and therefore likely plays a role in catalytic turnover by βPGM, and another which is implicated in the folding stability of the protein. These processes are characterised in detail, and attempts to perturb them by mutation are made. The role of dynamics occurring on a faster timescale in βPGM is also explored.

It is concluded that in βPGM and all other enzymes, dynamics on the timescales most amenable to study by NMR (ms–μs and ns–ps timescales) are not directly involved in the chemical step. Instead they may be involved in allowing the enzyme to traverse the complex energy landscape from substrate to product, a role of equal importance.

Acknowledgements

I must first thank my supervisor Jon Waltho for correctly assessing during my interview that I was game for a challenge and choosing my PhD project accordingly. Despite my luck rarely matching my imagination, Jon has always allowed me to explore my ideas and spend the necessary time getting to grips with theory and experimental procedure so as to achieve the best results possible. There were plenty of times — whilst undertaking certain unspecified duties — that he would welcome any legitimate procrastination and was always able to provide insight no matter how cloudy the view. I have also been allowed to attend a near-excessive number of conferences during my studies, which has exposed me to many new concepts and people, as well as plentiful free alcohol. For all this I am very grateful.

Matt has been absolutely indispensable during my time in the NMR group. It is rare to find someone to be so experienced and carefully considered in both practical procedures and data handling that you can have 100% confidence that whatever they have told you is right. It is also rare to find someone equally cynical and committed to doing things "the right way" as I have tried to be throughout. As a result Matt has been my sounding board throughout my PhD. I have come to trust and rely on his advice on all matters scientific, and even for things concerning life outside of the lab. I will try hard to resist the urge to continue sending frequent e-mails once I have left the group, but his validation will be missed!

I never ceased to be amazed by the seemingly endless catalogue of βPGM experiments & data that Nicky could pull out of her many lab books, making her an immensely valuable consultant for all things βPGM. Her keen eye for detail and ability to write the most fantastically transparent and unambiguous e-mails even when describing extremely complex data are amongst several traits that I hope have rubbed off on me. Despite being a physicist by training, Jeremy was always able to really grasp where the difficulties lie in understanding NMR and other more mathematical problems from the point-of-view of a biochemist, which makes him a talented and effective teacher. His scepticism and ability to see through jargon and ill-conceived ideas were also extremely useful providing they weren't exercised whilst I was giving a talk!

Mike has an unparallelled knowledge of all things protein-related and would always humour me when I asked about NMR techniques that I will likely never perform but would nonetheless like to understand. The same goes for Andrea who would always answer questions about how the magnets are built, how much they cost to run, what happens during a quench, etc, etc.! Her help for setting up complicated pulse programs, particularly at the start of my PhD, was invaluable. I particularly appreciated Maya's guidance on setting up and processing the data from the relaxation experiments, I'm glad I was able to take your dispersion pulse program on tour! I should also thank Laz for managing to install Relax on one of our machines remotely, even though I couldn't do it locally!

As fellow βPGM sufferers (although in actual fact it is a wonderful little protein), Jo & Jin Yi have made excellent company, always keen not to take things too seriously and provide guidance in the lab. Recovering from injury and having not been near a wet lab for the best part of 18 months, to describe me as rusty for the first protein prep would be a huge understatement! I wish the best of luck to Luke, with whom it seems the project is in safe hands. With decreasing numbers of people around, I have come to rely more and more on Pete to cheer me up and engage in a bit of (mostly) harmless mischief. His donation to the wet lab of some computer speakers has been instrumental in allowing me to retain some sanity during dreaded buffer exchanges, albeit at the expense of everyone else's. The latest batch of ice cream was particularly good, and "Falafel Tuesdays" is a recent tradition that I will definitely miss. You can have my Yttrium Chloride, Pete, use it wisely! (Don't use it at all!)

Shrimp deserves a mention even if just for the note she once left on my bench: "the indian people got your sample", which will never not be absolutely hilarious. I also thank Yasuko for preventing the complete destruction of my lab bench at the hands of a rogue bunsen burner. Ultimately how much I have enjoyed the time in the lab has been largely down to the people in it, and so in no particular order I would also like to thank: Abi, Rob, Caz, Clare, Dave, Andy, Amie, Jenny, James & Farhat for their company. At risk of going off on a tangent, I am also indebted to the NMR Discussion Group, being invited to give 2 talks and having them so well received was an enormous confidence boost for me last year.

Not wanting to waste more tree than is necessary (good job I'm printing double sided as this already resembles an oscar acceptance speech), I would like to briefly acknowledge a few key groups of people outside of the lab who have made the last 4 years what they have been. Firstly, the housemates I have been lucky to live with whilst here in Sheffield, particularly Cooney & John who have lived in Sheffield throughout. Secondly, anyone with whom I've been lucky enough to play music, but of course special mention goes to Sam B., Sam K., Tom & Scott. Thirdly, the members of the Postgrad Society & PG Café Forum, and especially the fantastic people who helped run the committees whilst I sat on them and those who have continued to do so since. As for everyone else, in the exceedingly unlikely event that you read this: you know who you are and I expect assistance with the further punishing of my liver when the time comes.

My parents have been the best support anybody could hope for. Not once did the words "why don't you get a *real* job?" leave their mouths, and I dare say it never entered their heads. It is easy to get caught up in the technicalities and details and what's going on around you, but they were always on hand to remind me to step back and get enthused about the things that mattered, often getting more excited than I was about whatever news I had to share. My Dad's relentless attempts to understand what on earth I was talking about were heroic (entirely my fault, not his!). The support that I have received from my parents over the last few months in particular has been

immense — well above and beyond the call of duty — and I am truly grateful. I also thank Soph & Mark for their support, and in particular putting me up so many times and often taking me to the airport to make travelling around to conferences or other European labs so much easier. I feel right at home in your spare room, and I hope this will continue!

Last, and certainly not least, Becks. I cannot begin to thank her for everything she has done for me, I don't want to write another thesis. Her unyielding faith in me borders on delusional I'm sure. Through the more difficult times during my PhD I have been awarded more patience than anyone deserves, and she always had the energy to give me a pep talk, even if I didn't have the energy to listen. If she was less generous and kind than she is, just seeing her face on a daily basis would have been support enough. Thank you.

Right. Apart from the bloody abbreviations list, I'm done. Phew.

Hugh Dannatt
10^{th} August 2012, 2:51 am

Table of Contents

List of Figures

List of Tables

Abbreviations and Symbols

‡ — Transition State
Å — Ångström
βG1P — β-Glucose-1-Phosphate
βG6P — β-Glucose-6-Phosphate
βG16BP — β-Glucose-1,6-Bisphosphate
βPGM — β-Phosphoglucomutase
δ — Chemical Shift
κ — Transmission Coefficient
ν_{CPMG} — Carr–Purcell–Meiboom–Gill Field Strength
5PAA — 5–Phosphoarabinonate
A_{280} — Absorbance at 280 nm
Adk — Adenylate Kinase
ADP — Adenosine Diphosphate
AIC — Akaike's Information Criterion
AMP — Adenosine Monophosphate
APS — Ammonium Persulphate
ATP — Adenosine Triphosphate
B_0 — Static Field
B_1 — Applied Field
CD — Circular Dichroism
C_p — Heat Capacity
CPMG — Carr–Purcell–Meiboom–Gill
CSA — Chemical Shift Anisotropy
CypA — Cyclophilin A
DD — Dipole–Dipole
DEAE — Diethylaminoethanol
DHFR — Dihydrofolate Reductase
DTT — Dithiothreitol
E. coli — Escherichia coli
EDTA — Ethylenediaminetetraacetic Acid
EXSY — Exchange Spectroscopy
F6P — Fructose-6-Phosphate
FID — Free Induction Decay
G — Gibbs Free Energy
GS — Ground State
GSA — Ground State Analogue
GTP — Guanosine Triphosphate
H — Enthalpy
HAD — Haloalkanoic Acid Dehalogenase
HEPES — 4–(2–Hydroxyethyl)–1–Piperazineethanesulfonic Acid

HSQC — Heteronuclear Single Quantum Coherence
hetNOE — Heteronuclear Nuclear Overhauser Effect
INEPT — Insensitive Nuclei Enhanced by Polarization Transfer
IPTG — Isopropylthiogalactopyranoside
J — Spectral Density
k_{cat} — Catalytic Turnover Rate
K_{eq} — Equilibrium Constant
k_{ex} — Rate of Exchange
KIE — Kinetic Isotope Effect
L. lactis — Lactococcus lactis
LBHB — Low-Barrier Hydrogen Bond
M6P — Mannose-6-Phosphate
MM — Molecular Mechanics
MOPS — 3–(N–Morpholino)Propanesulfonic Acid
M_r — Molecular Mass
NAC — Near-Attack Complex
$NADP^+$ / NADPH — Nicotinamide Dinucleotide Phosphate
NMR — Nuclear Magnetic Resonance
NOE — Nuclear Overhauser Effect
NOESY — Nuclear Overhauser Effect Spectroscopy
NtrC — Nitrogen Regulatory Protein
OD_{600} — Optical Density at 600 nm
PAGE — Polyacrylamide Gel Electrophoresis
PCR — Polymerase Chain Reaction
PDB — Protein Data Bank
PGI — Phosphoglucose Isomerase
PGK — Phosphoglycerate Kinase
P_i — Inorganic Phosphate
PNP — Purine Nucleoside Phosphorylase
QM — Quantum Mechanics
R_1 — Longitudinal (Spin–Lattice) Relaxation Rate
$R_{1\rho}$ — Spin–Lattice Relaxation Rate in the Rotating Frame
R_2 — Transverse (Spin–Spin) Relaxation Rate
$R_{2(eff)}$ — Effective Transverse Relaxation Rate
RDC — Residual Dipolar Coupling
R_{ex} — Relaxation due to Exchange
RF — Radio-Frequency
RMSD — Root Mean Square Deviation
RT — Room Temperature
S — Entropy
S^2 — Generalised Order Parameter
SDM — Site-Directed Mutagenesis

SDS — Sodium Dodecyl Sulphate
SEC — Size-Exclusion Chromatography
SEXSY — Selective Exchange Spectrosopy
S^2_f — Generalised Order Parameter for Faster Motion
SIIS — Solvent-Induced Isotope Shifts
S^2_s — Generalised Order Parameter for Slower Motion
T_1 — Longitudinal (Spin–Lattice) Relaxation Time
T_2 — Transverse (Spin–Spin) Relaxation Time
τ_e — Effective Correlation Time
TEMED — Tetramethylethylenediamine
τ_f — Effective Faster Correlation Time
τ_m — Molecular Correlation Time
T_m — Mixing Time
TROSY — Transverse Relaxation Optimised Spectroscopy
τ_s — Effective Slower Correlation Time
TS — Transition State
TSA — Transition State Analogue
UV — Ultra-Violet
WATERGATE — Water Suppresion by Gradient-Tailored Excitation
WT — Wild Type

1 Introduction and Literature Review

1.1 Introduction

β-Phosphoglucomutase catalyses the interchange of β-Glucose-1-Phosphate and β-Glucose-6-Phosphate; it is a phosphoryl transfer enzyme. Phosphoryl transfer underpins the regulation of chemical processes within the cell, yet it is extraordinarily slow without the presence of an enzyme catalyst. The mechanism by which enzymes are able to accelerate phosphate transfer reactions — or indeed any other biological reaction — remains poorly understood. Improvements in biophysical techniques have led to great interest in enzyme dynamics, and much speculations as to their potential role in catalysis. Nuclear magnetic resonance (NMR) is proving to be an extremely powerful technique, providing detailed information on the nature of protein dynamics on a wide range of timescales.

The work described herein centres around the use of NMR to study the dynamics of β-Phosphoglucomutase and how the dynamics may relate to its function as a phosphoryl transfer enzyme. This literature review first discusses the current theories as to the mechanism of function of enzymes, critically addresses the literature concerning the potential role of enzyme dynamics in catalysis, introduces the system under study and finally details the methods used throughout this thesis.

1.2 Enzyme Catalysis

1.2.1 Substrate Specificity of Enzymes

A conceptually useful way of explaining the high substrate specificity of enzymes has been the "lock and key" model, which was proposed by Emil Fischer in 1894[1]. It states that the enzyme and substrate are complementary in three-dimensional shape to one another, in much the same way that a key is designed to have complementary grooves and notches to the locks it opens. This model is depicted in Figure 1.

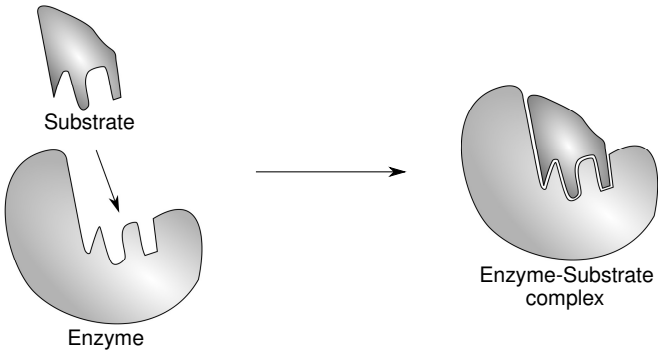

Figure 1: Diagram of the "lock and key" hypothesis, whereby the enzyme and substrate are exactly complementary in shape.

As the nature of enzyme catalysis was further studied, it was soon realised that the

"lock and key" model was too simplistic to be true. It was found that some enzymes have no activity against molecules which are identical to their native substrate but for the absence of a few atoms, e.g., a purine ring[2]. In addition, for many reactions it is necessary to exclude water, something that is not easily visualised if the substrate is designed to sit beautifully on top of its catalyst in aqueous solution.

To allow for these observations, in 1958 it was proposed that the free enzyme is not entirely complementary to its substrate, but is able to use the energy derived from binding its substrate to drive a conformational change, the result of which is complementarity between the molecules. Put differently, this model says that the free enzyme does not properly "fit" its substrate, but that binding of the substrate induces the enzyme to fit. As such, this model is called the "induced fit" model[3], and is represented in Figure 2.

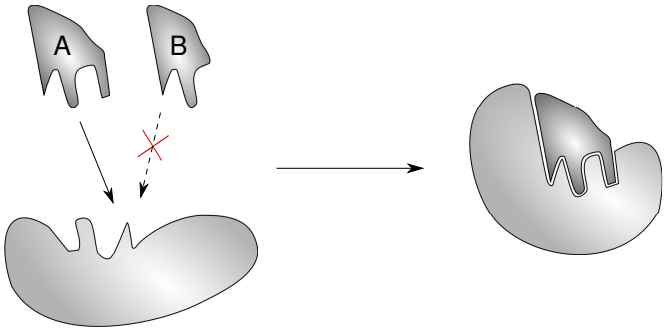

Figure 2: Diagram of the Induced Fit hypothesis, whereby binding of the native substrate (A) to the enzyme induces complementarity. Similar molecules which lack crucial interactions (such as B) may bind but will not induce the enzyme to fit, and will therefore not be catalysed.

The conformational change implicit in the induced fit model allows for water exclusion, for the enzyme to bring substrates together, and for the correct arrangement of all the catalytic groups to allow the chemistry to occur. The necessity for the substrate to induce a structural change in the enzyme also explains why, for example, it is possible to discriminate between the amino acids isoleucine and valine (which has one fewer methyl group but is otherwise identical in structure) during protein synthesis.

The "induced fit" model was conceived from theoretical work on protein synthesis before it was realised that the ribosome catalyses the synthesis of proteins — a discovery that completely revolutionised the understanding of this process — and it is therefore a credit to the model that it still persists today. However, one of its conditions is that the "active" conformation of the enzyme is found only after substrate binding; that the "active" conformation should not be present in the absence of the substrate. In the last 15 years, advances in biophysical methods have allowed studies which have shown that, in many cases, this condition is not met. Instead it is frequently found that a low population of free enzyme in a conformation similar to that when its substrate is bound exists in solution, and that the enzymes frequently fluctuate between "open"

(substrate free) and "closed" (substrate bound) conformations[4, 5, 6].

Instead of inducing a new conformation in the enzyme then, substrate binding just results in the shift of this equilibrium, such that the vast majority of the enzyme is in the "closed" conformation in the presence of ligand. It can therefore be said that the substrate is "selecting" the "closed" conformation to which it has greater affinity out of the conformations that already exist in solution. This model, proposed in 1995 to explain the kinetics of peptide binding to antibodies[7] — and first applied to the binding of enzymes to their substrates in 1999[8] — has since become known as the "conformational selection" model[9].

There have been many arguments over whether "induced fit" or "conformational selection" better represents particular scenarios, and more recently a hybrid model has been proposed[10]. In this model, a conformation of the enzyme which resembles the substrate-bound conformation already exists in solution, but when the substrate is actually added it induces further changes in the enzyme. This is likely to be nearer the truth in many cases, with pure "induced fit" and pure "conformational selection" occupying extreme ends of the spectrum of possibilities.

1.2.2 Free Energy in Enzyme Catalysis

All chemical reactions involve the reconfiguration of bonding electronic orbitals, where certain configurations are more stable than others. Transition between the stable configurations of the substrate and product during a chemical reaction will transiently sample an intermediate electronic configuration which is unstable. This is known as the transition state (TS) of the reaction, and the stable substrate and/or products are known as ground states (GSs). The relative amounts of Gibbs free energy (G) between the ground state(s) and the transition state determines the rate of reaction. Gibbs free energy is a measure of the amount of energy that can be transferred in a thermodynamic system; an unstable chemical which readily reacts is more capable of transferring energy than a highly stable unreactive chemical, and as such it has higher free energy.

By definition the transition state is an unstable entity, which therefore has high free energy. To proceed along the reaction coordinate from the substrate to the transition state there must therefore be an input of free energy, which is referred to as the "energy barrier", or activation energy ($\Delta G^{\ddagger}$), of a reaction. For reactions which occur rapidly this barrier is low, and so can be overcome by the thermal energy of the substrate(s). For other reactions, few substrates have sufficient thermal energy to pass over the energy barrier and the conversion of substrates to products is therefore slow.

The central paradigm that underpins most of modern enzyme catalytic theory was first postulated by Linus Pauling in 1948. He proposed that enzymes are not complementary to the ground state (as suggested by the models designed to explain the high substrate specificity of enzymes) but in fact complementary to the structure of the

transition state[11]. This would result in favourable interactions between the transition state and the enzyme, partially offsetting the activation energy through a decrease in the free energy of the TS. As such more substrate molecules would have sufficient thermal energy to overcome the energy barrier. Therefore, by selectively stabilising the transition state of a reaction, an enzyme is able to significantly increase the rate of reaction. This is depicted in the free energy diagram in Figure 3.

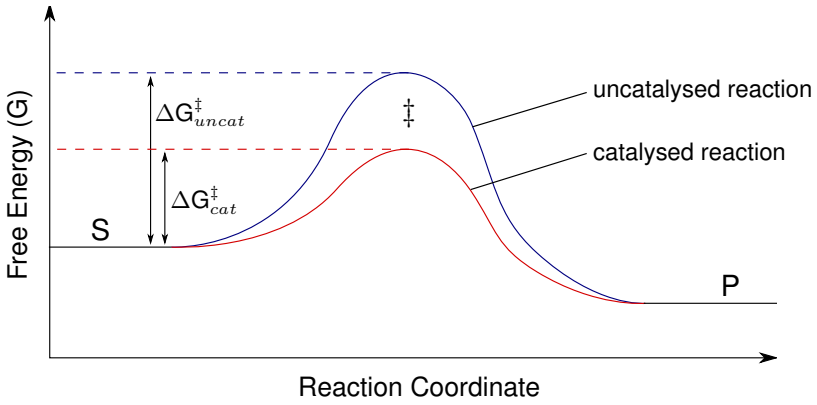

Figure 3: Free energy diagram showing the reduction in the activation energy ($\Delta G^{\ddagger}$) for a catalysed reaction (red) compared to the reaction proceeding uncatalysed (blue). The substrate, product, and transition states are represented by S, P, and the symbol ‡. It can be seen that by stabilising the transition state of a reaction — thereby lowering its free energy — the activation energy for the reaction is much lower. A larger proportion of substrates will therefore possess the thermal energy necessary to pass over the barrier and react.

Pauling's "transition state theory" was validated experimentally in 1986 from the study of so-called "catalytic antibodies". Antibodies are selected to bind to particular ligands, stabilising them in the process. Antibodies raised against transition state analogues (TSAs), which are stable chemicals which have a similar structure to the transition state, should therefore be capable of stabilising the transition state of a reaction and provide catalysis[12]. It was found that such antibodies are indeed capable of catalysing the reaction which corresponds to the TSA against which they were raised[13], although the rate enhancement is often much less than the corresponding enzyme[14].

1.2.3 Free Energy in Enzyme Conformation

Whilst 1-dimensional free energy plots such as that shown in Figure 3 are useful for illustration, they are a gross oversimplification. The "reaction coordinate" could refer to one of many parameters: electronic distributions, bond orders, angles between substrates, etc. In reality, there are many parameters which affect the transition from one molecule to another, and as such the free energy "landscape" is highly dimensional. Transition state theory dictates that in enzyme-catalysed reactions, the height of the energy barrier is dependent on the specific interactions formed between the enzyme and the transition state and therefore depends on the conformation of the enzyme.

The idea of free energy landscapes can also be applied to the conformational states of enzymes. Contrary to the static images provided by X-ray diffraction, proteins are highly mobile on a wide range of timescales, through from bond vector fluctuations on the ns–ps timescale to large-scale domain reorientations on the ms–μs timescale[15]. These dynamics give rise to a range of conformational substates which can be occupied by the enzyme, which may differ in hydrogen bonding lengths, stacking of hydrophobic rings, ionic interactions, etc. Some conformational substates may have more favourable interactions between residues than others, and as such have lower free energy. The various conformations adopted by enzymes (and the interchange between them) can be therefore be thought of as a protein sampling "conformational space" on a free energy landscape[16].

Language such as "conformational space" and "free energy landscapes" has proved particularly useful for the field of protein folding[17], where unfolded proteins are seen to sample conformational space as they progress downhill on the free energy landscape (often referred to as a "folding funnel") until they arrive at the global minimum which represents the folded state with the lowest free energy[18]. Exactly analogous to discussion of chemical reactions, the rate at which this occurs depends on the height of the energy barriers that must be overcome during the folding process.

Although for simplicity the folding funnel models tend to assume that there is a single global minimum — a single conformational state that the folded protein adopts — of course this is not true. There may be many conformational substates of a folded enzyme with similar free energies, and hence there may be many local free energy minima, or "wells" in the free energy surface, with individual transition states between each of them. Often the energy barriers between conformers are low compared to the thermal energy in the system and so various conformational substates of folded proteins exist in equilibria. These equilibria can be easily perturbed by changes in conditions like substrate binding, but also things such as temperature, pH, pressure, etc[19].

The "conformational selection" model — despite being easily understood by a simple change in the equilibrium position between two states — was the result of applying the free energy landscape approach to ligand binding[8]. The model is a useful advance on the "induced fit" model as it emphasises that folded enzymes cannot be thought of as static machines that go through discrete structural changes during catalysis. Instead enzymes are seen as soft, highly dynamic molecules which sample conformational space. On ligand binding, the relative free energies of the conformational substates is altered: those conformations which have favourable interactions with the ligand (i.e., more closely resemble a "closed" conformation) will be stabilised, and therefore more populated. This is represented in Figure 4.

1.2.4 Summary of General Catalytic Theory

Substrate Destabilisation/Strain. This theory represents the other side of the coin to transition state stabilisation, that $\Delta G^{\ddagger}$ can be decreased not only by decreasing the

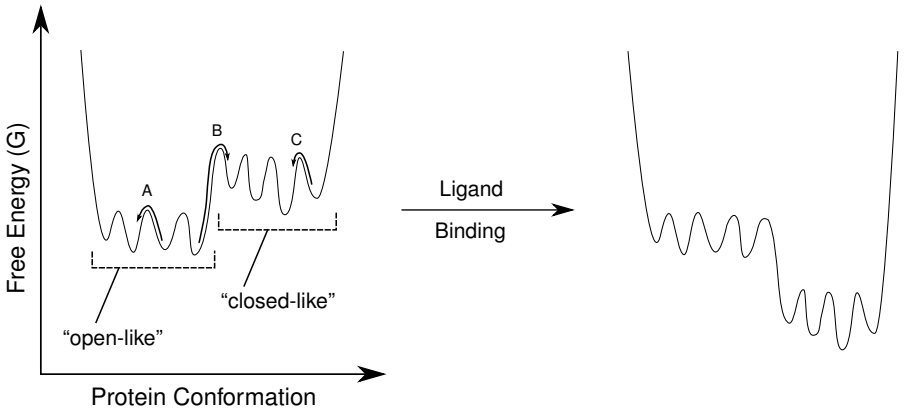

Figure 4: Conformation Selection model expressed as energy landscapes before and after ligand binding. Prior to binding, those conformational substates which are "open" have lower free energy than those which more closely resemble the "closed", ligand-bound form. It can be seen that transitions between substates of the open-like or closed-like conformations (A & C) is fast due to the small barriers between them, but transition from an open-like to a closed-like conformation has a large activation energy (B) and as such is slow. On ligand binding, the closed-like conformations have lower free energy, and so have higher population than the open-like structures.

free energy of the transition state but also by increasing the free energy of the ground states via substrate strain or destabilisation[20]. By imposing strain, the ground state can be made to adopt a more transition state-like conformation, which makes the transition between states less unfavourable[21].

The X-ray crystal structure of apo hen egg-white lysozyme was used to model the substrate binding site which showed that the substrate did not perfectly fit in the active site cleft[22]. It was only if the conformation of the fourth sugar ring was distorted from its ground-state "chair" shape towards a strained "half-chair" shape could the peptidoglycan backbone be modelled into the active site. The previously proposed carbonium ion intermediate of the reaction adopted a "half-chair" conformation, so by straining the substrate to more closely match this, the reconfiguration required for the formation of the intermediate is less, as shown in Figure 5[23]. Substrate-bound crystal structures have since revealed that the role of strain in lysozyme is more subtle: the sugar is distorted to facilitate nucleophilic attack at the C1 atom, rather than to force a more TS-like conformation in the substrate[24].

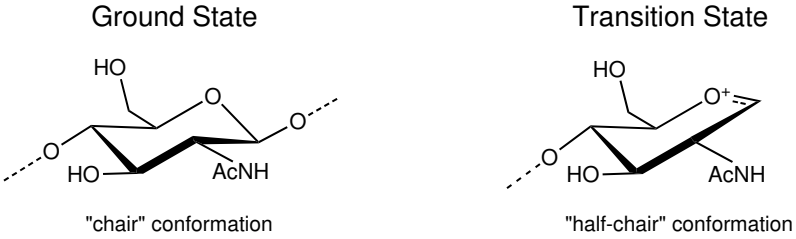

Figure 5: Substrate strain in the "old" mechanism of Lysozyme. For clarity only the sugar ring which occupies the "D" site in the Lysozyme active site is shown in both cases. The "D" ring of the substrate peptidoglycan molecule occupies the stable "chair" conformation, whilst the transition state of the reaction was proposed to be a carbonium ion in a less stable "half-chair" conformation.

Computational studies have suggested that enzymes are too flexible to impose significant strain on the substrate, and hence the contribution of strain energy to lowering the activation energy is low[25]. However, the diversity of conformations adopted by ligands bound to different enzymes demonstrates that substrate conformation can indeed be dictated by enzymes[26], and substrate strain has been invoked in more recent studies[27, 28].

Electrostatic Preorganisation. This theory suggests that the charge distribution of the TS is stabilised relative to the substrate by permenant dipoles and hydrogen bonding networks within the active site[25]. The crystal structure of hen egg-white lysozyme was again used as a model for this idea as there is an aspartate residue 3 Å away from the C1 atom of the sugar ring that was proposed to form a positively charged carbonium ion in the transition state. It has since been shown that catalysis by this enzyme proceeds not through a carbonium ion as previously proposed but through a covalent intermediate[24], though examples of the contribution of electrostatics to catalysis by enzymes are ubiquitous[29].

Although water — being a dipolar solvent — is also able to stabilise ionic charges, enzymes are suggested to be able to achieve higher stabilisation for two reasons. Firstly, the active site is predicted to have a lower dielectric constant than free solution, so the strength of electrostatic forces is larger[30]. Secondly, formation of an ordered solvation shell around charged molecules by water requires a decrease in degrees of freedom, and hence is entropically unfavourable. In proteins however, the dipoles and hydrogen bond donors/acceptors are already preorganised (an entropic cost paid for during the folding process), and therefore this cost is reduced[31].

Desolvation. For many chemical reactions, it has been found that reactions progress much faster in other solvents, such as ethanol, rather than water[32], and that solution reactions are often much slower than the equivalent reaction in the gas phase[33]. It has been proposed that similar effects could occur in enzymes where the environment of the active site differs from that of free solution. Recent evidence suggests that the reason enzymes show proportionally higher affinity for the transition states rather than the substrates of reactions is not solely due to stabilisation of the TS; also a significant effect comes from the fact that formation of the TS in bulk water requires reorganisation of water molecules which is unfavourable[34]. The active site is proposed as an environment in which the TS is shielded from the 'retarding' effects of bulk solvent, and as such the role of desolvation can be seen as a subset of electrostatic stabilisation. It has been suggested that this effect would be better described as "solvent substitution"[35], as the substrates are effectively "solvated" by active site hydrogen bond donors and acceptors.

Low Barrier Hydrogen Bonds. Normal hydrogen bonds involve a bond between the hydrogen and the hydrogen bond donor which is essentially covalent, and an electrostatic interaction between the hydrogen and the hydrogen bond acceptor. Low barrier

hydrogen bonds (LBHBs) are hydrogen bonds with high enthalpy, where the donor–acceptor distance is shorter and the character of the hydrogen–acceptor bond is much more covalent in character[36]. As a result, the energy barrier between the states corresponding to the hydrogen being proximal to either donor or acceptor is low, and the average position of the hydrogen is equally spaced from both, as shown in Figure 6[37]. Such bonds are common between molecules in organic solvents, and have been shown to occur in proteins[38].

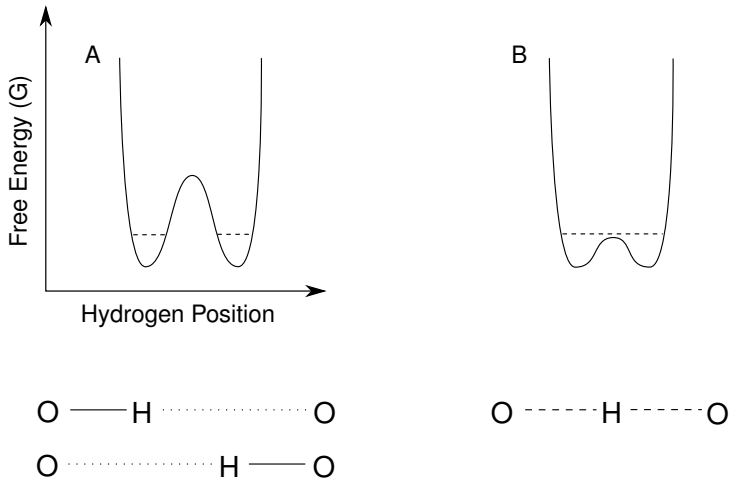

Figure 6: Low barrier hydrogen bonds represented as free energy profiles. The free energy profiles represent the energetic barrier to changes in bond lengths[40]. The horizontal dotted lines on the energy profiles represent the zero-point energy for a hydrogen atom. In usual hydrogen bonds (A), the hydrogen is covalently attached to the hydrogen bond acceptor and electrostatically interacting with the acceptor. In LBHBs (B), the donor–acceptor distance is shorter and the energy barrier between "ground state" positions of the hydrogen is lower, and on average it resides equally spaced between donor and acceptor.

The role of LBHBs in enzymes is proposed to be that the degree of LBHB formation may change throughout the reaction coordinate and therefore provide a mechanism for preferential transition state stabilisation[39]. It has also been suggested that LBHB formation could alter the pK_a of acid/base catalysts to make this process more efficient (see Section 1.2.5)[40]. However, it has been suggested that LBHBs would result in an increase in the solvation of the transition state which would disfavour its formation[41].

Substrate Entropic Effects. Enzymes bind substrates such that the correct chemical groups are positioned for reaction. The bringing of substrates together (approximation) accelerates the rate of reaction by increasing the "effective concentration" of the substrates to each other. This can be thought of as the loss of translational entropy from the substrates. The "effective concentration" of a substrate (calculated by comparing analogous inter- and intra-molecular reactions) can be very large, far exceeding 55 M[42]. This can be ascribed to favourable alignment of the necessary chemical groups in the intra-molecular case, i.e., when the reactants are tethered to one another, as often there is only a small fraction of the surface of molecules over which the reaction

can occur[43]. This idea has been referred to as "orbital steering"[44], and represents the loss of both translational and rotational components of substrate entropy, as is shown in Figure 7.

It has been suggested that the binding event may also serve to lower the activation entropy of the reaction itself[45]. However, quantum mechanical simulations have found that the formation of the transition state does not require the loss of many degrees of freedom (i.e., it is not strongly immobilised within the enzyme active site), implying that the activation entropy is not a significant determinant of catalytic rate[46].

Near-Attack Conformations. Near-Attack Conformation (NAC) theory follows from discussion of substrate entropy effects, and suggests that the activation barrier of reactions can be partitioned into two independent components: the energy required to bring the substrates within Van der Waals distance and the correct orientation, and the energy required for the chemistry to occur to form the TS[47]. Rather than the loss of substrate entropy driving the reaction as discussed above, the formation of an NAC — a conformation where all the appropriate chemical groups are aligned and positioned for reaction — is a sufficiently rare event in solution that the energy barrier to do so is comparable with the activation energy of the reaction itself. This was supported by molecular dynamics studies on chorismate mutase which showed that the enzyme exhibits little transition state stabilisation and therefore the rate enhancement must be purely a result of the formation of an NAC. Subsequent studies seem to contradict this observation though, stating the energy required to form NACs is probably low[48], and attributing the effects to electrostatic stabilisation instead[49].

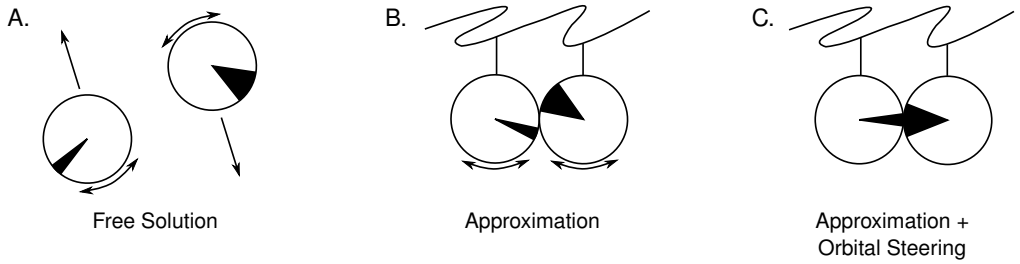

Figure 7: Diagrams to demonstrate substrate entropy effects and "orbital steering". The circles reflect molecules and the black shaded parts show the surface area on the molecule over which the reaction can occur. In free solution (A), the two molecules are unlikely to collide into each other with the correct geometry. When bound to the enzyme (B), their translational entropy is lost and so the "effective concentration" of the reactants is 55 M, leading to a rate enhancement. If the enzyme also optimises the geometry of the substrates by loss of the rotational entropy components (C) there is a further rate enhancement which is proportional to the fraction of surface area which is available for the reaction on both substrates[50]. The enzyme-substrate complex shown in C could be referred to as a near-attack conformation.

1.2.5 Direct Participation of Enzymes in Chemical Steps

All the above theories describe molecular mechanisms by which the size of the activation energy of a reaction can be decreased. In certain cases, enzymes can also

accelerate the rate of reaction by either providing alternative mechanisms that are not available in free solution, or by favouring the existing chemical mechanism by participating directly in the chemistry of the reaction.

In covalent catalysis, an intermediate which is covalently linked to the enzyme is formed, and as such an alternative energy barrier that may have lower activation energy but is not available in the uncatalysed case can be traversed[51]. However, it is not always clear why it is advantageous for the enzyme to form an intermediate rather than catalyse the reaction in a single step[42]. In addition, only a small subset of enzyme reactions are found to form covalently-linked intermediates[43]. Covalent catalysis cannot therefore be seen as a general effect, and does not describe the reduction in the height of the relevant activation barrier.

More ubiquitous is general acid/base catalysis. Acid catalysis describes the protonation of substrate by the enzyme, this makes the substrate more electrophilic and therefore more susceptible to nucleophilic attack. This often coincides with base catalysis whereby the removal of a proton from the attacking group makes it a more effective nucleophile. This is referred to as concerted acid-base catalysis[37]. Acid/base catalysis requires that the general acid/base should be able to occupy both protonated and deprotonated forms with little difference in free energy, and therefore histidine residues are commonly used as acid/base catalysts in enzyme reactions. This is because the pK_a of the hisitidine N^{ϵ} atom is close to neutral and can be "tuned" as necessary by the local environment in the active site[51]. Acid/base catalysis also enables the enzyme to avoid highly-unstable charged intermediates along the reaction coordinate[42]. Electrophiles can also be made more susceptible to nucleophilic attack by the removal of electron density by nearby metal ions[52]. The effect of acid/base catalysis on the rate of reaction is considered to be significant.

1.2.6 Need of Further Study

Despite decades of research, elucidation of the underlying mechanisms of enzyme catalysis is complicated by many issues. In many cases the mechanism of an enzyme-catalysed reaction differs from the mechanism of reaction in free solution, making direct comparison and calculation of the "enhancement factor" of an enzyme impossible[42]. In addition, the entropic and enthalpic processes involved are often impossible to measure within the microenvironment of the enzyme active site, or even in free solution[53]. Therefore, there is much debate as to how much the various chemical and environmental differences between enzymatic and uncatalysed reactions contribute to the rate difference, based on direct experimental evidence as well as quantum mechanical and molecular dynamical calculations[54]. Despite strong evidence against many of the historical theories described above, Koshland & Neet calculated in 1968 that the predicted effect of all of the theories summed together still falls short of accounting for the rate enhancement achieved by some enzymes[43]. Despite several decades of research conducted since, it is still agreed that this mystery is far from solved[55].

1.2.7 An Introduction to Enzyme Dynamics

Historically, it has been common to describe the dynamic properties of a protein as entirely separate from its structure, and as such at this point it is useful to outline exactly what is meant by "dynamics". Protein dynamics can be described as any way in which the atomic coordinates of a protein vary with respect to time. Any conformation change by a protein — however large or small — can be referred to as protein dynamics. It is therefore an extremely broad term which describes small-scale random uncoordinated atomic fluctuations, as well as concerted large-scale events such as interdomain movement. It has been suggested that large-scale concerted motions are the result of many faster uncoordinated motions which are occurring co-incidentally[56]. As the constituent faster fluctuations are uncoupled and only occur simultaneously by chance, large-scale motions are rare events and occur on slower timescales, necessitating separate methods of study.

Energy landscapes are another helpful way to visualise the difference between fast and slow timescale motions. Fast timescale motions (e.g., thermal fluctuations of backbone amide bond vectors) involve the interchange of protein conformers which are only slightly different in structure from each other, and therefore likely have similar free energy. As the structural rearrangements necessary to change between similar conformations are only very small, the activation energy of these transitions are low, and hence these fluctuations occur very frequently. Interdomain motions and other large-scale structural rearrangements, on the other hand, reflect the interchange between protein conformations which are significantly different from one another. Because of this, the interchanging conformers likely have different total free energy, and there is likely to be a large free energy barrier between them due to the number of non-covalent interactions which must first be broken before they can be formed in a different configuration. As a result of this, the interchange between, e.g., enzyme molecules which are opened and closed around a hinge region, is slow[57].

Often for discussion of a hinge opening or closing event, it is useful to only imagine two distinct protein conformations: that of the open and closed states. However, in the open state for example, there are a whole range of conformational substates which, though all being more similar to the "open" conformation consensus than a "closed" state, all differ from one another. In order to represent the true state of protein dynamics in a free energy diagram, this hierarchy of motions must be taken into account. The free energy diagrams in Figure 4 on Page 6 are something of an approximation of this idea, although in reality there is much more heterogeneity in the free energy of the conformational substates and the energy barriers separating them.

Protein dynamics have been described by some as a mechanism by which proteins are able to sample conformation space on their energy landscape; that dynamics are an intrinsic property of proteins designed to allow them to jump between alternative conformational states[58]. This discussion can be unhelpful as it divorces protein dy-

namics from the energy landscape on which they are determined. It is more useful to think of protein dynamics as the *result* of the energy landscape: if there are two minima in free energy which are the result of two distinct protein conformations, providing the energy barrier between them is low enough, they will interchange due to thermal fluctuations. Protein dynamics (like protein ground-state structure) are a result of the conformational energy landscape, not the other way around[59].

As with the other enzyme catalytic theory (as discussed in Section 1.2.4), the desire is to characterise how the enzyme alters the free energy over the reaction coordinate to accelerate the rate of reactions. This can be done by observing the free energy of different states along the catalysed reaction coordinate. The window into the variation in enzyme conformation as it cycles through free, substrate-bound and product-bound states provided by x-ray diffraction and other methods has been extremely useful in this regard but this view is too simplistic, as there are likely to be a great many other steps involved in the catalytic cycle that may not be amenable to study due to being relatively rare conformations. Stated differently: there are likely to be many other conformational states adopted by the enzyme–substrate and enzyme–product complexes which are of catalytic importance but cannot be sampled by these methods which rely on the study of the ground state. The measurement of enzyme dynamics can be seen as a method by which the free energy landscape can be explored experimentally in a way that cannot be achieved merely by the study of ground states.

The improvement in biophysical techniques over the last 20 years has enabled atomic resolution studies of protein dynamics, and the studies of enzyme dynamics have led many to speculate about their role in catalysis. In some cases enzyme dynamics are proposed to couple directly to the chemical step[60], and in other studies dynamics are seen as a means for the protein to change between the various conformations present in the catalytic cycle[61]. As these ideas have vastly different consequences and rely on motions on different timescales, they are now discussed separately.

1.2.8 Roles for Fast-Timescale Dynamics in Enzyme Catalysis

As the chemical step necessarily involves the fluctuations of relative atomic positions, it has been suggested that protein dynamics on the same timescale as the bond vibrations in bond-breaking and bond-forming steps (ps–fs timescale) may couple to crossing of the energy barrier for catalysis[62]. Hydride transfer steps are often used as examples, but these will be discussed separately due to the specific quantum-mechanical nature of the reaction. For such fast vibrations to play a role in catalysis, they would need to be directional, and therefore can be defined as non-Boltzmann coherent motions. If protein motions are linked to passage over the barrier, and the character of the transition state is determined by the exact position of the barrier that is traversed during the reaction, then protein motions should determine the character of the transition state.

This is a difficult hypothesis to test due to the extremely short lifetime of the

transition state. However, it is possible to characterise the transition state by studying the extent to which bonds in the substrate which are absent in the product are broken, and the extent to which new bonds are formed at the mid-point of the reaction, by measuring Kinetic Isotope Effects (KIEs). To take an example, suppose a C–H bond was broken during a reaction, and that some of this bond-breaking process occurs by the transition state. The bond would increase in length during passage from the substrate to the TS (according to the relationship between bond order and interatomic distance), and this increase in bond length requires energy. The zero-point energy of a deuterium atom (^{2}H, or D) is lower than that of a hydrogen atom (^{1}H), which means that if the substrate is isotopically labelled such that it is instead a C–D bond which is broken during the reaction, this process will be slower, as the input of energy required to form the TS will be larger.

How much slower the chemistry proceeds when this example hydrogen atom is substituted for deuterium gives a measure of the extent to which this example C–H bond is broken at the mid-point of reaction. This can also be done for heavier atoms (substituting ^{12}C for ^{13}C, ^{14}N for ^{15}N, ^{16}O for ^{18}O, ^{31}P for ^{33}P, etc.), and by systematically measuring KIEs at every position in the substrate, one can infer bond orders in the transition state[63]. A proof of principle for this technique has been provided by the use of KIEs measured on human and bovine Purine Nucleoside Phosphorylases (PNPs) to design transition state analogues which bind with extremely high (picomolar) affinity to their cognate enzymes[64]. Despite extremely high homology of the enzymes, their KIEs — and therefore transition states — differ. This is reflected in the relative affinities of the TSAs designed based on the KIE data for their cognate enzyme rather than the homologue.

More striking is that a chimeric enzyme (in which 3 residues distal from the active site of the human PNP are replaced with their bovine PNP equivalents) has a transition state which differs from the two parent enzymes[65]. It is therefore suggested that protein motions on the bond-vibration timescale are subtly altered by this mutation, and these dynamics are coupled to passage over the energy barrier, and thereby the structure of the transition state. It has been suggested that proteins may be evolved to display the particular dynamics which are necessary for barrier crossing. It has even been suggested that collisions of water molecules on the surface of the enzyme is channelled through the protein and coupled to catalysis-promoting motions in the active site[66].

This view of fast protein dynamics is that the enzyme undergoes a fast stochastic search of the free energy landscape until, by chance, all the necessary motions are coincident, pushing the substrate(s) along the reaction coordinate[67]. It has also been suggested that similar motions may have a more predetermined role in pushing the substrate(s) over the barrier by increasing the likelihood of conversion of the TS into product(s), rather than dropping back down the energy barrier to reform the substrate(s)[68]. The likelihood of the high-energy TS decaying into the product(s)

rather than the substrate(s) is called the Transmission Coefficient (κ), and is represented in Figure 8.

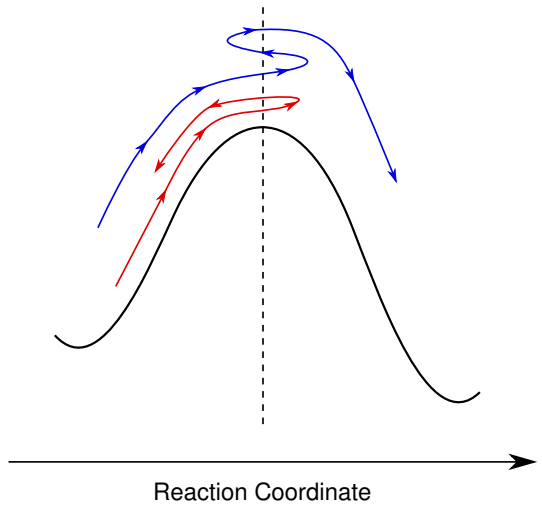

Figure 8: Representation of the transmission coefficient, κ, which is related to the number of times the barrier must be crossed before the product is generated. In the trajectory shown in red, two barrier crossings are ultimately unproductive, whereas the trajectory in blue generates the product but only after three barrier crossings.

Theoretical calculations recently demonstrated that increasing the overall kinetic energy of every vibrational mode of the cis-trans proline isomerase Cyclophilin A (CypA) gives no increase in reaction rate; increasing the kinetic energy of specific vibrational modes however does yield an increase in catalytic rate, arguably through change in κ since the directionality is so crucial[66]. Experimental evidence for this idea is lacking however, and even in the extreme case of a reaction with a low transition coefficient, the increase in the rate of reaction which could be achieved by an increase in κ is very small compared to the rate acceleration achieved by enzymes.

Another interesting example of the role of fast timescale dynamics in catalysis — although distinct from conventional "over the barrier" reactions — is that of hydride-transfer enzymes which operate by a quantum tunnelling mechanism[69], as represented in Figure 9. The rate of these reactions is more strongly dependent on the shape (particularly the width) of the energy barrier rather than its height. A high-pressure kinetic study of the enzyme Morphinone Reductase has shown that compression of the energy barrier by reduction in physical separation between the hydrogen donor and acceptor sites leads to an increase in enzymatic rate[70].

The unusual temperature-dependence of the large kinetic isotope effects (KIEs) of many of these enzymes imply that the reaction is dependent on "rate-promoting vibrations" [72], which may serve to compress the reaction barrier in a similar way to high pressure. Alternatively, enzyme motions may sample conformational space in order to find configurations with degenerate quantum states which allow tunneling[15]. Although a topic of interest, it is believed that tunnelling only applies to species of very

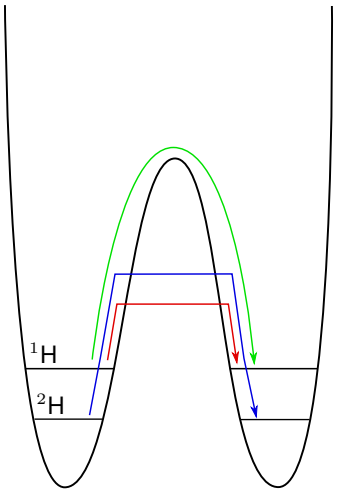

Figure 9: The role of barrier width and kinetic isotope effects in quantum tunnelling reactions. The free energy profile represents that for a hydride transfer reaction, where the left-hand well represents the hydrogen atom attached to the hydride donor molecule, and the right-hand well represents the products where the hydrogen atom is attached to the acceptor molecule. The x-axis represents the hydride position and the width between the wells therefore directly corresponds to the hydride donor–acceptor distance. The lines labelled ^{1}H & ^{2}H represent the zero-point energy of a hydride & deuteride respectively. The green arrow shows a "classical" over-the-barrier trajectory, whilst the red and blue trajectories represent quantum tunnelling for ^{1}H and ^{2}H atoms. The origin of the large KIE can be seen if the barrier width at which ^{2}H can tunnel must be smaller than that for a ^{1}H atom[71]. As the reaction rate for a tunnelling reaction depends on the width of the barrier, the rate can be increased if the substrates are pushed together by the enzyme.

low mass, and it is currently unclear as to whether or not the promotion of quantum tunnelling by enzymes would actually result in rate enhancement as a similar process is likely to occur in solution[41].

Although there is much evidence that fast protein dynamics are in some way linked to enzyme catalysis, it has been suggested by simulations that similar motions on the ps–fs timescale occur both in enzyme active sites and in free solution and even if a coupling between protein motions and barrier crossing exists, it is not a source of rate acceleration in enzymes[41]. Regardless of the role of individual fast motions, the sum of these disordered motions contributes to the overall conformational entropy (and therefore thermodynamics) of the system. Changes in the conformational entropy in proteins have been shown to play a role in the affinity of proteins for their ligands, and has also been suggested to contribute to the energetics of enzyme catalysis[73].

Nuclear Magnetic Resonance (NMR) studies have found that unlike the protein backbone, amino acid side-chain groups in proteins are often highly disordered. Study of changes in dynamics between proteins in an apo and ligand-bound state have shown that changes in overall conformational entropy make a significant contribution to ligand affinity. Interestingly, some proteins show increased side-chain and/or main chain entropy on ligand binding[74], whilst others have shown decreases[75], and in many cases the response from regions of the protein proximal to the ligand binding site and those further away are opposite[76]. These observations suggest compensation of

entropy between regions and a fine-tuning of the overall ligand affinity.

Changes in conformational entropy have been of particular interest in the study of the origins of binding energy. Despite large differences in binding enthalpy, for hydrophobic and hydrophilic ligands alike it has been shown that changes in conformational entropy between free and bound states make a significant contribution to the free energy of binding[74, 77]. A study on the carbohydrate domain of Galectin-3 demonstrated that the increase in conformational entropy on ligand binding is of the same order as the enthalpy of ligand binding. The role of conformational entropy in catalysis has also been speculated. As with ligand binding, changes in the conformational entropy of an enzyme may potentially play a stabilising role for enthalpically-unfavourable transitions during the catalytic cycle[73].

1.2.9 Roles for Slow-Timescale Dynamics in Enzyme Catalysis

E. coli Dihydrofolate reductase (DHFR) has been extensively used a model system to study the contributions of enzyme dynamics to catalysis[78]. It uses nicotinamide adenine dinucleotide phosphate (NADPH) to reduce dihydrofolate (DHF) to tetrahydrofolate (THF), which is a necessary cofactor in the production of nucleotides. DHFR shows large conformational changes between the Michaelis complex (NADPH and DHF bound) and the product complex ($NADP^+$ and THF bound), indicating that dynamics are necessary for flux through the reaction coordinate. Recently, a "dynamic knockout" of this enzyme was described whereby a N23PP/S148A mutant shows similar structural properties to the WT enzyme but a complete loss of active site ms–μs dynamics in the modelled Michaelis complex[60]. As this is coincident with a reduction in the hydride transfer rate of the mutant enzyme, it was proposed that the dynamics are necessary for the chemical step itself; that ms–μs dynamics are somehow coupled to the chemistry.

This study is only the latest in a series that assert that enzyme ms–μs dynamics can be directly coupled to the chemical step, although this has been the subject of intense debate[79]. This recent high-profile DHFR study therefore stimulated two major counter-studies. The first counter-study demonstrated that although protein dynamics are affected by the viscosity of the solvent, the chemical step is not, and is therefore not influenced by ms–μs dynamics[80]. The second rebuttal shows that despite the similar ground-state structure of the double mutant reported, the free energy landscapes of the WT and double-mutant DHFR are sufficiently different to ascribe the decrease in catalytic rate in the mutant to perturbation of active site electrostatics[59]. In contrast to fast timescale motions, it is generally agreed that there is no evidence for a mechanism of coupling between slow timescale enzyme dynamics and vibrations within the transition state during catalysis[81].

Despite its controversial conclusions, this DHFR study is not isolated in its observations. Mutagenesis experiments have frequently found that it is possible to mutate certain residues which lie far from the active site such that the enzyme is no longer

active yet the fold and active site geometry is maintained[82, 83, 84], and often this is coincident with a change in enzyme dynamics on the ms–μs timescale[85]. The complex effect of mutations which lie far apart from each other on the catalytic turnover rate has also been cited as evidence for dynamic contributions[86]. The problem with many such studies is that they often describe "dynamic mutants" of enzymes, where dynamics only (not structure) is affected by the mutation. These studies tend to ignore the fact that as the protein is sampling another conformation, it is not necessary to be able to rationalise the structural relationship between amino acids based on the ground-state structure visible from x-ray diffraction.

A more useful way to describe "dynamic mutants" is that although the ground-state structure is not affected by the mutation, the structure of another conformational species — with which the ground-state conformation interchanges during the dynamics — *is* affected. The result of this can be that the higher energy conformational species is destabilised, and therefore the population of this conformational species is decreased, such that the interchange can no longer be detected by experimental methods. As such, what appears to be a mutation which solely affects "enzyme dynamics", can actually be rationalised based on structural perturbation, as shown in Figure 10.

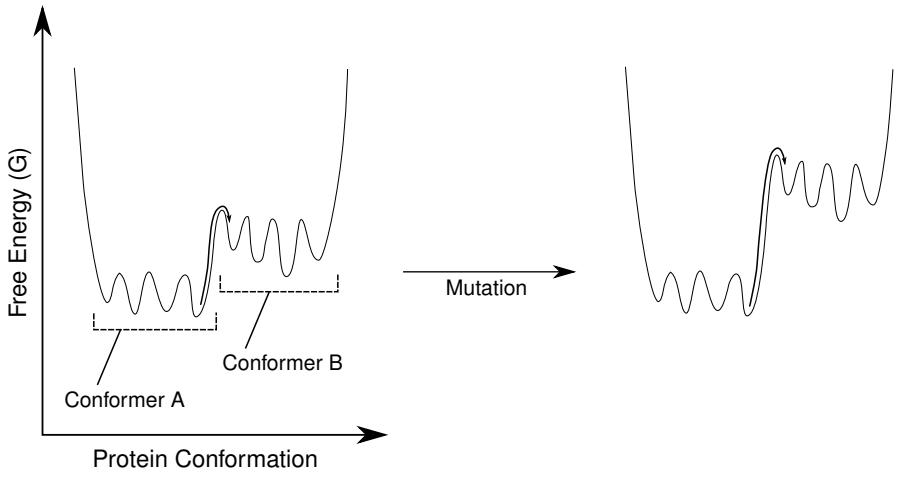

Figure 10: Structural rationale behind "dynamic mutations": if a mutation destabilises a higher energy conformer (B) by structural perturbation then although the ground state structure (A) is not affected, the interchange between the ground state and the higher energy conformer will be affected due to changes in the height of the energy barrier between them. Although this is often expressed as a mutation causing perturbation in dynamics and not structure, clearly dynamics and structure cannot be separated in this way. In this case, as the larger arrow in the case of the mutant demonstrates, the rate of conversion between low-energy and high-energy conformers would be much slower than the WT protein.

Such observations which cannot be rationalised on the basis of ground-state structures have been explained by the idea of coupled networks of amino acids in proteins which are involved in conformational changes, i.e., they interact in higher-energy conformations. It has been suggested that such networks are responsible for maintaining the appropriate dynamic motions responsible for catalysis[87]. To identify networks

of interacting amino acids involved in maintaining enzyme dynamics, there has been interest in the use of statistical coupling analyses to assess evolutionary coupling between amino acids that may be relevant to dynamics[88], and QM/MM simulations on DHFR have provided support for the idea of coupled networks of amino acids, even explaining the long-range effects of some mutations[89].

Studies of enzyme dynamics which show co-incident rates of conformational change and enzyme catalytic turnover rates have provided some insight[90, 91, 92] as to the true role of ms–μs dynamics. In the case of Adenylate Kinase (Adk), which catalyses the reversible conversion of ATP + AMP to two ADP molecules, opening of the nucleotide-binding lid (intimately linked with product release) was found to be rate limiting in both the mesophilic (*E. coli*) and hyperthermophilic (*Aquifex aeolicus*) enzymes[91]. Another example is provided by Cyclophilin A (CypA). Being a prolyl cis-trans isomerase, this enzyme is required to move around its substrate during the reaction. Whether unliganded or bound to model peptides, the enzyme undergoes ms–μs dynamics which match the rate of catalysis[90].

Both of these studies suggest that the limit of the catalytic rate is not the chemical step, but the ability for the enzyme to move around its substrate(s) and/or product(s). It may even be possible to accelerate protein dynamics by mutation such that the catalytic turnover rate is raised. Although this view of ms–μs dynamics may imply they are a hindrance to catalysis, both Adk and CypA are clearly evolved for their functions. Dynamics on the ms–μs timescale occur in Adk particularly in the nucleotide-binding lids which must open to enable product-release, and CypA is adapted such that the conformations of the enzyme required to bind the cis- & trans- form of the substrate have similar free energy, enabling bidirectional exchange between them. It has even been shown that non-native contacts exist which stabilise the transition state between ground states, therefore bringing the conformational exchange on to a biologically-relevant timescale[93].

This idea of matching of ms–μs dynamics to function is well illustrated by another study of DHFR. Although the co-incidence of rate constants is not convincing given the range of conditions under which the various kinetics and dynamics measurements have been carried out, it is suggested that DHFR has evolved to facilitate the flux of substrates and products through the catalytic cycle[92]. There are 5 distinct complexes that form during the catalytic cycle of DHFR which feature many of the possible combinations of substrates and products. The presence of the 2 products ($NADP^+$ and THF) bound to DHFR stimulates the release of $NADP^+$, but THF remains bound to DHFR until a molecule of reduced NADPH has been replaced, after which it dissociates and is replaced by DHF, forming the Michaelis complex once again and allowing the reaction to proceed. ^{15}N relaxation dispersion measurements have been done on all the possible intermediates (or in the case of the Michaelis complex, a model of the intermediate), and show that at each stage in the catalytic cycle, the conformational dynamics vary such that there is a “minor” conformer which resembles the “major”

conformer of either the preceding or following step[94]. The dynamics are therefore seen as promoting substrate recruitment and product release at each stage, in a manner which could be described as stepwise conformational selection.

These examples, as well as others, demonstrate that the role of an enzyme extends far beyond merely reduction of $\Delta G^{\ddagger}$, and that the interchange between conformational states with similar free energy is required for binding and chaperoning the substrate(s) along the reaction coordinate, before finally releasing product(s)[81]. It has been suggested that the free energy landscape of an enzyme is evolved such that each of the conformational changes which must occur during the catalytic cycle involve crossing of similarly sized barriers, such that no single exchange process represents a catalytic "bottleneck"[95]. Additionally it has been suggested that the necessity for enzyme to be stably folded puts a limit on their dynamic capabilities, which may explain why conformational exchange steps have been found to be the rate-limiting step in catalysis[56]. Stated differently, if an enzyme has high thermal stability (low free energy), it will require a large input of free energy to occupy a higher-energy conformation. This theory could potentially explain why the variation in the enzyme-catalysed rates of reaction is so much smaller than the variation in the rates of uncatalysed reactions.

1.3 Phosphoryl Transfer and β-Phosphoglucomutase

1.3.1 Cellular Role and Importance

Phosphoryl transfer has widespread and important roles in cellular metabolism and signalling and is a fundamental mechanism of control in biochemistry. Phosphoester bonds are extremely stable due to strong ionic repulsion of any potential nucleophile, necessitating an enzyme for their addition or removal. As such they make very effective molecular switches and phosphates are commonly used to modify proteins or metabolites. Phosphates can significantly alter the conformational free energy landscape of a protein such that it can be either activated or inactivated by phosphorylation due to their negative charge and ability to form hydrogen bonds[29]. Indeed, "kinase cascades" in which an upstream kinase phosphorylates its target kinase, thereby activating it and allowing it to phosphorylate and activate its downstream kinase target form an important part of many fundamental signalling pathways. Furthermore, the dephosphorylation of the so-called "energy currency" of the cell — Adenoside Triphosphate (ATP) — is used as an energy input for a wide range of cellular processes, and dephosphorylation of metabolites can be used to drive unfavourable reactions to proceed.

Although the targets of phosphorylation are numerous, kinases and phosphatases acting on other proteins alone are estimated to comprise 2% of the eukaryotic genome [96], and a third of all cellular proteins are estimated to be phosphorylated at any given time. Unsurprisingly given their fundamental role in cell biochemistry, the actions of kinases and phosphatases are implicated in the formation of many diseases and they are commonly targetted in drug screens. The disease most commonly as-

sociated with phosphotransfer enzymes is cancer as phosphorylation governs many of the "checkpoints" in the cell division cycle[29], although kinases and phosphatases are drug targets for a wide range of conditions[97].

1.3.2 Chemical Mechanism of Phosphoryl Transfer

It has previously been discussed that most enzyme catalytic events occur on a similar timescale, whereas uncatalysed reactions span several orders of magnitude in timescale. It is therefore argued that the most proficient catalysts are those enzymes which catalyse the reactions which occur slowest in free solution, as the rate enhancement — the difference between the uncatalysed and catalysed rates — is largest. To this end, phosphoryl transfer is a useful model to study enzyme catalysis as the hydrolysis of phosphate monoesters has a half-time on the order of 10^{12} years. Phosphate transfer enzymes must therefore achieve a rate enhancement on the order of 10^{21} to bring phosphoryl transfer onto a biologically-relevant timescale, in fact it is the largest enhancement factor so far identified[98]. It is therefore reasonable to expect that enzymes which catalyse these reactions make use of many of the available strategies that have evolved for rate enhancement by enzymes.

As was mentioned earlier, the primary reason for the high stability of phosphate monoesters is due to anionic repulsion of the incoming nucleophile which disfavours nucleophilic attack[99]. The electrostatics of the active site of phosphoryl transfer enzymes are therefore likely to play an important role in catalysis, and indeed the active site pockets tend to feature positively charged residues and metal ions such as Mg^{2+}, Mn^{2+} Ca^{2+}, Co^{3+}, Zn^{+} or a combination[100, 101]. Additionally, acidic residues in the active site often provide base catalysis on the attacking nucleophile.

The use of double isotopic labelling allowed the production of chiral [^{16}O,^{17}O,^{18}O] phosphate[102] whose stereochemistry could be followed either by Mass Spectrometry [103] or ^{31}P NMR[104]. In all cases, phosphoryl transfer is found to occur with an inversion of stereochemistry, whereby the phosphorus atom passes through the equatorial oxygen atoms becoming planar at the mid-point of reaction. Phosphoryl transfer occurs through "in-line" nucleophilic attack, whereby the attacking nucleophile atom, phosphorus atom, and leaving group atom are all precisely linear[99] along the vector of phosphate migration from donor to acceptor molecule, as shown in Figure 11. Substrate approximation and orbital steering (NACs) therefore likely play a role in the enzyme catalysis of phosphoryl transfer.

The order of events in a phosphate transfer step dictates the nature of the transition state, and therefore the nature of its stabilisation by enzymes. The elimination of the leaving group may precede the attack of the nucleophile, i.e., the reaction may occur through an S_N1 "dissociative" or "loose" mechanism where the transition state is a metaphosphate ion which is trivalent at phosphorus. Alternatively, nucleophilic attack may occur before leaving group dissociation in an S_N2-type "associative" or "tight" mechanism, yielding a phosphorane transition state which is pentacoordinate

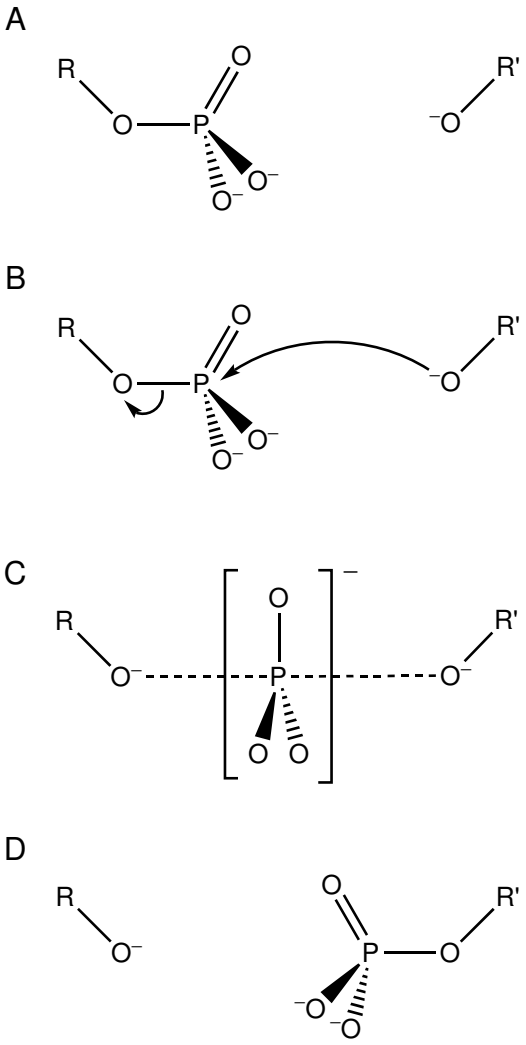

Figure 11: Example of phosphate transfer process showing: A. the substrates, B. the electron transfer steps which occur during the nucleophilic substitution mechanism, C. the transition state, D. the products. It can be seen that the transfer of the phosphate proceeds through in-line transfer with an inversion of stereochemistry. Note that the bonding and charge distributions in the transition state represent that of a dissociative phosphate transfer mechanism. The donor/acceptor oxygen atoms are referred to as the "bridging" oxygens, or the "apical" oxygens, whereas the oxygen atoms which transfer with the phosphorus atom are referred to as "non-bridging" oxygens, or the "equatorial" oxygens.

at phosphorus. These two scenarios are shown in Figure 12. Alternatively, a "concerted" mechanism may occur where bond formation and bond elimination are coincident. In reality these three mechanisms are just three scenarios on a spectrum of possibilities[105].

If the phosphoryl transfer mechanism is dissociative (S_N1-like), then the rate of reaction depends on the ability of the leaving group to accept electrons from the dissociating phosphorus atom. Deprotonation of ionisable groups depends on the ability of the group to accept electrons from their attached hydrogen atom, thereby forming a hydrogen ion (proton, H^+). As such, the ability of the leaving group to accept electrons from the dissociating phosphorus atom is reported in the pK_a of the leaving group.

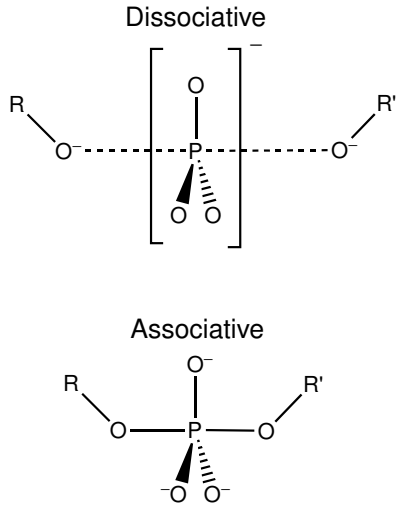

Figure 12: The transition states for dissociative and associative mechanisms of phosphate transfer. In the extreme dissociative case, the bond order between the phosphorus atom and the non-bridging oxygens is 5/3, and the bond order between the phosphorus atom and the bridging oxygens is 0. In the extreme associative case, the bond order between the phosphorus and all oxygen atoms is 1. In accordance with the differences in bond orders in the two mechanisms, the bond lengths are also different. The mechanisms also differ in charge distribution, with all negative charge accumulating on the non-bridging oxygens in the associative case, and negative charge also being localised to the bridging oxygens in the dissociative mechanism.

Alternatively, if the phosphate transfer step occurs through an associative (S_N2-like) mechanism, then the rate of reaction is dependent on the pK_a of the attacking nucleophile, which reports on its ability to donate electrons to phosphorus. This is called the Brønsted relationship.

Other experimental parameters which are often used to determine the character of the transition state (as discussed earlier in Section 1.2.8) are KIEs. If the "bridging oxygen" of a phosphate, i.e., that which links the phosphorus atom and the leaving group is substituted for ^{18}O, the average length of the P–O bond will decrease, due to lower zero-point energy of the heavier isotope of oxygen. As such, the increase in the length of this bond required to break it will require more energy, and will therefore occur slower. Therefore, if the KIE of substituting the bridging oxygens for ^{18}O is large, this demonstrates that the leaving group is largely eliminated in the transition state, which is therefore dissociative in nature.

It is also possible to label the non-bridging oxygens with ^{18}O and measure KIEs. If the transition state is dissociative in character with a trivalent phosphorus atom, then the bonds between the phosphorus atom and the non-bridging oxygen atoms will have some double-bond character, and will therefore be shorter. This is actually favoured by the heavier isotope of oxygen, and as such "inverse KIEs" would be measured for such a transition state, with the reaction occurring faster with the ^{18}O-substituted substrate. Conversely, if the ^{18}O KIE for the bridging oxygens is low but the non-bridging oxygens is not inverse, then the transition state is associative in character.

In non-enzyme-catalysed hydrolysis of phosphate monoesters, the dependence of the reaction rate on the pK_a of the leaving group, but not on that of the nucleophile, along with the large KIEs resulting from ^{18}O-labelling of the bridging oxygen are consistent with a reaction mechanism that is dissociative in character[106, 105]. There is, however, no evidence of the formation of a metaphosphate intermediate: instead a metaphosphate-like species is expected to correspond to the transition state of the reaction in solution[106].

As was mentioned earlier, there is no reason to assume that the reaction mechanism in solution will be the same as that which occurs in the enzyme-catalysed case. It has been suggested by quantum mechanical simulation that the activation barrier for either dissociative or associative mechanism is comparable in height[107], such that an enzyme could easily select an associative pathway. There has been the suggestion that the positively charged active sites of phosphoryl transfer enzymes may be better at stabilising associative, phosphorane-like, transition states as this mechanism increases the negative charge on the non-bridging oxygens[108, 105]. As such, there is great interest in investigating whether the enzymatic mechanism of phosphoryl transfer is associative or dissociative in character. Unfortunately, this is much more difficult than establishing the behaviour in the solution.

Establishing the Brønsted relationships for the nucleophile and/or leaving group relies on being able to use a variety of molecules with differing pK_as, and is impossible in enzymes which show high substrate specificity. The analysis of KIEs for the non-bridging oxygen atoms is also clouded by changes in ionisation of the phosphate during catalysis[109]. As such the mechanism of enzymatic phosphoryl transfer remains unsolved, and what data is available suggests that different enzymes use different mechanisms, with some reflecting the dissociative character of the solution mechanism and others forming more associative transition states[106].

1.3.3 Metal Fluorides as TSAs of Phosphoryl Transfer

For any enzyme-catalysed reaction, the nature of transition state analogues and the interactions formed in complexes with their cognate enzymes can give information on the way in which the reaction proceeds. Transition state analogues of phosphoryl transfer enzymes have potential value in being able to discriminate between associative and dissociative mechanisms, which differ in terms of bond order and charge distribution of the transition state[110].

It has been known for many years that F^- ions can inhibit the activity of phosphoryl transfer enzymes[111], and that G-proteins become activated in the presence of F^- ions[112]. It was subsequently shown that the mechanism of F^- ions is metal-dependent with Al^{3+} ions being the common metal[113], forming AlF_x ions which mimic the γ phosphate when bound to G-proteins, thereby activating them[114]. It was shown by X-ray crystallography that $AlF_4{}^-$ ions form octahedral complexes, and thereby mimic the geometry of the transferring phosphate ion during hydrolysis and not the ground-

state GTP molecule[115].

Structures of TSAs of phosphate-transferring enzymes with AlF_x ions bound are ubiquitous in the Protein Data Bank (PDB), although there is some ambiguity over the coordination of the Al^{3+} ion. In some crystal structures an octahedral AlF_4^- ion is reported, in others a trigonal bipyramidal AlF_3^0 ion. As has been discussed previously, the conservation of charge is highly important in the active sites of phosphoryl transfer enzymes, and so ambiguity of the charge on a transition state mimic is unusual. It was noticed that the crystal structures which contained neutral AlF_3 ions were formed at high pH values[110], at which Al^{3+} ions are known to precipitate out of solution as $Al(OH)_3$[116].

The ability of F^- ions to activate G-proteins had previously been shown to be dependent on the concentration of Mg^{2+} ions[117], and inclusion of the Al^{3+}-chelator deferoxamine to crystallisation buffer demonstrated that it is indeed possible to form MgF_3^- TSAs of phosphoryl transfer in the absence of Al^{3+} ions[118]. It is therefore suggested that many if not all of the reported AlF_3-bound protein structures in the PDB solved at high pH are infact MgF_3^--based TSA structures instead. Unlike AlF_4^- ions, MgF_3^- ions are both isoelectric and isosteric with a transferring PO_3^- moiety, and are therefore better transition state mimics, as shown in Figure 13. The fact that AlF_4^- ions bind to phosphate transfer enzymes as TSAs at all demonstrates that much of the stabilisation of the transition state of the reaction arises from electrostatic rather than steric effects.

Although MgF_3^- ions are a truer mimic of the transferring phosphate than AlF_4^- ions, AlF_4^- TSA complexes tend to dominate over MgF_3^- TSA complexes. The concentration of Al^{3+} ions required to form a AlF_4^- TSA complex is much lower than the concentration of Mg^{2+} ions required to form a MgF_3^- TSA, and aqueous F^- ions have the ability to leach Al^{3+} ions from glass. The reason for the seemingly poorer affinity of MgF_3^- ions is that the formation constant for MgF_3^- ions in aqueous solution is very low, meaning that they are only present at extremely low concentrations. It is proposed that the precise environment of a phosphate transfer enzyme active site is required for its formation, and that the complex ion assembles on-enzyme[119].

As well as geometry and overall charge, another important consideration in the validity of these TSAs for the phosphoryl entity they are designed to mimic is the bond lengths. The bond lengths from phosphorus to the non-bridging (equatorial) oxygens in a ground-state tetrahedral phosphate are 1.6 Å. The Al–F bonds in an AlF_4^- moiety are 1.7 Å, and the Mg–F bond lengths in a MgF_3^- moiety are even longer, at 1.9 Å. The P–O bond lengths for a phosphate mid-flight depend on the character of the transition state: an associative mechanism would result in longer P–O bonds, whereas a dissociative mechanism would yield P–O bonds which are more double-bond-like in character, and therefore shorter[105].

The fact that phosphoryl transfer enzymes bind these metal fluoride ions whose

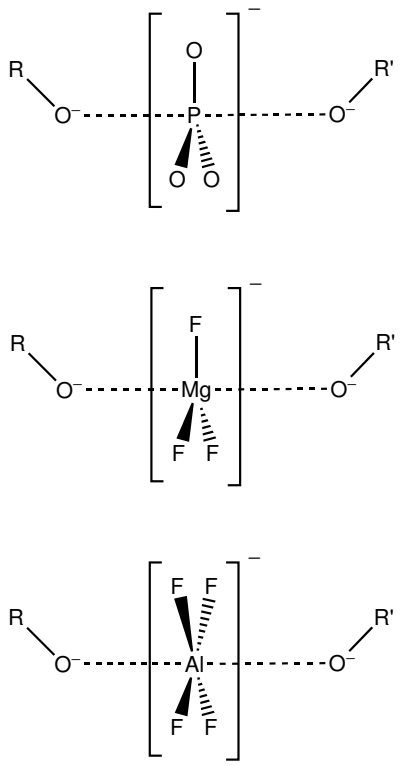

Figure 13: MgF_3^- and AlF_4^- ions as TSAs of phosphoryl transfer. It can be seen that although both MgF_3^- and AlF_4^- ions are isoelectric with the transferring phosphate, only MgF_3^- ions are isosteric, and are therefore truer transition states. Note that a dissociative mechanism is assumed in this diagram.

metal–fluoride bond lengths are longer than that of a ground state phosphate P–O bond may be evidence that the enzymatic mechanism of phosphoryl transfer is more associative than its solution counterpart. This is also supported by the bond lengths between the metal and the leaving group oxygen in the crystal structures solved of metal fluoride TSA complexes so far. In all cases, the bonds are relatively short, implying high degree of bond formation between the bridging oxygens and the phosphorus at the transition state, i.e., an associative mechanism generating a phosphorane-like transition state. Although it is of course the case that any structures of these TSA complexes will be the product of the ground-state energies of the analogue rather than the true transition state, it is particularly striking in a crystal structure of a MgF_3^- TSA complex of the enzyme Phosphoglycerate Kinase (PGK) that the O–Mg–O bonds appear distorted from linearity, suggesting that the enzyme is forcing the donor and acceptor oxygens closer together than the Mg–O bond lengths will allow, and hence distoring the MgF_3^- moiety[120].

In addition to AlF_4^- and MgF_3^- complexes, there have been reports of scandium fluoride (ScF_x) and gallium fluoride (GaF_x) transition state analogue complexes of phosphoryl transfer enzymes[121]. Metal fluorides — specifically trifluoroberyllate (BeF_3^-) anions — can also be used to mimic ground state complexes: as beryllium cannot access 3d orbitals, BeF_3^- ions have obligate tetrahedral geometry which mimics ground state phosphate ions[122]. As a result, BeF_3^- ions have been used as tools

to form mimics of phosphorylated proteins, which often then change into an "active" conformation[123], or the γ phosphate of ATP or GTP in a similar way to that which was first proposed for the activity of AlF_4^- ions for G-proteins[124].

Although metal fluoride analogues have been primarily used in structural studies by X-ray crystallography, the introduction of ^{19}F nuclei into the active sites of phosphate transfer enzymes makes ^{19}F NMR an extremely powerful tool for the study of GSA and TSA complexes. ^{19}F is a spin-$1/2$ nucleus with 100% natural abundance. It is an extremely sensitive nucleus, only slightly less sensitive than ^1H ($\gamma_F/\gamma_H = 0.941$). It also has a chemical shift range of nearly 1000 ppm[125], far exceeding that of ^1H, which is only 15 ppm. As a result, ^{19}F δs are extremely sensitive to their environment, providing excellent probes for surrounding electron and proton densities, and therefore electrostatics in the active site.

1.3.4 β-Phosphoglucomutase

β-Phosphoglucomutase (βPGM), shown in Figure 14, is a member of the Haloalkanoic Acid Dehalogenase (HAD) superfamily taken from *Lactococcus lactis*. βPGM interconverts β-Glucose-1-Phosphate (βG1P) and β-Glucose-6-Phosphate (βG6P) via a β-Glucose-1,6-Bisphosphate (βG16BP) intermediate. This interconversion is an important process which links glycolysis to synthesis/breakdown of exopolysaccharides[126], and also forms part of the maltose degradation pathway[127]. Although most members of the HAD superfamily act as hydrolases, βPGM is unusual in that it functions as a mutase, forming a stable phosphoenzyme intermediate allowing subsequent transfer of the phosphate group to an acceptor hexose sugar[128]. It is a 24 kDa protein consisting of 2 domains: the core domain (residues 1–14 & 93–221) and the cap domain (residues 15–92), which move relative to each other about a hinge region to enclose the substrate within the active site.

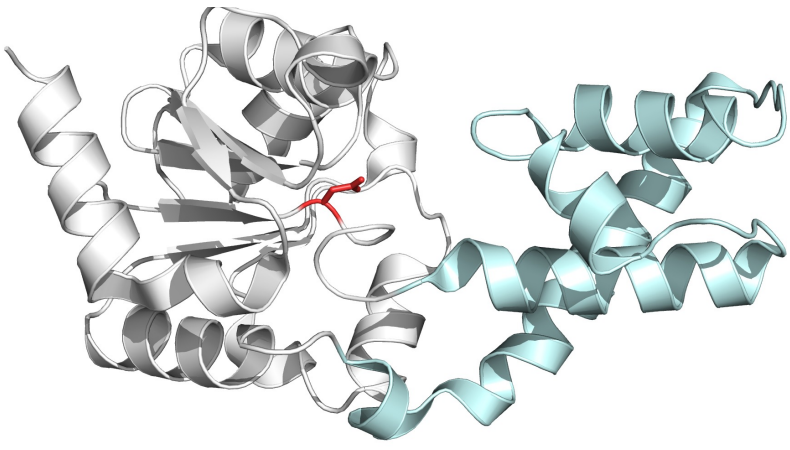

Figure 14: Image of apo βPGM showing the core (white) and cap (blue) domains. The active site is located at the domain interface, and displayed by the aspartate residue which becomes phosphorylated (D8), which is coloured red. This image was made using the co-ordinate set with PDB code 2WHE.

The catalytic cycle of βPGM is best considered with the phospho-βPGM as a starting point, and although βPGM can catalyse the conversion of βG1P and βG6P in either direction, for illustration the βG1P $\rightarrow$ βG6P direction is considered, and illustrated in Figure 15. βPGM is phosphorylated at conserved residue D8 which acts as a phosphate donor to βG1P, forming the bisphosphate intermediate. The $1'$ phosphate is then removed by the enzyme, regenerating the phospho-βPGM and forming βG6P. In kinetic analyses, the βG16BP intermediate has been shown to be required for the activation (phosphorylation) of βPGM[129], although the cellular source of bisphosphate is unknown. In principle the enzyme could also be phosphorylated (albeit at a slower rate) by βG1P in vivo[130].

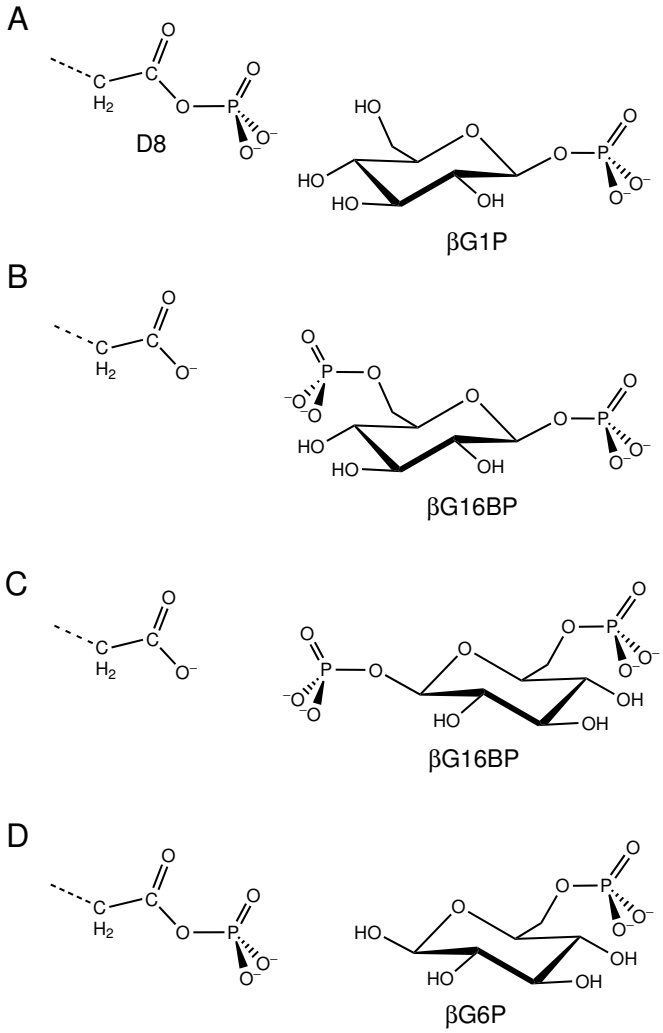

Figure 15: The catalytic of βPGM, operating in the βG1P $\rightarrow$ βG6P direction. A. βPGM, phosphorylated at conserved aspartate D8 binds βG1P. B. nucleophilic attack on the phosphate by βG1P leads to phosphoryl transfer (as depicted in Figure 11), generating the bisphosphate intermediate: βG16BP. C. βB16BP dissociates and rebinds in an alternate conformation, presenting the $1'$ phosphate to the nucleophile D8. D. After phosphoryl transfer which regenerates phosphorylated βPGM, and forms βG6P.

The apparent symmetry of the active site residues in crystal structures of βPGM

initially gave rise to speculation that the enzyme has two regions capable of catalysing phosphoryl transfer at the 1′ and 6′ positions respectively[128]. This has since been found to be false by kinetic analysis which demonstrates that, in the βG1P $\rightarrow$ βG6P direction, the formation of βG6P is dependent on the concentration of βG16BP[129]. Therefore, after generation of βG16BP this intermediate is released by βPGM, allowing it to bind in an alternative conformation which presents the 1′ phosphate to the phosphate acceptor residue D8, to which it is transferred. It has since been shown by x-ray crystallography by Dr. Matt Bowler that βG1P and βG6P do indeed bind to βPGM in different conformations.

Much of the detail of the catalytic mechanism of βPGM has come from crystal structures with various ligands bound, mimicking various points in the catalytic cycle. βPGM is proposed to use acid-base catalysis by another conserved active site aspartate residue — D10 — to activate the nucleophile to enable phosphoryl transfer. Against the background of two essential acidic residues in the active site (D8 & D10), βPGM retains a net positive charge in the active site by the presence of K145 and — in line with many phosphoryl transfer enzymes — a catalytic Mg^{2+} ion. Glucose phosphates are coordinated by residues both from the cap and core domains, and as such substrate binding is linked to active site closure.

Closure of the active site around its substrate requires rearrangement of the hinge regions between the two domains, which in turn disrupts a hydrogen bonding network between residues in the hinge and the side chain of the base catalyst D10[130]. It is seen that in open, unliganded structures of βPGM, D10 points away from the active site and towards the hinge. On closure, D10 rotates into the active site where it can promote catalysis. In this way, substrate binding is directly linked to a conformational change that is essential for catalysis. This means that unlike other members of the HAD family which function as phosphatases, the acid-base catalyst is not available to activate water molecules when βPGM is phosphorylated, which results in a slow rate of hydrolysis of the phospho-enzyme, around 0.05 s^{-1}[130]. This conformational switch is vital in the function of βPGM as a mutase, and not as a phosphatase, i.e., its ability to discriminate between water and βG1P/βG6P as a phosphate acceptor.

As with all enzymatic phosphoryl transfer reactions, the character of the transition state is the subject of intense discussion. A crystal structure of βPGM was solved which seemed to contain a pentacoordinate phosporane species in the active site, strongly suggesting that the reaction mechanism was associative in character[131]. However, a crystal structure recorded at cryogenic temperature represents the conformation at a global energy minimum, and the proposed phosphorane species would represent the transition state of the reaction; despite the network of strong hydrogen bonds proposed to stabilise the structure[132], this remains an inherently unstable species. Given the crystallisation conditions used in the study, which included NH_4F and $MgCl_2$, it was suggested that the electron density attributed to a phosphorane species in the crystal structure of βPGM was instead the result of an MgF_3^- transition state analogue[133].

Although this re-assignment of the electron density was rebutted repeatedly[134, 135, 136], there is overwhelming evidence from crystallography, solution-state NMR, and computational studies that it is correct. ^{19}F NMR experiments performed under the same conditions used by Lahiri and co-workers unambiguously demonstrate the presence of three non-exchanging protein bound F^- ions, which have been assigned as occupying the positions of the oxygen atoms in the proposed phosphorane structure by transfer NOE methods[119]. Quantum mechanical simulations also demonstrate that an MgF_3^- species would be stable in the active site, but the proposed phosphorane would not[137, 108]. It has even been shown by re-analysis of the electron density recorded in the original study that it is more consistent with an MgF_3^- entity than a phosphorane[138, 139]. It is also likely that a previously published structure which was described as phosphorylated βPGM[128] is more likely to be the result of an AlF_4^--bound species instead[138].

As well as giving insight into the nature of the mechanism of catalysis by phosphoryl transfer enzymes, these metal fluoride complexes allow for the use of ^{19}F NMR to probe the metal fluoride moieties. It has been shown that there are many orthogonal measurements which can be made which give access to extremely sensitive information about the geometry, electrostatics, and hydrogen bonding arrangements of the active site through measurements of chemical shifts, solvent-induced isotope shifts (SIISs), and J-couplings[138]. As such the use of metal fluorides opens up new methods to study the active site in detail in the solution-state, and the variety of metal fluorides available allows a detailed description of the structural and dynamical changes during the catalytic cycle of βPGM.

1.4 NMR in the Study of Protein Dynamics

1.4.1 Introduction to NMR

Nuclear Magnetic Resonance (NMR) is the phenomenon which results from the property of nuclear spin; when nuclei which possess spin are held in a magnetic field they begin to precess. This precession can be detected by the application of radiofrequency pulses and is measured by NMR spectroscopy. Since its discovery in 1945[140, 141] NMR has proved a useful tool with a wide variety of applications. NMR in structural biology is an extremely powerful technique with which it is possible to determine protein structure and also observe time-dependent phenomena such as dynamics, molecular binding events, and catalysis.

All nuclei with a non-zero spin number (I) possess nuclear spin, and can therefore be detected by NMR. The nuclei typically studied in solution are spin-$^1/_2$ which possess two spin states (commonly referred to as "spin-up" and "spin-down"). Examples of such nuclei are ^{1}H, ^{13}C, ^{15}N, ^{19}F, and ^{31}P. In a magnetic field there is a slight excess of nuclei in the "spin-up" state compared to the "spin-down" state, and the size of this polarisation depends on the gyromagnetic ratio (γ) of the nucleus, as well as the strength of the external magnetic field, B_0. Nuclei which have a non-spherical charge

distribution have a coupling to the surrounding electronic field gradient, and as a result of this there are many more non-equivalent spin states that they can occupy. Such a nucleus is the deuteron (^{2}H) which — being a spin-1 nucleus — has three spin states it can occupy. As in the "spin-up" and "spin-down" case of spin-$^1/_2$ nuclei, these spin states are not equivalent in energy which gives rise to polarisation.

Although NMR-active nuclei precess in a magnetic field, they precess out of phase with each other, resulting in no net magnetisation and no observable signal. In order to observe an NMR signal, radiofrequency (RF) pulses applied perpendicular to the external magnetic field is used to force precession of all like spins to become in phase, which then induces a current in the surrounding detection coil. The rate at which spins precess depends on their γ, the B_0 field, and also their chemical shift (δ). The chemical shift arises because of the ability for surrounding electrons to shield nuclei from the static B_0 field. As a result, the external field that is experienced by the nucleus, and therefore its precession frequency, is dependent on the electron density which surrounds it. As a result, the chemical shift (δ) is an extremely useful parameter which reports on the chemical environment of the nucleus under study.

1.4.2 NMR to Study Proteins

By bacterial expression of protein in the presence of labelled nitrogen and carbon sources, it is possible to produce uniformly labelled proteins in which NMR signals originating from many atomic positions in the backbone ($^1H^N$, $^1H^\alpha$, ^{15}N, $^{13}C'$, $^{13}C^\alpha$) and also the side-chains of amino acids can be studied[142]. However the number of distinct chemical environments — resulting from, e.g., different hydrogen bonding — makes signal overlap a significant problem, and the assignment of these signals to their position in the protein difficult. The feasibility of protein NMR studies was greatly improved by the development of heteronuclear multi-dimensional NMR. This allows for correlation of δs from one nucleus to that of their neighbour. For example, in the ubiquitous ^{1}H–^{15}N Heteronuclear Single Quantum Correlation (HSQC) experiment, a two-dimensional spectrum with ^{1}H δ on one axis and ^{15}N δ on the other is generated[143].

^{1}H–^{15}N HSQC spectroscopy is extremely useful for the study of ^{15}N-labelled proteins, as most amino acids yield a single signal from their backbone amide group. The exceptions to this are Proline residues, which due to their cyclic side chain lack an amide proton and therefore produce no signal, and also amino acids which have side-chain amides, guanidinium or indole groups, which produce two signals. There is also a multitude of 3D experiments which allow for the correlation of ^{1}H and ^{15}N δs with those from $^{13}C'$, $^{13}C^\alpha$, and/or $^{13}C^\beta$ sites. In addition, it is possible to study aliphatic side-chains by the use of ^{1}H–^{13}C correlation spectroscopy.

^{15}N and ^{13}C δs provide information on backbone and side-chain dihedral angles and therefore give structural constraints. Therefore, by measuring chemical shifts at each of these sites, it is possible to calculate protein structures. For high-resolution

structure calculations, the chemical shifts are often used in concert with values of the Nuclear Overhauser Effect (NOE, discussed later) which gives distance restraints, and also Residual Dipolar Couplings (RDCs, not discussed in this thesis) which give long-range angular restraints. Using NMR it is possible to solve protein structures in solution to a resolution which is comparable to X-ray diffraction of protein crystals.

1.4.3 Introduction to Relaxation

When nuclei are pulsed in order to generate transverse magnetisation which can be detected, two separate effects occur: the polarisation, i.e., slight excess in "spin-up" spins, is abolished; secondly, the precession of all like nuclei is brought into phase. After the RF field is switched off, both of these effects reverse with time. The restoration of the polarisation is referred to as longitudinal relaxation as it concerns the bulk magnetisation along the axis of the external field. It occurs with rate constant R_1. The dephasing of precessing spins is referred to as transverse relaxation and causes NMR signals to decay with rate constant R_2. In order to restore equilibrium polarisation, nuclei must interchange between spin states, a process that necessarily causes dephasing of like spins. Therefore, R_2 is at least as fast as R_1, and in the case of proteins, it is much faster.

The ^{15}N–$\{^1$H$\}$ Heteronuclear NOE (hetNOE) measurement consists of saturating the ^{1}H signals, i.e., pulsing them repeatedly such that polarisation is lost but there is no transverse magnetisation, and observing the effects on the polarisation of the ^{15}N nuclei. The restoration of the equilibrium ^{1}H polarisation through changes between spin states can occur through the simultaneous changes in the spin states of the adjacent ^{15}N nuclei.

Because the precession frequency of a nucleus is directly related to the differences in energy between its spin states, the processes that give rise to R_1 and R_2 relaxation and also the NOE depend on local fluctuations in the magnetic field on the timescale of the precession frequency of the nucleus under study: the ns–ps timescale. Fluctuations in the magnetic field experienced by a nucleus can result from various effects. One such effect is chemical shift anisotropy (CSA) which is the result of non-spherical distribution of electrons around the nucleus, the result of which is that the chemical shift of a nucleus depends on the orientation of the surrounding electrons with respect to the external field.

Another source of fluctuation is dipole-dipole effects (DD): because the local magnetic field of one nucleus is influenced by the direction of polarisation of another, a change in relative orientation of these two nuclei is another source of fluctuation. For nuclei with spin number $I > 1/2$, fluctuations in the relative orientation of the nucleus with the surrounding electric field gradient (to which it is coupled) are another source of relaxation, called quadrupolar relaxation. Finally, dephasing of spins giving rise to transverse relaxation can also result from chemical exchange phenomena in which a nucleus is alternating between 2 or more environments with different preces-

sion frequencies. Relaxation due to exchange (R_{ex}) is useful as it provides a method of quantifying exchange processes occuring on a slower timescale to nuclear precession (the s–μs timescale), but it complicates analysis of fast-timescale dynamics.

1.4.4 Relaxation Measurements for Fast Timescale Dynamics

Although there is a wide range of choice between spin systems in proteins whose relaxation could be studied, analysis of relaxation data becomes more difficult as there are more contributions to the relaxation process. Backbone amide ^{15}N relaxation is dominated by DD interactions with the $^{1}H^{N}$ nucleus and the CSA of the ^{15}N nucleus in the N–H bond, and can therefore be treated as an isolated two-spin system. This greatly simplifies the interpretation of ^{15}N relaxation rates and as a result, the study of fast-timescale protein backbone dynamics is routinely done on ^{15}N relaxation[144].

Although cross-correlated relaxation processes can cloud the analysis of ^{13}C relaxation data, the measurement of $^{13}C'$–$^{13}C^{\alpha}$ dipolar relaxation — which is dependent on fluctuations in the C'–C^{α} bond vector — allows the detection of motion perpendicular to that measured by ^{15}N relaxation, and these methods therefore complement each other. Discrepancies between order parameters derived from ^{15}N and $^{13}C'$ relaxation have been used to infer local anisotropic motion, e.g., rotation of an α-helix about its axis[145]. Although potentially useful, the study of $^{13}C'$ relaxation is relatively rare and is not discussed in this thesis.

To quantify dynamics in the N–H bond vector, the ^{15}N R_1 and R_2 rates as well as the hetNOE are measured at at least 2 different field strengths. As the precession frequencies of the nuclei under study specify the exact timescale of motions to which these rates are sensitive, the same relaxation measurements at different field strengths inform on different dynamics. The variation in a bond vector with time due to dynamics can be described by a correlation function, the Fourier transform of which is called a spectral density function. The values of spectral density at the frequencies which correspond to the energies required for the processes which cause relaxation can be used to calculate the relaxation rates, as shown in Figure 16.

The most routinely studied probe for the measurement of side-chain dynamics has been ^{2}H relaxation in CH_2D methyl groups. Because the relaxation of the ^{2}H nucleus is dominated by the quadrupolar interaction, other effects can be ignored and so the analysis of this data is much simpler than ^{13}C relaxation of CHD_2 methyl groups which was the previously preferred method[146]. ^{2}H relaxation is dominated by variation in the orientation of electric field gradient with respect to the ^{2}H nucleus, and therefore B_0. As the ^{2}H is quadrupolar, possessing three spin states, there is greater opportunity for changes in spin state, and therefore more relaxation measurements that can be performed: a total of five. two of these are analogous to the R_1 and R_2 measurements performed on spin-1/2 nuclei[146], but the other three are more complicated[147]. Unlike the case of ^{15}N relaxation as shown in Figure 16, ^{2}H relaxation is dominated by the quadrupolar interaction and is relatively insensitive to DD interactions. It is therefore

$$R_1 = (d^2/4)\left[J(\omega_H - \omega_X) + 3J(\omega_X) + 6J(\omega_H + \omega_X) + c^2 J(\omega_X)\right]$$
$$R_2 = (d^2/8)\left[4J(0) + J(\omega_H - \omega_X) + 3J(\omega_X) + 6J(\omega_H) + 6J(\omega_H + \omega_X)\right] + (c^2/6)\left[4J(0) + 3J(\omega_X)\right] + R_{ex}$$
$$NOE = 1 + (d^2/4R_1)(\gamma_X/\gamma_H)\left[6J(\omega_H + \omega_X) - J(\omega_H - \omega_X)\right]$$

where:

$d = (\langle r_{XH}^3 \rangle / 8\pi^2)\mu_0 h \gamma_X \gamma_H$

$c = \omega_X \Delta\sigma / \sqrt{3}$

μ_0 is the permeability of free space

h is Planck's constant

γ_H & γ_X are the gyromagnetic ratios of ^{1}H and X (e.g. ^{13}C or ^{15}N), respectively

ω_H & ω_X are the Larmor frequencies of ^{1}H and X spins, respectively

r_{XH} is the X-H bond length

$\Delta\sigma$ is the chemical shift anisotropy of the X spin

Figure 16: R_1, R_2, and NOE expressed in terms of spectral density functions[140].

not sensitive to the spectral density at frequencies corresponding to any other spins, and only sensitive to spectral density at J(0), J(ω_D), and J($2\omega_D$)[147].

1.4.5 Model Free Analysis

Molecular tumbling in solution is a source of relaxation of both ^{15}N and ^{2}H nuclei, as it provides a mechanism for variation in bond vectors with respect to B_0. As a result, for the analysis of both backbone and side-chain dynamics it is necessary to have a detailed description of the rotational motion of the molecule of interest, so that the contribution to the relaxation rates from brownian diffusion in solution can be separated from the contributions of more interesting intramolecular dynamics. There have been many ways proposed to do this, and the most frequently used is the "Model-Free" formalism[148, 149].

Analysis of relaxation data involves the fitting of "correlation times", which are a measure of the way in which the direction of a vector (e.g., an amide bond) changes with time. Importantly, it is not necessary to impose an explicit model on the nature of the motion to fit a correlation time. In analyses of relaxation data using the "Model-Free" formalism, two correlation times are fitted: one which refers to variation in a vector by rotational diffusion of the molecule (τ_m), and the other which refers to internal motion, i.e., dynamics, of that vector (τ_e). A third parameter — the "generalised order parameter" (S^2) — specifies to what extent the relaxation of the probe nucleus can be explained solely by rigid-body diffusion in solution. It is therefore a measure of flexibility, with a value of 1 indicating no internal dynamics, and a value of 0 indicating that its motion is completely independent of diffusion in solution.

Spectral density functions can be expressed using these three parameters (as shown in Eqn. 1 of Figure 17), allowing them to be directly related to experimental relaxation rates. The Model-Free analysis was later extended to allow for deconvolution of internal

motions on two timescales[150], as shown in Eqn. 2 of Figure 17. This features order parameters and correlation times for a faster (S_f^2 and τ_f respectively) and slower (S_s^2 and τ_s respectively) internal motion.

$$J(\omega) = \frac{2}{5}\left[\frac{S^2.\tau_m}{1+(\omega.\tau_m)^2} + \frac{(1-S^2).\tau}{1+(\omega.\tau)^2}\right] \quad (1)$$

$$J(\omega) = \frac{2}{5}\left[\frac{S^2.\tau_m}{1+(\omega.\tau_m)^2} + \frac{(1-S_f^2).\tau_f'}{1+(\omega.\tau_f')^2} + \frac{(S_f^2-S^2).\tau_s'}{1+(\omega.\tau_s')^2}\right] \quad (2)$$

where:
$\tau^{-1} = \tau_m^{-1} + \tau_e^{-1}$
$\tau_f'^{-1} = \tau_m^{-1} + \tau_f^{-1}$
$\tau_s'^{-1} = \tau_m^{-1} + \tau_s^{-1}$

Figure 17: Spectral density function expressed in terms of the Model Free parameters (1), and extended Model Free parameters (2)[148, 150].

Although by using the equations in Figures 16 & 17 it is possible to calculate theoretical relaxation rates from Model Free parameter values, complex fitting procedures must be employed to calculate the parameter values from measured relaxation rates due to the non-linear relationship between the spectral density functions and the Model Free parameters. This is discussed further in Section 2.10.5.

1.4.6 Calculation of Conformational Entropy

The order parameter which is fit during Model Free analysis is effectively a measure of the amplitude of the internal motion on which it reports. As such it describes the angular distributions of the bond vector under study (e.g., N–H bond). This can be related to a partition function which describes the energy of all the substates of a particular bond, and this can in turn be used to calculate the conformational entropy at that position. Although it is not necessary to impose a model of the internal motion in order to fit order parameters for each residue from relaxation data, it is necessary to impose a model of the nature of the internal motion in order to calculate the conformational entropy from the fitted order parameters.

Despite the linearity of the values of the partition function with respect to S^2 for values below 0.5 (regardless of the chosen model), the total energy summed over an entire protein is inherently dependent on the model of diffusion used in the calculation[151]. Furthermore, inherent in the entropy calculation is an assumption that all motions present are uncorrelated, and motions on a timescale which are not accessible by NMR relaxation (s–ns and ps–fs) which also contribute to the overall conformational entropy are ignored. Although the absolute value of the conformational entropy may not be accurate, it is possible to make qualitative comparisons between proteins, or between different states of a protein providing the same model of internal motion has been used.

1.4.7 Relaxation Measurements for Slow Timescale Dynamics

"Slow timescale" refers to anything slower than the precession frequency of a nucleus. There are a variety of techniques available to study dynamics in different time windows, although the μs–ns timescale is difficult to study by solution-state NMR with the only insights coming from fitting S^2 values from measurements of RDCs[152]. Dynamics occurring on the minutes-seconds timescale can be studied by hydrogen exchange experiments whereby the protein of interest is solubilised in deuterated buffer and the rate at which the ^{1}H–^{15}N HSQC signals disappear is followed[153]. These rates are used to calculate "protection factors": the extent to which the hydrogen bonding of an amide prevents solvent access, and therefore preventing the exchange of the ^{15}N–^{1}H spin pair for the undetected ^{15}N–^{2}H spin pair. Those residues with low protection factors undergo dynamics which open up the hydrogen bond and enable the amide proton to exchange to a deuteron.

For events on the s–ms timescale, both sets of resonances are often visible in the spectra and exchange spectroscopy (EXSY) methods can be used. 1D and 2D methods have been used to follow reaction kinetics and chemical exchange processes[154, 155] and ^{1}H–^{15}N HSQC-based pulse programs can also be used to monitor protein conformational exchange. Frequently they are referred to as ZZ-exchange experiments[156] as the exchange of I_zS_z magnetisation is monitored. If there is chemical exchange during the mixing time (after the ^{15}N frequency labelling period), the ^{15}N δ of one conformational state is correlated to the ^{1}H δ of another state, producing a cross peak. The rate at which the cross peaks build up depends on the exchange rate between the two species. This method is useful but in large systems signal overlap is a severe problem[157].

For events on the ms–μs timescale, there are two experiments which quantify the dependence on the R_{ex} component of the R_2 — or, indeed, $R_{1\rho}$ — relaxation rate on the strength of the applied magnetic field, the B_1 field. The Relaxation Dispersion experiment[158] uses Carr–Purcell–Meiboom–Gill (CPMG) spin-echo pulse sequences to average out the differences in precession frequencies between the spins of interest in the conformational states which are exchanging. If there are few spin-echo pulses during the fixed relaxation period, the effects of the chemical exchange are most pronounced and the effective R_2 is high; if there are many spin-echo pulses then the effective R_2 is low, reflecting only phenomena on the ns–ps timescale, as discussed earlier.

The off-resonance $R_{1\rho}$ experiment[159] uses spin-locking to average the effects of chemical exchange during a fixed relaxation period. If the applied spin-lock field is relatively weak, then exchange is a significant component on the $R_{1\rho}$ relaxation rate; if the applied field is strong then the effects of exchange are averaged better, leading to smaller R_{ex} components. As the spin-lock is off-resonance from the spin whose exchange is studied, there is a residual B_0 component to the spin-lock field which this enables measurements at higher applied field strengths than the dispersion experiment

which is limited by the rate at which 180° pulses can be applied to the spins under study. The off-resonance $R_{1\rho}$ experiment can therefore be used to study faster exchange processes than the dispersion experiment, although it is less commonly used[156].

The power of the relaxation dispersion and off-resonance $R_{1\rho}$ methods lies in the fact that the R_{ex} component of the relaxation rates depends on many experimentally useful parameters: the populations of the exchanging states, the rate of the exchange between them, and the difference in resonant frequency of the nucleus under study in the different states. As such, data acquired using these methods can give a picture of the thermodynamics, kinetics, and structure of the "minor", often referred to as the "invisible" or "hidden", conformational state[160].

As is the case with the fast-timescale dynamics, the most common nucleus to study relaxation dispersion of is ^{15}N, although it is possible to study $^{1}H^{N}$, $^{13}C'$, $^{13}C^{\alpha}$, and $^{1}H^{\alpha}$ sites[161] in order to measure all the δs necessary for a structure calculation of the "invisible" state[162]. This can be supplemented with relaxation dispersion measurements performed in partial alignment media which allow for the measurement of RDCs[163] and residual chemical shift anisotropies (RCSAs)[164]. This approach was first applied to a folding intermediate of the SH3 domain from the Fyn tyrosine kinase: despite the intermediate being present at approx. 1% in solution and therefore being inaccessible by traditional structural determination methods, it was possible to solve the structure and compare it to the natively folded structure[165].

2 Materials and Methods

2.1 Reagents and Recipes

2.1.1 Source

Chemical reagents including those for bacterial growth media were purchased from Sigma-Aldrich, Fisher Scientific, or Melford. Reagents for making and running SDS-PAGE gels were purchased from Bio-Rad. Chromatography resins were purchased from GE Healthcare. The kits for Site-Directed Mutagenesis and Plasmid Minipreps were purchased from Agilent Technologies and Qiagen respectively. Primers for Site-Directed Mutagenesis were purchased from Eurofins Scientific. Competent cells were originally purchased from Invitrogen or Novagen. Isotopically labelled chemicals were purchased from Goss Scientific Instruments Ltd. or from Cambridge Isotope Laboratories Inc. directly. Milli-Q grade deionised water was supplied by a Sartorius Arium 611VF Ultrapure Water System.

2.1.2 Bacterial Growth Media

Luria-Bertani agar (LB agar) was made as follows:

- 10 g/l Tryptone
- 5 g/l Yeast Extract
- 10 g/l NaCl
- 15 g/l Bacto-Agar

The pH was adjusted to 7.0 before sterilisation by autoclave.

Luria-Bertani (LB) media was made as follows:

- 10 g/l Tryptone
- 5 g/l Yeast Extract
- 10 g/l NaCl

The pH was adjusted to 7.0 before sterilisation by autoclave.

M9 minimal media was made as follows:

- 6 g/l Na_2HPO_4
- 3 g/l KH_2PO_4
- 0.5 g/l NaCl

The pH was adjusted to 7.4 before sterilisation by autoclave. After autoclaving, the following was added to each 1 l of media:

- 1 ml 1 M $MgSO_4$ (autoclaved)
- 100 μl 1 M $CaCl_2$ (autoclaved)
- 650 μl Trace Elements (autoclaved, below)
- 100 μl 10 mg/ml Thiamine (filter-sterilised)
- 2 ml 50% (w/v) $(^{15}NH_4)_2SO_4$ (filter-sterilised)
- Carbon Source (filter-sterilised, see Table 1)

In the cases where D_2O was used in the growth media, the media was filter-sterilised

Labelling Scheme	Carbon Source added per litre of M9 media	D_2O content
^{15}N	15 ml 20% (w/v) unlabelled Glucose	0%
^{15}N & ^{13}C	10 ml 20% (w/v) Uniformly ^{13}C-labelled Glucose	0%
^{15}N & ^{2}H	10 ml 20% (w/v) Uniformly ^{2}H-labelled Glucose	100%
^{15}N, ^{13}C & ^{2}H	10 ml 20% (w/v) Uniformly ^{13}C- & ^{2}H-labelled Glucose	100%
^{15}N, ^{13}C & 60% ^{2}H	10 ml 30% (w/v) Uniformly ^{13}C-labelled Acetate	60%

Table 1: The carbon source and percentage of D_2O used in M9 minimal media for different isotope labelling schemes.

rather than autoclaved to avoid exchange. In high levels of D_2O, all additives to the media were made up in D_2O and filter sterilised to prevent dilution, with the exception of Trace Elements. pD was adjusted to 7.4 by use of the equation $pD = measured\ pH + 0.4$. For the 60% D_2O growth, the deuteration needed to be as random as possible, and therefore Acetate was used as the carbon source as it is more effectively scrambled by the Krebs Cycle.

Trace Elements was made up by dissolving the following in 80 ml of Milli-Q water:

- 550 mg $CaCl_2.2H_2O$
- 220 mg $ZnSO_4.7H_2O$
- 140 mg $MnSO_4.H_2O$
- 45 mg $CoCl_2.6H_2O$
- 40 mg $CuSO_4.5H_2O$
- 40 mg H_3BO_3
- 26 mg $Na_2MoO_4.2H_2O$
- 26 mg KI

The pH was adjusted with acetic acid to pH 8.0 before adding:

- 500 mg EDTA

The pH was re-adjusted with acetic acid to pH 8.0 before adding:

- 375 mg $FeSO_4.7H_2O$

The solution was then made up to 100 ml and sterilised by autoclave.

2.1.3 Solutions and Buffers

RF1 solution was made as follows:

- 160 mM KCl
- 50 mM $MnCl_2.4H_2O$
- 30 mM K–Acetate
- 10 mM $CaCl_2.2H_2O$
- 15% (v/v) Glycerol

The pH was adjusted to 5.8 with Acetic Acid before the solution was filter-sterilised.

RF2 solution was made as follows:

- 2% (v/v) 0.5 M MOPS Buffer pH 6.8 (filter-sterilised)
- 20 mM KCl
- 75 mM $CaCl_2$
- 15% (v/v) Glycerol

The pH was adjusted to 5.8 with NaOH before the solution was filter-sterilised.

βPGM K–HEPES Buffer was made as follows:

- 50 mM HEPES
- 5 mM $MgCl_2$
- 2 mM NaN_3

The pH was adjusted to 7.2 before the solution was filter-sterilised.

βPGM K–HEPES Buffer w/ Salt was made as follows:

- 50 mM HEPES
- 5 mM $MgCl_2$
- 2 mM NaN_3
- 1 M NaCl

The pH was adjusted to 7.2 with KOH before the solution was filter-sterilised.

βPGM Tris–HCl Buffer was made as follows:

- 10 mM Tris
- 5 mM $MgCl_2$
- 2 mM NaN_3

The pH was adjusted to 8.0 with HCl before the solution was filter-sterilised.

4X SDS-PAGE Stacking Gel Buffer was made as follows:

- 0.5 M Tris
- 0.4% (w/v) SDS

The pH was adjusted to 6.8 with HCl before the solution was filter-sterilised.

4X SDS-PAGE Resolving Gel Buffer was made as follows:

- 1.5 M Tris
- 0.4% (w/v) SDS

The pH was adjusted to 8.8 with HCl before the solution was filter-sterilised.

SDS-PAGE Stacking Gel (4.5% BisAcrylamide) was made as follows:

- 2.5 ml 4X SDS-PAGE Stacking Gel Buffer
- 1.125 ml 40% (w/v) BisAcrylamide (37.5:1)

This solution was then diluted to 10 ml before adding:

- 110 μl 10% (w/v) Ammonium Persulphate (APS) (filter-sterilised)
- 11 μl Tetramethylethylenediamine (TEMED)

The solution was thoroughly mixed before pouring into the gel apparatus to set.

SDS-PAGE Resolving Gel (16% BisAcrylamide) was made as follows:

- 2.5 ml 4X SDS-PAGE Resolving Gel Buffer
- 4 ml 40% (w/v) BisAcrylamide (37.5:1)

This solution was then diluted to 10 ml before adding:

- 100 μl 10% (w/v) Ammonium Persulphate (APS) (filter-sterilised)
- 10 μl Tetramethylethylenediamine (TEMED)

The solution was thoroughly mixed before pouring into the gel apparatus to set.

SDS-PAGE Running Buffer was made as follows:

- 25 mM Tris
- 250 mM Glycine
- 0.1% (w/v) SDS

The pH was adjusted to 8.3 with HCl before the solution was filter-sterilised.

2X SDS-PAGE Loading Buffer was made as follows:

- 100 mM Tris
- 200 mM DTT
- 4% (w/v) SDS
- 0.2% (w/v) Bromophenol Blue
- 20% (v/v) Glycerol

The pH was adjusted to 6.8 with HCl before the solution was filter-sterilised.

Gel stain and destain were made as follows:

- 10% (v/v) Acetic Acid
- 45% (v/v) Methanol
- 45% (v/v) Milli-Q Water
- 0.25% (w/v) Coomassie Brilliant Blue R250 (stain only)

2.2 Basic Techniques and Equipment

2.2.1 UV/Vis Absorbance Spectrophotometry

Readings of Optical Density at 600 nm (OD_{600}) were used to assess cell density of bacterial culture and measurements taken using WPA Lightwave S2000 UV/Vis spectrophotometer, with the sample contained in 1 cm path length plastic cuvettes. Sterile media (either LB or M9 media depending on the culture) was used to provide reference absorbance. As absorbance is a measurement derived from the difference in transmission between the reference and the sample, particularly high or low transmission by the sample leads to erroneous absorbance readings. Therefore, if necessary, the culture being measured was diluted appropriately such that the absorbance reading fell into the range 0.1–1.0.

The absorbance of UV light at 280 nm (A_{280}) is used to measure protein concentration, for the benefits of relatively high sensitivity and little interference from signals originating from other molecules present in the sample, such as buffer molecules. Absorbance at this wavelength is due to the presence of hydrophobic amino acid side

chains in the protein. This method is more accurate than the Bradford assay which measures protein concentration by binding of the chromaphore Coomassie Blue which binds to certain amino acids better than others, therefore skewing the results based on primary sequence.

As this method is so sensitive, 5 μl of protein sample was typically added to 500 μl of buffer (which was first used to measure the background absorbance of the buffer) in a quartz cuvette with a path length of 1 cm. The A_{280} was then measured using a Varian Cary 50 Bio UV/Vis spectrophotometer, the value converted into the concentration of βPGM by use of the Beer–Lambert law, which is shown in Figure 18. WT βPGM has an extinction coefficient (ϵ) of 19440 l/mol/cm.

$$A = \epsilon c l$$

where:
A is absorbance
ϵ is the extinction coefficient (l/mol/cm)
c is the concentration (mol/l)
l is the path length (cm)

Figure 18: Beer–Lambert law for determining protein concentration.

2.2.2 Centrifugation

A variety of centrifuges were used, depending on the volume and nature of the sample. For spinning cells out of culture, the culture was transferred into 500 ml plastic pots and centrifuged using a Beckman Avanti J-25I centrifuge fitted with a JLA-10,500 rotor. The cells were spun at 10 krpm for 20 mins at 4 °C. As this produced cell pellets in each of the containers (12 pellets for cultures of $>$ 4 l), all of the cells were then resuspended in a smaller ($<$ 50 ml) volume. They were then centrifuged at 9 krpm at room temperature (set to 21 °C) for 20 mins using a Sigma 3–15 Centrifuge to produce a single pellet.

Following sonication of the cells, the smaller volume and requirement for pelleting of small cell fragments led to the use of a Beckman Avanti J-25I centrifuge fitted with a JA-25,50 rotor. The solution was centrifuged at 24.5 krpm for 20 mins at 4 °C. For other applications where the volume is less than 50 ml but the required centrifugal force was lower (e.g., protein concentration and buffer exchange), a Heraeus Labofuge 400R Centrifuge which was capable of a maximum speed of 4.5 krpm was used, refrigerated to 4 °C. Protein concentration and buffer exchange were achieved by the use of 20 ml Sartorius VivaSpin columns. Spin columns with a 10 kDa molecular weight cut-off were used to ensure that βPGM (24 kDa) could not pass through the membrane during concentration. Buffer exchange consisted of repeated steps of concentration, then dilution into the new buffer, until the extent of the old buffer was reduced to less

than 0.1%.

Finally, samples smaller than 1 ml were centrifuged using an Eppendorf Minispin Centrifuge, typically run at 13 krpm for 1 min. This centrifuge was transferred into a fridge when required.

2.2.3 SDS-PAGE

Polyacrylamide Gel Electrophoresis in the presence of Sodium Dodecyl Sulphate (SDS-PAGE) was used to assess expression levels and monitor purity of βPGM during purification, and was carried out by use of Bio-Rad Mini-Protean II equipment. Gels were made in-house using the gel-forming apparatus and recipes described in Section 2.1.3. The resolving gel layer was poured first, and covered with a layer of isopropanol to ensure a uniform transition between the gel layers. Once this had set, the isopropanol was removed, and the stacking gel was poured. A comb was immediately inserted to create wells for sample insertion. Once both layers were set, the gel was removed from the casting apparatus and any extraneous polyacrylamide removed before wrapping the gel in lab roll. This was then saturated with SDS-PAGE running buffer before being wrapped inside tin foil and left in the fridge. Although 1 day was sufficient to ensure complete polymerisation, under these conditions the lifetime of the gels was in excess of 1 month.

Samples were prepared for SDS-PAGE by mixing 10 μl of sample with 10 μl of 2X SDS-PAGE loading buffer and boiling at 95 °C for 10 mins to ensure denaturation. A prepared gel was removed from the fridge and inserted into the gel tank before being immersed in SDS-PAGE running buffer. The comb was removed to reveal 15 wells of approx. 20 μl capacity, but to ensure no transfer between wells, only 15 μl of sample was ever loaded. The 2 outermost lanes were never used as they can run unevenly, and 1 lane was reserved for 6 μl of pre-stained molecular weight markers (see Table 2), meaning that a maximum of 12 samples were run on each gel.

The gels were run for 5 mins at 50 V to ensure maximal stacking of the protein being entering the resolving gel, giving maximum resolution. After this, the gels were run at 180 V for 45–60 mins, until the layer of Bromophenol Blue from the loading buffer had reached the bottom. At this point, the gels were removed from the tank, and gently shaken in Stain for 1 hour, before being transferred into Destain and left shaking overnight. It was often necessary to exchange the Destain for fresh Destain the following morning and leave for another few hours to ensure complete destaining of the gel. The Destain was recycled by passing through activated charcoal to remove the Bromophenol blue.

Scans of gels was made using an Epson DX3800 Printer/Scanner and the Epson Scan software.

Protein	M_r (kDa)
Myosin	194.9
β-Galactosidase	104.3
Bovine Serum Albumin	59.3
Ovalbumin	41.9
Carbonic Anhydrase	27.9
Soybean Trypsin Inhibitor	20.8
Lysozyme	15.3
Aprotinin	6.5

Table 2: Molecular weight markers used for SDS-PAGE.

2.3 DNA Manipulation

2.3.1 DNA Purification and Sequencing

Plasmid DNA was purified using the QIAprep Spin Miniprep Kit (Qiagen), following the manufacturer's instructions. A 5 ml overnight culture of transformed XL1-Blue cells (see Sections 2.4.2 & 2.4.3) was grown from which to purify the plasmid. The plasmid was eluted in Milli-Q water, rather than the Elution Buffer provided with the kit, to prevent inhibition of subsequent sequencing reactions. For 5 ml of overnight cell culture, 100 μl of typically 100 ng/μl plasmid DNA could be purified.

After plasmid miniprep, the sequence of the plasmid was verified by DNA sequencing, in both 5′ and 3′ directions. For each direction, 10 μl of plasmid were sent to the Genetics Core Facility at the Medical School, University of Sheffield. The plasmids were sequenced using T7 promoter and terminator primers by the dye-terminator sequencing method, using a ABI 3730 Capillary Sequencer. The sequence data was analysed using the software FinchTV (Geospiza).

2.3.2 Site-Directed Mutagenesis

Site-Directed Mutagenesis (SDM) works by using the Polymerase Chain Reaction (PCR) using primers which contain a deliberate mismatch, causing any resultant plasmids to contain that mutation. Primer design is based on the need for the primer to anneal to the template DNA at relatively high temperatures (therefore needing 40–60% G–C base pairs and sufficient base pairs for a relatively strong interaction), whilst having as little ability to form secondary structures as possible. For annealing it is particularly important that the 1–2 base pairs at either end are either G or C nucleotides. All primers were designed using PrimerX (`www.bioinformatics.org/primerx/index.htm`), had their propensity to form secondary structures calculated by OligoCalc (`www.basic.northwestern.edu/biotools/oligocalc.html`, and their melting temperatures calculated by the Poland server (`www.biophys.uni-duesseldorf.de/local/POLAND/poland.html`).

SDM was performing using the QuikChange II SDM kit (Stratagene) with Progene Thermal Cycler (Techne) PCR apparatus. 2.5 ng of each primer (5′ & 3′) and 1 ng of the template plasmid were added to the reaction mixture, and 14 cycles of annealing, extension and melting — rather than the recommended 12 — carried out. Following

completion of the reaction and degradation of the methylated template strand (as per manufacturer's instructions), the newly-mutagenized plasmid was used to transform some competent XL1-Blue cells (see Sections 2.4.2 & 2.4.3) which were then cultured and used for purification of the plasmid, after which the sequence was determined to confirm success of the mutagenesis (see Section 2.3.1).

2.4 Bacterial Culture

2.4.1 βPGM Plasmid

The gene for *Lactococcus lactis* β-Phosphoglucomutase is contained within the pET-22b(+) vector (shown in Figure 19), cloned in by use of the Nde1 and Xho1 restrictions sites as decribed previously[119]. The plasmid features the β-lactamase gene, which confers resistance to Ampicillin which is therefore used to select transformants by adding 100 μg/ml to all growth media. The transcription of the βPGM gene is under the control of the T7 promoter, and is activated by the addition of Isopropyl-β-Thiogalactoside (IPTG) to the growth media. The plasmid also contains a C-terminal His tag sequence, as well as a Pectate Lyase B (pelB) leader sequence to enable transfer of the transcribed protein into the periplasmic space. As the inserted gene contains a Stop codon, and because of the choice of restriction sites used to insert the βPGM gene in to the plasmid, neither of these features are used.

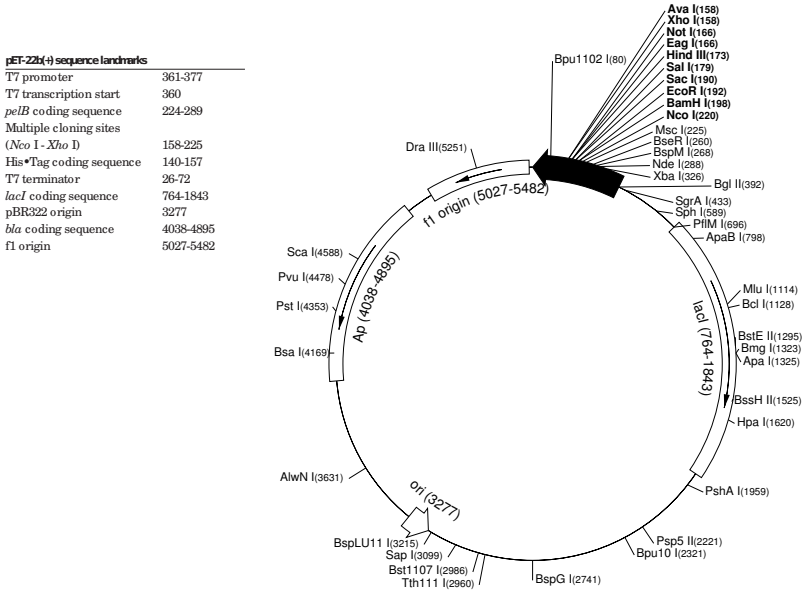

Figure 19: pET-22b(+) Vector Map (Novagen), showing the location of restriction sites and coding regions.

2.4.2 Production of Chemically Competent Cells

The cells used for production of βPGM were *Escherichia coli* BL21 (DE3) cells, which have been optimised for protein expression.*Escherichia coli* XL1-Blue cells were used to prepare plasmid for sequencing or subsequent transformations, as these have

lower germline mutation rates, preserving the sequence of the plasmid. In both cases, a sterile loop was used to scrape the current frozen glycerol stock of cells, and streaked across an LB Agar plate with no antibiotics, with contamination prevented by aseptic technique. The plate culture was incubated at 37 °C overnight.

The following day, a single colony from this plate was used to incubate 50 ml of LB media the following morning, which was then shaken at 200 rpm at 37 °C, until the OD_{600} reached 0.3, using sterile LB media for reference absorbance. The cells were then incubated on ice for 10–15 mins, then centrifuged at 4.5 krpm for 20 mins at 4 °C. The supernatant was discarded, and the cells resuspended in 16.6 ml ($1/3^{rd}$ original volume) of chilled RF1 solution, before incubating on ice for 15 mins. The cells were then again centrifuged at 4.5 krpm for 20 mins at 4 °C, the supernatant discarded and the cells resuspended in 4.1 ml ($1/12^{th}$ original volume) of chilled RF2 solution, and incubated again on ice for 15 mins. The cells were then divided into 200 μl aliquots and stored at −80 °C.

2.4.3 Transformation of Competent Cells

Cells were transformed using the heat-shock method. For each plasmid with which the cells were to be transformed, as well as another for a negative control, a 200 μl aliquot of chemically competent cells were removed from the −80 °C freezer and thawed on ice. Typically, 1 μl of 100 ng/μl plasmid DNA was added to the cells, which were then incubated on ice for 30 mins. After this time, the cells were heat-shocked at 42 °C for 90 secs before being returned to the ice for 3 mins. 0.35 ml of LB media (preheated to 37 °C) was then added to each aliquot of cells, which were then incubated at 37 °C for 1 hour.

Various volumes of each of the cells (typically 1, 10, and 100 μl) were then spread onto LB agar plates containing Ampicillin for selection of transformants. The plates were then sealed and incubated at 37 °C. As no plasmid was added to one aliquot of cells, a lack of colonies on this plate was always confirmed in the morning as a negative control to check that the antibiotic was active.

2.4.4 Bacterial Expression of Unlabelled WT βPGM

Following successful transformation of BL21 (DE3) cells, a single transformant was picked from the plates and used to innoculate a starter culture (75 ml) of LB media containing Ampicillin. This was shaken at 200 rpm overnight at 37 °C. These cultures grew to an OD_{600} of 4–6. The next day, enough of this culture was used to innoculate large scale cultures (2–5 l of LB media with Ampcillin, divided into 750 ml fractions in 2 l baffled conical flasks to aid aeration) such that the starting OD_{600} was 0.05. These flasks were shaken at 200 rpm at 37 °C until the OD_{600} reached 0.6, which represents mid-log growth phase for the bacterial cells. This increase in cell density represents 3.5 doubling periods, but with the initial lag phase after innoculation, typically took 3–4 hours. Due to slight variation in the growth rates of the cultures in different

flasks, each was monitored individually. When the OD_{600} reached 0.6, 1 mM IPTG was added to induce expression of βPGM. The flasks were then shaken at 200 rpm at 25 °C overnight.

Following overnight expression, the cells were transferred into 500 ml centrifuge pots and spun at 10 krpm for 20 mins at 4 °C. The supernatant was then discarded, and all the cell pellets resuspended in small amounts of media, which were then pooled into a 50 ml Falcon tube. This was centrifuged at 9 krpm at room temperature before the supernatant was discarded, the cell pellet dried, and then frozen at −20 °C.

2.4.5 Bacterial Expression of Isotopically-Labelled WT βPGM

For simple ^{15}N-labelled βPGM, the protocol is exactly as above expect for the use of M9 minimal media (containing a source of ^{15}N, but unlabelled glucose as detailed in Table 1) for the expression of βPGM. LB media was still used for the overnight starter culture with which to innoculate the large-scale growth for reasons of simplicity. For the production of βPGM which was also ^{13}C-labelled, there was an additional step of a second small-scale starter culture of M9 minimal media with all the isotope labels present prior to the large-scale culture. This was to ensure that the incorporation of ^{13}C into βPGM was as high as possible, which is important as ^{13}C NMR often consists of ^{13}C–^{13}C magnetisation transfer steps whereby adjacent carbon nuclei must be labelled.

For the expression of perdeuterated proteins, it is necessary to condition the bacterial cells to growth in heavy water prior to expression. This is a tedious process as the immediate transfer of bacteria from 100% H_2O media to 100% D_2O media results in cell death. Instead, following the LB media starter culture, the cells were transferred into M9 minimal media containing 100% H_2O and no isotopic labelling. This was to allow the bacteria to adjust to growing on glucose as the sole carbon source before exposing them to the additional stress of growing in heavy water. Following successful overnight growth, a small-scale growth of M9 minimal media containing 90% D_2O but no other isotopic labelled was carried out. The next day, a small-scale growth of M9 media containing 100% D_2O and all the isotopically labelled nitrogen and/or carbon sources was innoculated and grown overnight. Finally, this was used to innoculate the large-scale M9 media for expression.

The fractional deuteration growth using acetate as the carbon source (see Table 1) provided another challenge. In this case, the metabolic strain induced by transfer to acetate as the sole carbon source dictated the need for an additional conditioning step, whereby a starter culture of M9 media using acetate as the carbon source but lacking any isotopic labelling or D_2O was carried out. This was then used to innoculate a second M9 media starter culture where all isotopic labels including 60% D_2O were present. The final OD_{600} readings suggested that the the acclimatisation to acetate took longer than the acclimatisation to 90% D_2O from 0%.

In all cases, small-scale growths were shaken at 200 rpm overnight at 37 °C. In contrast to the use of relatively large (>5 ml) volumes of starter cultures required to innoculate the large-scale media for protein expression, much smaller volumes (0.1 ml) of starter culture was used to innoculate the next starter culture in the series to reduce carry-over of unlabelled carbon sources and/or H_2O. The large scale growths were carried out as before: starting OD_{600} of 0.05, induction with 1 mM IPTG at $OD_{600} = 0.6$, and expression at 25 °C.

The presence of acetate and/or D_2O in the media significantly slowed bacterial growth: the fractional deuteration growth took 12 hours for the cell density to reach OD_{600} of 0.6 (the doubling time was approximately 3 hours), and therefore the expression time at 25 °C was increased to 24 hours. Post-hoc expression tests confirmed this strategy to be effective, with amounts of soluble βPGM reaching a plateau at this time. The expression time of other perdeuterated growths was also increased depending on the doubling time of the bacteria prior to induction.

2.5 βPGM Purification

2.5.1 Cell Lysis

The frozen cell pellet was thawed and thoroughly resuspended in approx. 40 ml K–HEPES buffer (See Section 2.1.3). To this was added small amounts of DNase, RNase, and Protease Inhibitor Cocktail tablets (Roche) to facilitate purification and prevent proteolytic degradation of βPGM. The Protease Inhibitor tablets used were the EDTA-free variety: although many proteases require metal ions and are therefore inhibited by EDTA, βPGM has a catalytic Mg^{2+} ion which could be unintentionally removed. The cells were broken open by sonication using a Soniprep 150 (MSE). As sonication creates a lot of heat which may denature βPGM, the cells were placed in icy water (which has greater thermal conductivity than ice alone) and sonicated in 5 cycles of 20 secs, separated by 1 min to allow for cooling.

Following this, the cells were centrifuged at 24.5 krpm at 4 °C for 20 mins. The supernatant which contained soluble βPGM was carefully separated from the pellet which contains cell fragments, insoluble proteins, and lipids. It was then filtered through a 0.2 μm filter to remove any small aggregated proteins or remaining cell fragments which may prove problematic for subsequent chromatography. As much as possible the supernatant was kept on ice to prevent proteolysis.

2.5.2 Anion Exchange Chromatography

βPGM has a theoretical pI (pH at which it would have exactly zero net charge) of 4.84, meaning that in contrast to many of the contaminating proteins in the cell lysate, at pH 7.2 it has significant negative charge. Anion exchange is therefore an efficient purification method as most of the contaminating proteins will not bind to the positively charged resin at neutral pH. For this step, a weakly-charged resin was used — a DEAE–Sepharose column (approx. 25 mm diameter, 70 mm length) — run using

an Äktaprime Plus which monitors the elution of proteins and DNA by measurement of UV absorbance. This column was equilibrated with 5–10 column volumes (150–300 ml) of K–HEPES buffer at 5 ml/min before the cell lysate added at a reduced flow rate of 3 ml/min to aid binding. The column was then washed with another 5–10 column volumes of K–HEPES buffer at 3 ml/min to remove any non-specifically binding contaminants from the column.

Following this rinsing step, a salt gradient was used to elute bound proteins according to their pI. This was accomplished by setting up a gradient between the K–HEPES buffer which contains no NaCl, and the K–HEPES buffer with 1M NaCl (Section 2.1.3). The gradient was run for 10 column volumes (300 ml), with an end point of 50% K–HEPES buffer with Salt, i.e., 0.5 M NaCl. 5 ml fractions were collected during the gradient. βPGM eluted at 0.18–0.20 M NaCl, typically across 75 ml. The success of this purification and the fractions containing βPGM were assessed by SDS-PAGE (Section 2.2.3), after which all other elutions collected during chromatography were discarded. Fractions which contained high concentrations of βPGM in comparison to other proteins present were pooled together and concentrated by use of a Vivaspin column, spun at 4.5 krpm at 4 °C.

2.5.3 Size Exclusion Chromatography

A pre-packed 350 ml Superdex G75 column (again run by an Äktaprime Plus system) was then used to separate βPGM from contaminants by molecular weight. The resolution of size exclusion chromatography depends to an extent on the volume of sample loaded. The βPGM liberated from the anion exchange purification step was therefore highly concentrated, typically to over 60 mg/ml, to allow small volumes (<1 ml) to be added to the column at a time. Unlike in the anion exchange chromatography, the success of gel filtration relies on the uniform flow of material down the column, and so the column was therefore equilibrated with buffer with a high solute concentration to reduce the affinity of proteins for the beads: 2X the included volume in de-gassed K–HEPES buffer containing 1M NaCl was pumped through the column at 1.5 ml/min.

A 40–60 mg aliquot of impure βPGM was then loaded onto the column, and buffer continued to be run at 1.5 ml/min. If a larger quantity of material was loaded onto the column then uniform flow of the material (and therefore separation) was impaired. It was not uncommon to need to run the column several times in order to purify all the material resulting from a large-scale expression of βPGM. Fractions were collected in 2 ml, which were again analysed by SDS-PAGE (Section 2.2.3). All fractions containing pure βPGM were pooled and concentrated by VivaSpin.

2.5.4 Buffer Exchange and Storage

Small charged ions or molecules diffuse in solution and therefore generate magnetic fields. This is a major source of noise in NMR, and as such the concentration of unwanted solutes (e.g., NaCl) is kept as low as possible. As the size exclusion chro-

matography step was done in the presence of 1 M NaCl, it was necessary to perform a buffer exchange of the protein immediately afterwards, whereby all of the K–HEPES buffer containing high salt is exchanged for buffer with no salt, either K–HEPES buffer or Tris–HCl buffer, depending on the desired pH of the sample (see Section 2.2.2).

Once the βPGM was more than 95% pure as determined by SDS-PAGE and successfully exchanged into the required buffer, its concentration was measured by A_{280}. βPGM is extremely soluble, but to avoid any potential problems with aggregation resulting from very high concentrations it was often diluted to exactly 2 mM (50 mg/ml) before being frozen at -20 °C.

2.6 Circular Dichroism Spectroscopy

All measurements were done on a Jasco J-810 spectropolarimeter at standard sensitivity, employing a quartz cuvette with a path length of 0.2 mm. A scanning rate of 20 nm/min and integration time of 2 secs — giving 1 measurements per $^2/_3^{\text{rds}}$ of a nm — were used. For the high-quality GuHCl-free spectra the spectrum was measured between 190–300 nm and averaged over 8 scans, for the denaturation study, only the 200–240 nm region was measured for 4 scans.

Due to the high absorption of far-UV light by HEPES buffer, but also the desire not to introduce phosphate ions or ions which may mimic phosphates (e.g., sulphates) into the βPGM samples, prior to circular dichroism (CD), βPGM was buffer exchanged into 10 mM Tris–HCl, pH 8.0 (see Figure 2.1.3). Tris does not strongly absorb far UV light but Cl^- ions do, hence the lower buffer concentration than used for NMR experiments.

50 μM βPGM was used for the unfolding study, as it was found that at wavelengths above 210 nm, reducing the concentration of βPGM or the buffer components resulted in no change in ellipticity or photomultiplier voltage. For good-quality CD spectra below these wavelengths in the absence of GuHCl, it was necessary to use 10 μM βPGM in 1 mM Tris–HCl (obtained by diluting the 10 mM Tris–HCl buffer ten-fold). Samples of Myoglobin and Lysozyme (whose CD spectra have been published) were made to validate the spectra recorded under these conditions.

2.7 NMR Samples

2.7.1 Sample Preparation

Unless stated otherwise, all NMR data was recorded using ^{15}N-labelled βPGM at a concentration of 1 mM in a 300 μl sample contained in a Shigemi tube. Shigemi tubes are designed to eliminate the step between the aqueous solution and the tube glass, and to eliminate the solution meniscus. These design features reduce sample convection and improve the homogeneity of the magnetic field, thereby improving data quality. 5% D_2O was in all samples included for the lock. The various complexes of βPGM were formed by the addition of specific ligands into the solutions. For all samples containing F^- ions it is necessary to add 2 mM of the Al^{3+}-chelator Deferoxamine to prevent

formation of the βPGM-AlF_4^--βG6P TSA complex due to the F^--dependent leaching of Al^{3+} ions from the NMR tube glass[119]. The apo βPGM samples were simply 1 mM βPGM and 5% D_2O within the standard HEPES buffer (see Section 2.1.3), and therefore contained the following:

- 1 mM ^{15}N-labelled βPGM
- 50 mM K–HEPES, pH 7.2
- 5 mM $MgCl_2$
- 2 mM NaN_3
- 5% D_2O

For study of the βPGM mutants, unlabelled protein was used.

The βPGM-MgF_3^--βG6P TSA complex samples consisted of the following:

- 1 mM ^{15}N-labelled βPGM
- 50 mM K–HEPES, pH 7.2
- 5 mM $MgCl_2$
- 2 mM NaN_3
- 10 mM NaF
- 10 mM G6P
- 2 mM Deferoxamine
- 5% D_2O

Deviations from this are as follows. For the ^{15}N R_1, R_2, and heteronuclear NOE measurements, 1 mM βPGM which was labelled both with ^{15}N & ^{2}H was used. For the ^{19}F NMR of the WT enzyme, 2 mM of unlabelled βPGM in 100 % D_2O buffer (pD 7.2) was used. For ^{19}F NMR of the βPGM mutants, unlabelled protein was used.

Mutant βPGM-BeF_3^--βG6P GSA complex samples were exactly as described for the MgF_3^- complex samples except for the addition of 3 mM $BeCl_2$. Mutant βPGM-AlF_4^--βG6P TSA complex samples were as described for the MgF_3^- complex but with the 2 mM Deferoxamine replaced with 2 mM $AlCl_3$.

The addition of metal chlorides as well as sugar phosphates lowered the pH value of the samples, which was then re-adjusted to 7.2 by addition of small amounts of concentrated (typically 2 M) KOH.

2.7.2 Sample Longevity

^{31}P NMR studies allowed the signals of α & β anomers of glucose-6-phosphate (present at a ratio of approximately 40:60 in solution) to be studied. After a few days, a few extra resonances develop: some of them presumably other sugar-6-phosphates due to their ^{31}P δs, and another assigned as inorganic phosphate (P_i) by the addition of phosphate buffer to one of the samples. ^{31}P spectra acquired at different time periods are shown in Figure 20.

The sugar-phosphate signals visible in the ^{31}P spectra were assigned by natural-abundance ^{1}H-^{13}C HSQC as originating from mannose-6-phosphate (M6P; both α &

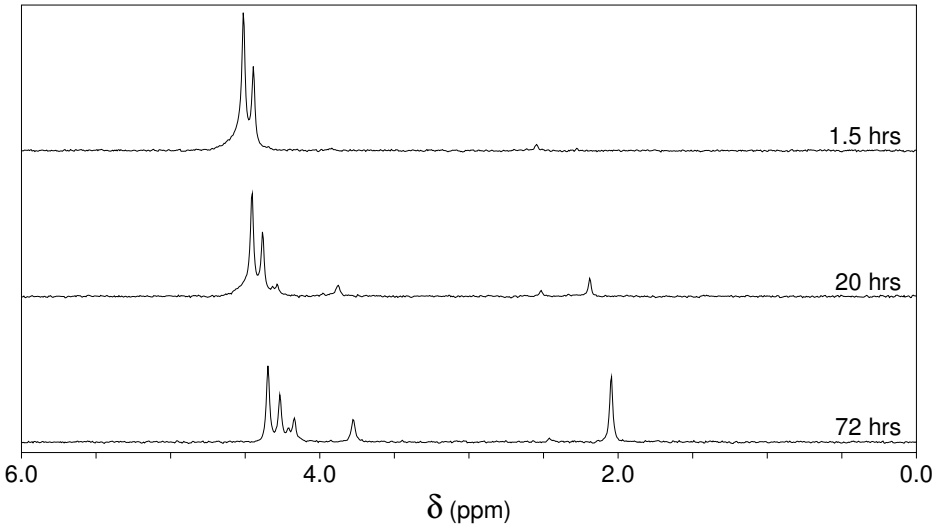

Figure 20: ^{31}P spectra of a βPGM-MgF_3^--βG6P TSA complex sample, shown after 1.5 hours, 20 hours, and 72 hours since the addition of βG6P. It can be seen that there are a range of peaks which slowly increase with time, whilst the G6P peaks — seen in isolation after 1.5 hours — decrease in intensity. The upfield shift of all signals with time is due to the drop in sample pH as a result of the accumulation of P_i (the signal at 2.1 ppm after 72 hours) from G6P.

β anomers are present in solution like G6P) and fructose-6-phosphate (F6P). These assignments were confirmed by ^{31}P NMR, which also showed that F6P accumulates first. Supporting ^{1}H–^{13}C HSQC and ^{31}P spectra are shown in Figures 21 & 22. The accumulation of the sugar-phosphate peaks was found to correlate with the concentration of βPGM stock added to the sample, however the activity persisted in SEC fractions from which βPGM was absent. The conversion of G6P to F6P was therefore ascribed to contaminating Phosphoglucose Isomerase (PGI) present in the protein prep, and the production of M6P was thought to be F6P-dependent. This was proven by the addition of a cis-enediol transition state analogue complex for the PGI reaction — 5-phospho-D-arabinoate (5PAA)[166] — which prevented the occurrence of both F6P and M6P ^{31}P resonances, as shown in Figure 22.

Because of this contaminating activity, the protocol for βPGM purification was altered. *E. coli* PGI is a much bigger enzyme than βPGM, a homodimer of 62 kDa subunits, which suggests that it should elute from the gel filtration column in a smaller volume than βPGM. Its failure to do so implies that its passage through the column is retarded due to non-specific binding to the dextran beads. Prior to the identification of PGI as a contaminant, gel filtration was routinely done in 0 M NaCl. Performing the purification in 1 M NaCl was found to significantly delay the accumulation of M6P and F6P in βPGM-MgF_3^--βG6P TSA complex samples, and all data shown in this thesis (except the ^{31}P spectra shown in this section) were acquired from βPGM samples which had been purified with this new protocol.

Although it is important to remove any time-dependence from samples, even after the production of M6P and F6P from G6P had reached equilibrium βG6P was still the dominant species in solution. Furthermore, no evidence that M6P or F6P could

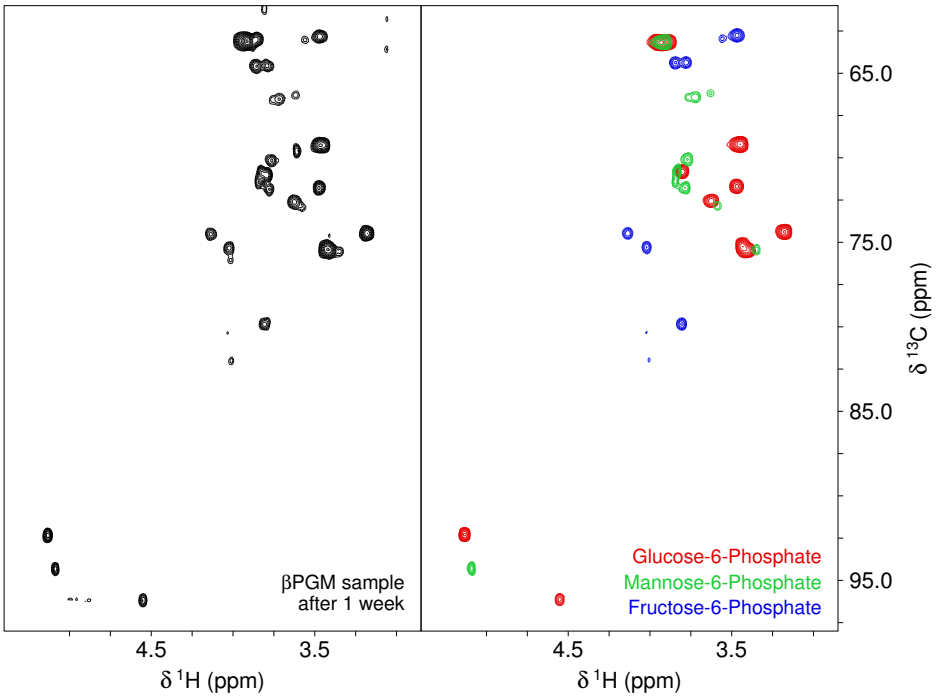

Figure 21: ^{1}H–^{13}C HSQC spectra for assigning sugar phosphate signals. On the left is a spectrum from a week-old βPGM-MgF_3^--βG6P TSA complex sample, with minor aberrations from incomplete water suppression at the bottom, and truncation artefacts from the buffer peaks at the top. On the right is an overlay of spectra acquired from samples of G6P (red), M6P (green), and F6P (blue), all in the absence of βPGM. These spectra have been individually scaled to better match the spectrum on the left, and it can be seen the agreement is excellent. The only peak which is unaccounted for occurs at a ^{13}C δ of 71 ppm. The origin of this peak is still unknown.

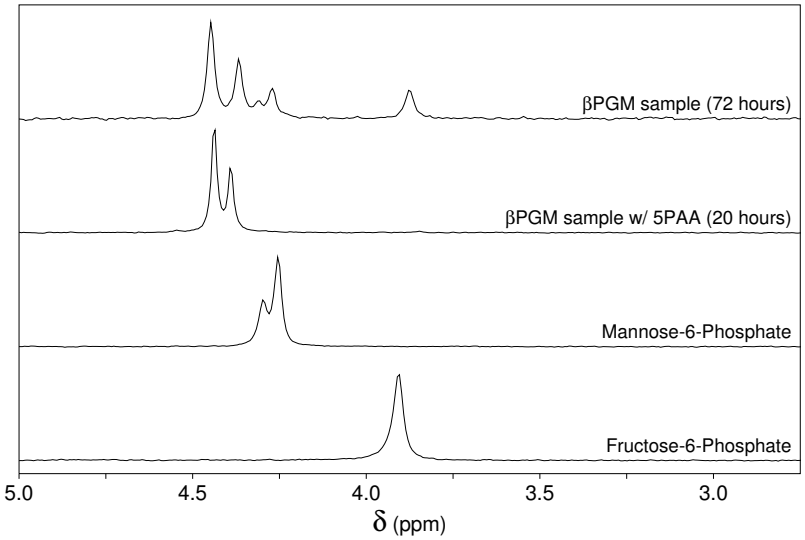

Figure 22: ^{31}P spectra for assigning sugar phosphate signals. From top-bottom: a βPGM-MgF_3^--βG6P TSA complex sample after 72 hours showing the presence of M6P and F6P, a βPGM-MgF_3^--βG6P TSA complex sample after 20 hours in the presence of 5PAA showing only G6P signals demonstrating the inhibition of PGI activity, a sample of M6P, and finally a sample of F6P. These spectra confirm the identity of the ^{31}P resonances seen in the βPGM samples as originating from M6P and F6P. Note the peaks do not exactly align due to small differences in pH between samples.

bind to βPGM has been seen by NMR. The presence of very small amounts of PGI even after gel filtration in higher NaCl concentrations was therefore little more than a nuisance. More of a concern was the slower dephosphorylation of βG6P into glucose and inorganic phosphate (P_i): this reaction goes to completion, effectively removing βG6P from the sample, leading to a loss of the βPGM-MgF_3^--βG6P TSA complex. The rate at which this occurred, though varied between samples, showed no dependence on the concentration of protein stock added to the sample and so is likely not the result of contaminating enzymes in the protein prep.

Instead this dephosphorylation activity was attributed to contaminating azide-resistant microbes present in the samples. This also explained the slow proteolysis of βPGM which could be seen by ^{1}H–^{15}N HSQC, even in the presence of protease inhibitors. Great effort was spent attempting to prevent these activities, by filtering the NMR samples immediately before transferring to thoroughly cleaned NMR tubes, and by the addition of a range of antiseptics, antibiotics, and antifungals. The use of antibiotics had some success with inhibiting bacterial growth on plates innoculated with old NMR samples, but this couldn't be replicated in NMR samples, suggesting that there are a range of bacteria present even in samples that have been made as carefully as possible. The dephosphorylation and proteolysis activities remains in βPGM samples, and so to avoid any impact of this on any measured results, none of the data presented in this thesis was acquired from samples which were more than a fortnight old.

2.8 NMR Data Acquisition

2.8.1 Instruments

NMR data was collected at a range of field strengths on a range of instruments. All ^{31}P spectra, the natural abundance ^{1}H–^{13}C HSQCs, and the lower-field apo βPGM ^{15}N relaxation data were collected on a Bruker Avance 500 MHz spectrometer (operating at 202.5 MHz for ^{31}P) equipped with either a 5 mm BBO Multinuclear Observe probe tuned to ^{31}P, or a 5 mm ^{1}H–^{13}C/^{15}N TXI probe. The ^{19}F spectra were recorded on a Bruker Avance 500 MHz spectrometer (operating at 470.4 MHz for ^{19}F) equipped with a 5 mm ^{1}H/^{19}F/^{2}H–^{13}C/^{15}N QXI probe. The lower-field βPGM-MgF_3^--βG6P TSA complex ^{15}N relaxation and relaxation dispersion data were collected on a Bruker Avance 600 MHz magnet equipped with a 5 mm ^{1}H–^{13}C/^{15}N/^{2}H CPTXI cyroprobe. The high-field ^{15}N relaxation and relaxation dispersion data sets (for both apo βPGM and βPGM-MgF_3^--βG6P TSA complex samples) were recorded on a Bruker Avance 800 MHz spectrometer equipped with a 5 mm ^{1}H–^{13}C/^{15}N/^{2}H PATXI probe. The higher-field ^{15}N relaxation dispersion data set of the βPGM-MgF_3^--βG6P TSA complex was recorded at the SONNMRLSF facility in Utrecht, The Netherlands on a 900 MHz Bruker Avance spectrometer equipped with a 5 mm ^{1}H–^{13}C/^{15}N/^{2}H CPTCI cryoprobe.

2.8.2 Temperature Calibration

As the temperature reported by each spectrometer is inaccurate and slowly drifts with time, temperature calibrations were done immediately prior to the acquisition of each dataset. This was done using fresh samples of neat methanol-d_4, by a method which exploits the temperature-dependence of residual protonated hydroxyl groups due to intermolecular hydrogen bonding[167]. The chemical shift difference between the OH and CD_2H resonances is input into a quadratic equation to calculate the actual temperature of the sample chamber, by use of the equation shown in Figure 23. The use of highly perdeuterated methanol dilutes the number of ^{1}H spins in the sample, which eliminates problems with radiation damping on cryoprobes.

After any change in the requested temperature (however small), the methanol-d_4 sample was left to equilibrate in the probe for at least 10 mins before a spectrum was acquired, although it was actually found to equilibrate much faster than this. The requested temperature was adjusted until the calibrated temperature was within 0.1 K of the appropriate temperature.

$$T = 419.1381 - 16.7467 \times \Delta\delta^2 - 52.5130 \times \Delta\delta$$

where:
T is the calibrated temperature (K)
$\Delta\delta$ is the measured chemical shift difference (ppm)

Figure 23: Formula for temperature calibration of NMR spectrometers

2.8.3 1D ^{1}H Spectroscopy

1D ^{1}H spectra were used throughout the study of βPGM to assess the condition of the sample and the conditions on the magnet, e.g., shimming, temperature, etc. All ^{1}H spectra were recorded using pulse programs which employ presaturation of the water signal for solvent suppression. The pulse programs also feature Hahn-echo refocusing to assist with digital removing of the residual water signal during data processing, necessitating the removal of the first 32 points in the FID during processing. In the case of unlabelled βPGM samples, the pulse program `_1dpe` was used, otherwise the analogous pulse program featuring broadband ^{15}N decoupling — `_1dpef3dec2` — was used. Typically 128 scans of 8192 points (4096 complex pairs) were recorded with a spectral width of 25 ppm, centred on the water signal at 4.70 ppm. This resulted in an acquisition time of 273 ms. The inter-scan delay was typically 1 s. To avoid RF heating of the cryoprobes, the number of pairs (and hence acquisition time) for datasets recorded at 600 and 900 MHz was halved in the cases where ^{15}N decoupling was used. A typical 1D ^{1}H spectrum of the βPGM-MgF_3^--βG6P TSA complex is shown in Figure 24.

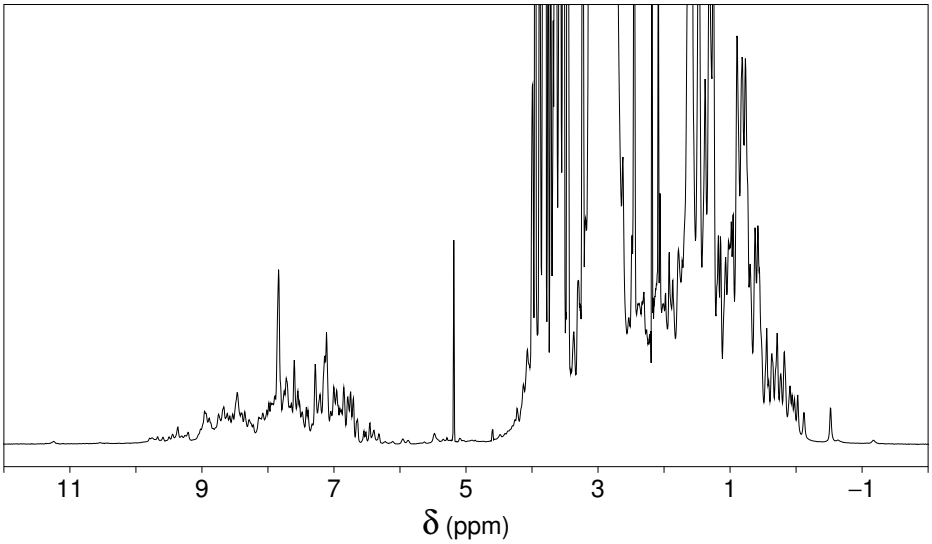

Figure 24: Typical 1D ^{1}H spectrum of the βPGM-MgF_3^--βG6P TSA complex recorded at 800 MHz. The peaks running off the top of the scale are those from small molecules, e.g., ^{1}H resonances from buffer molecules. The peaks between 11.5 and 6 ppm come from ^{1}H nuclei attached to electron-withdrawing nitrogen or oxygen atoms, whereas all peaks upfield of 4.5 ppm are from carbon-bound protons. Some peaks from methyl protons occur upfield of 0 ppm due to ring-current effects of nearby aromatic side-chains, and can be used as an indicator for folding. The sharp peak at 5.2 ppm is from the anomeric proton of βG6P and was used as an indicator for how much βG6P had been lost from the sample (see Section 2.7.2). A representative ^{1}H–^{15}N HSQC spectrum of this complex is shown in Figure 26.

2.8.4 1D ^{31}P Spectroscopy

All ^{31}P spectra were recorded simply with a 90° pulse–acquire type "pulse program" `zg`. Typically 2048 transients of 4096 complex pairs were recorded with a spectral width of 50 ppm, centred at −10 ppm. The acquisition time was 403 ms, and the inter-scan delay was 1 s. All ^{31}P δs were referenced using an internal standard of 85% phosphoric acid sealed in a glass capillary (0 ppm). Due to the absence of z-gradients in the broadband probe used to detect ^{31}P, all shimming was done manually.

2.8.5 1D ^{19}F Spectroscopy

As with ^{31}P NMR, simple 1D ^{19}F spectra were recorded with the `zg` "pulse program". Owing to poor signal:noise, typically 32768 transients of 16384 complex pairs were recorded with a spectral width of 200 ppm, centred at −140 ppm. The acquisition time was 174 ms, and the inter-scan delay 1 s. Presumably due to the materials used to construct the probe, there is a large, broad solid-state signal present in all ^{19}F spectra. Attempts at removing it by the use of Hahn-echo refocusing as is done for ^{1}H NMR were unsuccessful, as the signal itself refocuses. Instead it is necessary to delete the first few (30–100) points after which the signal has relaxed away, and reconstruct the beginning of the FID by backwards linear prediction. This is incompatible with digital data acquisition and so the FIDs are recorded in analogue mode, and as such the digital filters cannot be used. All ^{19}F spectra are therefore recorded with a very large spectral width of 200 ppm to avoid folding the noise let through from the ana-

logue filter, which has a minimum width of 200 ppm. Backwards linear prediction also introduces significant first-order phasing and baselining errors into the spectra, and so careful processing is required to minimise these artefacts.

2.8.6 ^{19}F SEXSY Spectroscopy

The selective exchange spectroscopy (SEXSY) method[168, 169] was used to assess the exchange rate between major and minor conformations of the MgF_3^- moiety in the βPGM-MgF_3^--βG6P TSA complex. SEXSY is identical to the 1D Nuclear Overhauser Effect SpectroscopY (NOESY) method whereby a selective 180° pulse is used to transfer the magnetisation of a single resonance to −z, before a mixing time, and then acquisition[170]. If there is exchange during the mixing time, molecules with inverted spins begin to cancel out the uninverted spins in the other exchanging species, giving rise to a loss of signal intensity of the other exchanging species resonance. The extent of the modulation in signal intensity is proportional to the mixing time and the exchange rate.

It is possible to measure the interconversion rate by traditional EXchange SpectroscopY (EXSY) methods[171] because the system is in slow exchange, i.e., resonances from both conformations are clearly visible in the spectra. Due to sensitivity limitations and the very large spectral width required in the indirect dimension due to the large signal dispersion, preliminary testing of ^{19}F–^{19}F 2D-EXSY experiments by Dr. Nicola Baxter did not yield satisfactory results, necessitating the use of the one-dimensional method.

The major F_c peak was chosen as the target for inversion. A significant cause of spin–spin relaxation of ^{19}F resonances at 500 MHz is the dipolar interaction[172], and with the largest dipoles being ^{1}H nuclei, the linewidth of the ^{19}F resonances correlates strongly with the surrounding proton density[138]. The F_c resonances are therefore the most intense in the spectrum due to having fewer hydrogen bond partners than F_a and F_b. Isolation of the major F_c peak in the spectrum also allowed for total specificity in inversion.

To reduce spectral artefacts (particularly important as the minor F_c resonance is nearby), F_c was inverted by an adiabatic Hyperbolic Secant pulse[173], the parameters of which were optimised in the ShapeTool in TopSpin 2.0. It was necessary to reduce the length of the pulse as much as possible to reduce the amount of exchange that could occur during this time. The Hyperbolic Secant pulse was therefore truncated such that a 1.5 ms pulse resulted in complete inversion of the F_c resonance, but that the minor F_c peak which is 2.5 kHz away was unperturbed. The response of the pulse used in the experiments is shown in Figure 25.

^{19}F SEXSY spectra were recorded with the pulse program `hd_1dexsy`, measuring 16384 scans of 16384 complex pairs were recorded with a spectral width of 200 ppm, centred at −140 ppm. The acquisition time was 174 ms, and the inter-scan delay 2

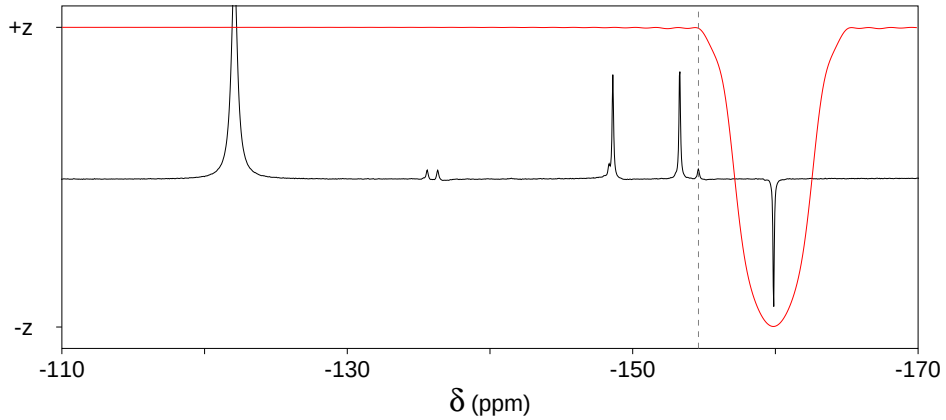

Figure 25: ^{19}F SEXSY spectrum of the βPGM-MgF_3^--βG6P TSA complex with a mixing time of 3 μs, showing the response of the Hyperbolic Secant pulse used for inversion. The red line represents the final position on the z-axis of magnetisation at different frequencies immediately after the pulse. As can be seen, the major F_c peak is inverted, but the minor F_c resonance (and all those further away) is unaffected by the pulse.

s. The Hyperbolic Secant pulse had a pulse length of 1.5 ms, a $B_{1(max)}$ of 2020 Hz, and was centred on the major F_c resonance, which occurs at -159.86 ppm at 298 K in 100% D_2O solvent. The mixing times used for each temperature were optimised to best sample the decay in intensity of the $F_{c(minor)}$ peak to fit the exchange rate, which was highly temperature-dependent. The mixing times at each temperature were as follows:

- 303 K: 3 μs, 1 ms, 2 ms, 4 ms, 7 ms, 12 ms.
- 298 K: 3 μs, 1 ms, 2 ms, 4 ms, 7 ms, 12 ms.
- 293 K: 3 μs, 2 ms, 4 ms, 8 ms, 15 ms, 26 ms
- 288 K: 3 μs, 4 ms, 8 ms, 16 ms, 30 ms, 60 ms.
- 283 K: 3 μs, 8 ms, 16 ms, 32 ms, 65 ms, 125 ms.

2.8.7 2D ^{1}H–^{13}C Correlation Spectroscopy

Natural abundance ^{1}H–^{13}C heteronuclear single-quantum coherence (HSQC) spectra to identify the contaminating sugar phosphates in the βPGM samples (see Section 2.7.2) were recorded using the `hsqcetgp` pulse program. 32 scans of 512 complex pairs in ^{1}H were each recorded for 200 complex pairs in ^{13}C. The ^{1}H dimension had a spectral width of 13 ppm, centred on the water signal at 4.70 ppm. The ^{13}C dimension had a spectral width of 60 ppm to avoid folding signals from the buffer over the signals from the sugar phosphates; the spectra were recorded at a ^{13}C carrier frequency offset of 80 ppm. The acquisition time for each scan was 77 ms, and the inter-scan delay 1.5 s. The coupling constant $^1J_{HC}$ of 145 Hz was used during the INEPT transfers. The pulse program uses z-gradients to select for coherence transfer and suppress the water signal[174].

2.8.8 2D ^{1}H–^{15}N Correlation Spectroscopy

All of the ^{15}N relaxation and relaxation dispersion experiments were acquired as 2D ^{1}H–^{15}N HSQC-type spectra. An HSQC recorded at 900 MHz on the βPGM-MgF_3^--

βG6P TSA complex is shown in Figure 26.

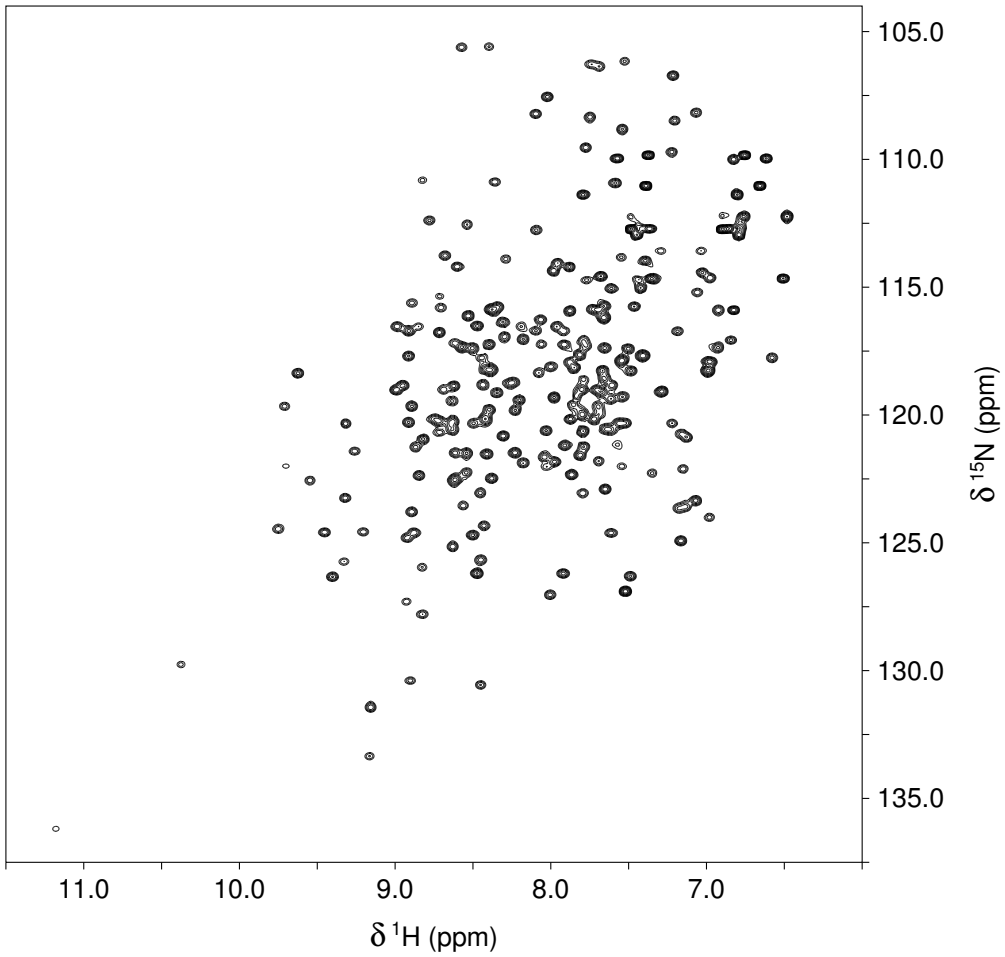

Figure 26: Example ^{1}H–^{15}N HSQC of the βPGM-MgF_3^--βG6P TSA complex, recorded at 900 MHz. It can be seen that signal dispersion is good, and signal overlap — which reduced further with TROSY-detection and/or perdeuteration — is minimal. The βPGM-MgF_3^--βG6P TSA complex gives a characteristic signal from the amide of K117, which is the most downfield signal in both ^{1}H and ^{15}N dimensions due to hydrogen bonding to the phosphate of βG6P.

^{1}H–^{15}N HSQCs were run using the `hsqcetfpf3gpsi` pulse program, which also uses gradient-selection for water suppression. In the case of the βPGM-MgF_3^--βG6P TSA complex relaxation (R_1, R_2, heteronuclear NOE) series spectra which were recorded as transverse relaxation optimised spectroscopy (TROSY)-detected experiments, ^{1}H–^{15}N TROSYs were instead run to monitor the state of the sample. For TROSY spectra on the 800 MHz spectrometer, the `trosyetf3gpsi` pulse program was used which again uses gradient-selection for water suppression. On the 600 MHz spectrometer however, the high Q-factor of the cryoprobe causes radiation damping, so better results were gained by the use of the `trosyf3gpphsi19` pulse sequence which uses the WATERGATE[175] (water suppresion by gradient-tailored excitation) method for solvent suppression. In all cases the coupling constant $^{1}J_{HN}$ for INEPT magnetisation transfer was set to 90 Hz.

TROSY-detection consists of selecting the J-coupling singlet for which the spin states of ^{1}H and ^{15}N are such that ^{15}N relaxation due to dipolar coupling with the amide proton and chemical shift anisotropy (CSA) cancel each other to a certain extent. The result of this is that the ^{15}N R_2 is slower, and as a result the linewidths of the peaks (and therefore the resolution in the indirect dimension) is improved. To increase the resolution in the direct dimension also, the βPGM-MgF_3^--βG6P TSA complex relaxation measurements were performed on ^{2}H,^{15}N-labelled βPGM. As the primary source of relaxation of the amide protons is the dipolar interaction with the H^{α} nuclei, the R_2 of the $^{1}H^{N}$ is also decreased with perdeuteration.

For the HSQC spectra 8–24 scans (depending on the required sensitivity) of 1024 complex pairs in ^{1}H were recorded for each of the 100 complex pairs in ^{15}N. The ^{1}H dimension had a spectral width of 17 ppm, centred on the water signal at 4.70 ppm. The ^{15}N dimension had a spectral width of 35 ppm, centred at 120.5 ppm. The acquisition time for each scan was in the range of 76–114 ms depending on the field strength, and the inter-scan delay was set to 1 s. In the TROSY spectra, 200 complex pairs in ^{15}N were collected rather than 100, to take advantage of the superior resolution offered by the slower ^{15}N transverse relaxation of the TROSY operators. Experimental set up was otherwise identical to that of the HSQC spectra.

For yielding similar sensitivity between spectrometers, it was found that 32 scans per increment at 500 MHz yielded approximately similar (slightly higher) signal:noise to 16 scans at 800 MHz. In contrast, spectra acquired on the 600 MHz spectrometer fitted with the cryoprobe with half the number of scans used at 800 MHz yielded significantly better sensitivity than those acquired at 800 MHz with a room-temperature probe.

2.8.9 ^{15}N R_1 and R_2 spectroscopy

All ^{15}N relaxation and relaxation dispersion experiments were recorded as ^{1}H–^{15}N correlation spectra with acquisition parameters as described above, except the interscan delay was increased to 2.5 s. The pulse programs used for the HSQC-type ^{15}N R_1 and R_2 measurements were written in-house by Dr. Maya Pandya. They feature pre-saturation, gradient selection and WATERGATE sequences for water suppression. Prior to the ^{15}N frequency labelling period there is a relaxation block of variable length during which the ^{15}N magnetisation is either longitudinal (R_1) or transverse (R_2). The R_1 measurements consists of transferring the ^{15}N magnetisation to the $-$z axis and monitoring the rate at which the bulk magnetisation returns to equilibrium along +z, although the phase cycling is such that the end point after infinite relaxation is zero recorded signal. The R_2 experiment measures the rate at which coherence is lost from like spins in the xy plane. It is necessary to employ ^{15}N Carr Purcell Meiboom Gill (CPMG) pulsing (a sequence of 180° ^{15}N pulses) to quench any contribution from ms–μs exchange to the transverse relaxation rate (R_2), which is referred to as R_{ex}. For this work a CPMG field strength (ν_{CPMG}) of approximately 500 Hz was used.

Due to the continued pulsing at ^{15}N radio frequencies, the sample is heated by an ex-

tent that is dependent on the length of the relaxation block. Therefore to avoid systematic errors, both R_1 and R_2 pulse programs use Duty-Cycle Heating Compensation[176] which is a method by which the amount of RF heating is kept constant throughout the relaxation series, independent of the relaxation block length. This is done by keeping the total relaxation block the same length, but changing the way that it is partitioned between the true relaxation block when ^{15}N is transverse and the duty-cycle compensation block which is located after each FID has been recorded and before the interscan delay. For example, in the case where the ^{15}N coherences are only allowed to relax for a short time, there is a long duty-cycle compensation block before the next scan, and vice versa. RF heating also affects the probe, and as such ^{15}N 90° pulses were 45 μs as a minimum (B_1 = 5555 Hz), although in many cases due to the age of the amplifiers and probes the ^{15}N 90° pulse length at maximum power was already longer than this. This is true for all spectrometers except the 900 MHz at the SONNMRLSF facility in Utrecht, The Netherlands: to avoid damaging the high-field cryoprobe it was necessary to use ^{15}N 90° pulses of 80 μs.

Another peculiarity of the R_2 pulse program is that because of the number of ^{15}N 180° pulses employed, the data are extremely sensitive to off-resonance effects. This means that the calibrated ^{15}N pulse length needs careful optimisation, and that over a ^{15}N spectral width of 2840 Hz at a spectrometer frequency of 800 MHz, it is necessary to record the data at several carrier frequencies. 3 offsets in ^{15}N were typically employed: 109.5, 120.5, and 131.5 ppm. This ensured that no resonance was further than 5.5 ppm (470 Hz at a B_0 of 800 MHz) away from the nearest carrier frequency. Computer simulations to address the same problem in the relaxation dispersion experiment have been published[177]. The conclusion was that for a ^{15}N 90° pulse and maximum ν_{CPMG} of 2000 Hz, a signal 500 Hz away from the ^{15}N carrier frequency offset would have a systematic error on its signal intensity measurements of approximately 3%. This supports our finding that for ^{15}N R_2 data recorded with a ν_{CPMG} of only 500 Hz, no systematic error was seen over the random error due to spectral noise.

For the apo βPGM relaxation measurements where the data was recorded without TROSY-detection (i.e., just as HSQC-type spectra), the pulse programs used for the ^{15}N R_1 and R_2 measurements were `mjp_n15_r1_pss_02_2005_3` and `mjp_n15_r2_cpmg_pss_02_2005_7`. For the measurements on the βPGM-MgF_3^--βG6P TSA complex, TROSY-detected versions of these programs were engineered by Dr. Matthew Cliff. The pulse programs are exactly as described above with Duty-Cycle Compensation and the same method for water suppression, but with a TROSY-selection block immediately before acquisition of the FID. The programs used for the ^{15}N R_1 and R_2 measurements on the βPGM-MgF_3^--βG6P TSA complex at 800 MHz were `mjc_n15_r1_pss _tr_phasecycle_1` and `mjc_n15_r2_cpmg_pss_tr_3`. For the measurements at 600 MHz the analogous programs `mjc_n15_r1_pss_tr_phasecycle` and `mjc_n15_r2_cpmg_pss_tr` were used.

2.8.10 ^{15}N–{^{1}H} Heteronuclear NOE spectroscopy

The ^{15}N–{^{1}H} Heteronuclear Nuclear Overhauser Effect (NOE) measurement consists of observing the effect of saturating ^{1}H resonances on the ^{15}N nuclei. Saturation is the phenomenon when there are equal number of spin-up and spin-down nuclei, with no transverse magnetisation, and therefore no observable signal. This is achieved by continued pulsing of the ^{1}H nuclei, typically with a chain of 120° pulses[178]. The population of spin-up and spin-down ^{15}N nuclei for each residue is then measured, but in order to do this the ^{1}H polarisation cannot be transferred to the ^{15}N spins by INEPT transfer as is usually done for HSQC-type experiments. Instead the first pulses are on the ^{15}N nuclei, and because the polarisation of ^{15}N nuclei is approximately 10-fold less than that of ^{1}H spins, this experiment yields poor signal:noise, and therefore demands many scans per increment.

The population of spin-up and spin-down ^{15}N nuclei in the absence of ^{1}H saturation is also measured in a reference experiment. It is essential that both ^{1}H and ^{15}N populations have returned to equilibrium for this measurement, and as such in some cases extremely long interscan delays are required[179]. There are many other practical problems with the running of this experiment, particularly: the unintentional saturation of the water signals which subsequently saturate the ^{15}N nuclei through amide proton exchange[180, 181], cross-correlation relaxation of the ^{15}N nuclei due to contributions from both the dipolar interaction and also CSA[182], and incomplete saturation of the amide protons[183].

The pulse program used to measure the ^{15}N–{^{1}H} hetNOE on apo βPGM was `hsqcnoef3gpsi` which, as before, uses gradient selection for water suppression. The pulse program records the experiments with and without ^{1}H saturation (hereby referred to as the NOE and NONOE experiment) interleaved as a pseudo-3D experiment. This reduces any differences in heating between the NOE and NONOE experiments due to the difference in pulsing between them. The interscan delay was set to 10 s to allow all resonances (particularly the water protons with their long T_1) to relax. An interscan delay any shorter than this yielded physically impossible NOE values, and longer than 10 s was impractical. In the NOE experiment, high-power 120° ^{1}H pulses were applied every 5 ms for 10 s to saturate the ^{1}H resonances. These pulses were absent in the NONOE experiment.

For the fast-timescale dynamics measurements on the βPGM-MgF_3^--βG6P TSA complex, a TROSY-based pulse program was instead used. A pulse sequence based on that published in Ferrage et al, 2008[184] was programmed by Dr. Matthew Cliff. This pulse program also uses 120° ^{1}H pulses for saturation (although it has been suggested that 180° pulses may be more effective[183]) applied every 10 ms for 5 s before measuring the ^{15}N polarisation. In the NONOE experiment, which is run separately in this case, an interscan delay of 10 s is used. This pulse program is unique in that there is a composite 180° ^{1}H pulse applied to all but water protons at the very start of

the pulse program during the NONOE experiment. This ensures that rather than the usual slowly-relaxing $2N_zH^{\alpha}$ operator being observed, the $2N_zH^{\beta}$ operator — whose longitudinal relaxation is faster and therefore will have decayed closer to equilibrium during the interscan delay — is selected.

In testing it was found that this pulse program run with an interscan delay of 10 s gave approximately the same NOE values as the other program described above when employing a 20 s recovery period, both of which give lower values than the old program when an interscan delay of 10 s is used. This means that the TROSY-based pulse program likely allows measurement of the true equilibrium ^{15}N magnetisation after a 10 s delay, and it also demonstrates that although the NOE and NONOE spectra are acquired separately, the differences in heating between the NOE and NONOE experiments is minimal. The TROSY-based pulse program is called `ah_trosy_15n1hhetnoe_1`.

2.8.11 ^{15}N Relaxation Dispersion spectroscopy

Relaxation dispersion consists of measuring the dependence of the transverse relaxation rate (R_2) on the CPMG field strength (ν_{CPMG}). The CPMG pulses vary the extent that transverse relaxation occurs due to chemical exchange (in the case of proteins this refers to conformational exchange where a particularly nucleus is exchanging between 2 different environments). When ν_{CPMG} is large, the effective transverse relaxation rate ($R_{2(eff)}$) is small as the exchange is quenched, whereas when there are few CPMG pulses during the relaxation block, $R_{2(eff)}$ is large to due contributions from the exchange process, R_{ex}.

Following the statistical justification for its use, all data presented herein is fit using a two-state conformational exchange model; it is assumed that the chemical exchange detected by relaxation dispersion is due to the interchange between two protein conformers. How $R_{2(eff)}$ varies with ν_{CPMG} is dependent on a number of parameters. In the two-state model these are: the rates of interchange between species, the populations of the species, and the chemical shift differences of the nuclei being studied between the species[185].

The pulse program used for ^{15}N relaxation dispersion was `mjp_n15_rd_cpmg_pss_6`, which is another in-house pulse program written by Dr. Maya Pandya, which has been described in Long et al, 2010[186]. There is a fixed 40 ms relaxation delay during which the frequency of CPMG pulses is varied. Because the relaxation rates of N_x and $2N_xH_z$ coherences are different, the evolution of $^1J_{HN}$ coupling during the relaxation delay can complicate the results. To overcome this problem, the pulse program use throughout this thesis is a "relaxation-compensated" pulse program which ensures that equal lengths of time are spent with the magnetisation in-phase and anti-phase (N_x and $2N_xH_z$) during the relaxation block[187]. More recently, a pulse program which features ^1H decoupling throughout the CPMG block, eliminating the need to average the time spent in N_x & $2N_xH_z$ states has been published[188], though these

changes have not been implemented. The pulse program also features duty cycle heating compensation[176], by means of another 40 ms duty cycle compensation block prior to the start of the pulse program; for every CPMG pulse omitted from the true relaxation block, one is added to the compensation block. This ensures that the total number of ^{15}N pulses employed is the same regardless of the ν_{CPMG}, and the RF heating is kept constant. At present a TROSY-detected version of this pulse program does not exist.

ν_{CPMG} values in the range 50-750 Hz were used in random order, with repeated points measured to aid with error analysis as described in Ishima & Torchia, 2005[177]. Due to the largest ν_{CPMG} value used for the dispersion experiment exceeding that used in the standard ^{15}N R_2 pulse program, the pervasion of off-resonance effects is expected to be greater. For this reason, data at more carrier frequency offsets were recorded at higher fields. At 600 MHz — as with the standard ^{15}N R_2 measurements — 109.5, 120.5 & 131.5 ppm ^{15}N carrier frequency offsets were used. At 800 MHz, four offsets were used to compensate for the larger signal dispersion: 109, 117, 125, & 133 ppm were selected to ensure maximum coverage. For the 900 MHz data collected at the SONNMRLSF facility, the ^{15}N power for the CPMG pulses had to be decreased to avoid damage to the cryoprobe, which will increase any off-resonance effects in the measured peak heights. Therefore five offsets were used: 109.05, 115.15, 121.25, 127.35, & 133.45 ppm. In addition to recording the data several times, it is possible to calculate (and therefore remove) the contribution of off-resonance effects in the data, but issues such as B_1 field inhomogeneity can make this difficult.

2.9 General Data Fitting & Statistical Methods

Unless otherwise stated, all data fitting was performing using "MattFit", an in-house Python program which uses a Levenberg–Marquardt non-linear least squares routine[189, 190] to minimise to the appropriate function. When error estimates on the peak height measurements were derived from baseline noise, they were adjusted after fitting such that the resulting total χ^2 value for the fit matched the expected value, where $\Sigma\chi^2 = \textit{no. of data points} - \textit{no. of fitted parameters}$. In the case of the ^{15}N relaxation dispersion data, the peak height error estimates were derived from the deviation of repeated measurements[177], a method which should take into account all sources of error, and as such the estimates were not adjusted to achieve the expected $\Sigma\chi^2$. The ^{15}N relaxation dispersion data was also separately fitted using Dr. Flemming Hansen's CATIA program to validate the resulting fit values from "MattFit".

The ^{15}N R_1 & R_2 values for apo βPGM were fit using "MattFit" with adjusted baseline noise estimates, but the fitting of the R_1 & R_2 values for the βPGM-MgF_3^--βG6P TSA complex was instead done using the software Relax[191, 192]. Relax uses the initial baseline error estimates to carry out 500 Monte Carlo simulations[193] for more accurate estimation of the errors on the fitted relaxation rates for each residue. It being a simple ratio of peak intensities, the error on the NOE value was calculated

by propagation of the baseline noise estimates. All Model Free[148, 149] fitting was done using Relax, as is detailed in Section 2.10.5, which also carries out 500 Monte Carlo simulations to ascribe error estimates to the fitted Lipari-Szabo parameters.

For all F-tests included those done by the fitting program Relax, Akaike's Information Criterion[194] was used. The "AIC" value is dervied by multiplying the number of parameters used in the fitting by 2, and adding it to the χ^2 value. The solution which gives the lowest AIC is thereby the most parsimonious. In order to assess correlations between parameters, as well as plotting them, Pearson's correlation coefficient[195] is calculated. It can be in the range -1–$+1$, where -1 is a perfect negative correlation, 0 is no correlation, and 1 is a perfect positive correlation. The formula for the correlation coefficient, r, is shown in Figure 27.

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2}\sqrt{\sum(y - \bar{y})^2}}$$

Figure 27: Formula for the correlation coefficient, r.

Another statistical device used throughout this thesis is the weighted mean. In all cases where this is used, the weights are derived from the uncertainties on the values which are being averaged. The formula for calculation of the weighted mean, as well as its uncertainty, are shown in Figure 28.

$$\bar{x} = \frac{\sum(x/\sigma x)}{\sum(1/\sigma x)} \quad (1)$$

$$\sigma\bar{x} = \frac{1}{\sum(1/\sigma x)} \quad (2)$$

Figure 28: Formula for calculating the weighted mean of a data set (1), and its error (2). x represents each data point value, and σx represents the uncertainty of each value.

2.10 NMR Data Processing

2.10.1 Processing Raw NMR Data

All NMR data has been processed using Felix 2007. Macros are written to transform raw FIDs acquired on the spectrometers into phased, Fourier transformed spectra. The macros are specific to each pulse program, an example macro for processing the FID shown in Figure 29 into the 1D ^{1}H spectrum shown in Figure 24 on Page 55 is shown in Figure 30.

The processing of 2D spectra follows the exact same procedure as the 1D processing

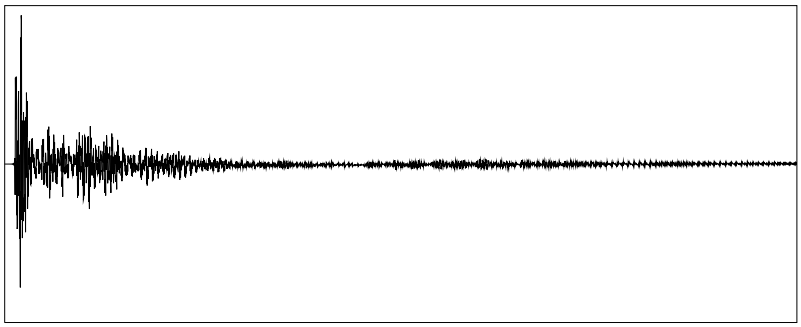

Figure 29: Unprocessed FID for 1D ^{1}H Spectrum shown in Figure 24 on Page 55. The macro used to process the raw data is shown in Figure 30. It can be seen that the first few points have no signal due to the Hahn-Echo used in the pulse sequence (as discussed in Section 2.8.3), and that the signals have mostly decayed by half-way through the acquired points.

```
def phase0 202.0    ; zero-order phase correction
def phase1 0
def gibbs 1
def swidth 20000    ; spectral width in Hz

rn /ld10/hugh/pgm298bb/raw/pgm298bb_4100_14apr10m4.dat

dft 8 12            ; for digital oversampling
ift
cnv 0 16            ; convolution to remove water signal
shl 32              ; deletion of first points from Hahn-Echo
sb 2048 90          ; sine-bell window function
zf 8192             ; zero-filling
ft                  ; fourier transform
ph                  ; phasing
red                 ; deleting imaginary data
abq                 ; baseline correction
pol 0
def sfreq 800.344   ; referencing
ref 4096 4.70
def axtype 3

wn /ld10/hugh/pgm298bb/processed/pgm298bb_4100_14apr10m4.ft

dr                  ; display spectrum
end
```

Figure 30: An annotated processing macro to convert the 1D ^{1}H FID shown in Figure 29 to the spectrum shown in Figure 24 on Page 55.

macro shown in Figure 30 except that the direct dimension processing is nested in a loop to process every increment in the same way, before the indirect dimension is itself processed and fourier transformed.

Different window functions were used to optimise the sensitivity:resolution ratio of different types of data. ^{31}P data were processed with 1 Hz line-broadening, whereas the very broad lines and poor signal:noise ratio of the ^{19}F data meant line-broadening of 20 Hz was used. As discussed in Section 2.8.5, ^{19}F NMR data requires backwards linear prediction to remove a solid-state signal from the spectrum. This is an error-

prone procedure and the parameters for linear prediction were optimised in each case to achieve the best possible baseline. Typically for ^{1}H 1D spectra, as is demonstrated in Figure 30 a sine-bell function was used to bring the intensities down to 0 half-way through each FID (as most interesting signals had relaxed by half-way through the FID). The angle that the sine-bell was truncated to was optimised for each spectrum.

For 2D processing, the window functions were typically sine-squared functions which bring the intensities down to zero less harshly than the sine-bell function, and hence reduce truncation artefacts in the resulting spectra. Again the angles that the functions were truncated to was optimised for each spectrum to achieve the narrowest lines without causing artefacts. To improve the resolution in the indirect dimension, typically as many points as were measured were linear predicted, and the window function then performed over twice as many points as were acquired on the spectrometer.

In the case of 1D spectra, the peak heights were measured and corrected for any baselining errors directly in Felix, and the intensity of the noise was estimated from regions of the spectrum devoid of signals. For ^{19}F spectra it was found that it was necessary to zero-fill to many times the number of acquired points so that the digitisation of the spectra did not affect the peak intensity measurements. For 2D spectra, the tools for peak picking in Felix were utilised, along with in-house macros which exported the peak heights of all assigned peaks in a series of spectra into a single table. For example, peak heights of all assigned, non-overlapped peaks from the ^{15}N R_1 spectra which had been acquired with a variety of relaxation delays were exported into a single table to facilitate later relaxation curve fitting. For an estimation of the baseline noise in 2D data sets, a macro which calculated the RMSD over an area of the spectrum which contained no signals was used.

As the peak height measurements required different treatment depending on the nature of the measurement, they are discussed separately below.

2.10.2 Fitting of ^{19}F SEXSY Measurements

As is outlined in Section 2.8.6, the ^{19}F method consists of inverting the $F_{c(major)}$ resonance and observing the intensity of the $F_{c(minor)}$ resonance, which decreases as the mixing time is increased as a result of exchange between the two spin states. The experiment relies on the fact that the z-magnetisation of both states is perturbed from equilibrium during the mixing time, and as such it was necessary to not only fit the rate of exchange between the states (k_{ex}), but also the rate of longitudinal relaxation (R_1). To minimise the number of fitted parameters it was assumed that the major and minor F_c resonances have equal R_1 values. In all cases studied, k_{ex} far exceeds R_1, and so the parameters are statistically distinct. Because of this, even if the two resonances have slightly different R_1 rates, the impact on the results is likely to be very low.

The peak heights of both the major and minor F_c resonances were measured and fit to the equations shown in Figure 31 using "MattFit". It was found for several

temperatures that the $F_{c(major)}$ peak intensity values deviated from the fit line more than was expected from the baseline noise estimates, due to some source of error which was unaccounted for in the baseline noise estimates. The minor F_c peak heights were not affected, but because the nature of the error source was unknown, there was no justification for adjusting the error estimates of the two measurements separately. Therefore, the errors on both the major and minor F_c resonance peak heights were adjusted together to satisfy the expected $\Sigma\chi^2$value.

$$I_{major} = I_0.F_a.\Big(1 - 2.F_a.e^{-R_1.t_m} - 2.(1 - F_a).e^{-(R_1+k_{ex}).t_m}\Big)$$

$$I_{minor} = I_0.(1 - F_a).\Big(1 - 2.F_a.e^{-R_1.t_m} - e^{-(R_1+k_{ex}).t_m}\Big)$$

where:
I_{major} is the intensity of the major F_c resonance after mixing time t_m
I_{minor} is the intensity of the minor F_c resonance after mixing time t_m
t_m is the mixing time (s)
I_0 is the sum of the intensities of both major & minor peaks at $t_m = 0$ s
F_a is the population of major conformer
R_1 is the spin-lattice relaxation rate of both resonances (s^{-1})
k_{ex} is the sum of forward & backward exchange rates (s^{-1})

Figure 31: Equations to which the ^{19}F SEXSY peak heights were fit.

Subsequent manipulation and fitting of the resulting k_{ex} values at different temperatures is discussed in Section 3.

2.10.3 Fitting of ^{15}N Relaxation Dispersion Data

All effective transverse relaxation rates $R_{2(eff)}$ values were derived by fitting an exponential decay curve between peak intensities at each ν_{CPMG} value, and the peak intensity from a reference experiment lacking the 40 ms relaxation block, as shown in Figure 32. Although there are many approximations of the Bloch–McConnell equations [196, 197, 198] which can be used to speed up the processing of relaxation dispersion data, they make assumptions on the nature of the exchange, and have been avoided in this case. Instead, relaxation dispersion curves were fit to the subsequent $R_{2(eff)}$ values by solving the full Bloch–McConnell[141, 199] equations. This was done as a function of the rate of interconversion of major conformer to minor conformer (k_a), the population of major conformer (p_a), and the difference in chemical shift between the two conformers ($\Delta\delta$).

For the lower-field βPGM-MgF_3^--βG6P TSA complex data of poor signal:noise it was necessary to fix k_a & p_a to the values derived from the ^{19}F SEXSY analysis. For the apo βPGM data it was necessary to fix the parameters to the values derived from the 900 MHz data collected on the βPGM-MgF_3^--βG6P TSA complex. For the βPGM-MgF_3^--βG6P TSA complex data, it was possible to derive the exchange rates

$$R_{2(eff)}(\nu_{cpmg}) = -\frac{1}{T_{relax}} . \ln\left(\frac{I(\nu_{cpmg})}{I_0}\right)$$

where:
$R_{2(eff)}$ is the effective transverse relaxation rate (s^{-1})
ν_{CPMG} is the CPMG field strength (Hz)
T_{relax} is the relaxation period (s)
I is the intensity of the peak after the relaxation period
I_0 is the intensity of the peak in the absence of the relaxation period

Figure 32: Calculation of $R_{2(eff)}$ values for each residue.

independently of the ^{19}F SEXSY data, but only after thorough statistical analysis. In all cases, subsequent processing or analysis of the relaxation dispersion data is discussed in Section 4.

2.10.4 Fitting of ^{15}N R_1 & R_2 and ^{15}N–{^{1}H} hetNOE Data

For the spin–lattice & spin–spin relaxation measurements, the signal intensity for each residue decays exponentially as the relaxation delay is increased, according to its individual R_1 & R_2 rate. The peak heights were therefore fit to a simple exponential decay equation such as that in Figure 33. For the apo βPGM data this was done using "MattFit", but for the βPGM-MgF_3^--βG6P TSA complex data it was done using the Relax suite.

$$I = I_0 e^{-kt}$$

where:
I is the intensity of the peak at each specific time point
t is the relaxation delay time for each specific point
I_0 (unknown) is the initial intensity of the peak
k (unknown) is the relaxation rate, R_1 or R_2

Figure 33: Formula for exponetial decay, to fit ^{15}N R_1 and R_2 data.

The ^{15}N–{^{1}H} hetNOE value is simply the ratio of peak heights in the spectra with and without proton saturation for each residue, as shown in Figure 34. The error is estimated simply by calculating the propagation of baseline noise in the two spectra. This error figure will therefore only give an impression of random experimental error, and not take into account any systematic errors, to which the NOE measurement is particularly susceptible, as was discussed in Section 2.8.10.

2.10.5 Model-Free Estimation & Fitting

Using the default settings, the program HydroNMR[200] was used to approximate anisotropic molecular tumbling in solution from crystal coordinates. The program uses bead-modelling procedures as a basis for hydrodynamic calculations, and can

$$NOE = \frac{Intensity\ with\ proton\ saturation}{Intensity\ without\ proton\ saturation}$$

Figure 34: Formula to calculate NOE value.

calculate theoretical relaxation rates of backbone ^{15}N nuclei assuming that all residues are perfectly rigid. Theoretical T_1, T_2, and hetNOE values can also be calculated from estimated τ_m, S^2, and τ_e values. Therefore, plotting the experimentally measured and the predicted relaxation values T_1, T_2, and/or hetNOE against each other is a useful method by which approximate Model Free parameters and also the agreement of experimental data with expected values can be assessed. R_1 & R_2 relaxation values predicted by HydroNMR for Apo βPGM at a field strength of 800 MHz are shown in Figure 35 plotted onto grids which represent the calculated relaxation values at a range in the values of the model free parameters. The equivalent grid showing the variation in T_1 values with respect to hetNOE values is shown in Figure 36.

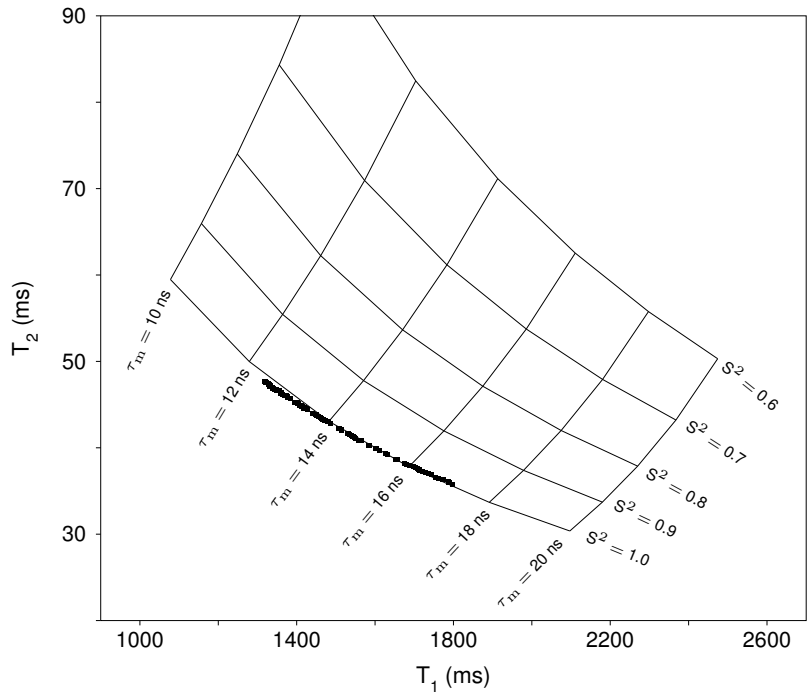

Figure 35: T_1 vs T_2 plot showing Model-Free grid lines (black lines) and predicted relaxation data for βPGM from HydroNMR (black squares) at a field strength of 800 MHz. The grid lines are labelled with the value which is constant over that line, i.e., the $\tau_m = 10$ ns line shows all the values of T_1 & T_2 where the value of τ_m is 10 ns and S^2 varies between 1.0 and 0.6. Likewise the $S^2 = 1.0$ line shows the expected values of T_1 & T_2 for τ_m values in the range 10–20 ns where there is no internal mobility. It is seen that all the HydroNMR predicted relaxation rates fit onto this line, showing an expected range in apparent τ_m due to anisotropic rotational diffusion of approximately 12.5–17.0 ns. For the plotting of the grid, τ_e is assumed to be 50 ps. Experimental relaxation rates would be expected to conform to the range in apparent τ_m values predicted from HydroNMR, but show variation in S^2 value.

There have been a number of programs designed for the specific purpose of fitting

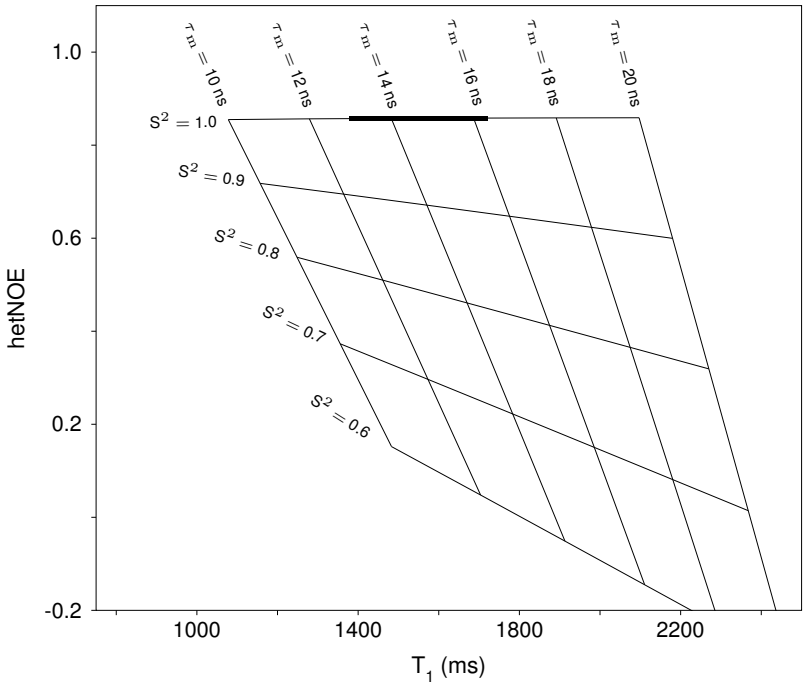

Figure 36: T_1 vs hetNOE plot showing Model-Free grid lines (black lines) and predicted relaxation data for βPGM from HydroNMR (black squares) at a field strength of 800 MHz. The grid lines are labelled with the value which is constant over that line, i.e., the $\tau_m = 10$ ns line shows all the values of T_1 & hetNOE where the value of τ_m is 10 ns and S^2 varies between 1.0 and 0.6. Likewise the $S^2 = 1.0$ line shows the expected values of T_1 & hetNOE for τ_m values in the range 10–20 ns where there is no internal mobility. For the plotting of the grid, τ_e is assumed to be 50 ps.

NMR relaxation data, such as Tensor[201, 202], ModelFree[203, 204], and Dasha[205]. More recently, after a more statistically-orientated treatment of the fitting procedures employed, the program Relax[191, 192] was written. It has been shown that Relax performs considerably better than the other available programs through a combination of factors including: avoiding the singular matrix failure of Levenberg-Marquardt minimisation procedures, avoiding bugs present in other programs, avoiding round-off error, and minimising χ^2 to much higher precision than the competing programs[191]. The subsequent discussion is therefore centred on the use of Relax.

The primary way in which Relax fitting differs from the fitting procedures of the other programs is that the model-free parameters are optimised for each spin prior to the optimisation of the global diffusion tensors. This eliminates the assumption that all the residues are perfectly rigid during fitting of the diffusion tensor. There are a number of diffusion models in Relax: the first assumes that each residue has its own "local" τ_m value, the second is an isotropic rotational diffusion model (spherical diffusion), and there are three anisotropic models. The first, "prolate spheroid" assumes that there are two equal sides of the molecule which are shorter than the remaining side, "oblate spheroid" assumes two equal sides which are longer than the remaining side, and the "ellipsoid" model is a fully anisotropic diffusion model with three diffusion tensors, one for each dimension.

Within each diffusion model, there are 10 model-free models which differ in the

parameters used to describe the internal motion. For example, in one model only S^2 is fit; in another model S_f^2, S_s^2, τ_f, τ_s, & R_{ex} are fit. A series of F-tests are then carried out to determine which of the models best satisfies the data with the fewest parameters. After this is done for each residue within a diffusion model, the total AIC values from the diffusion models are then compared to determine the solution[192].

For the fitting, a ^{15}N CSA value of -172 ppm was fixed, and the N–H bond length was fixed at 1.02 Å. These values are interdependent, and this combination of values has been shown to satisfy ^{15}N relaxation data well[206], as opposed to the combination of using 1.02 Å as the N–H bond length and -160 ppm for the ^{15}N CSA value, a combination which has been widely used for the analysis of relaxation data[207]. (The -160 ppm value for the ^{15}N CSA was derived on the basis of solid-state NMR data from peptides[208, 209].) It has in fact been shown that the ^{15}N CSA varies over the course of the primary sequence, and is related to the secondary structure at each backbone amide site. In one study, the average ^{15}N CSA value was measured as -172 ± 13 ppm[210], although even the mean value has been shown to vary between proteins[211]. As the effect of CSA on spin relaxation is related to B_0^2, this variation in CSA value and orientation can account for a difference in relaxation rates between otherwise identical residues of outside experimental error at high (>600 MHz) field strengths. In theory it is possible to fit the ^{15}N CSA value during the Relax fitting, but for reasons of parsimony this has not been persued. In most cases the lower field data sets were more sensitive than the higher field data sets, and so the impact of this potential problem is likely insignificant. It has also been shown that the orientation of the CSA tensor deviates from the N–H bond vector, though this was not taken into account during the relaxation data fitting.

For the anisotropic diffusion models within Relax, it is necessary to use a high resolution structure of the protein in order to calculate the individual N–H bond vectors. The apo βPGM data set were processed using the published crystal structure of apo βPGM (PDB code: 2WHE) which was solved to 1.55 Å resolution. The published crystal structure of the βPGM-MgF_3^--βG6P TSA complex(PDB code: 2WF5) does not feature the final three residues due to high B factors. As such, for analysis of the βPGM-MgF_3^--βG6P TSA complex relaxation data, an unpublished crystal structure which included all residues was used. This crystal structure was produced from the identical diffraction data as the published structure (1.30 Å resolution), but processed such that the C-terminus is included. There are no significant differences in the N–H bond vectors between the two structures.

3 Characterisation of Slow Timescale Dynamics of the βPGM-MgF_3^--βG6P TSA Complex by ^{19}F NMR

3.1 ^{19}F NMR Spectra of the βPGM-MgF_3^--βG6P TSA complex

3.1.1 Introduction

As was briefly discussed in Section 1.3.3, through the presence of ^{19}F nuclei in the active site of βPGM, the βPGM-MgF_3^--βG6P TSA complex allows for investigation of signals originating from the transition state analogue molecule itself; from the site at which the chemistry occurs. This allows for measurement of changes in the electrostatics of the active site on the catalytic timescale from the most relevant position in the active site. This is done by measuring the perturbation of the ^{19}F spins under different conditions. Before the differences in the ^{19}F NMR spectra of the βPGM-MgF_3^--βG6P TSA complex recorded under varying conditions can be discussed, the appearance of the ^{19}F NMR spectra must first be described.

3.1.2 ^{19}F NMR Spectra at 298 K

Figure 37 shows a typical ^{19}F NMR spectrum of the βPGM-MgF_3^--βG6P TSA complex. The largest, most downfield peak (-122.1 ppm) is that resulting from free F^- ions in solution. The three smaller peaks between -145 and -165 are the peaks corresponding to the MgF_3^- moiety, with each peak resulting from a F^- ion in a distinct chemical environment, with different numbers of hydrogen bond donors from the enzyme. They are referred to as (left–right) F_a, F_b, and F_c. Transfer NOE experiments were used to assign these to specific ions in the crystal structure[119]. In addition to these, there are three very small peaks which correspond to a minor conformer of the MgF_3^- TSA. By saturation transfer experiments, these have been shown to correspond to minor conformers of each of the larger peaks: again F_a, F_b, and F_c (left–right)[119]. These assignments are summarised in Table 3. The presence of these minor peaks, and their distance from their corresponding major peaks, indicates that there is significant electrostatic perturbation taking place on the ms–μs timescale which is felt by the atoms in the very heart of the active site where the chemistry occurs.

^{19}F δ (ppm)	Assignment
-122.1	Free F^-
-135.6	F_a(minor)
-136.3	F_b(minor)
-148.6	F_a
-153.3	F_b
-154.6	F_c(minor)
-159.9	F_c

Table 3: Assignments of the resonances in the ^{19}F NMR spectrum of the βPGM-MgF_3^--βG6P TSA complex as shown in Figure 37.

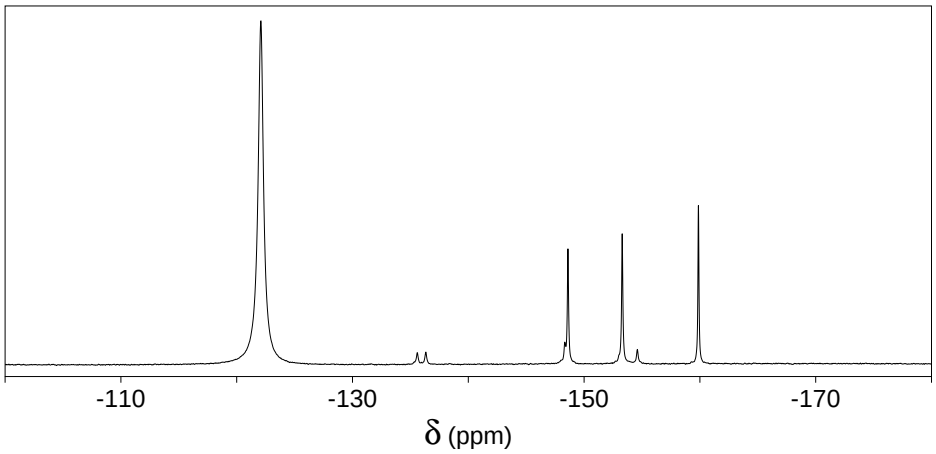

Figure 37: ^{19}F NMR spectrum of the βPGM-MgF_3^--βG6P TSA complex, showing the presence of two distinct sets of resonances for the MgF_3^- moiety. The assignments of the peaks are shown in Table 3. It can be seen in some spectra that there is the F_a peak has a small downfield shoulder, which is the result of the well-protected amide of L9 which is one of its hydrogen bonding partners. In the βPGM molecules where the amide proton from L9 has not exchanged to a deuteron during the buffer exchange, there is an isotope shift on the F_a peak.

3.2 ^{19}F SEXSY NMR

3.2.1 Results at 298 K

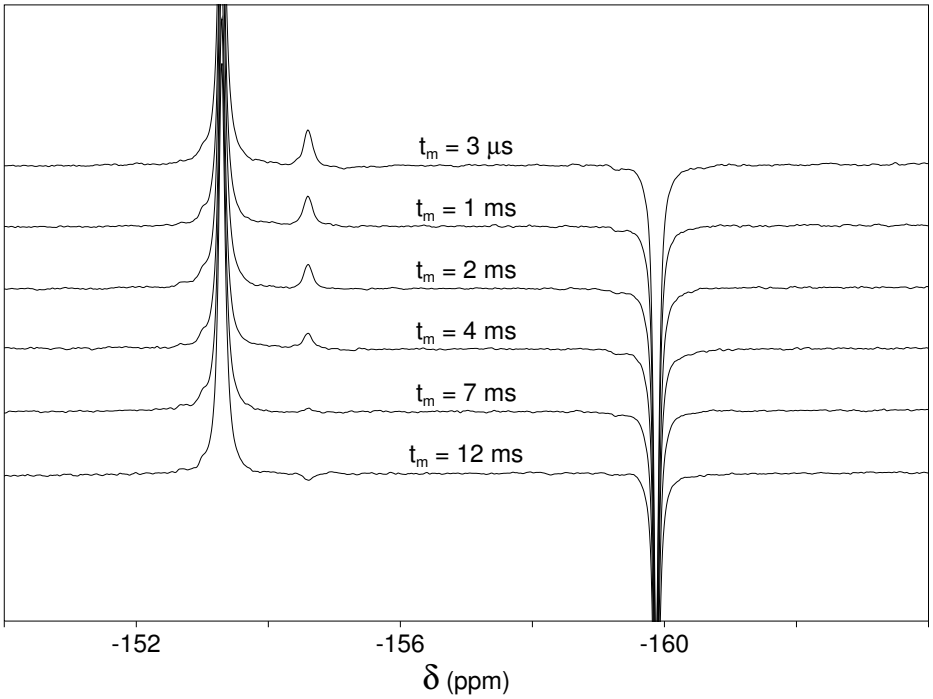

Figure 38: Cropped ^{19}F SEXSY spectra recorded at 298 K, showing the decrease in minor F_c peak intensity with increasing mixing times (t_m). These spectra demonstrate that at 298 K, k_{ex} is sufficiently large that after 12 ms the minor resonance is actually inverted, in spite of longitudinal relaxation (R_1).

The SEXSY spectra recorded at 298 K (calibrated) with a range of mixing times are shown in Figure 38. The measured peak heights of both major and minor F_c reso-

nances are shown, with fit lines corresponding to the equations shown in Figure 31, in Figure 39. From the agreement of the points with the fitted curves it is obvious that there is an additional source of noise on the F_c major peak intensity measurements. This was mirrored in the processing of ^{19}F SEXSY data recorded at different temperatures, though the source of error is unknown. As the differences between the fitted line and the measurements correspond to very small fractional changes, it is possible that inconsistency in the linear prediction algorithm is to blame. However, as the source of error remains unidentified, it cannot reasonably be assumed that this source of error affects the major intensities only, therefore measurements on both major and minor peak intensities had their errors adjusted together.

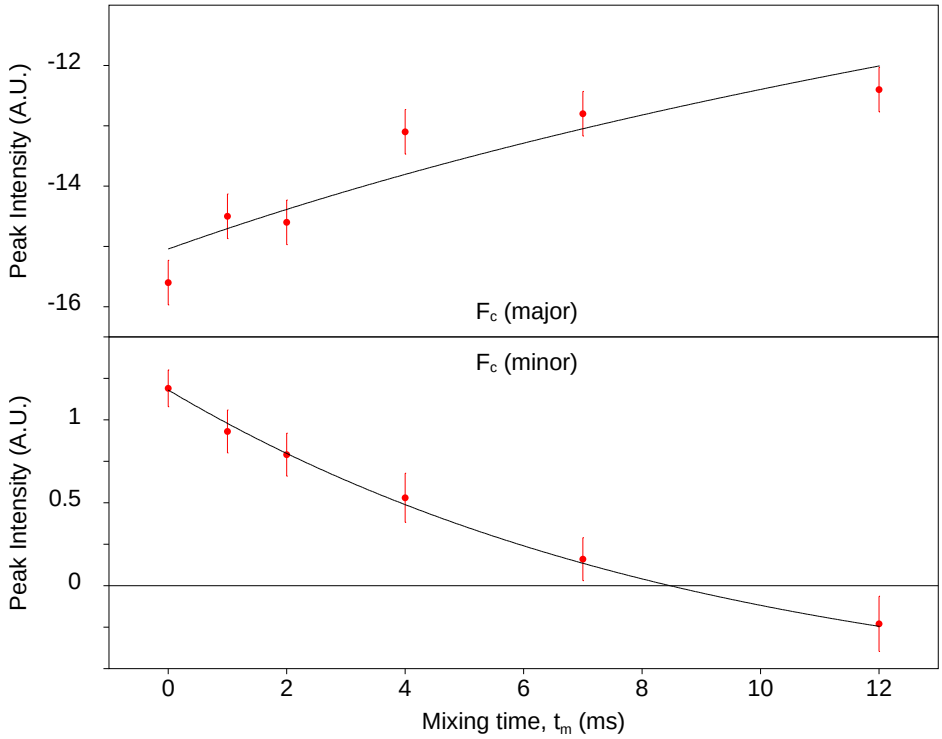

Figure 39: F_c major and minor resonance peak heights measured from ^{19}F SEXSY spectra recorded at 298 K plotted against mixing time t_m. The black curves are the fit lines resulting from fitting these data with the equations in Figure 31.

The resultant fitted parameter values were:

- $I_0 \quad = 16.2 \pm 0.2$
- $F_a \quad = 0.927 \pm 0.005$
- $R_1 \quad = 5 \pm 1\ s^{-1}$
- $k_{ex} \quad = 97 \pm 11\ s^{-1}$

As was mentioned earlier, k_{ex} is the sum of the forward and backward rate constants. Using the populations of the Major and Minor conformers, it is possible to split this number into components. Due to differential linewidths between the 2 species the F_a number cannot be directly used for this. The linewidths must first be measured: this was done from spectra which had been processed using the same window function

as those spectra from which the peak heights were measured (20 Hz linebroadening). Repeated measurements were used to generously estimate the errors. The linewidth of F_c major under these conditions was measured as 46 ± 2 Hz, whereas the minor peak was broader with a linewidth of 75 ± 4 Hz. If the proportion of minor conformer (given by $1 - F_a$) is multiplied by the ratio between these linewidths, then the population of minor conformer present at 298 K is estimated to be 12 ± 1%.

This population figure allows the rate of exchange, k_{ex}, to be split into forward and backwards rate constants. The slower rate — the rate at which the major conformer is transformed into the minor conformer — is 12 ± 2 s^{-1}. The faster of the two rate constants — the rate at which molecules in the minor conformer exchange to the major conformer — is 86 ± 10 s^{-1}.

3.2.2 Similarity to Timescale of Catalysis

The rate of catalytic turnover, k_{cat}, has been measured in the βG1P → βG6P direction by coupled-assay under steady-state conditions as 65 ± 1 s^{-1}[129]. Unfortunately access to the reverse reaction (βG6P → βG1P) is difficult as no coupled assay to detect the production of βG1P exists, prohibiting spectrophotometric measurement. It is also important to note that for many enzyme reactions, and particularly for the Ping Pong Bi Bi reaction mechanism of βPGM, k_{cat} is a combination of many rate constants from the individual reactions steps and affinities of the intermediates. It is therefore not necessarily a useful figure for comparison.

Instead, the most pertinent rate constant measured in the study detailed in Goličnik et al, 2009[129] is the rate at which the bisphosphate (βG16BP) is converted to βG6P, as the βPGM-MgF_3^--βG6P TSA complex mimics the mid-point of either this or the reverse reaction. Whereas the k_{cat} of βPGM is 65 ± 1 s^{-1}, this individual reaction step is reported to occur at 105 ± 3 s^{-1}. Although neither is exactly matched by the faster of the two conformational exchange rate constants measured by ^{19}F SEXSY NMR, it is notable that this rate occurs on the same timescale as catalytic events during substrate turnover. As discussed in Section 1.2.9, this has been seen for many enzyme systems, and suggests that the conformational exchange seen by ^{19}F NMR may have some catalytic significance.

3.2.3 Nature of the "Minor" Conformer

It is impossible to get a detailed picture of the nature of the minor conformation from the ^{19}F NMR alone, though some useful information can be extracted. The δs indicate that there is substantially lower electron density around each of the fluorides in the minor conformation. Measuring the $\Delta\delta_{(\text{major}-\text{minor})}$ figures for each of the F^- ions provides site-specific information. For F_a, F_b, and F_c, the $\Delta\delta_{(\text{major}-\text{minor})}$ figures are −13.0, −17.0, and −5.3 ppm respectively. This shows that the electron density around F_b undergoes the largest change during the conformational dynamics.

Solvent-induced isotope shifts (SIISs) — the $\Delta\delta$s for each of the ^{19}F nuclei when

the spectra are recorded in 100% H_2O vs. 100% D_2O — provide a measurement of proton density around each of the nuclei[212], as shown in Figure 40. Measurements of the $\Delta SIIS_{(major-minor)}$ for each of the ^{19}F nuclei were done by Dr. Joanna Griffin[213] and show that, in keeping with lower electron density, in general there is higher proton density around the F^- ions in the minor conformer, and that the largest change in proton density between the two conformers is experienced by F_a. According to both measurements the F_c site is the most similar in both conformers. This information is shown in Table 4. These observations are consistent with a shortening in length of the hydrogen bonds from the protein to the MgF_3^- moiety.

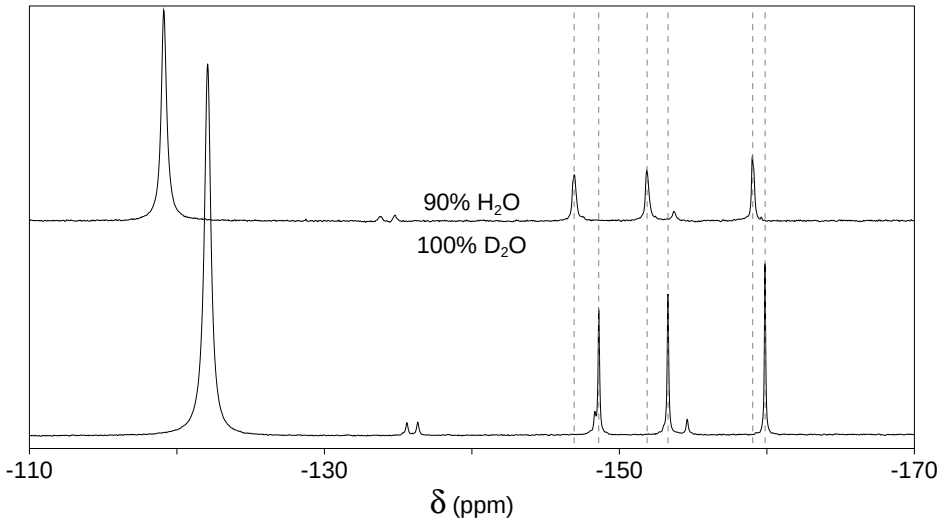

Figure 40: ^{19}F spectra of the βPGM-MgF_3^--βG6P TSA complex in H_2O and D_2O, showing SIISs for the major conformer peaks, where the SIIS for F_a is largest, and F_c is the smallest. It can be seen here that the free F^- peaks moves the most as it is surrounded solely by water molecules. The advantage of working in 100% D_2O for ^{19}F NMR is also clear: the peaks are far sharper due to the lack of dipolar interactions with ^1H nuclei, thereby increasing signal:noise.

^{19}F nucleus	**$SIIS_{(major)}$**	**$SIIS_{(minor)}$**	**$\Delta SIIS_{(major-minor)}$**	**$\Delta\delta_{(major-minor)}$**
F_a	-1.59	-1.77	0.18	-13.0
F_b	-1.43	-1.49	0.06	-17.0
F_c	-0.96	-0.89	-0.07	-5.3

Table 4: $\Delta\delta$s and SIISs between the Major and Minor conformation of the βPGM-MgF_3^--βG6P TSA complex. All measurements have units of ppm.

3.3 ^{19}F SEXSY at Other Temperatures

Whilst the co-incidence of the rate of conformational exchange with the rate of catalysis is suggestive of a catalytic role for the active site dynamics, it is not proof. What would be convincing is if the activation enthalpy ($\Delta H^‡$) and entropy ($\Delta S^‡$) of the two processes were similar. The ^{19}F SEXSY method was therefore applied to βPGM-MgF_3^--βG6P TSA complex samples at range of temperatures (283–303 K, in increments of 5 K) such that these parameters could be fit for the conformational

exchange. The range of temperatures studied would ideally be wider, but βPGM is not stable at higher temperatures and at temperatures lower than 283 K the solvent viscosity increases such that the correlation time of the molecules — and therefore linewidths and sensitivity — suffer. The selective R_1 and k_{ex} fitted at each temperature as well as the population of minor state (corrected for linewidths as detailed above) for each temperature are shown in Table 5.

T (K)	$\mathbf{k_{ex}}$ **(s^{-1})**	**Minor pop^n (%)**	$\mathbf{R_1}$ **(s^{-1})**
303	211 ± 17	7 ± 1	7 ± 1
298	97 ± 11	12 ± 1	5 ± 1
293	52 ± 3	16 ± 1	4 ± 1
288	29 ± 1	19 ± 1	3.4 ± 0.1
283	15 ± 1	22 ± 2	3.5 ± 0.1

Table 5: ^{19}F SEXSY results at different temperatures.

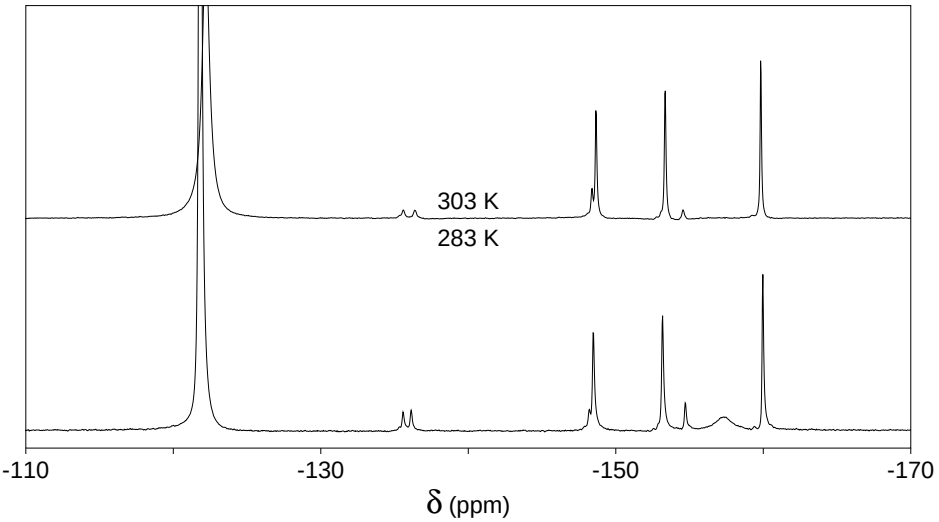

Figure 41: ^{19}F spectra of the βPGM-MgF_3^--βG6P TSA complex recorded at 303 K and 283 K, to demonstrate the temperature-dependence of the minor conformer population. The spectra have been scaled such that the height of the peaks in the major conformation are matched. At lower temperature, a broad peak for MgF^+ ions appears at −157.3 ppm.

The results shown in Table 5 show that the population of minor conformer is highly temperature-dependent, increasing at low temperature, as shown by the comparison of ^{19}F spectra recorded at 303 K and 283 K shown in figure Figure 41. This indicates that the minor conformer is entropically disfavoured relative to the major conformer, and may become the dominant species at cryogenic temperatures, such as those used in X-ray diffraction. This is discussed later. The rate of exchange between the two conformers is also very temperature-dependent, increasing roughly by a factor of 2 for every 5 K increase in temperature. This shows that the $\Delta G^{\ddagger}$ for this transition is large, demonstrating significant changes in bonding between the two conformers and therefore implying large changes in the protein backbone also occur as part of this transition.

The k_{ex} numbers become better defined at lower temperature for two reasons: firstly the population of minor conformer increases at low temperature, meaning that the signal:noise on the peak height measurements is better, and secondly the delays were better chosen to sample more of the decay of the minor peak intensity. For example, there is more curvature in the decays recorded at 283 K shown in Figure 42 than those recorded at 298 K shown in Figure 39.

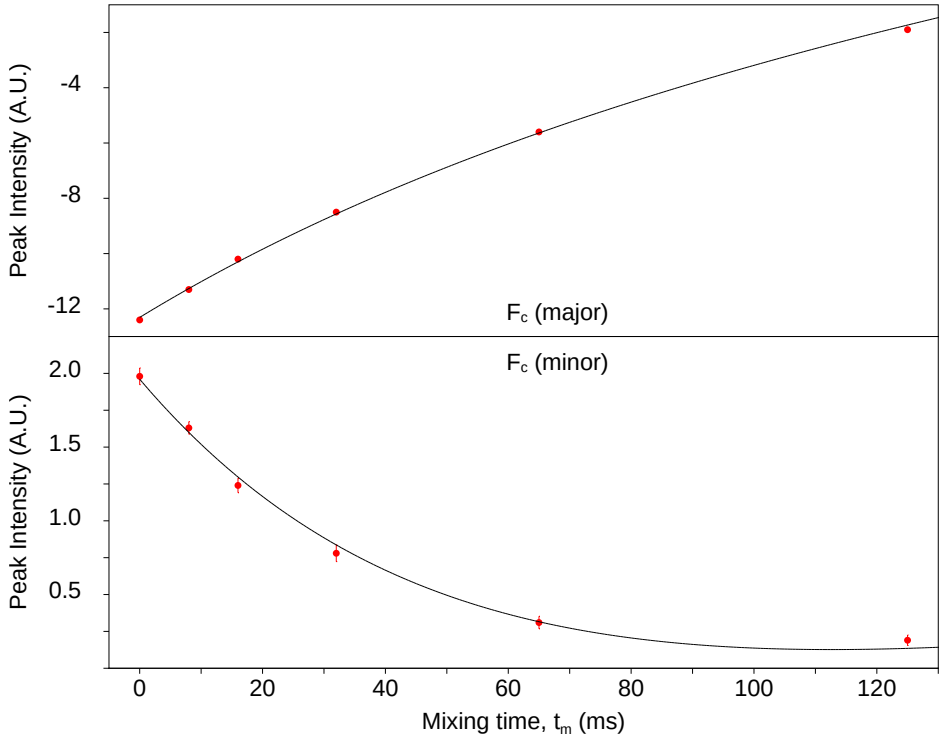

Figure 42: F_c major and minor resonance peak heights measured from ^{19}F SEXSY spectra recorded at 283 K plotted against mixing time t_m. The black curves are the fit lines resulting from fitting these data with the equations in Figure 31. Compared to the decay curves recorded at 298 K (see Figure 39 on Page 75), there is more curvature due to more appropriately chosen delay times, meaning that k_{ex} is better defined.

Although the selective R_1 figures which are fit in this analysis are not of interest mechanistically, it is remarkable that they are not temperature dependent at low temperature, but become increasingly so as the temperature is raised. Presumably there are molecular motions present at low temperature which occupy a region of the spectral density function $J(\omega)$ on which the ^{19}F spin–lattice relaxation rate is not dependent (i.e., not ω_F, ω_H, or $\omega_H \pm \omega_F$). An increase in temperature presumably increases the rate of these motions such that they occur at frequencies closer to those to which the relaxation of ^{19}F is sensitive.

3.4 Thermodynamic Analysis

The rate of a reaction or conformational exchange at a given temperature depends on the activation energy of the transition, as shown by the Arrhenius equation in Figure 43. Therefore, the temperature dependence of the exchange rates can be used

to fit the activation energy of the transition. The fitting of the activation energy (E_a or $\Delta G^{\ddagger}$) can itself be split into enthalpic ($\Delta H^{\ddagger}$) and entropic ($\Delta S^{\ddagger}$) contributions by use of the Eyring–Polanyi equation shown in Figure 44.

The equation makes two assumptions: the first is that the transmission coefficient (κ) is equal to 1, and the second is that there is no change in heat capacity (ΔC_p) during the reaction. The re-arranged linear form of the equation shown in Figure 44 facilitates plotting of the data to assess for ΔC_p. The general pre-exponential factor in the Arrhenius equation is also specified as ${}^{k_B T}/_{h}$ in the Eyring–Polanyi equation. The validity of this figure, derived from gas-phase studies, has been widely debated, particularly for the application of thermodynamic analyses to the field of protein folding in solution[214].

$$k = A \exp\left(\frac{-E_a}{RT}\right)$$

where:
k is the exchange/reaction rate (s^{-1})
T is the temperature (K)
E_a is the activation energy (J/mol)
A is a pre-exponential factor
R is the gas constant

Figure 43: The Arrhenius equation, describing the dependence of reaction/exchange rates on the activation energy.

$$k = \left(\frac{k_B T}{h}\right).\exp\left(-\frac{\Delta H^{\ddagger}}{RT}\right).\exp\left(\frac{\Delta S^{\ddagger}}{R}\right)$$

$$\ln\left(\frac{k}{T}\right) = -\frac{\Delta H^{\ddagger}}{R}.\frac{1}{T} + \frac{\Delta S^{\ddagger}}{R} + \ln\left(\frac{k_B}{h}\right)$$

where:
k is the unidirectional exchange rate (s^{-1})
T is the temperature (K)
k_B is the Boltzmann constant
h is Planck's constant
R is the gas constant
$\Delta H^{\ddagger}$ is the activation enthalpy (kJ/mol)
$\Delta S^{\ddagger}$ is the activation entropy ($kJ/mol/T$)

Figure 44: The Eyring–Polanyi equation (top), and re-arranged to its linear form (below).

As in the analysis of the 298 K dataset, by use of the population figures k_{ex} can be split into individual forward and backward rate constants for every temperature. These numbers are shown in Table 6. If $1/T$ is plotting against $\ln(k/T)$ as in the linearised form of the Eyring–Polanyi equation, the points should lie on a straight line

if there is no significant change in heat capacity between the ground states and the transition state. These plots are shown in Figure 45. Although there is a non-random distribution of the points around the lines of best-fit for both forward and backwards rates, the points do indeed fit to a straight line within the errors on the measurements.

T (K)	$k_{(major \rightarrow minor)}$ s^{-1}	$k_{(minor \rightarrow major)}$ s^{-1}
303	15.6 ± 3.7	195 ± 17
298	11.6 ± 3.2	86 ± 10
293	8.2 ± 1.5	44 ± 2
288	5.5 ± 0.7	23.4 ± 0.7
283	3.3 ± 0.6	11.3 ± 0.4

Table 6: Unidirectional rate constants calculated from ^{19}F SEXSY results at different temperatures.

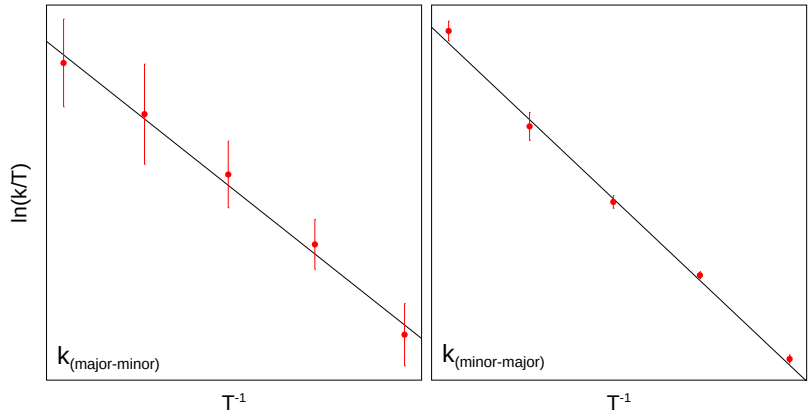

Figure 45: Eyring plots of forward and backward exchange rates showing that in both cases, despite the appearance of curvature, the points can be satisfied by a straight line. The black straight lines are derived from fitting the points to $y = mx + c$.

A similar process can be taken for looking at the temperature-dependence of the equilibrium constant, K_{eq}, which is easily calculated from the population figures at different temperatures. This can be used to calculate the difference in enthalpy and entropy between the two ground states (major conformer and minor conformer) by use of the integrated Van't Hoff equation shown in Figure 46. As ΔH and ΔS can be fit between all three species (two ground states and transition state between them), the results can be tested for self-consistency. Like the Eyring–Polanyi equation, the integrated Van't Hoff equation can be linearised to allow for plotting to assess for ΔC_p between ground states, as shown in Figure 47.

As there was no definable ΔC_p between either ground state and transition state, ΔC_p between the two ground states should equal 0. However as is shown in Figure 47, there is very apparent curvature in the plot of $1/T$ against $\ln K_{eq}$, meaning that ΔC_p between the ground states is significant. Perhaps the reason why this is not reflected in the Eyring plots in Figure 45 is because in the studied temperature range, $\Delta G^{\ddagger}$ is dominated by $\Delta H^{\ddagger}$, whereas the curvature introduced by ΔC_p is manifested in $\Delta S^{\ddagger}$.

$$K_{eq} = \exp\Big(-\frac{\Delta H}{RT}\Big).\exp\Big(\frac{\Delta S}{R}\Big)$$
$$\ln K_{eq} = -\frac{\Delta H}{R}.\frac{1}{T} + \frac{\Delta S}{R}$$

where:

K_{eq} is the equilibrium constant

T is the temperature (K)

R is the gas constant

ΔH is the difference in enthalpy (kJ/mol)

ΔS is the difference in entropy $(kJ/mol/T)$

Figure 46: The integrated Van't Hoff equation (top), and re-arranged to its linear form (below).

The dependence of ΔH and ΔS on ΔC_p is shown in Figure 48.

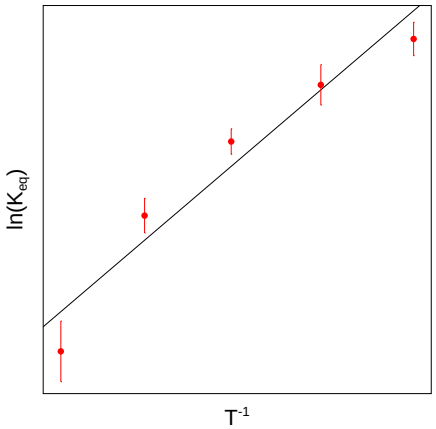

Figure 47: Van't Hoff plot of the equilibrium constant between major and minor conformer, showing that ΔC_p is significant. The black straight line is derived from fitting the points to $y = mx + c$.

Although it is possible to fit the individual forward and backward rate constants to the Eyring–Polanyi equation, to allow for simultaneous fitting of both rates and to minimise the manipulation of the data prior to fitting, k_{ex} and the population of minor conformer were fit instead. When the fitting was done assuming $\Delta C_p = 0$, the fit lines did not well match the points, and the resulting χ^2 was 28.2. When ΔC_p is allowed to vary, the resulting χ^2 is 11.4 showing a clear improvement in fit quality. The pre-exponential constant was also allowed to vary — effectively allowing the transmission coefficient, κ, to vary — and the resulting χ^2 value was 10.2, which is insufficient an improvement to the fitting quality to justify the introduction of another variable according to Akaike's Information Criterion[194]. Figure 49 shows that it is only the k_{ex} at 303 K that is not satisfied by the fitted curves when ΔC_p is allowed to vary, but otherwise the points are well satisfied by the fitted curves.

Fitting of the K_{eq} values to the integrated Van't Hoff equation shown in Figure 50 was also much improved by fitting ΔC_p, with χ^2 dropping from 9.2 to 1.9 when it is

$$\Delta H = \Delta H_0 + (\Delta C_p.(T - T_0))$$
$$\Delta S = \Delta S_0 + \left(\Delta C_p . \ln\left(\frac{T}{T_0}\right)\right)$$

where:

T is the temperature

T_0 is a reference temperature, eg. 298 K

ΔH is the difference in enthalpy at a given value of T

ΔH is the difference in enthalpy at T_0

ΔS is the difference in entropy at a given value of T

ΔS is the difference in entropy at T_0

ΔC_p is the difference in heat capacity

Figure 48: The dependence of enthalpy and entropy changes on heat capacity. It can be seen that the dependence of ΔS on ΔC_p is non-linear, and as such ΔC_p will only give rise to curvature in plots such as those shown in Figure 45 where ΔS makes a significant contribution to the overall ΔG.

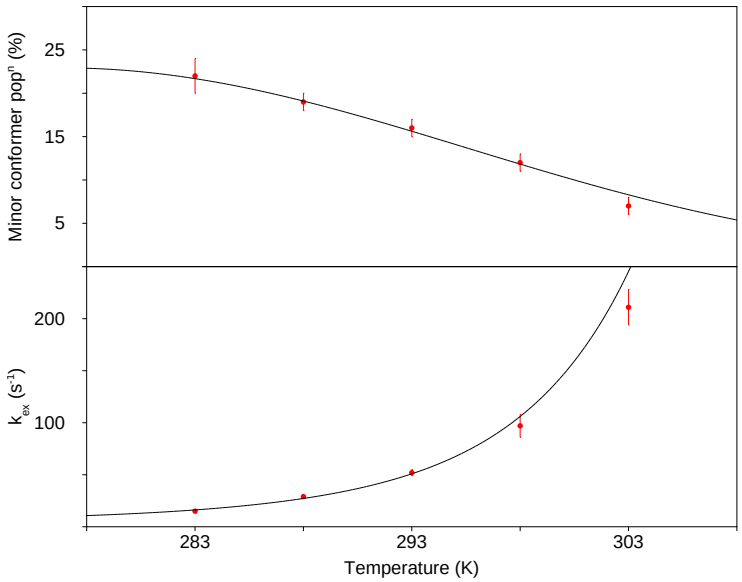

Figure 49: Fitting of the population and k_{ex} values to the Eyring–Polanyi equation.

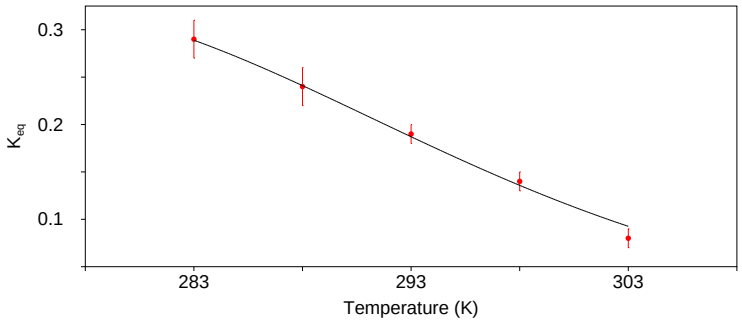

Figure 50: Fitting of the K_{eq} values to the integrated Van't Hoff equation.

introduced. As with the Eyring–Polanyi fitting results, the 303 K value is least satisfied by the fitted curve. Table 7 shows the resultant ΔH, ΔS, and ΔC_p values between both ground states and transition state. The displayed ΔH and ΔS values refer to the values at 298 K.

Transition	**ΔH (kJ/mol)**	**ΔS (J/mol/K)**	**ΔC_p (kJ/mol/K)**
Major $\rightleftharpoons$ Minor	-52 ± 8	-190 ± 30	-2.2 ± 1.2
Major $\rightleftharpoons$ TS	67 ± 10	0 ± 30	1.3 ± 1.6
Minor $\rightleftharpoons$ TS	120 ± 6	200 ± 20	3.8 ± 1.0
Residual	1 ± 14	10 ± 41	0.3 ± 2.2

Table 7: Fit values for ΔH, ΔS and ΔC_p from the Eyring–Polanyi and Van't Hoff fitting. The "residual" numbers refer to the difference between the values fit for, e.g., the Minor→TS transition, when compared to the values derived from a combination of the other two, e.g., Major→Minor − Major→TS. It can be seen that these differences are much smaller than their uncertainty, showing excellent agreement between the Eyring–Polanyi and Van't Hoff fitting.

The magnitude of the $\Delta H^{\ddagger}$, $\Delta S^{\ddagger}$, and ΔC_p values are all very large: on the order of those figures derived from thermodynamics analyses of protein folding[215]. As with the large ^{19}F $\Delta\delta$ values and the implied changes in hydrogen bonding as indicated by SIISs, the size of these values suggest that the conformational exchange between major and minor conformer of βPGM in the βPGM-MgF_3^--βG6P TSA complex is very large, leading to widespread changes in intramolecular interactions. As with the ^{19}F NMR spectral data, unfortunately this conclusion does not allow the elucidation of the finer details of the structure of the minor form, and therefore speculation as to how it may relate to catalytic turnover by βPGM. NMR experiments studying protein resonances are necessary to unravel structural detail, and the results from such experiments are discussed in Section 4.

These data also for allow comment on the agreement of the minor→major exchange rate from ^{19}F SEXSY (86 ± 10 s^{-1}) with the figure for the dephosphorylation of βG16BP derived from the kinetic analysis (105 ± 3 s^{-1})[129]. With the $\Delta H^{\ddagger}$ for the exchange process being so large, a difference in the temperature at which the NMR and the kinetics were performed at by as little as 0.5 K would bring these numbers within error. Although the NMR spectrometer temperature control was carefully calibrated, the temperature control of the kinetics assays was rather more crude: the experiments were performed at ambient temperature with the air-conditioning of the room set to the desired temperature. The error in the set temperature in the kinetics data is likely much larger than $\pm$ 0.5 K, which means that within experimental error, the exchange rate and dephosphorylation step occur at precisely the same rate.

4 Characterisation of Slow Timescale Dynamics of the βPGM-MgF_3^--βG6P TSA Complex by ^{15}N NMR

4.1 ^{15}N Relaxation Dispersion Experiments

4.1.1 Minor Resonances in ^{1}H–^{15}N TROSY Spectra

Section 3 showed that there is a minor conformation of the βPGM-MgF_3^--βG6P TSA complex which transforms into the major conformer at a rate remarkably similar to k_{cat}. The thermodynamic parameters of the exchange implied that the change involved large scale rearrangement of the protein structure. Accordingly, as well as being visible in ^{19}F spectra of the βPGM-MgF_3^--βG6P TSA complex, minor conformation peaks can also be seen in ^{1}H–^{15}N TROSY spectra of the complex, and many of them have been assigned to their corresponding major peak by Dr. Joanna Griffin[213]. The population of the minor species as seen by ^{1}H–^{15}N TROSY approximately matches the population of the minor conformer seen in ^{19}F spectra, and the populations also have a similar temperature-dependence, as demonstrated in Figure 51. It is therefore tempting to suggest that the minor resonances seen by ^{19}F and ^{15}N NMR are different observations of the same exchange phenomenon.

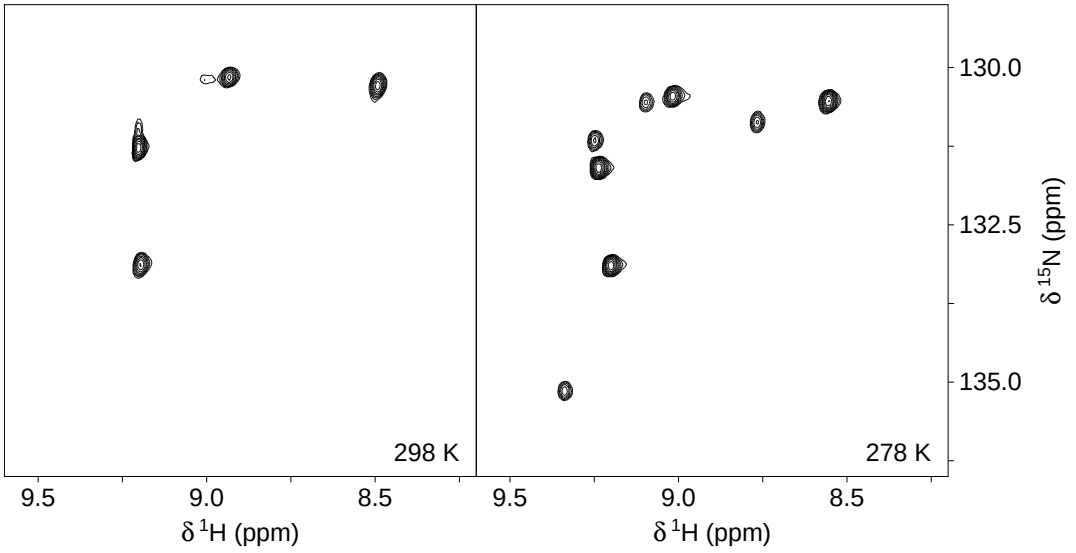

Figure 51: Section of ^{1}H–^{15}N TROSY spectra of the βPGM-MgF_3^--βG6P TSA complex recorded at 298 K and 278 K by Dr. Joanna Griffin. The low-temperature spectrum clearly shows the existence of minor peaks, where the higher temperature spectrum does not. The peaks are visible in the low-temperature spectrum due to higher population of minor species and lower exchange rate and therefore less exchange broadening of the minor peaks.

Although the similar temperature-dependence of the populations is compelling, similar exchange rates measured by both ^{19}F and ^{15}N NMR would be convincing. Relaxation dispersion provides a method by which the exchange rates between conformations and their populations can be measured, as well as changes in chemical shifts ($\Delta\delta$s) of each spin system[185]. ^{15}N Relaxation dispersion is therefore a way of assess-

ing the link between the minor conformer assignment of the ^{1}H–^{15}N TROSY spectrum peaks and the exchange studied by ^{19}F NMR.

4.1.2 Preliminary ^{15}N Relaxation Dispersion Data

^{15}N Relaxation dispersion data was collected at 600 MHz with a cryoprobe to assess the viability of the experiment. There were a subset of resonances which displayed some dispersion behaviour, although the signal:noise on these measurements was poor, preventing k_{ex} from being fitted. However, relaxation dispersion curves for all residues could be satisfied by fixing the rate of exchange (k_{ex}) and populations to those derived from the ^{19}F SEXSY analysis, and allowing only ^{15}N $\Delta\delta$ to vary for each residue. These data are therefore consistent with the rates derived from the ^{19}F analysis. Example data from residue D137 to illustrate the axes used to display subsequent dispersion curves in this chapter are shown in Figure 52, and data from all residues which show dispersion behaviour at 600 MHz are shown in Figure 53.

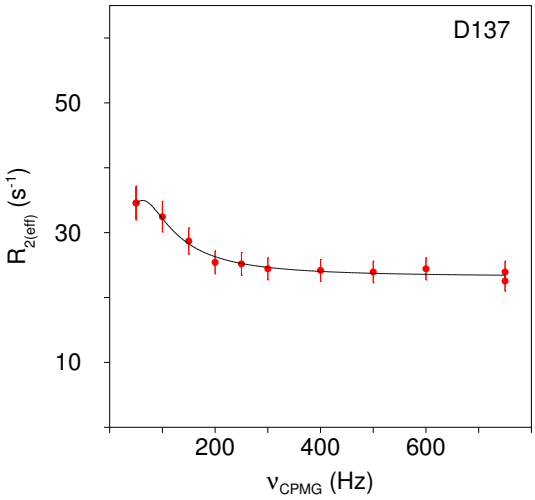

Figure 52: Example ^{15}N relaxation dispersion profile from residue D137, recorded at 600 MHz at 298 K. At low values of ν_{CPMG}, there are significant R_{ex} contributions to the effective R_2 relaxation rate, $R_{2(eff)}$. The limits for the axes are used in subsequent plots and are 0–800 Hz & 0–65 s^{-1} for the x & y axes, respectively.

Whilst this agreement supports the idea that the protein minor peaks seen in the ^{1}H–^{15}N TROSY originate from the same minor conformer that gives rise to the peaks in the ^{19}F spectrum, these relaxation dispersion data do not define the exchange rates. The relaxation dispersion method works by quantifying variation in the chemical exchange (R_{ex}) terms of each spin system as a function of the refocussing pulse frequency. The R_{ex} contribution to R_2 therefore needs to be sufficiently large such that this variation can be quantified. Therefore, signal:noise could be increased not only by more sensitive data, but also by increasing the R_{ex} terms on the spin relaxation, such that the difference between points is larger. The relaxation dispersion method is only sensitive to chemical exchange occurring on a particular timescale: in the range of 100–3000 s^{-1}[216, 156], which the major⇌minor exchange is on the very edge of. Because the

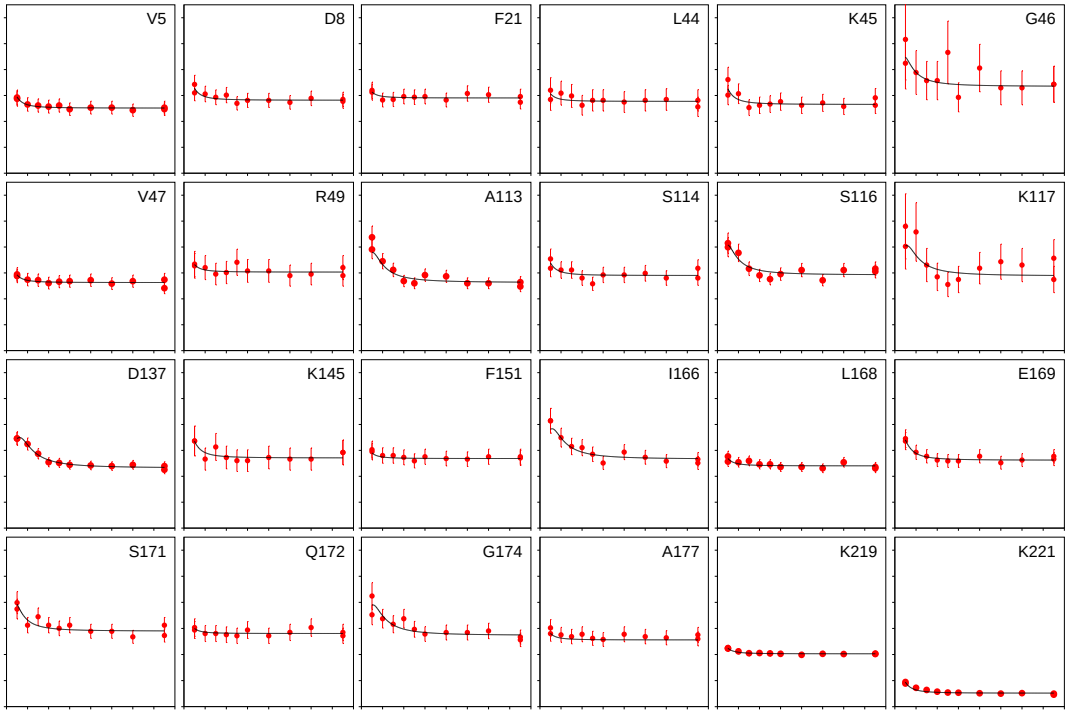

Figure 53: ^{15}N relaxation dispersion profiles recorded at 600 MHz at 298 K. The signal:noise on the $R_{2(eff)}$ values is generally poor. The fit lines are those resultant from fixing the forward and backward rates of exchange to those measured from the ^{19}F SEXSY analysis, and satisfy the points well. Axes and limits as shown in Figure 52.

exchange is so slow, only measurements with the lowest ν_{CPMG} field strength deviate from the baseline, and as only a few points define the exchange parameters these exchange parameters are not well defined. The amplitude of the dispersion curves and number of points which deviate from the baseline can both be improved by either increasing the temperature, or increasing the external magnetic field.

Data sets at higher temperatures were collected and did increase the signal:noise on the dispersion curves considerably, as shown for 2 residues in Figure 54. However, the aim of these measurements was to compare the exchange rates to that measured by ^{19}F SEXSY experiments and therefore the catalytic rate (currently measured only at 298 K), and also to the fitted ^{15}N $\Delta\delta$s of each spin system to the minor form assignment (known only up to 298 K). Therefore, despite the improvement in data quality at higher temperature, in order to allow comparisons to other data sets it was decided to record the data at 298 K and improve the sensitivity by collecting data at higher field instead. Not only is high field NMR intrinsically more sensitive due to higher Boltzmann polarisation of the nuclei, but R_{ex} terms also increase with field strength. The field-dependence of the dispersion curves measured for 2 representative residues is shown in Figure 55. ^{15}N relaxation dispersion experiments were conducted at 800 MHz but due to the RT probe used to record the data, the exchange parameters were defined no better than for the 600 MHz data which had been recorded using a cryoprobe. Only when ^{15}N relaxation dispersion data was collected at 900 MHz with a cryoprobe at the SONNMRLSF facility at Utrecht University were these sensitivity issues overcome.

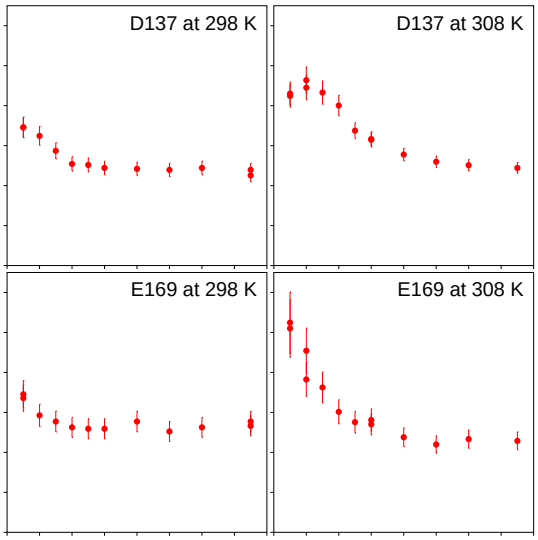

Figure 54: Temperature-dependence of ^{15}N relaxation dispersion profiles recorded at 600 MHz at 298 K and 308 K for 2 residues. Increased R_{ex} components at higher temperature lead to loss of sensitivity, but this is far outweighed by the increase in amplitude of the dispersion curves. The number of points which deviate from the baseline is also increased, allowing more sensitive fitting of exchange parameters. Axes and limits as shown in Figure 52.

4.1.3 900 MHz ^{15}N Relaxation Dispersion Data

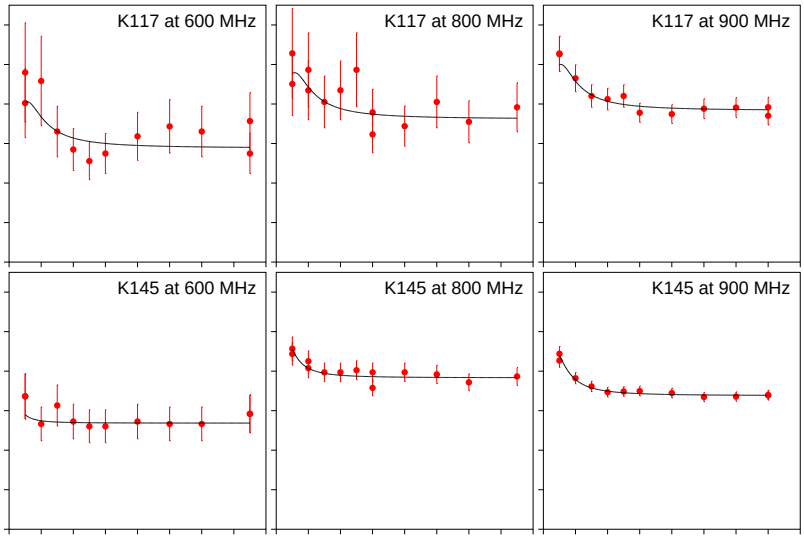

Figure 55: Field-dependence of ^{15}N relaxation dispersion profiles recorded at 600 MHz, 800 MHz, and 900 MHz at 298 K for two residues. Increased R_{ex} components at higher field increase the amplitude of the dispersion curves, and increase the number of points which deviate from the baseline. The sensitivity benefit from collecting data at high field with a cryoprobe is also demonstrated. The black lines are the result of fitting the data with the forward and backwards exchange rate constants fixed to those values derived from the ^{19}F SEXSY analysis. Axes and limits as shown in Figure 52.

Although the sensitive 900 MHz relaxation data allowed better ability to discriminate between fitted parameter values, most residues could still be satisfied by minimising while fixing the exchange rate and populations at those values derived from the ^{19}F SEXSY analysis, strongly implying that the major⇌minor exchange event studied

by both nuclei are the same. However, there is a small subset of residues which clearly are not satisfied by this exchange regime: their fitted curves either do not reach sufficiently high $R_{2(eff)}$ values or flatten off too quickly, such as in residues V5, I166, L217 and Q220. This implies that there is a faster exchange regime present in a small subset of the enzyme. The relaxation dispersion curves recorded at 900 MHz for all residues which give ^{15}N $\Delta\delta$ of at least 0.3 ppm when the SEXSY exchange rates are fixed are shown in Figure 56.

4.2 ^{15}N Relaxation Dispersion Data Fitting

4.2.1 Elucidation of Exchange Parameters

The presence of two separate exchange rates emphasises the need to use the ^{15}N relaxation dispersion data to define the rates independent of any other data. Despite the sensitive $R_{2(eff)}$ measurements, fitting rates was still difficult due to the exchange being too slow and the $\Delta\delta$s generally being quite small, particularly compared to the many studies of transient folding intermediates using this technique[217]. In particular, because the system is in slow exchange, the faster rate (minor→major) is poorly defined by the dispersion data[198].

To separate out the rates of exchange and differences in precession frequency between conformers during the fitting — the latter of which is field-dependent — it is necessary to simultaneously fit dispersion data which has been recorded at multiple field strengths[218]. Residues for which data at three fields (600, 800, and 900 MHz) had been measured and which gave significant dispersion at 900 MHz, i.e., $\Delta R_{2(eff)}$ between the ν_{CPMG} = 50 Hz & ν_{CPMG} = 700 Hz points of more than 5 s^{-1}, were selected for this fitting. There were 19 such residues: D8, D15, L44, K45, V47, L112, A113, S114, S116, K117, D137, K145, F151, I166, E169, S171, G174, A177, and K221. These residues were fit allowing the rate of major→minor exchange (k_a), population of major conformer (p_a) and $\Delta\delta$ to vary individually for all residues.

The average error on the k_a figures of >2500% emphasises how non-ideal the exchange parameters of this system are. Despite most residues poorly defining the k_a rate, the rates could still be extracted from the data. The "consensus" values could be estimated by plotting a gaussian curve for each residue, with the centre defined by the fitted k_a value and the width defined from its fitting error. The sum of these gaussians can then be plotted, as in Figure 57, which indicates the likelihood of each particular k_a value existing in the system. Figure 57 shows that although a range of k_a values have been fit by the data, there are 2 distinct regions that these values cluster around: 10.6 $s^{-1}$and 25.2 s^{-1}.

To sort the residues into the exchange regime by which they are affected, the multi-field fitting was repeated twice, with the k_a values being fixed at the values derived from the "gaussian analysis": 10.6 and 25.2 s^{-1}. According to Akaike's Information Criterion (AIC), if the introduction of a parameter does not improve χ^2 by at least

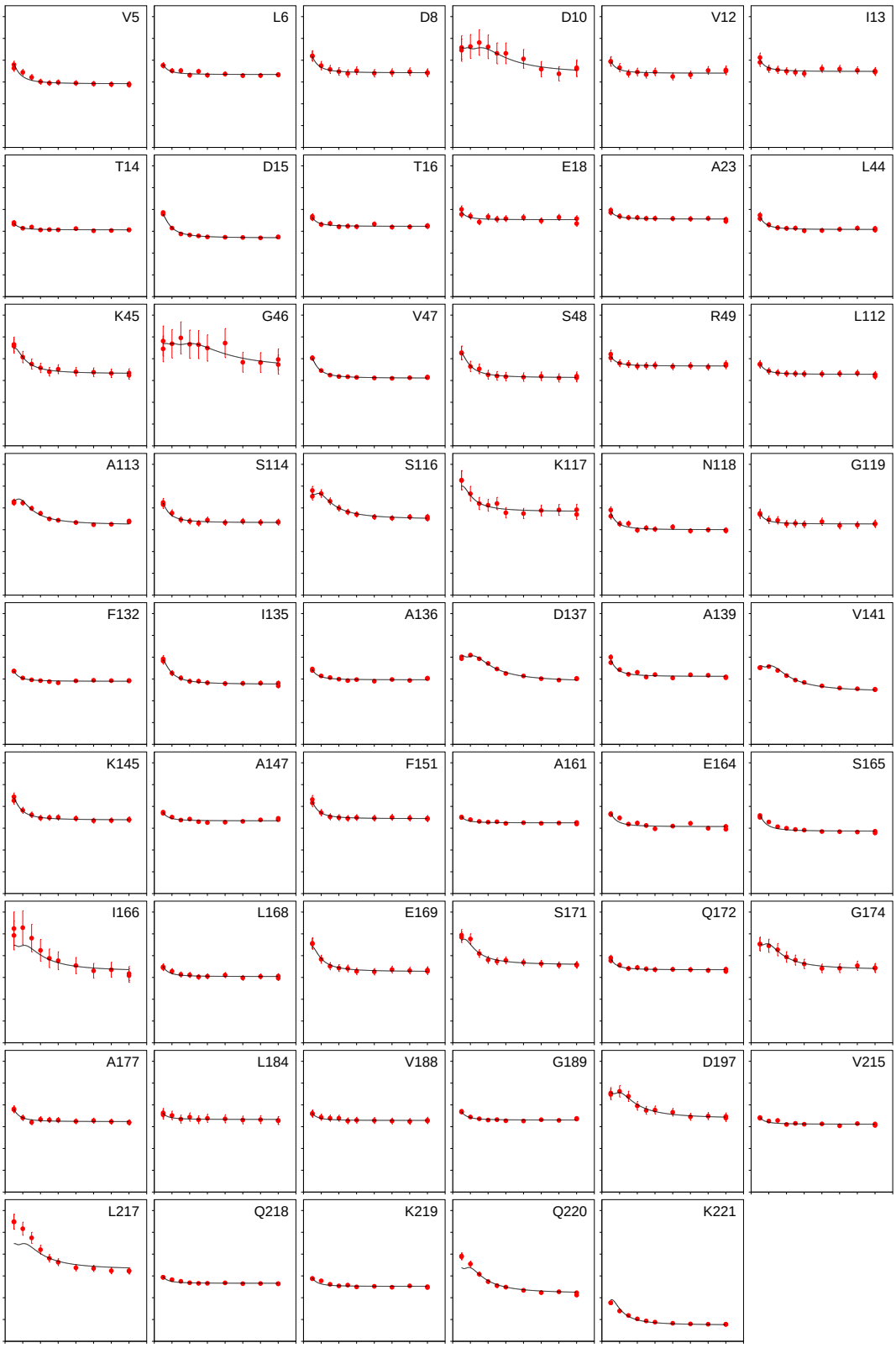

Figure 56: ^{15}N relaxation dispersion profiles recorded at 900 MHz at 298 K. Only residues with a fitted $\Delta\delta$ of more than 0.3 ppm are shown. Fitted curves are the result of minimisation with the exchange rates fixed to those derived from the ^{19}F SEXSY analysis, and despite the very sensitive measurements, satisfy the points well in most cases. There are a few residues which cannot be satisfied with these rates, such as V5, I166, L217 and Q220. Axes and limits as shown in Figure 52.

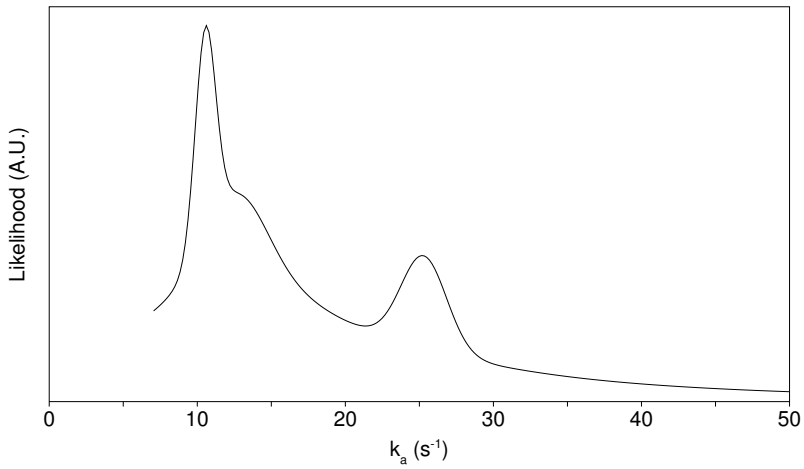

Figure 57: Likelihood of k_a values represented in ^{15}N relaxation dispersion data, obtained by plotting a gaussian curve for every residue, with the centre on the k_a fit value, and with a width defined by the fitting error. k_a values clearly cluster around 2 regions: 10.6 s^{-1} and 25.2 s^{-1}.

2, the data do not justify its inclusion in the fitting procedure. Stated differently, if fixing a parameter at a particular value increases χ^2 by less than 2, the data do not define that parameter differently than the value to which it has been fixed. When the k_a value was fixed during the fitting at 10.6 s^{-1}, the $\Delta\chi^2$ value for most residues was <2, showing them to be consistent with this rate.

On the other hand, I166 and K221 each gave $\Delta\chi^2$ of >9 when the rate is fixed at 10.6 s^{-1}, but $\Delta\chi^2$ of <2 when the rate is fixed as 25.2 s^{-1}, indicating that they belong to the faster regime. D15 also fit better to the fast regime, but less convincingly so, with an increase in χ^2 of >5 when the exchange parameters are fixed to the fast regime, compared to >21 when the parameters are fixed to the slower regime. Its χ^2 values were unusually high, suggesting that the data at different fields do not fully agree with each other. Nonetheless, D15 was treated as a rapidly-exchanging residue in the subsequent multi-field analysis.

When the "gaussian analysis" used to elucidate the k_a estimates was repeated with the two sets of residues, the efficiency of this separation was clear as the peak with a centre at 25.2 s^{-1} results from residues D15, I166 and K221 only. This is shown in Figure 58.

Following the successful separation of residues which belong to the two different exchange regimes, the weighted averages of the individual fitted k_a values were calculated for a more reliable estimate of the global k_a value. The faster exchanging residues give a weighted average $k_a = 25 \pm 2$ s^{-1}. For the slower exchange regime, the weighted average value of k_a is 11.7 ± 0.4 s^{-1}, which is identical to the forward rate derived from the ^{19}F SEXSY analysis, where $k_a = 12 \pm 2$ s^{-1}.

For the two sets of residues, this process was repeated for arriving at a weighted

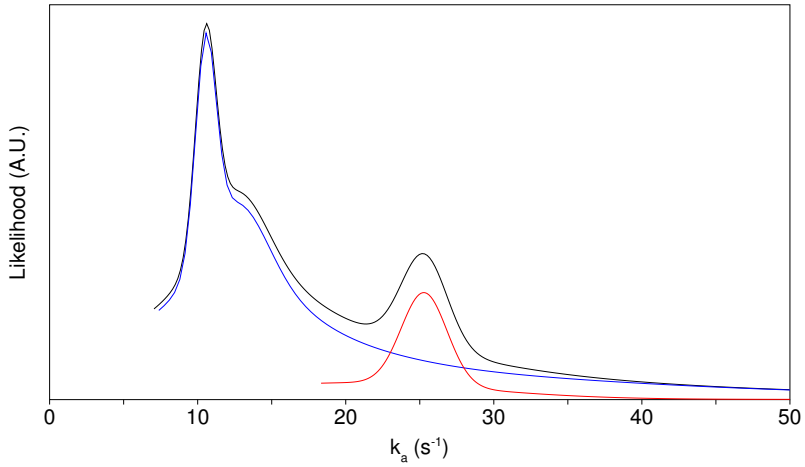

Figure 58: Likelihood of k_a values represented in ^{15}N relaxation dispersion data, split into different exchange regimes. The black line is the sum of the gaussians resulting from all the residues, as in Figure 57, the blue line is the sum of the gaussians resulting from residues which fit to the 10.6 s^{-1} rate, and the red line is the sum of the gaussians resulting from D15, I166, and K221, which fit to the 25.2 s^{-1} rate. The peak at 25.2 s^{-1} clearly results from only these 3 residues.

average p_a value. The "gaussian analysis" of the p_a values resulting from the residues in the slower exchange regime gave a single peak, so no separation based on χ^2 was necessary in this case, as shown in Figure 59. The weighted average p_a value was 0.90 $\pm$ 0.02, again in agreement with the 0.88 $\pm$ 0.01 figure derived from the ^{19}F SEXSY analysis, which confirms that the ^{15}N and ^{19}F exchange data are measuring the same process. The ^{15}N relaxation dispersion data defines the backward rate as 105 $\pm$ 24 s^{-1}, which is also in agreement with the rate of dephosphorylation of the βG16BP intermediate measured by steady-state kinetics[129].

For the faster regime, it is only K221 which gave data sufficiently sensitive to define p_a, with the fitted value being 0.950 $\pm$ 0.005. This defines the sum of exchange rates (k_{ex}) as 490 $\pm$ 60 s^{-1}, which means anything with $\Delta\omega < 500$ Hz is in fast exchange. The minor resonance assignment was done at 800 MHz meaning that any residues with a $\Delta\delta$ of over 6 ppm are in fast exchange. As none of the residues studied up to this point have $\Delta\delta$ values this large, all residues which belong to the fast exchange regime are likely in fast exchange and do not give rise to any minor resonances. This explains why, despite the very favourable relaxation properties of the flexible C-terminal residues, a minor peak for K221 was not found during the assignment of minor conformer resonances by Dr. Joanna Griffin[213].

4.2.2 Fitting of 900 MHz ^{15}N Relaxation Dispersion Data

Following the definition of the rate of exchange and populations for both exchange regimes, these were used to fit the data from the 900 MHz measurements only, as this data set is much more sensitive and complete than those measured at lower fields. To allow for the elimination of residues which did not display significant dispersion, as

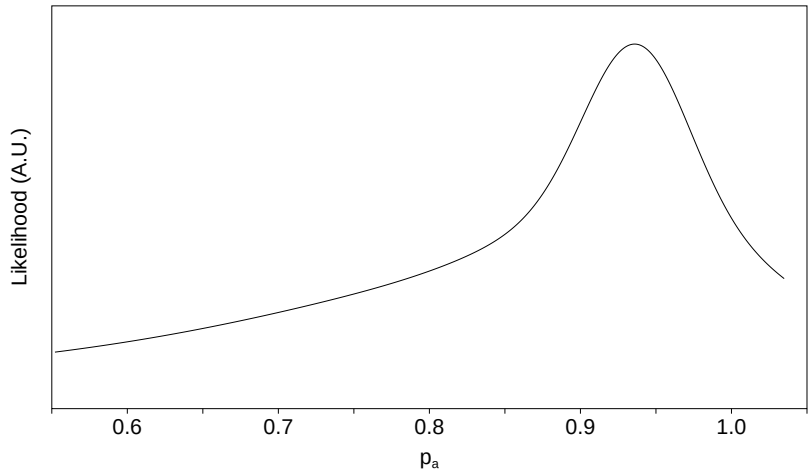

Figure 59: Likelihood of p_a values represented in ^{15}N relaxation dispersion data of slow-regime residues. As there is only 1 peak, all residues can be satisfied by a single global p_a figure.

well as identify to which of the two exchange regimes particular residues belonged, the data were fit under the following conditions, and their χ^2 values analysed:-

- Freely fit: k_a, p_a, nor $\Delta\delta$ constrained
- Slow regime: $k_a = 11.7\ s^{-1}$, $p_a = 0.90$, $\Delta\delta$ not constrained
- Fast regime: $k_a = 25\ s^{-1}$, $p_a = 0.95$, $\Delta\delta$ not constrained
- No exchange: $\Delta\delta = 0.0$ ppm, k_a and p_a not constrained

On the basis of AIC, any residues for which χ^2 did not increase by >2 when $\Delta\delta$ was fixed at 0.0 ppm compared to entirely unconstrained fitting were removed from the analysis, as these residues displayed statistically insignificant dispersion. For differentiating between the two exchange regimes in the remaining residues, because there were only 12 data points per residue in the 900 MHz data a difference in the χ^2 value of >1 between fitting to the two different sets of parameters was treated as significant. For example, if a particular residue gave $\chi^2 = 5.3$ when fixed to the slow-regime parameters and $\chi^2 = 6.4$ when fixed to the fast regime parameters, the residue was concluded to belong to the slow exchange regime, and vice versa. A $\Delta\chi^2$ value of 1 is equivalent to one of the measurements lying 1σ away from the true value.

Given the difficulty in extracting the exchange parameters from the multi-field data, it came as no surprise that for many residues, the $\Delta\chi^2$ between the two exchange regimes was <1, meaning that the residue could not define to which regime it belonged. The exceptions to this — along with χ^2 for each scenario listed above — are listed in Table 8 for residues in the slow regime, and in Table 9 if they define the faster rates of exchange. An example of two residues which are involved in the faster exchange is shown in Figure 60 to illustrate the improvement in satisfaction of the data points by the fitted curves if the fast exchange rates and populations are fixed in the fitting, compared to use of the incorrect, slower exchange rates and population values.

In validation of these results, the AIC values for the residues which can define their exchange regime do not increase when k_a and p_a are fixed to their appropriate values

when compared to entirely free fitting, with the sole exception of D15. The reason for this is unclear, though in contrast to the multi-field fitting, the χ^2 values clearly indicate that this residue is better satisfied by the slow exchange regime parameters than the faster regime. For all the other residues, the very similar χ^2 values (within 1) which result from fitting when exchange parameters are fixed to their appropriate values and completely unconstrained fitting is a clear indication that the exchange parameters calculated using the gaussian plotting and weighted average techniques are accurate.

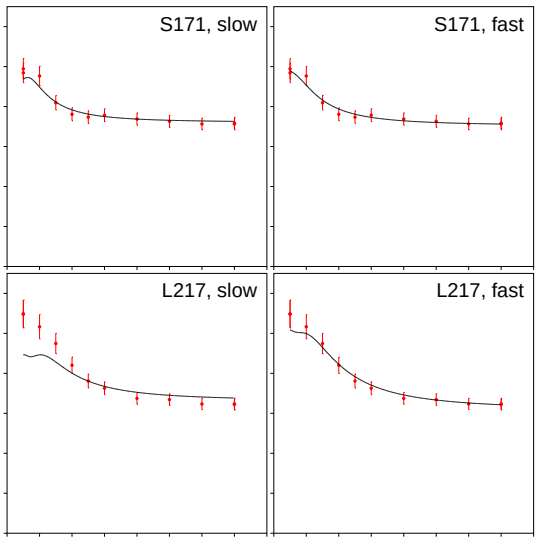

Figure 60: Comparison of fitted curves resulting from slow and fast exchange parameters in ability to satisfy ^{15}N relaxation dispersion profiles from two residues assigned as part of the faster exchange regime. The data from S171 is only slightly more satisfied when fit to the faster exchange regime parameters ($\Delta\chi^2 = 1.7$), whereas the fit of the data from L217 is significantly improved ($\Delta\chi^2 = 28.7$). Axes and limits as shown in Figure 52.

Whilst 49 residues were able to statistically distinguish between the two different exchange regimes, 47 were not. At this point, the minor resonance assignment by Dr. Joanna Griffin was studied. As explained earlier, any of the residues in the faster exchange regime with $\Delta\delta < 6$ ppm will be in fast exchange, and therefore will not have minor resonance peaks. The largest $\Delta\delta$ value for the whole of the minor conformer assignment is less than 3 ppm, which means that any of the residues for which a minor peak is assigned belong to the slower exchange regime. This was the case for 28 residues, meaning that a total of 64 residues are known to be part of the dynamics felt by the ^{19}F nuclei in the active site.

13 residues showed improvements in χ^2 on being fit with the faster exchange parameters. 19 residues were left unassigned to either exchange regime, but with the largest $\Delta\delta$ value being 0.5 ± 0.1 ppm, it can safely be assumed that these residues do not play a significant role in either regime. It is entirely possible that some of these residues feel contributions from both regimes which has prevented full minimisation to a two-state model, and indeed some of the uncertain residues are sandwiched between

	χ^2 Value			
Residue	**Free fitting**	**Slow regime**	**Fast regime**	**No exchange**
D10	1.07	1.40	6.92	1008.41
G11	6.98	7.15	8.34	9.76
I13	2.50	2.45	3.73	16.21
D15	2.68	10.16	51.23	791.67
T16	2.37	2.38	3.74	28.28
Y19	1.78	1.87	3.16	14.63
D37	20.32	20.39	22.44	33.74
L44	2.50	2.53	4.18	50.37
G46	1.17	1.24	4.63	685.21
V47	0.85	1.25	12.44	201.80
E71	7.57	7.49	8.84	8.88
N77	7.54	7.55	8.96	9.01
S88	3.68	3.61	4.86	4.89
D91	11.92	11.51	12.87	16.05
V92	4.80	4.80	6.52	6.59
K109	5.00	4.97	6.07	6.11
A113	5.56	6.94	17.87	320.87
S114	1.51	1.59	3.45	54.07
S116	1.79	2.23	6.83	176.70
L123	2.29	2.45	4.13	21.69
F132	4.60	4.72	10.97	116.97
I135	1.48	1.70	4.87	134.69
A136	11.92	12.43	17.53	198.72
D137	5.72	7.85	117.40	1225.77
A139	8.84	9.09	11.70	123.56
V141	1.21	3.67	126.12	1674.92
K145	1.31	1.33	2.84	74.70
A147	20.59	21.13	24.46	115.96
F151	0.69	0.81	3.49	34.15
E169	0.68	2.04	4.39	62.55
G174	1.47	0.77	2.83	3050.58
A177	1.81	1.86	3.77	25.82
G189	2.59	2.57	4.45	38.45
D193	2.01	1.94	3.15	3.19
D196	1.42	3.56	4.57	4.63
D197	2.73	1.19	6.39	96.61

Table 8: χ^2 values from ^{15}N relaxation dispersion fitting for residues which define the slow exchange regime, fit under different conditions as listed on Page 93.

	χ^2 Value			
Residue	**Free fitting**	**Slow regime**	**Fast regime**	**No exchange**
V5	1.11	4.26	1.28	79.35
A143	2.44	4.36	2.71	23.38
A161	2.71	6.44	2.80	72.92
E164	9.16	14.04	9.56	89.27
S165	4.10	28.79	4.72	441.45
I166	11.77	4.98	0.55	909.24
S171	2.87	4.01	2.31	89.79
V215	11.15	5.97	3.89	36.60
L217	141.57	32.82	4.16	173.14
Q218	143.70	14.24	2.91	212.16
K219	10.71	30.49	11.23	304.04
Q220	332.26	35.10	3.01	481.28
K221	56096.60	1036.00	106.45	91689.40

Table 9: χ^2 values from ^{15}N relaxation dispersion fitting for residues which define the fast exchange regime, fit under different conditions as listed on Page 93.

residues belonging to both slow and fast exchange regimes.

4.2.3 Validation of Fitting Results by use of CATIA

Although all data processing and statistical methods used in order to elucidate the exchange parameters from the ^{15}N relaxation data were entirely reasonable, the fitting results were validated by the use of Dr. Flemming Hansen's fitting program CATIA (CPMG, Anti-Trosy, and Trosy Intelligent Analysis). To differentiate the two analyses, data at all 3 field strengths from all the residues which had been previously identified as belonging to the "slow" (major⇌minor) regime were fit to global k_{ex} and global p_b values, as opposed to individually fitting k_a and p_a values.

D15 had to be excluded in order to properly satisfy the data from all the other residues. Although this suggests that D15 is unrelated to the major⇌minor dynamics, it is not clear that this is not simply a product of a bug in the fitting program as D15 yields faster exchange rates when fit individually yet when all the data are fit together, its inclusion causes the global exchange rates to decrease. After the exclusion of D15 and the adjustment of the uncertainties on the $R_{2(eff)}$ values such that the $\Sigma\chi^2$ value matched that expected, the global parameters agreed with those which had been derived by the statistical methods detailed earlier: $k_{ex} = 93 \pm 19\ s^{-1}$, and $p_b = 13 \pm 2$ %. This further proves that the data processing used to tease out the exchange parameters was reasonable and reliable.

4.2.4 Location of Exchange Regimes

In addition to many residues being statistically able to distinguish between the regimes, plotting the residues involved in both regimes on the crystal structure of the βPGM-MgF_3^--βG6P TSA complex shows them to also be structurally distinct. In Figure 61, those residues involved in the slower exchange process are shown in red, the residues involved in the fast exchange are shown in blue, and the intensity of the colours reflects the magnitude of the fitted $\Delta\delta$ value. The slower dynamics are much more widespread, with the largest changes in chemical shift clustering around the active site, whereas the faster dynamics are localised to the C-terminus and particularly residues which are unstructured. The locations of these dynamics support the interpretation that the slower exchange regime seen also by ^{19}F NMR may be catalytically relevant, whilst the involvement of unstructured residues in the faster dynamics leads to the suggestion that these dynamics may represent a transient local unfolding event. This hypothesis is tested in Section 5.

There are two residues which χ^2 refines as belonging to the fast regime which are surrounded by slower residues when plotted on the crystal structure, implying that they in fact belong to the slower exchange regime: A143 and S171. As described earlier, a $\Delta\chi^2$ value of 1 is equivalent to one of the data points being 1σ away from its true value, which is in fact a reasonably likely scenario. It is notable that the dispersion curves fitted to the S171 data from both regimes shown in Figure 60 do not differ by nearly as

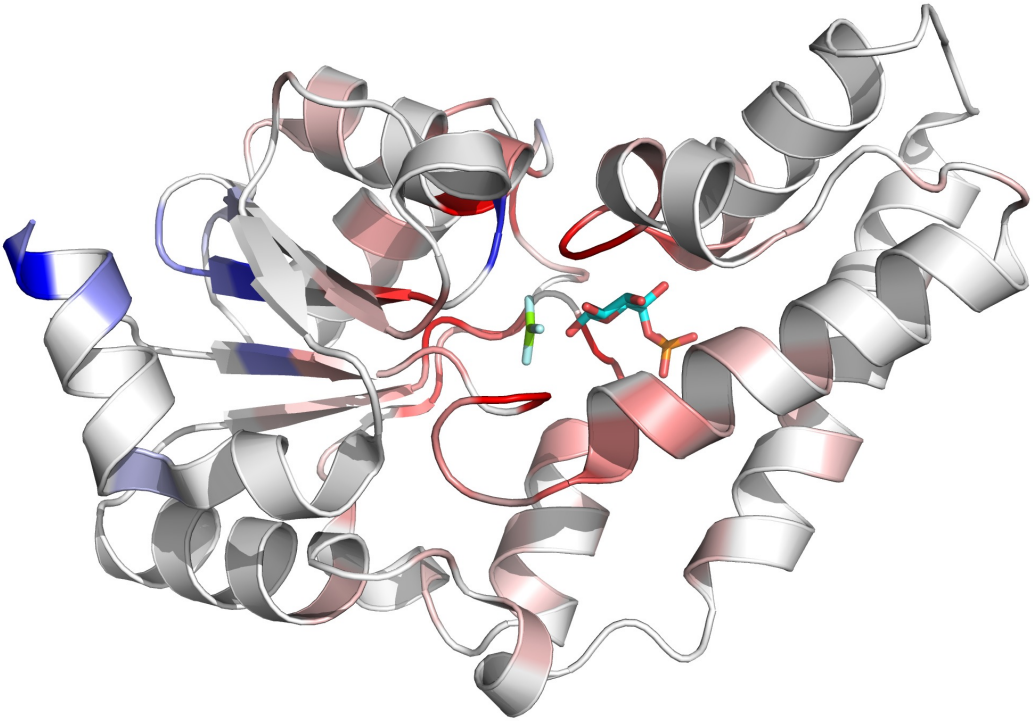

Figure 61: $\Delta\delta$ values from both exchange regimes plotted on βPGM-MgF_3^--βG6P TSA complex crystal structure. βPGM-MgF_3^--βG6P TSA complex co-ordinates represented as a cartoon of the backbone atoms and the MgF_3^- moiety and βG6P molecule are shown as sticks. All residues involved in the slower (major$\rightleftharpoons$minor) exchange regime are shown in red, whilst the residues involved in the faster regime are shown in blue. The colour intensity indicates the magnitude of $\Delta\delta$ for each residue. The structural separation of residues involved in different exchange processes shows that these processes are distinct.

much as 1σ for any of the points, suggesting perhaps the changes are not significant. However, as shown in Figure 61, the separation of the two exchange regimes based on χ^2 has succeeded in separating out dynamic processes occurring in different regions of the protein in most cases, and as so few data points define the exchange parameters in many residues, it is impossible to scrutinise every residue in this way. Instead it must simply be acknowledged that the system based on $\Delta\chi^2$ is not entirely perfect, and there may be a handful of residues which have been mis-assigned. The χ^2 analysis is the only objective method to separate the residues.

The residues which cannot be assigned to either regime based on χ^2 are not shown in Figure 61. In some cases these residues are surrounded only by residues in one regime or the other and so could be tentatively assigned (e.g., V12), but in many cases these residues occupy the space between residues involved in different exchange regimes, supporting the idea that many of these residues may feel contributions from both dynamic processes. As already stated, any further analysis than the $\Delta\chi^2$ analysis will be inherently subjective and has been avoided.

4.2.5 Faster Dynamics Also Occur in Apo βPGM

To investigate whether or not the faster dynamics (hypothesised as a local unfolding event) are a property unique to the βPGM-MgF_3^--βG6P TSA complex, preliminary ^{15}N relaxation dispersion data were collected on Apo βPGM at 800 MHz. Due to acquisition with an RT probe, the quality of the data is very poor (as with the 800 MHz data on the βPGM-MgF_3^--βG6P TSA complex if judged alone). Furthermore, due to exchange broadening and signal overlap only five residues which had been identified from the statistical analysis as belonging to the "fast" exchange regime could be studied in apo: D5, A161, S165, I166, and K221.

The signal:noise of these data is far too poor to define the exchange rates, except for the C-terminal residue K221. Although the errors are extremely large, the resultant values are strikingly similar to those measured from the βPGM-MgF_3^--βG6P TSA complex data, with $k_a = 25\ s^{-1}$, and $k_b = 450\ s^{-1}$. Data from the five residues were then fitted whilst fixing the exchange rates at the values derived from the βPGM-MgF_3^--βG6P TSA complex data ($k_a = 25\ s^{-1}$ and $k_b = 492\ s^{-1}$) and the resulting χ^2 values compared. Even for the relatively high sensitivity data from K221, this resulted in no increases in χ^2 of more than 0.5. The ^{15}N $\Delta\delta$ values were then compared (see Table 10) and although they do not perfectly match, they are similar enough to demonstrate that a very similar — if not identical — process at the C-terminus as measured in the βPGM-MgF_3^--βG6P TSA complex also takes place in apo βPGM. This shows that the C-terminus dynamics are not related to the TSA complex, and are not strongly affected by the ligation state of βPGM.

Residue	**Apo βPGM ^{15}N $\Delta\delta$**	**βPGM-MgF_3^--βG6P ^{15}N $\Delta\delta$**
V5	0.7 ± 0.2	0.77 ± 0.07
A161	0.6 ± 0.4	0.40 ± 0.03
S165	0.4 ± 0.1	0.70 ± 0.02
I166	2 ± 1	2.2 ± 0.5
K221	0.82 ± 0.03	0.876 ± 0.003

Table 10: ^{15}N $\Delta\delta$s from apo ^{15}N relaxation dispersion data, compared to those derived from data on the βPGM-MgF_3^--βG6P TSA complex. It can be seen that although two values are outside of error, the overall agreement between the two sets of values is good.

4.3 Analysis of ^{15}N $\Delta\delta$ Values

4.3.1 Refinement of Minor Form Assignment

It has been shown that the ^{15}N relaxation dispersion data and the ^{19}F SEXSY data measure coincident rates, and therefore are measuring the same exchange process. The $\Delta\delta$ figures derived from the dispersion data can be compared to the $\Delta\delta$ figures measured from the minor resonance assignment by Dr. Joanna Griffin to see if the minor peaks are indeed the result of this same dynamic process. Initial correlations showed that many of the residues did give $\Delta\delta$ values from both techniques that agreed with each other, but there were many cases where the $\Delta\delta$ measured from the spectra

was extremely small, but the value fit from the dispersion data was large.

To understand the cause of this disparity, the assignment data was revisited. To take advantage of the larger population of minor conformer at lower temperature, the minor resonance assignment by Dr. Joanna Griffin was done at 278 K. A series of spectra recorded at 5 K temperature increments were then used to follow the minor peak to its final position at 298 K. As is clearly shown in Figure 51, because the exchange rate is so temperature-dependent, when $\Delta\delta$ between major and minor resonance is relatively small the resonance is in slow exchange at 278 K but fast exchange at 298 K. It was a common scenario that at 298 K, the minor peak was in fast exchange with respect to ^{15}N frequencies, but still slow exchange for ^{1}H. In these cases, the minor peaks were picked as they appeared in the spectrum at 298 K, giving the correct minor ^{1}H δ, but a weighted average value for the ^{15}N δ. In these cases, the ^{15}N $\Delta\delta$ figures between the major and minor resonance is anomalously low compared to the $\Delta\delta$ value fit from the relaxation dispersion data.

To overcome this issue, the ^{15}N chemical shifts were extrapolated from the spectra recorded at 278 K and 283 K (and, where possible, 288 K) back up to 298 K, giving the position where the peak would appear if the system was still in slow exchange. During this process, the entire minor conformer resonance assignment was reviewed, linking major and minor resonances using temperature coefficients of ^{1}H–^{15}N TROSY peaks (changes in $^{1}H^{N}$ and ^{15}N δs with temperature) and/or $^{13}C^{\alpha}$ δs from HNCA spectra recorded at both 278 K and 298 K. $^{13}C^{\alpha}$ δs are predominantly affected by backbone dihedral angles[219, 220], whereas $^{1}H^{N}$ and ^{15}N δs are also sensitive to hydrogen bond lengths[221, 222], which are the primary determinant of the temperature coefficients in proteins[223].

These analyses are complementary as the changes in ^{15}N δs between the major and minor conformer are likely to reflect either changes in backbone dihedral angles (preserving hydrogen bonding and therefore temperature coefficients), changes in hydrogen bonding (preserving dihedral angles and therefore $^{13}C^{\alpha}$ δs), or both. In the cases where both change significantly minor peaks may have been missed, though there were many residues for which the changes were relatively small and transfer of assignment was possible on the basis of both the $^{13}C^{\alpha}$ δs and the temperature-dependence of $^{1}H^{N}$ and ^{15}N δs.

It was possible to validate the assignment of most of the residues by this method. Assignments for a few resonances with small $\Delta\delta$s were deleted as they were not sufficiently convincing on review, particularly as the small value of $\Delta\delta$ implies insufficiently large changes in backbone dihedral angles and/or hydrogen bonding for the method of comparison to become flawed. Some new assignments were made based on the predicted ^{15}N $\Delta\delta$ values from the relaxation dispersion data. For example, a minor peak with corresponding temperature coefficients to the major G11 resonance was found 1.8 ppm away in ^{15}N, matching the fitted ^{15}N $\Delta\delta$ value from the dispersion data of 2.0 $\pm$

0.3 ppm. The errors in $\Delta\delta$ values derived from the spectra were taken as the average ^{15}N linewidth, 0.15 ppm. Following this review of the assignment, 106 (50%) of the 211 non-proline residues had a minor resonance assigned.

4.3.2 Magnitude and Location of ^{15}N $\Delta\delta$s

The chemical shift changes measured from the assignment match those fit from the dispersion data extremely well, as shown in Figure 62, confirming that the minor peaks seen in the ^{1}H–^{15}N TROSY spectra are indeed the result of the major⇌minor exchange seen also by ^{19}F NMR. Note that only the magnitude, not the sign, of the $\Delta\delta$ values is known from the ^{15}N relaxation dispersion data, although there are methods by which the sign of the change can be elucidated[224]. The statistics for this correlation are very convincing: the correlation coefficient is 0.974, and when fit to a $y = mx$ equation the resulting gradient (m) of these points is 1.02 ± 0.04. The only major outlier on Figure 62 is a case where the measured $\Delta\delta$ from the spectra is very low, i.e., within the linewidth of the major peak, and it is likely that this peak has been mis-assigned. Peak picking in spectra is inherently more subjective than χ^2 minimisation of relaxation dispersion data, and the problems with changes in exchange regime have been discussed already.

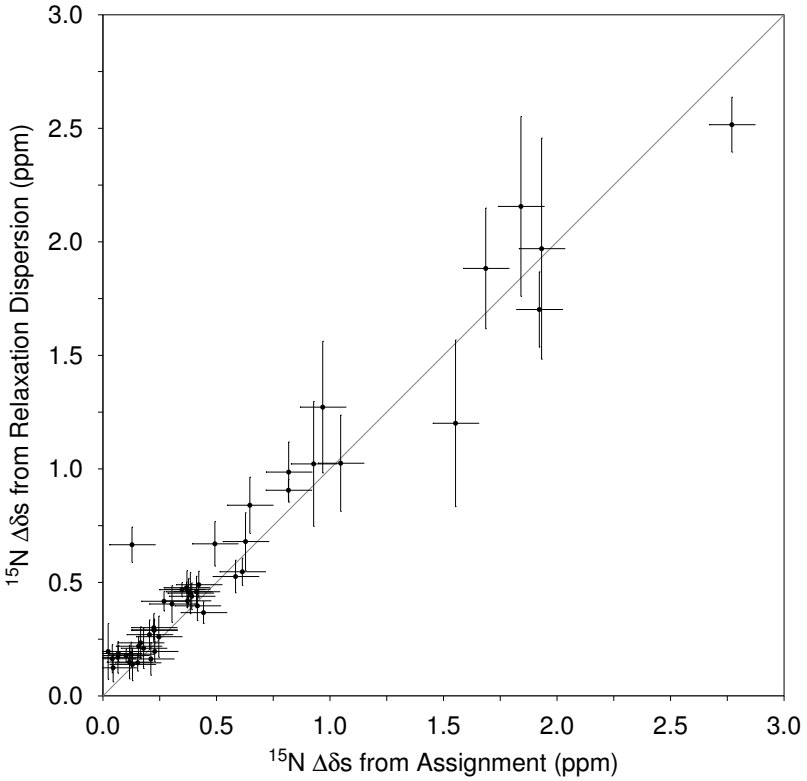

Figure 62: Comparison of $\Delta\delta$ values measured from assignment and relaxation dispersion, showing excellent agreement between values. The grey line has a gradient of 1, with these data fitting to a gradient of 1.02 ± 0.04, and having r = 0.974.

The location of the chemical shift changes with respect to the primary structure is shown in Figure 63, giving an impression of the location of the major structural

changes. It shows that most of the changes occur in the loops between secondary structure elements. The comparison of ^{15}N $\Delta\delta$s derived from relaxation dispersion and assignment again show good agreement, but it is immediately apparent that many of the larger changes have been missed by the assignment. This is probably due to changes in hydrogen bonding and/or backbone dihedral angles preventing transfer of assignment from the major resonances as discussed earlier. Also, in the case of D10 ($\Delta\delta = 3.6 \pm 1.4$ ppm), the ^{1}H–^{15}N TROSY peak is split due to ^{19}F-coupling through its hydrogen bond to F_a[138], making the peak appear very weak in the spectrum, preventing the assignment of a minor due to inadequate signal:noise. It is testament to the quality of the ^{15}N relaxation dispersion data that exchange parameters could be fit to this residue. Note that on the dispersion chart, only the residues which have been proven to be part of the slower (major⇌minor) exchange regime are plotted.

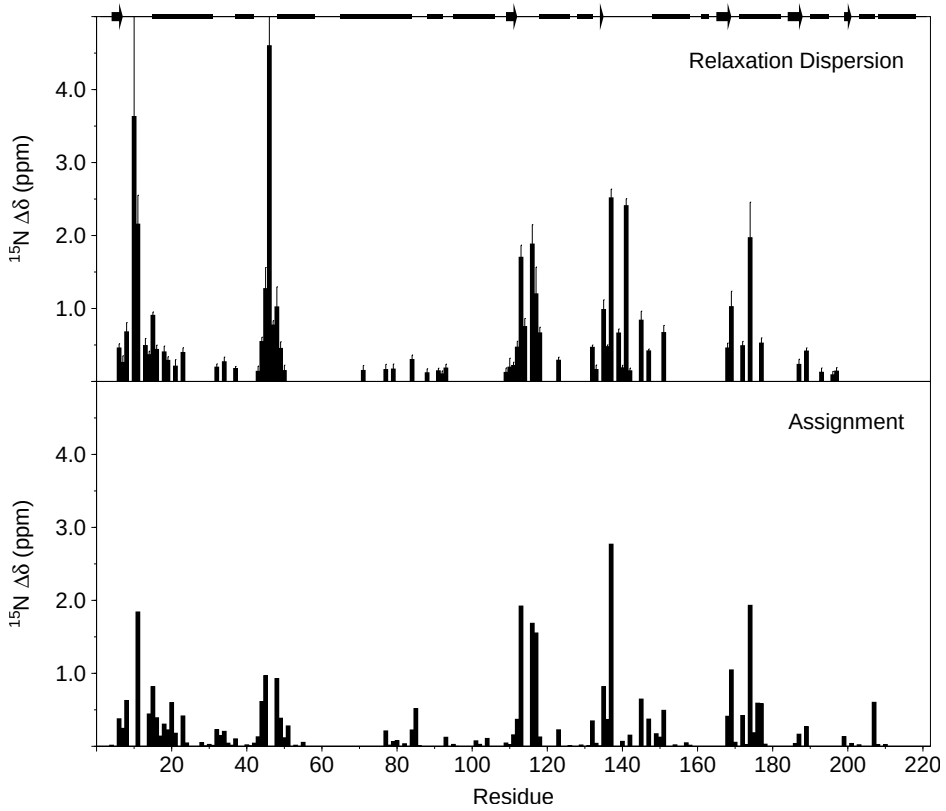

Figure 63: $\Delta\delta$ values measured from relaxation dispersion (top) and assignment (bottom) plotted against the primary sequence, showing the location of the largest changes. The topology diagram at the top indicates where the protein is in α-helical (bar) or β-sheet (arrow) configuration. It can be seen that most of the changes are found in the loops between secondary structure elements.

Although the quality of these data is extremely high given the difficulty in acquiring it, as has been discussed, a set of ^{15}N $\Delta\delta$s alone do not allow for elucidation of structure. These data have served to prove beyond reasonable doubt that the slower of the two exchange regimes seen by ^{15}N relaxation dispersion is also that affecting the ^{19}F resonances as discussed in Section 3 and also that which gives rise to some minor peaks in the ^{1}H–^{15}N TROSY spectrum of this complex, as initially assigned by Dr.

Joanna Griffin. Due to the reliance on other data to gain insight as to any structural details of the minor conformer, it is further discussed in Section 7.

5 Characterisation of βPGM Mutants

5.1 Introduction

Thermodynamic analysis of ^{19}F NMR data on the βPGM-MgF_3^--βG6P TSA complex in Section 3 determined that the equilibrium constant between the major and minor conformer of the complex is 0.16 ± 0.01 at 298 K. Based on this value of K_{eq}, use of the equation shown in Figure 64 gives the difference in free energy between the conformers as 4.5 ± 0.2 kJ/mol. This energy difference is very small when compared to the enthalpy of a single hydrogen bond, which is estimated to be between the values of 12–38 kJ/mol[225]. Given the high activation energy for the transition between conformers (see Section 3) the intramolecular contacts likely differ significantly between them, i.e., there will likely be some hydrogen bonds present in one conformer that are absent in the other. In principle it is possible to selectively remove hydrogen bonding partners which only exist in the major conformer of the βPGM-MgF_3^--βG6P TSA complex by site-directed mutagenesis (SDM). The result of this would be that the enthalpy of the major conformer only would decrease, preferentially stabilising the minor conformer. Because the ΔG between the states is relatively low, the removal of such a hydrogen bond would cause the population of the minor conformer to increase to between 96–100%.

$$\Delta G = -RT \ln K_{eq}$$

where:
ΔG is the difference in free energy
K_{eq} is the equilibrium constant
T is the temperature (K)
R is the gas constant

Figure 64: Equation relating difference in free energy to equilibrium position.

In reality, the relationship between all the non-covalent interactions in a folded protein is much more interdependent than this, with many residues engaged in several hydrogen bonds, meaning that even if a hydrogen bond which was absent in the minor conformer could be found and removed, the shift in the equilibrium constant would likely be much smaller due to perturbation of interactions elsewhere in the structure. Additionally, although there has been progress in the incorporation of non-natural amino acids in proteins[226], with the ability to swap between only 20 amino acids, site-directed mutagenesis can rarely be as delicate as removing a single hydrogen bond from a position in the protein.

Nonetheless, there are some successful examples of similar strategies in the literature. In the work by Dr. Dorothee Kern's lab at Brandeis University. Mutating a serine residue to a threonine, i.e., addition of a methyl group, in the Proline Isomerase

Cyclophillin A (CypA) succeeded in stabilising a previously hidden minor conformation through a steric clash[227]. In another example, a transient non-native hydrogen bond in Nitrogen Regulatory Protein C (NtrC) was removed by mutation, having the effect of destabilising the transition state for — and thereby increasing the activation energy of — a conformational exchange process. This was reflected in the decrease in the rate of exchange between the conformers from 14,000 s^{-1} to 3,000 s^{-1}, whilst the populations of the two interconverting species was unaffected[93].

Similar strategies could potentially produce a mutant of βPGM which primarily occupied the "minor" state, which could then be studied by traditional structural biology methods, allowing critical assessment of its potential role in catalysis. However, knowledge of where the main changes in the protein conformation are (as provided by the minor form assignment and ^{15}N relaxation dispersion in Section 4) do not immediately assist in planning such mutations, as the nature of the changes at these sites during the conformational exchange is unknown. Any mutations at these positions therefore have equal chance of destabilising the major and minor conformations. It is instead necessary to have a model for what the exchange may involve.

5.2 Choice of Mutations

A potential model was provided by the study of near-attack conformations (NACs) which are occupied by the βPGM-BeF_3^--βG6P GSA complex[228]. Two distinct structures of this complex were solved by X-ray crystallography: these conformations are referred to as "NAC I" and "NAC II", and the crystal structures have PDB codes 2WF9 and 2WF8, respectively. The structures differ with respect to the orientations of several of the active site residues, particularly the catalytic base D10 and the phosphosugar. NAC II is a similar conformation to that adopted by the βPGM-MgF_3^--βG6P TSA complex, suggesting that the NAC I structure may be similar in conformation to the minor conformer of the βPGM-MgF_3^--βG6P TSA complex. This suggestion is compelling as D10 and G46 (which co-ordinates the 2′ sugar hydroxyl) are the residues which display the largest ^{15}N $\Delta\delta$s during the major⇌minor exchange as measured by ^{15}N relaxation dispersion and discussed in Section 4. The differences between the structure of the two NACs may therefore provide opportunities to create mutations which will shift the equilibrium between the two conformers of the βPGM-MgF_3^--βG6P TSA complex. This is be easily assessed by ^{19}F NMR spectra of the resulting mutant βPGM TSA complexes.

The primary difference between the two conformations is that NAC I is a partially open structure (approx. 75 % closed as determined by D15 δs), whereas NAC II is fully closed, strongly resembling the conformation of the βPGM-MgF_3^--βG6P TSA complex. In the NAC I structure, as well as in the fully open structures of either the βPGM-BeF_3^- complex or apo βPGM, D10 is pointing towards the hinge region between cap and core domains, and away from the active site. In the NAC I structure, the side-chain of D10 is hydrogen-bonded to the backbone amide of T16, the 2′ OH of

βG6P, and a water molecule. In the NAC II crystal structure, as well as that of the βPGM-MgF_3^--βG6P TSA complex, the side-chain of D10 is hydrogen-bonded to the 1′ OH of βG6P, i.e., the nucleophile, and to a water molecule that is also co-ordinated by the side-chains of residues H20, Y80, and S116. It may also form a hydrogen bond directly to the Nϵ of H20, though the distance is a little long (3.1 Å), and the geometry non-ideal. Additionally in these structures the side-chain of D15 no longer N-caps the helix that runs from residues D15–H20. Instead it points out into the solvent with no hydrogen bonding to any other residues. Diagrams showing differences in interactions between residues in both NAC structures are shown in Figure 65.

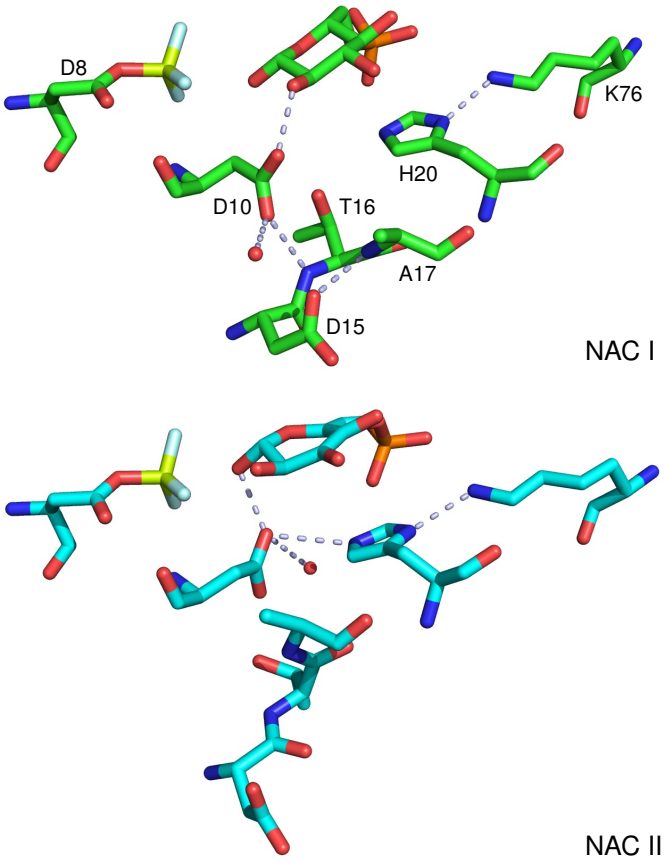

Figure 65: Diagrams to show the differences in interactions in NACs I & II of the βPGM-BeF_3^--βG6P GSA complex. For clarity, only residues D8, D10, D15, T16, A17, H20, and K76 are shown, as well as any hydrogen bonding between them. Residue labels are shown on the NAC I diagram only. The figures are aligned on D8 and the BeF_3^-moiety, with C atoms shown in green (NAC I) and cyan (NAC II), O atoms shown in red, N atoms shown in blue, P atoms shown in orange, Be atoms shown in yellow, and F atoms shown in light blue. Water molecules are shown as red spheres. The dotted light blue lines represent hydrogen bonds. The differences in interactions formed by these residues between the 2 NACs are shown in Table 11.

The large $\Delta\delta$ value for D10 between major and minor conformers of the βPGM-MgF_3^--βG6P TSA complex measured from ^{15}N relaxation dispersion fits with the significantly different orientations in which it is found in various crystal structures of βPGM. Together these observations suggest that in the minor conformer, D10 may

be in its "out" position, pointing towards the hinge and away from the active site. To attempt to perturb this equilibrium, differences in the interactions which it makes with other residues in the two conformations may be exploited, as well as differences in the interactions made by the important hinge residue, D15. These are summarised in Table 11.

Residue	**Interactions in NAC I**	**Interactions in NAC II**
D10 Sidechain COO^-	βG6P 2′ OH Water T16 Backbone NH	βG6P 1′ OH Water (H20 Sidechain Nϵ)
D15 Sidechain COO^-	A17 Backbone NH	None

Table 11: Interactions made by the sidechains of D10 and D15 in NACs I & II. The interaction between D10 and H20 in NAC II is in brackets as the proposed hydrogen bond between them is relatively long and of imperfect geometry. This table demonstrates why D15, T16, and H20 are proposed sites for mutation, due to variation in interactions in the two NAC structures.

Given that the side-chain of D15 seems to play no stabilising role in NAC II (D10 "in"), mutagenesis to an alanine should stabilise this conformer of the βPGM-BeF_3^--βG6P GSA complex. This may also be possible by mutating T16 to a proline which lacks an amide proton and so cannot be a hydrogen bond donor to D10 in its "out" conformation. A crystal structure of apo T16P βPGM shows D10 pointing into the active site, and the enzyme shows greatly accelerated unproductive hydrolysis of the phosphate from D8[229], suggesting that this strategy may work.

Both of these mutations are expected to preferentially stabilise the major conformer of the βPGM-MgF_3^--βG6P TSA complex, which though potentially useful, is not the intention of this mutagenesis. Mutations expected to destabilise the major conformer are sought instead. As the sugar nucleophile is clearly necessary for MgF_3^- TSA complex formation and removing a single water molecule is a difficult task, the only option was mutants of H20, despite the nature of any bond between H20 and D10 being unclear from the crystal structures.

As the removal of a single hydrogen bond is desired, it is necessary to study any other interactions formed by H20. The imidazole side chain also forms a hydrogen bond with K76 via its N^δ atom. The proposed hydrogen bonds from H20 to both K76 and D10 could likely be preserved by the mutation to a glutamine. In this scenario, the N atom of the guanidino group could act as the hydrogen donor to the D10 side chain (mimicking the N^ϵ atom of H20), and the O atom of the guanidino group could accept the hydrogen bond from the side chain of K76 (thereby mimicking the N^δ atom of the H20).

Mutation of H20 to an asparagine (with a side chain which is one carbon atom shorter than glutamine) would probably mean the interaction with D10 would be lost, but the hydrogen bond to K76 would likely persist. Due to the uncertainty in bonding between the residues, it was decided that a H20A mutant would also be produced to

eliminate any chance of an interaction between this residue and D10. The proposed bonding arrangements of WT βPGM, as well as the result if H20 was mutated to a glutamine, asparagine, and alanine are shown in Figure 66. This reasoning is backed up by kinetic study of various mutants of H20, which show that mutation to a glutamine results in a mere 8-fold reduction in k_{cat}. The reductions get larger with the shorter side chain substitutions, with H20N and H20A slowing the rate of catalysis by 300-fold and 7000-fold respectively[229]. All proposed mutation strategies are summarised in Table 12.

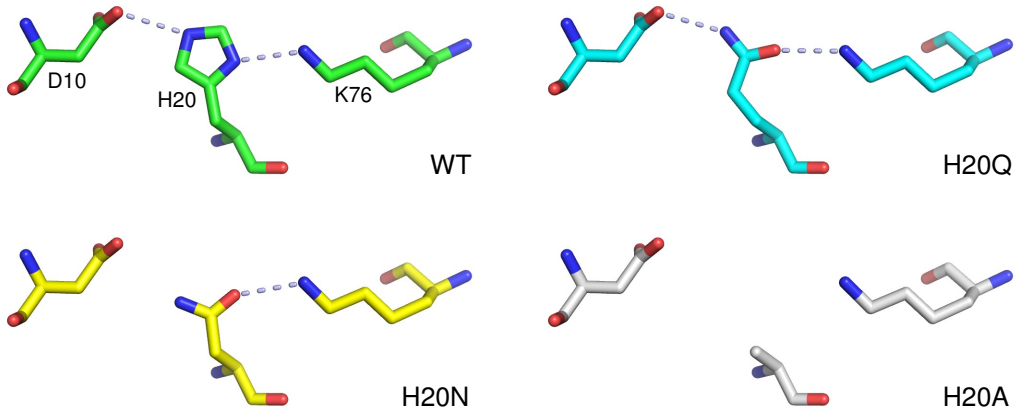

Figure 66: Diagrams to show how interactions between H20, D10, and L76 may be selectively disrupted through mutation of H20 to different residues. For clarity, only residues D10, H20, and K76 are shown, and residue labels are shown on the WT diagram only. C atoms are shown in green, cyan, yellow and white for WT, H20Q, H20N, and H20A respectively. N atoms are blue, O atoms are red, and the dotted light blue lines represent hydrogen bonds. H20N and H20A mutations are expected to remove a hydrogen bond from H20 to D10 in its "in" position, and therefore stabilise the minor conformer, proposed D10 "out". Note that only the H20 side chain was changed to make these diagrams, the side chains of D10 and K76 were not repositioned. The coordinates used are from the WT βPGM-MgF_3^--βG6P TSA complex crystal structure.

Mutant	**Interactions lost from NAC I**	**Interactions lost from NAC II**	**Preferred conformer**
D15A	D15 SC – A17 MC	None	NAC II (Major)
T16P	T16 MC – D10 SC T16 SC – Water	T16 MC – T14 SC	NAC II (Major)
H20N	None	H20 SC – D10 SC	NAC I (Minor)
H20A	H20 SC – K76 SC	H20 SC – K76 SC H20 SC – D10 SC	NAC I (Minor)

Table 12: All interactions in the two NACs of the βPGM-BeF_3^--βG6P GSA complex which will be perturbed by the proposed mutations to βPGM. "MC" and "SC" are shorthand for "Main Chain" (backbone) and "Side Chain", respectively. The last column shows which NAC of the βPGM-BeF_3^--βG6P GSA complex— and therefore which conformer of the βPGM-MgF_3^--βG6P TSA complex— is expected to be preferentially stabilised as a result of the mutation. In all cases it can be seen that this corresponds to the conformer which is expected to lose the least interactions, and thereby lose the least enthalpy.

In addition to the proposed mutations based on the NACs formed by the βPGM-

BeF_3^--βG6P GSA complex, it was decided to test the hypothesis that the faster motion involving the C-terminal residues (uncovered by the ^{15}N relaxation dispersion study in Section 4) is a partial unfolding event as predicted. As the last three residues are not particularly structured (as evidenced from crystallographic B factors and order parameters fit in Section 6), there are only two obvious ways of significantly perturbing them through changes in protein conformation. Either they are structured in the minor conformer of the faster exchange process, or more preceding residues become unstructured. Given that other residues show larger ^{15}N $\Delta\delta$ values for this exchange process, the latter possibility seems more likely. In either case, removal of the last three residues is expected to perturb this equilibrium. If this conformational exchange process is a true local unfolding event, a truncation mutant would show lower folding stability than WT βPGM. In addition to this, similarity between βPGM-MgF_3^--βG6P TSA complexes formed by WT and K219Δ βPGM would prove that the faster dynamics are truly independent of the more functionally-relevant slower dynamic exchange between major and minor conformers of the βPGM-MgF_3^--βG6P TSA complex as discussed in Section 4.

5.3 Production of βPGM Mutants

5.3.1 Primer Design

Designing primers to mutagenize βPGM proved to be a difficult task, as the entire gene has an average percentage of G–C base pairs of just 38%, below the required minimum (40%). As a result of this, the melting temperatures of the primers were initially too low for use in the Polymerase Chain Reaction (PCR). To compensate for this, relatively long primers were made, but this had the undesired effect of increasing the propensity for secondary structure formation. The ability to self-anneal was, where possible, tuned by trimming the ends of the primers.

Because of these difficulties, each mutation was done with as few base substitutions as possible, meaning that some of the codons used for the mutated residues are relatively rare. Rare codons have the potential to lead to poor yield of the protein from bacterial expression as the low concentrations of the appropriate tRNA molecules can become rate-limiting. In the case of the K219Δ mutation, the primers had lots of potential for secondary structure formation, and so it was decided to do this mutation in 2 steps, first performing a silent K219K mutation, which facilitated the subsequent K219Δ mutation. Table 13 shows the preferred codon (i.e., most common in *E. coli*) for each of the mutations and the codon which was actually used in a bid to facilitate the SDM. Table 14 shows the primers which were used for all the mutations.

5.3.2 Bacterial Expression of βPGM Mutants

Following mutagenesis of the WT βPGM gene to produce all of the proposed mutants, since even conservative mutations can substantially alter the expression of a gene it was necessary to optimise the conditions for the bacterial expression of each of them.

Mutation	Codon in WT	Ideal Mutant Codon (Usage, %)	Preferred Mutant Codon (Usage, %)
D15A	GAT	GCG (34)	GCT (19)
T16P	ACC	CCG (55)	CCC (10)
H20A	CAT	GCG (34)	GCT (19)
H20N	CAT	AAC (61)	AAT (39)
K219K	AAG	AAA (76)	AAA (76)
K219Δ	AAA	TAA (62)	TAA (62)

Table 13: Codons used for βPGM mutations. The table illustrates the compromise between mutations requiring few base substitutions and usage of the resulting codon. Note that for the K219Δ mutation, the K219K mutant gene was used as the template.

Mutation	Direction	Primer	Secondary Structure
D15A	Forward	GATGGTGTAATTACAGCTACCGCAGAGTATCATTTTAG	None
	Reverse	CTAAAATGATACTCTGCGGTAGCTGTAATTACACCATC	None
T16P	Forward	GGTGTAATTACAGATCCCGCAGAGTATCATTTTAGAGC	None
	Reverse	GCTCTAAAATGATACTCTGCGGGATCTGTAATTACACC	None
H20A	Forward	CAGATACCGCAGAGTATGCTTTTAGAGCTTGGAAAGC	Hairpin
	Reverse	GCTTTCCAAGCTCTAAAAGCATACTCTGCGGTATCTG	Hairpin
H20N	Forward	CAGATACCGCAGAGTATAATTTTAGAGCTTGGAAAGC	Hairpin & Two Self-Annealing Sites
	Reverse	GCTTTCCAAGCTCTAAAATTATACTCTGCGGTATCTG	Hairpin & Two Self-Annealing Sites
K219K	Forward	GTTTGGCTTCAAAAACAAAAATAACTCGAGCACC	Three Hairpins
	Reverse	GGTGCTCGAGTTATTTTTGTTTTTGAAGCCAAAC	None
K219Δ	Forward	GAAGTTTGGCTTCAATAACAAAAATAACTCGAGC	Three Hairpins
	Reverse	GCTCGAGTTATTTTTGTTATTGAAGCCAAACTTC	None

Table 14: Both forward and backward primers used for βPGM mutations, as well as any potential secondary structure. Hairpins are structures where one primer molecule folds over and anneals to itself, self-annealing sites are sites where two of the same codon molecules are able to anneal to each other. Note that for the K219Δ mutation, the K219K mutant gene was used as the template.

βPGM is normally expressed in *E. coli* BL21 (DE3) cells, induced at $OD_{600} = 0.6$, with 1 mM Isopropyl-β-Thiogalactoside (IPTG) and left at 25 °C overnight. It was assumed that due to the use of relatively rare codons, the expression of mutant βPGM would be slower. Initial test conditions were therefore expression in the same cell type, induced at $OD_{600} = 0.6$, with 1 mM IPTG and left at either 25 °C or 37 °C overnight. The culture volume in each case was 25 ml. WT βPGM was also tested under the same conditions for comparison. The resulting SDS-PAGE gels of these initial expression tests are shown in Figure 67.

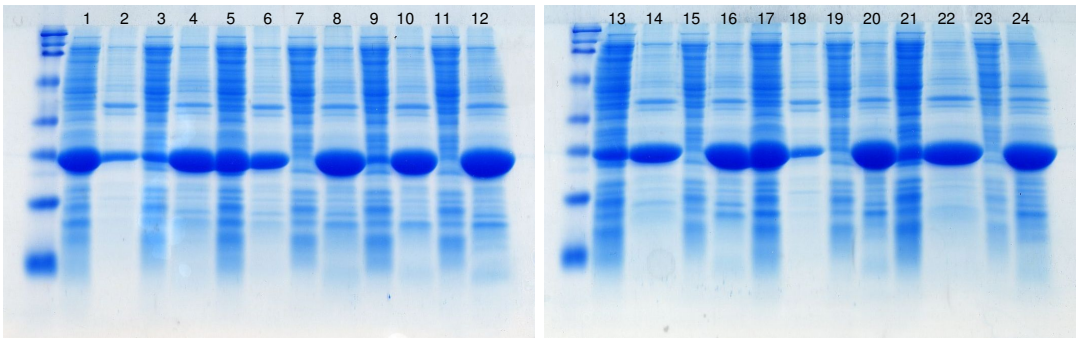

Figure 67: Gels from initial expression testing of βPGM mutants. Lanes 1–4 consist of WT βPGM and are arranged as: Lane 1 = soluble fraction, 25 °C; Lane 2 = insoluble fraction, 25 °C; Lane 3 = soluble fraction, 37 °C; Lane 4 = insoluble fraction, 37 °C. This pattern is repeated for all mutants, which are ordered as: Lanes 5–8 = D15A; Lanes 9–12 = T16P; Lanes 13–16 = H20A; Lanes 17–20 = H20N; Lanes 21–24 = K219Δ. Molecular weights of the markers are shown in Table 2 on Page 43

These gels show that all mutants express to a high yield despite the introduction of rare codons by SDM. D15A and H20N both show reasonable levels of soluble protein when expressed at 25 °C, but there is little or no soluble yield for many of the mutants, with all the expressed βPGM forming into inclusion bodies and therefore coming out in the insoluble fraction after sonication. There are many reasons why proteins can form inclusion bodies in the cell, but the most common view is that the production of the recombinant protein is too fast, such that the demand on molecular chaperones is too high, resulting in many of the molecules mis-folding and aggregating[230, 231]. As such there are two main ways of increasing the soluble expression of proteins: either the expression conditions can be changed such that the expression of the protein is slower, or the protein can be co-expressed with the molecular chaperones, preventing aggregation[231, 232]. It is also possible to recover (often highly pure) proteins from inclusion bodies, but this requires unfolding and refolding of the protein, which often has low yield[230].

By far the simplest option is to try to vary the expression conditions such that soluble protein is produced. Four main parameters can be changed to achieve better soluble expression of mutant βPGM, these are: OD_{600} at which expression is induced, the concentration of IPTG used to induce expression, the temperature, and also the length of time over which the expression occurs. As the gels in Figure 67 clearly show that the results at 37 °C are substantially worse than at 25 °C, expression at 20 °C was tested for D15A, T16P, H20A and K219Δ. This was found to work well for D15A, H20A and K219Δ.

The remaining mutant, T16P, was found to be difficult to express in high quantities, and the best condition found was to induce at OD_{600} = 0.6 and leave the culture shaking at 20 °C for 6 hours. The reduced time for expression results in lower final cell densities, and therefore even under these conditions the yield was low. After a large-scale (1 l) growth failed to produce enough T16P βPGM for ^{19}F NMR, work on this mutant was discontinued. It may be that the introduction of a proline residue into the hinge region disrupted the structure of βPGM sufficiently that it had a greater propensity to misfold and/or aggregate. However, this project would likely be feasible if time was spent producing larger bacterial growths, particularly as the crystal form of this mutant has been solved[229].

In contrast, 2 l cultures of LB media to express all the other mutants provided plenty of material, with the expression levels under the optimised conditions being approximately that for WT βPGM in LB media: around 50 mg/l of culture. All mutants were then purified by precisely the same protocol as WT βPGM (see Section 2), with no discernible differences in behaviour except for K219Δ, which predictably eluted later from the DEAE–Sepharose column due to the removal of two positive charges from the protein surface causing a reduction in the pI of the mutant enzyme.

5.4 NMR of Mutant βPGM complexes

5.4.1 D15A βPGM

A D15A βPGM-MgF_3^--βG6P TSA complex was formed under the usual conditions as for WT βPGM. A ^{19}F NMR spectrum of this complex is shown, in comparison to WT βPGM, in Figure 68. The spectra are identical, demonstrating that the D15A mutation has not purturbed the major⇌minor equilibrium at all. This implies that there are no differences in the free energy of the interactions made by the sidechain of D15 in either conformer. This could be because the N-capping of the D15–H20 helix by D15 in NAC I confers no stabilisation, or alternatively, it may be that the position of D15 in NAC I of the βPGM-BeF_3^--βG6P GSA complex is not representative of the minor conformer of the βPGM-MgF_3^--βG6P TSA complex. The assumption that the NAC I structure of the βPGM-BeF_3^--βG6P GSA complex resembles the minor conformer of the βPGM-MgF_3^--βG6P TSA complex underlies the strategy behind all the mutations made to βPGM, and so it was attempted to form a D15A βPGM-BeF_3^--βG6P GSA complex in order to test its validity. The ^{19}F NMR spectrum of this complex is compared to the WT complex in Figure 69.

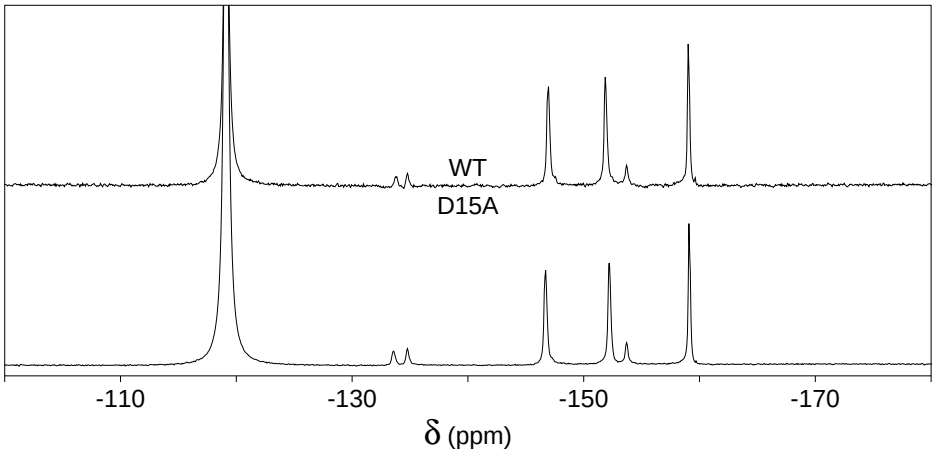

Figure 68: ^{19}F NMR spectrum of the βPGM-MgF_3^--βG6P TSA complex of WT and D15A βPGM, showing the spectra to be identical. As the equilibrium position between major and minor conformers is the same in the WT and mutant βPGM, there are therefore no differences in the interactions of the D15 sidechain between major and minor conformers of the βPGM-MgF_3^--βG6P TSA complex.

The comparison of the WT and D15A βPGM-BeF_3^--βG6P GSA complexes shows that the D15A mutation has destabilised the BeF_3^- complex, as it has been outcompeted by the βPGM-MgF_3^--βG6P TSA complex. Destabilising of the BeF_3^- complex is expected as the behaviour of this complex in solution is believed to be closer to NAC I[228]: the conformation which has the sidechain of D15 forming an interaction with the backbone amide of A17. Although the peaks originating from the free BeF_x species do not match between the spectra — suggesting differences in pH and/or Be^{2+} ion concentration — this suggests that the D15A mutation has achieved the desired effect: destabilised the NAC I complex of the βPGM-BeF_3^--βG6P GSA complex. However,

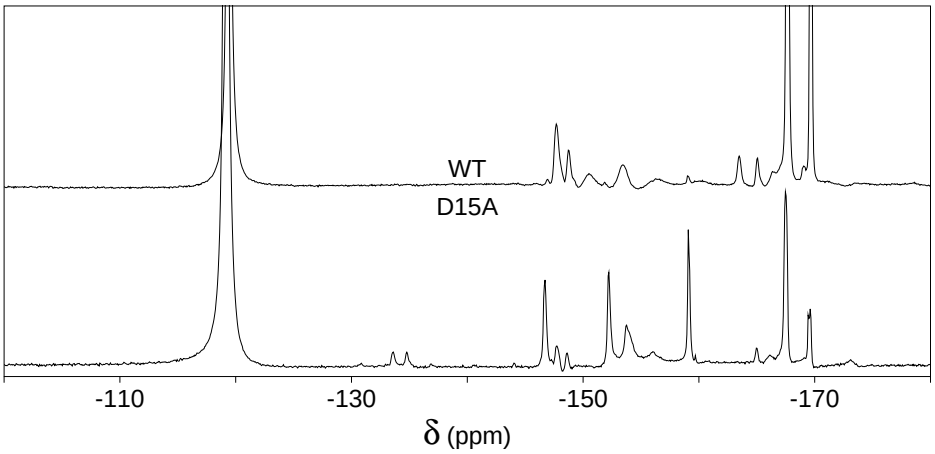

Figure 69: ^{19}F NMR spectrum of the βPGM-BeF_3^--βG6P GSA complex of WT and D15A βPGM. The peak at -163.5 ppm and the large peaks between -165 and -170 ppm originate from free BeF_x species in solution. As before the most downfield peak is the free F^- ion peak. All other peaks originate from enzyme–metal fluoride complexes. Due to the large number of signals present, backwards linear prediction has resulted in poor baselines. In the WT spectrum, there are a mixture of sharper peaks — originating from a βPGM-BeF_3^--βG1P complex[213] — and broader peaks which originate from a βPGM-BeF_3^--βG6P GSA complex. Although these same peaks are still present in the D15A βPGM spectrum, it shows mostly βPGM-MgF_3^--βG6P TSA complex peaks, which demonstrates that the affinity for the BeF_3^- complex has been lowered by this mutation. The upfield peaks from free BeF_x species likely look different due to inconsistencies with pH-adjustment between samples, but there may also be unintentional differences in Be^{2+} ion concentration.

as the mutation has had no effect on the major⇌minor equilibrium, the assumption that the minor conformer of the βPGM-MgF_3^--βG6P TSA complex resembles NAC I of the βPGM-BeF_3^--βG6P GSA complex was likely invalid.

5.4.2 H20N βPGM

A H20N βPGM-MgF_3^--βG6P TSA complex was formed under normal conditions and run at 298 K. It is shown in comparison to a WT βPGM spectrum in Figure 70. It appears that in comparison to the WT βPGM, the population of minor form is approximately half in H20N, which is the opposite result from what was expected. This mutant was thought to be the best candidate to destabilise the major conformer (based on the structures of the NAC complexes of the βPGM-BeF_3^--βG6P GSA complex), however as the D15A mutant demonstrates, the minor form conformation of the βPGM-MgF_3^--βG6P TSA complex probably does not particularly resemble the NAC I structure.

The most striking thing about the comparison between the two spectra is the broadening of F_b and F_c. To test whether this was exchange broadening, another ^{19}F 1D spectrum was run at 283 K, where the exchange rate (and therefore broadening) should be less. This proved to be the case, and so the two ^{19}F resonances are broadened due to ms–μs exchange. Although it is perhaps not a surprise that F_a — with the most hydrogen bonding to protein residues — is not affected as much as F_b and F_c, this does imply that far from the MgF_3^- moiety operating as a complex ion, the F^- ions

are independent of one another.

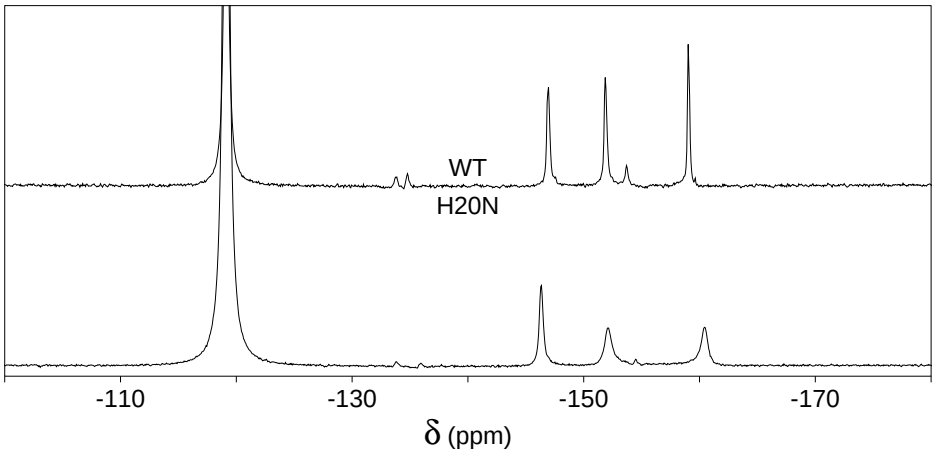

Figure 70: ^{19}F NMR spectrum of the βPGM-MgF_3^--βG6P TSA complex of WT and H20N βPGM, showing the lower population of minor conformer in the H20N mutant, and the broadening of the F_b and F_c resonances.

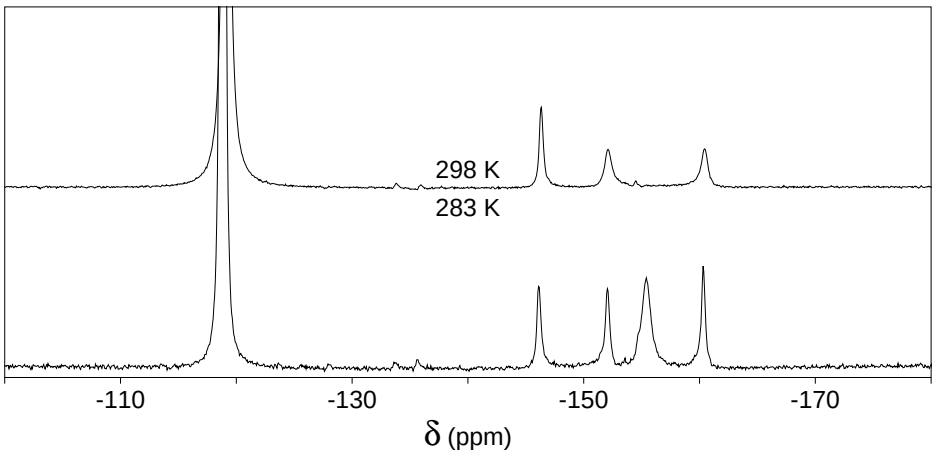

Figure 71: ^{19}F NMR spectrum of the βPGM-MgF_3^--βG6P TSA complex of H20N βPGM recorded at 298 K and 283 K. The sharper F_b and F_c resonances at 283 K confirms that they are broadened at 298 K due to conformational exchange.

The nature of the conformational exchange process affecting the F_b and F_c resonances is unclear: such broadening could be the result of F^- ions rapidly alternating between the F_b and F_c sites, or the result of both F^- ions exchanging with free F^- ions. Although N20 is expected to retain a hydrogen bond to K76, the side chain of asparagine is still much smaller than that of a histidine, as shown in Figure 66. As such, it is possible that this mutation has resulted in poor binding of the sugar (allowing the F^- ions to swap sites) or increased solvent access which would allow the F^- ions to exchange with free ions. This interpretation is further supported by the apparent role of H20 and Y80 in positioning the sugar for catalysis[228].

The lower temperature spectrum makes it more obvious how poor the signal:noise of the protein-bound ^{19}F peaks are: the broad MgF^+ peak at -155.4 ppm has much

lower intensity compared to the protein-bound ^{19}F peaks in WT spectra recorded at 283 K. Given that in both cases the concentration of βPGM was 1 mM, the H20N mutant must only be partially occupied by the TSA complex, showing the affinity for the MgF_3^- moiety, the substrate, or both, to be much lower than the WT enzyme. The lower temperature spectrum also more clearly demonstrates the lower population of minor conformer, as it is clearly much less than the 22 $\pm$ 2 % measured from ^{19}F SEXSY NMR on the WT enzyme at 283 K.

The other thing that can be asserted from these spectra is that the ^{19}F $\Delta\delta$s between WT and H20N are quite small. This shows that the electron density around the ^{19}F nuclei is not affected much by the mutation. This is surprising, as removal of the (likely protonated) N^{ϵ} would be expected to increase the pK_a of D10, thereby increasing the likelihood of the 1′ hydroxyl proton of the sugar being positioned on D10. This would in turn cause significant changes in the electron density around the MgF_3^- moiety, and therefore would alter the chemical shifts of the ^{19}F nuclei. The fact that this has not occurred suggests that any interaction between D10 and H20 is weak.

Given that this mutation had also been designed on the basis of the NACs formed by the βPGM-BeF_3^--βG6P GSA complex, an H20N βPGM-BeF_3^--βG6P GSA complex was formed to assess its effect. The resultant ^{19}F spectrum closely resembles spectra measured on the WT βPGM when complexed with BeF_3^- in the absence of βG6P (the comparison between these spectra are shown in Figure 72). This result is significant as it demonstrates that the affinity for the βG6P substrate is severely weakened by this mutation. This is supported by the kinetics study of this mutant which shows a 6-fold increase in K_m when compared to WT βPGM. The spectra do not exactly match however, as the linewidths in the mutant are larger, which again implies the existence of ms–μs dynamics in this complex. The source of this conformational exchange is unknown; as the βPGM-BeF_3^- complex is open, the exchange could be due to the substrate transiently binding to the enzyme. This could be easily tested by forming the complex in the absence of βG6P.

Finally, the formation of a H20N βPGM-AlF_4^--βG6P TSA complex was attempted. The resulting ^{19}F spectrum showed there to be a mixture of complexes in the presence and absence of βG6P, which supports the interpretation that the affinity for the substrate is lower in the mutant than in the WT enzyme.

5.4.3 H20A βPGM

Under the same conditions as for WT βPGM, an H20A βPGM-MgF_3^--βG6P TSA complex was formed, as verified by the ^{19}F spectrum shown in Figure 73. This spectrum shows that the exchange broadening seen in the ^{19}F resonances in the H20N MgF_3^- TSA complex is even more severe when residue 20 has a shorter side chain. F_b and F_c are completely exchange broadened, and there is also an R_{ex} contribution to the linewidth of F_a. The exchange contribution on F_a suggests that the conformational exchange occurring is not simply the alternating of F^- ions between the F_b and F_c

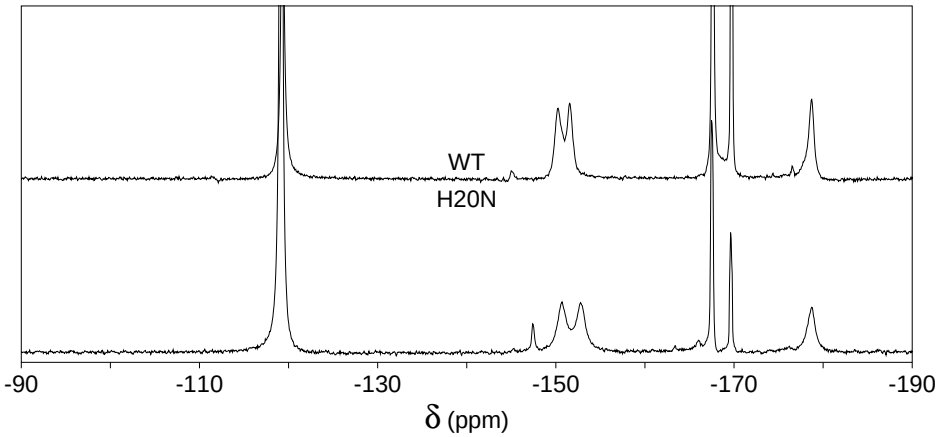

Figure 72: ^{19}F-NMR spectrum of the H20N βPGM-BeF_3^--βG6P GSA complex, compared to the WT βPGM-BeF_3^- complex. The similarity between these spectra suggests that the substrate is not significantly bound to the mutant. The additional exchange broadening in the mutant could arise from transient substrate binding. Only F_b has a slightly different chemical shift in the mutant to wild-type βPGM, supporting the conclusion that H20 and D10 do not significantly interact.

sites, and reflects a more global effect, such as increased solvent exposure in the active site. The signal:noise of the mutant spectra is also very poor, suggesting that the mutant is only very slightly occupied by the complex, as was the case for the H20N complex.

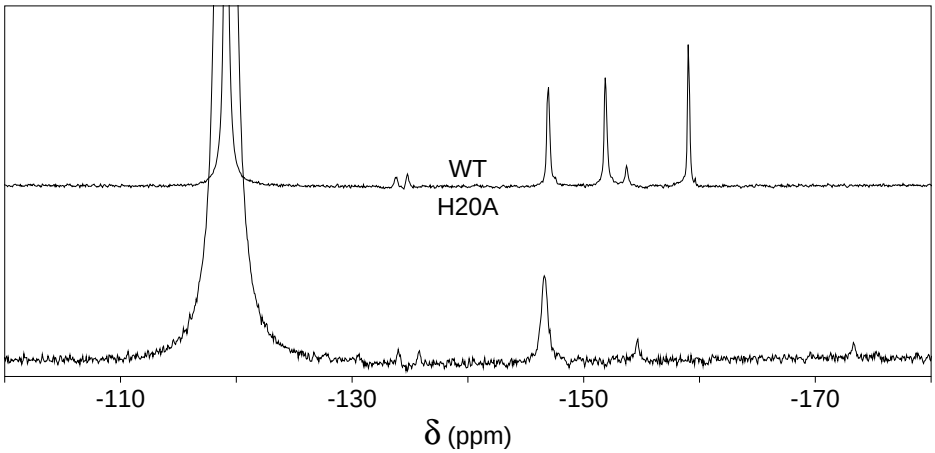

Figure 73: ^{19}F NMR spectrum of the H20A βPGM-MgF_3^--βG6P TSA complex, compared to the WT complex. As with the H20N complex, the population of minor conformer is actually decreased in the mutant, and there is exchange broadening on the major ^{19}F resonances.

This spectrum is consistent with the interpretation that the affinity for the sugar substrate is lowered by mutation of residue 20. In contrast to the H20N case however, the kinetic data from the study of Dai et al, 2009 is not consistent with this as the K_m for H20A is very similar to WT[229]. However, for a Bi Bi Ping Pong reaction, the apparent K_m value is the product of multiple catalytic rate constants and equilibrium constants between the different substrates, and is therefore not necessarily a good indicator of the true dissociation constant K_D. For example, if the relative affinities

of mono- and bis-phosphorylated sugars for βPGM has changed as a result of the mutation, this will have an effect on the measured kinetic parameters.

As with the H20N βPGM-MgF_3^--βG6P TSA complex, the exchange broadening of the ^{19}F resonances is alleviated with a drop in temperature, although in the case of H20A F_b and F_c remain very broad even at 283 K, suggesting that whatever the nature of the conformational exchange event causing this exchange broadening, the exchange rate is increased in this mutant compared to H20N. This would also be consistent with increased solvent exposure as the shorter side chain substitution will create a larger cleft in the active site, allowing more water and unbound F^- ions to diffuse in. The temperature dependence of the ^{19}F resonances is shown in Figure 74.

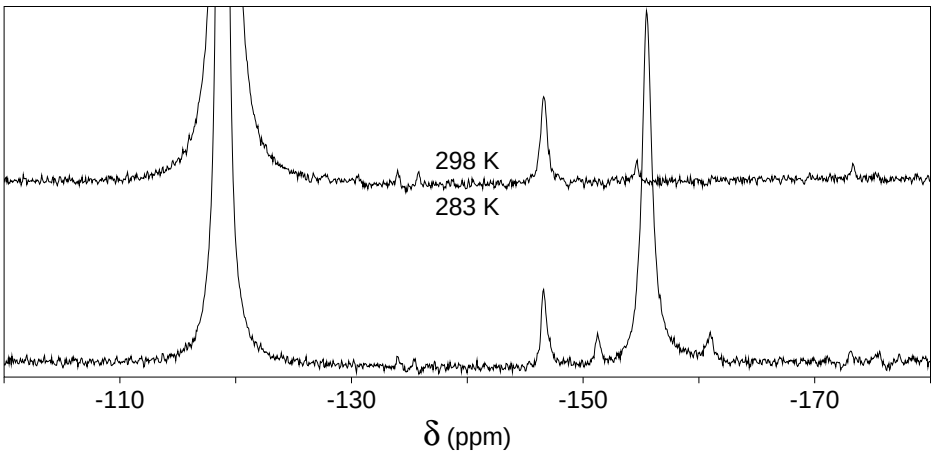

Figure 74: ^{19}F NMR spectrum of the H20A βPGM-MgF_3^--βG6P TSA complex recorded at 298 K and 283 K, showing partial alleviation of exchange broadening on F_b and F_c at lower temperature due to a drop in the rate of exchange.

An intriguing observation — which is more obvious in the H20A mutant due to the extreme exchange broadening of the major conformer peaks — is that despite the low signal:noise on the minor peaks, it is apparent that they are not affected by the same exchange broadening process that is affecting the major resonances. This is curious, as it implies that the F^- ions are more tightly bound by the Mg^{2+} ion in the minor form. It is not obvious why this would be the case.

5.4.4 K219Δ βPGM

Finally, a βPGM-MgF_3^--βG6P TSA complex was made using the K219Δ mutant which proved to be identical to WT, as measured by ^{19}F NMR (spectra shown in Figure 75). The removal of three residues from a protein is a relatively severe mutation, and so for this to have no effect of any kind on the βPGM-MgF_3^--βG6P TSA complex supports the view that the faster exchange process as measured by ^{15}N relaxation dispersion in Section 4 is completely unrelated and separate from the major$\rightleftharpoons$minor exchange process, and thus further validates the statistical analyses used to analyse the ^{15}N relaxation dispersion data in Section 4.2.

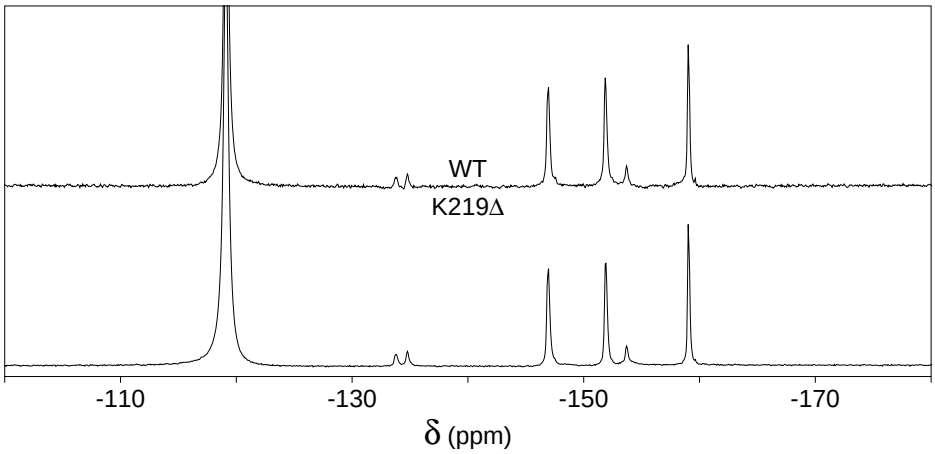

Figure 75: ^{19}F NMR spectrum of the βPGM-MgF_3^--βG6P TSA complex of WT and K219Δ βPGM, showing the spectra to be identical. As the equilibrium between major and minor conformers are the same in the WT and mutant βPGM, there are therefore no differences in the interactions of any of the last three residues of βPGM between the major and minor conformers of the βPGM-MgF_3^--βG6P TSA complex.

To gain insight into any differences in the fold of the K219Δ mutant compared to WT βPGM, ^{1}H 1D spectra were studied. The overlap of ^{1}H signals in the 1D spectra of large proteins make it impossible to be quantitative about how large or widespread these changes are, hence the widespread use of ^{1}H–^{15}N HSQC for such tasks, but the lack of ^{15}N-labelled K219Δ prevented this from being done. Although changes cannot be analysed in a site-specific fashion, with so many signals overlapping in ^{1}H 1D spectra if a few resonances move such that they now overlap and interfere with different signals, the resulting spectra can look significantly different despite only few changes being responsible.

It is obvious from the spectra shown in Figure 76 that there are a subset of ^{1}H nuclei in the K219Δ βPGM-MgF_3^--βG6P TSA complex which are in different chemical environments to the TSA complex of the wild-type protein. The differences cannot be accounted for by loss of resonances from the three truncated residues, and given that all three have aliphatic side-chains, the effects also cannot be due to abolished ring-current effects on the remaining residues nearby. Furthermore, there are too many differences between the spectra to be explained solely by the loss of the hydrogen bonding from or to the three removed residues. However, there are still many signals which overlay between the proteins, implying the differences in tertiary structure are relatively localised to a particular region within the structure.

These spectra suggests that the mutation — which was designed to perturb the equilibrium of conformational substates at the C-terminus — may have been successful. To further investigate whether the faster exchange process does in fact reflect a local unfolding event, the folding stability of the K219Δ and WT βPGM was tested by a guanidinium hydrochloride titration, using Circular Dichroism to report on the breakdown of secondary structure elements during unfolding.

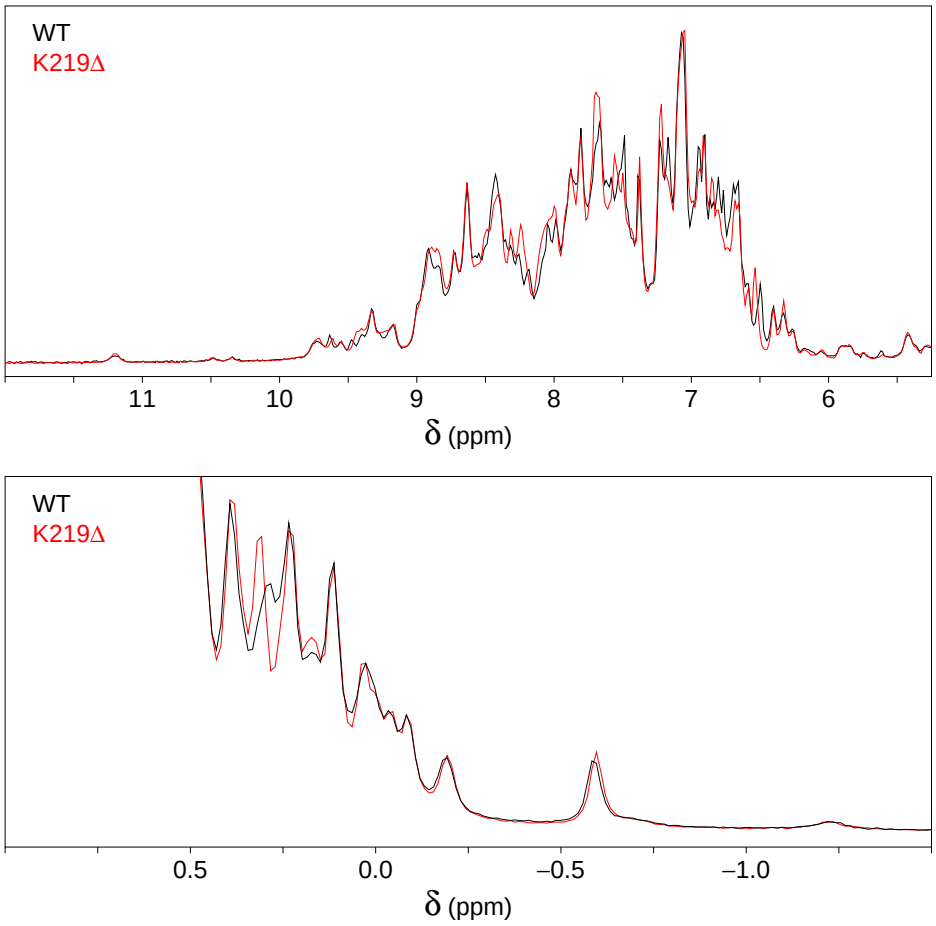

Figure 76: Overlay of ^{1}H NMR spectra of the βPGM-MgF_3^--βG6P TSA complex of WT (black) and K219Δ (red) βPGM. A few of the ^{1}H signals in the amide and methyl regions have changed chemical shift, whilst many others are identical, suggesting that there are a few localised changes in the tertiary structure of the mutant compared to WT. Importantly, the methyl region upfield of 0 ppm which is indicative of the packing of hydrophobic delocalised ring structures against methyl groups look identical, as does the signal from the amide proton of K117 at 11.2 ppm, which is dispersed from the main envelope due to packing against the phosphate group of the substrate.

5.5 Circular Dichroism of K219Δ βPGM

Although the ^{1}H NMR unambiguously shows that the tertiary structure of the K219Δ mutant has some small differences to WT in the βPGM-MgF_3^--βG6P TSA complex, circular dichroism — a reporter of the secondary structure of proteins — shows the two apo proteins to look identical within error, as shown in Figure 77. This is expected as the residues which have been truncated in the mutant did not comprise any distinct secondary structure elements. As expected from the crystallographic data on βPGM, the spectra show both proteins consist mostly of α-helix. As the same C-terminus dynamics measured in the βPGM-MgF_3^--βG6P TSA complex are also present in apo βPGM (as demonstrated in Section 4.2.5) it is expected that the perturbation of the equilibrium between conformers in the βPGM-MgF_3^--βG6P TSA complex will be reflected in the apo proteins. A modest change in the tertiary structure as reported by ^{1}H NMR and no perceivable change in secondary structure show that the K219Δ mutant is not significantly unfolded at the C-terminus.

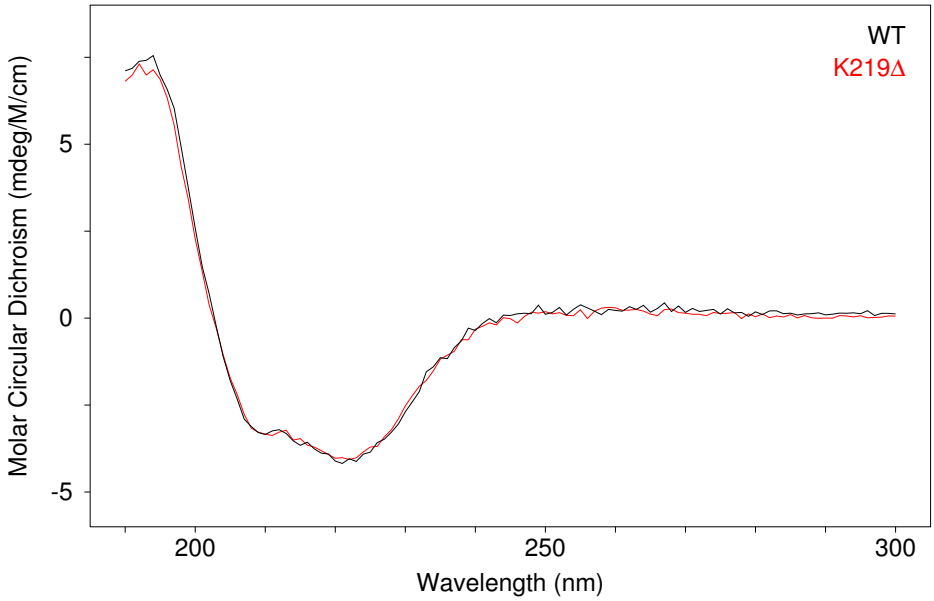

Figure 77: CD spectra of the WT (black) and K219Δ (red) βPGM, unligated. Due to inaccuracies in the determination of protein concentration, the K219Δ spectrum has been scaled to 105%. It can be seen that within error the spectra are identical, showing that the secondary structural elements are unchanged by the K219Δ mutation, and are largely α-helical.

Figure 78 shows the dependence of the WT βPGM CD spectra on the denaturant concentration. It can be seen that the negative peak at 222 nm reduces as the concentration of denaturant (GuHCl) increases due to the breakdown of α-helical structure. The spectrum at 0 M GuHCl does not exactly match that in Figure 77 due to different βPGM concentrations and buffer conditions. It is clear from Figure 78 that the midpoint of the unfolding transition in WT βPGM takes place at close to 1 M GuHCl, and that once βPGM has unfolded there is still some residual secondary structure which is slowly lost at higher denaturant concentrations. This "slope" of CD readings at higher GuHCl concentrations was taken into account in the fitting of these data, the equations for which are shown in Figure 79.

For greater sensitivity on the fitted parameters for both WT and K219Δ βPGM unfolding, each wavelength between 210–230 nm was fit simultaneously with global D_{half} and m values. For representation, the dependence of the CD value at 220 nm for both WT and K219Δ is shown in Figure 80. It is clear that not only does the mutant become 50 % unfolded at a lower GuHCl concentration, but the unfolding curve is also much shallower, which indicates that both D_{half} and m values are decreased by the K219Δ mutation. This demonstrates that the folding stability of the mutant is considerably lower than for the WT protein, supporting the hypothesis that the C-terminal dynamics represent a local unfolding event. The folding stability of both proteins can be quantified using the equations shown in Figure 81, and the final calculated figures for the free energy of folding (ΔG_{fold}) are shown in Table 15.

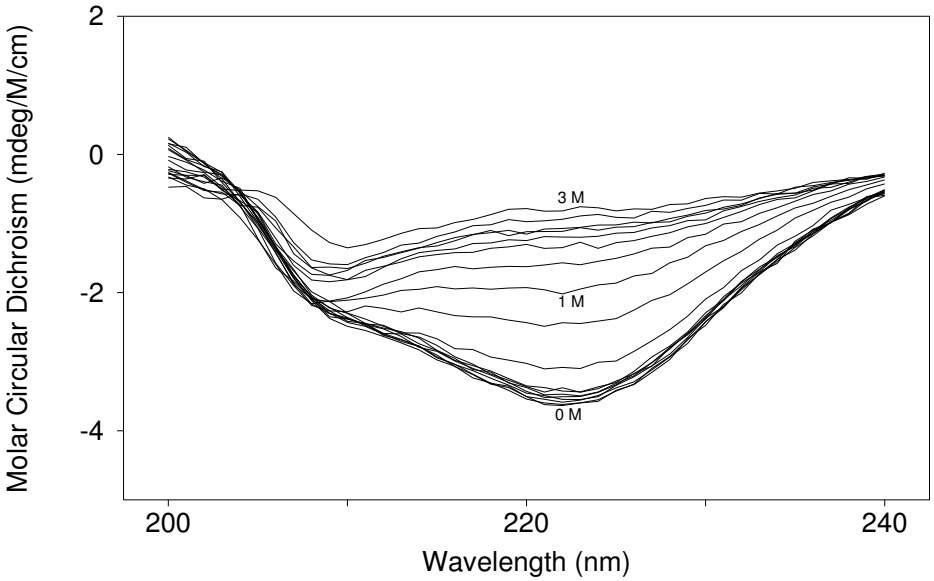

Figure 78: CD spectra of WT βPGM with various denaturant concentrations. The spectra become less negative around the minimum of 222 nm on increasing denaturant concentrations due to the unfolding of the primarily α-helical secondary structure. For clarity, the spectra corresponding to 0, 1, and 3 M GuHCl are labelled. It can be seen that βPGM does not unfold at all at low GuHCl concentrations, then undergoes a rapid unfolding transition with a mid-point close to 1 M GuHCl. The protein still has some residual secondary structure at the end of this transition, which is slowly unfolded at higher GuHCl concentrations.

$$K = \exp\left(m.(D - D_{half})\right)$$

$$CD = I_u + (slope.D) + \left(I_f - (I_u + (slope.D)) * \left(\frac{K}{1+K}\right)\right)$$

where:

D is denaturant concentration (M)

D_{half} is the value of D at which $K = 1$ (M)

K is the equilibrium constant between folded and unfolded states

m is the "m value": the dependence of K on D

I_u is the CD value of the unfolded state

I_f is the CD value of the folded state

slope is the dependence of I_u on D, due to residual secondary structure

Figure 79: Equations to which the CD data was fit: representing the folded⇌unfolded equilibrium in terms of the denaturant concentration, and the CD values as a function of the equilibrium position. The "slope" is a mechanism of taking into account the residual secondary structure in the unfolded state.

The large differences in both the D_{half} and m values between WT and K219Δ βPGM are surprising. χ^2 analysis confirms that the differences are very significant: fixing the m value at the weighted average figure results in an increase in χ^2 of 21 and 49 for the mutant and the WT protein, respectively. The result of the removal of the three, predominantly unstructured, residues from βPGM is the loss of half the folding stability of the protein. It is difficult to justify why the change is so large, but it is clear that the

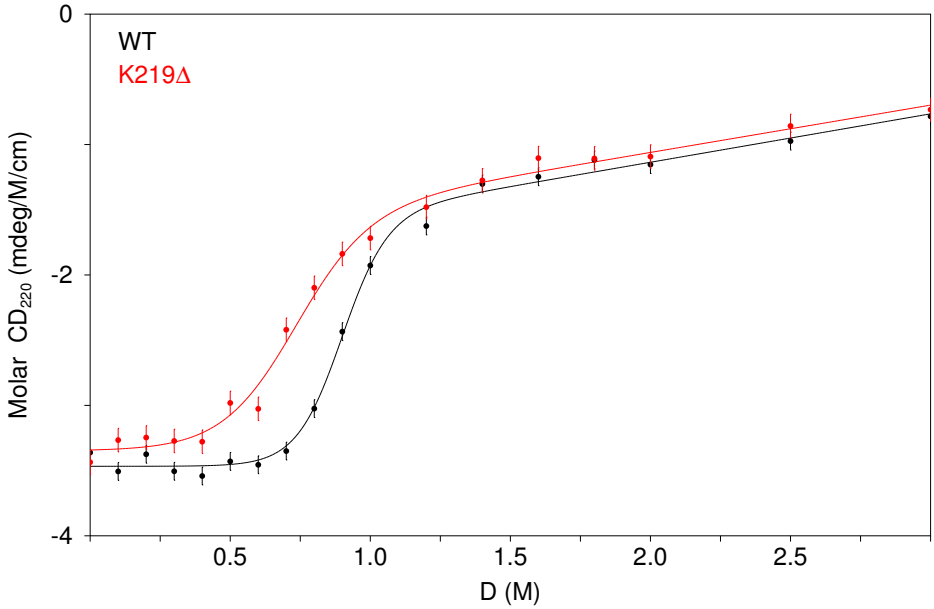

Figure 80: CD_{220} of WT and K219Δ βPGM as a function of denaturant concentration. This shows that the K219Δ mutation causes βPGM to become 50% unfolded at lower denaturant concentrations, and have a smaller slope during unfolding (m value), than WT βPGM. This demonstrates that the K219Δ mutation significantly impairs the folding stability of βPGM.

$$K_{fold} = 10^{-mD_{half}}$$
$$\Delta G_{fold} = RT \ln K_{fold}$$

where:

D_{half} is the denaturant concentration when protein is 50 % unfolded (M)

m is the "m value"

K_{fold} is the equilibrium constant between folded and unfolded states at 0 M GuHCl

T is the temperature (K)

R is the gas constant

ΔG_{fold} is the free energy of folding

Figure 81: Equations to calculate ΔG_{fold} from CD data. The "m value" (which is effectively the dependence of the equilibrium constant between folded and unfolded states on the denaturant concentration) and concentration of denaturant at the 50 % unfolded point can be used to calculate the equilibrium constant at 0 M GuHCl, when can then be used to calculate the free energy of folding.

mutant is substantially less stable than the WT enzyme even without quantification. NMR samples of WT βPGM — although eventually becoming proteolysed and the sugar ligand eventually being lost to dephosphorylation as discussed in Section 2.7.2 — are extremely stable. Fresh samples were always used to ensure reproducibility of results, but it is possible to acquire data from WT βPGM samples which are several months old. In contrast, 1 mM K219Δ βPGM samples aggregate over the course of a few weeks, even at 4 °C. Although the difference in folding stability between the enzymes could be smaller than that measured by CD, it is clearly significant.

β**PGM Type**	**WT**	**K219Δ**
D_{half} (M)	0.902 ± 0.003	0.723 ± 0.006
m value	−12.1 ± 0.4	−7.7 ± 0.3
$slope_{220}$	2.6 ± 0.3	2.7 ± 0.3
ΔG_{fold} (kJ/mol)	62 ± 1	32 ± 1

Table 15: Folding stability parameters for WT and K219Δ βPGM. It can be seen that both the "m value" and the concentration of GuHCl at the half-way point in the unfolding transition have changed in the mutant, leading to a very large change in the overall folding stability ΔG_{fold}.

6 Characterisation of Fast Timescale Dynamics of βPGM by ^{15}N NMR

6.1 Introduction

To explore the variation in fast timescale dynamics of βPGM as it progresses through its catalytic cycle, relaxation of backbone ^{15}N and methyl side-chain ^{2}H nuclei have been studied on Apo βPGM, the βPGM-BeF_3^--βG6P GSA complex, and the βPGM-MgF_3^--βG6P TSA complex. This work is ongoing, and as such the ^{2}H relaxation is not discussed, nor is the ^{15}N relaxation study of the βPGM-BeF_3^--βG6P GSA complex complex which is yet to be completely recorded.

6.2 Apo βPGM

6.2.1 Exchange Broadening

It has previous been documented[119] that several peaks in the apo βPGM ^{1}H–^{15}N TROSY spectra could not be assigned due to exchange broadening, i.e., large R_{ex} terms in the R_2 relaxation rates due to ms–μs timescale exchange phenomena. Of those which have been successfully assigned, in many cases the exchange broadening is sufficient to prevent any relaxation analysis. Unfortunately many of these exchange-broadened resonances correspond to active site residues which are of most interest from a functional viewpoint. In addition to this, as the Apo βPGM data was acquired without TROSY detection on fully protonated βPGM, signal overlap is also an issue for many residues. Nonetheless, there were a total of 129 residues whose relaxation rates could be studied at both 500 MHz and 800 MHz. As the HSQC of apo βPGM recorded at 800 MHz shown in Figure 82 demonstrates, there is much heterogeneity in the peak intensities, a further indication of ms–μs exchange. The locations of the residues which could not be studied are shown in Figure 83.

6.2.2 Relaxation Data

R_1 and R_2 measurements were carried out at 500 MHz and 800 MHz, but due to experimental difficulties the heteronuclear NOE was measured at 800 MHz only. These experiments were carried out and analysed as described in Sections 2.8 and 2.10. The average percentage errors for the fitted T_1, T_2, and hetNOE values from each dataset are shown in Table 16. For illustration, selected R_1 relaxation decays recorded at 500 MHz are shown in Figure 84. To counteract the lower intrinsic sensitivity of the lower field experiments, the 500 MHz data was carried out with twice the number of scans per increment as were used to acquire the 800 MHz data, resulting in slightly better signal:noise on the fitted values.

To evaluate if the fitted T_1, T_2, and hetNOE values at both field strengths were theoretically reasonable, graphs were plotted to compare the experimental values to those predicted by HydroNMR, as is explained in Section 2.10.5. In the T_1 vs. T_2 plots at both field strengths (Figures 85 and 86) it is immediately apparent that the range in

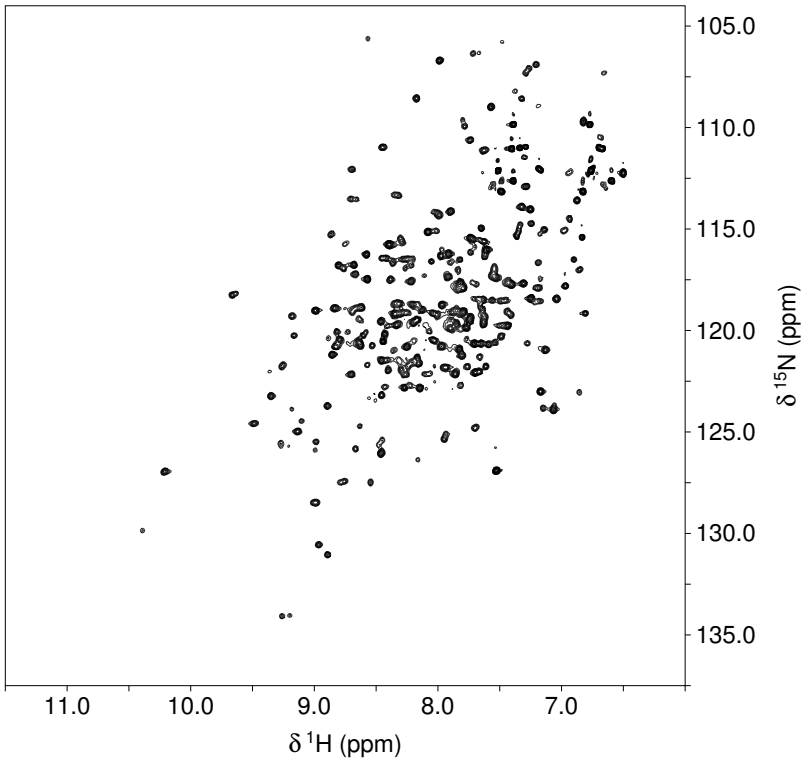

Figure 82: ^{1}H–^{15}N HSQC of apo βPGM recorded at 800 MHz. The heterogeneity in peak intensities between residues is indicative of chemical exchange, although there are many well-dispersed intense peaks which can be studied.

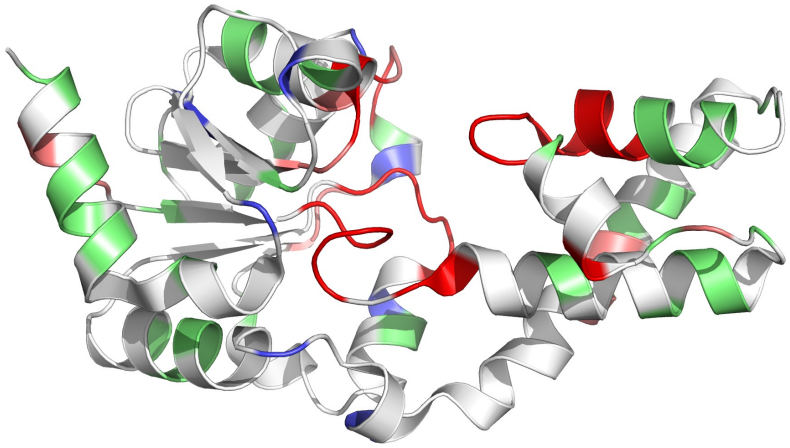

Figure 83: Apo βPGM co-ordinates represented as a cartoon of the backbone atoms. Residues which could not be assigned due to chemical exchange are in red, whilst residues which were subsequently excluded from relaxation analysis due to chemical exchange are shown in pink. Unassignable (i.e., proline residues and the N-terminal methionine) residues are shown in blue, and those residues whose relaxation could not be studied due to signal overlap are green. It can be seen that the most severe exchange broadening is localised to the active site.

molecular tumbling times reported by the experimental data — the range in effective molecular tumbling times reported by residues with N–H bond vectors which point in different directions with respect to the anisotropic diffusion tensors — fit exactly into the range expected from the crystal structure. The exceptions to this include residues

Field Strength	Measurement	Mean Value	Mean Error
500 MHz	T_1	883 ms	4.0%
	T_2	58.6 ms	3.3%
800 MHz	T_1	1570 ms	6.5%
	T_2	45.1 ms	3.8%
	hetNOE	0.724	3.8%

Table 16: Mean values and errors for relaxation measurements on apo βPGM.

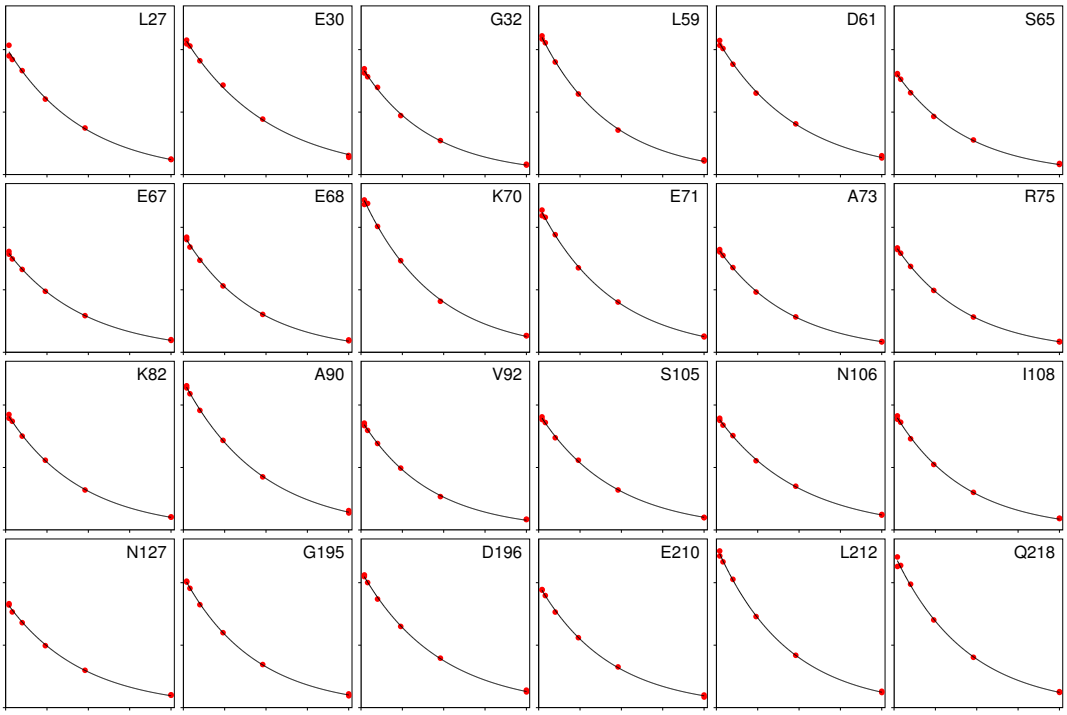

Figure 84: Selected ^{15}N R_1 decay curves recorded at 500 MHz on apo βPGM. As previously, the red filled circles are experimental peak intensity measurements, and the black lines correspond to fitted exponential R_1 decays for each residue. Error bars corresponding to the peak intensity values are included, but in all cases are contained within the red circles and cannot be seen. Despite the differences in signal intensity between residues, all points are well satisfied by the fit lines.

with extremely short T_2 constants likely due to R_{ex} terms, or those with very long T_2 constants due to flexibility on the ns–ps timescale (supported by a low hetNOE value for those residues). The points with short T_2 constants due to R_{ex} are more obvious in the 800 MHz dataset, as the R_{ex} components scale with the external field strength.

At both field strengths, the position of the T_1 and T_2 values relative to the grid suggest an average τ_m value of just over 14 ns, and most S^2 values in the range 0.9–0.85. The T_1 vs. hetNOE plot from the 800 MHz data (Figure 87) shows that the NOE values suggest higher S^2 values than the T_1 vs T_2 plots, with an average of nearer 0.95. Given how susceptible the hetNOE measurement is to systematic errors (as discussed in Section 2.8.10), it is likely that the hetNOE values are perhaps all a little higher than they should be. This is supported by the existence of hetNOE values which are physically impossible, i.e., outside of the grid. It is worth mentioning that the apparent S^2 does depend on the value of τ_e used to plot the grids, particularly true for the T_1 vs. hetNOE plot. In all cases an internal correlation time of 50 ps was assumed.

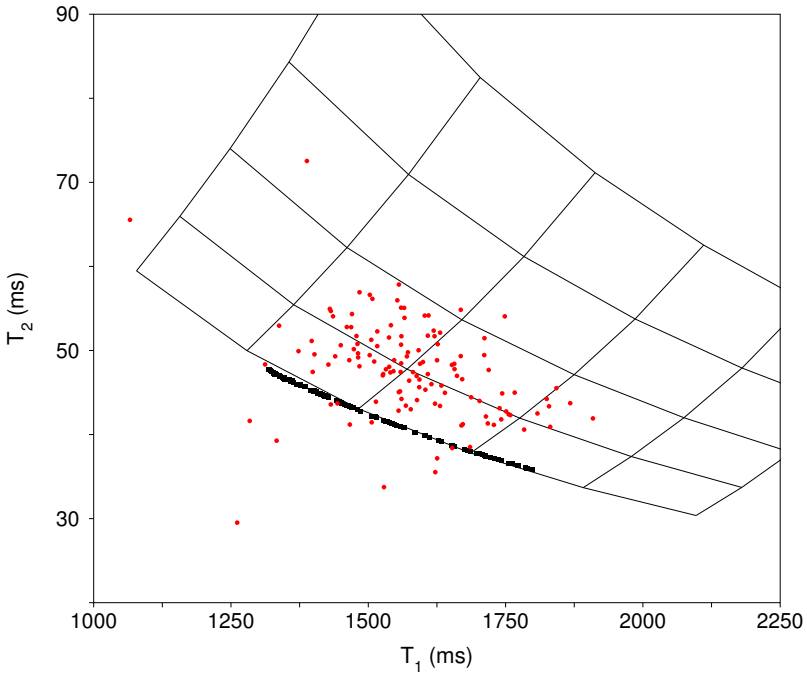

Figure 85: T_1 vs. T_2 plot for apo βPGM data recorded at 800 MHz (red circles), showing comparison of data predicted from HydroNMR (black squares). The grids represent expected values for a range of τ_m and S^2 values, whilst the τ_e value is fixed at 50 ps. It can be seen that the range in values measured experimentally agrees with the values predicted by HydroNMR except for variation in S^2.

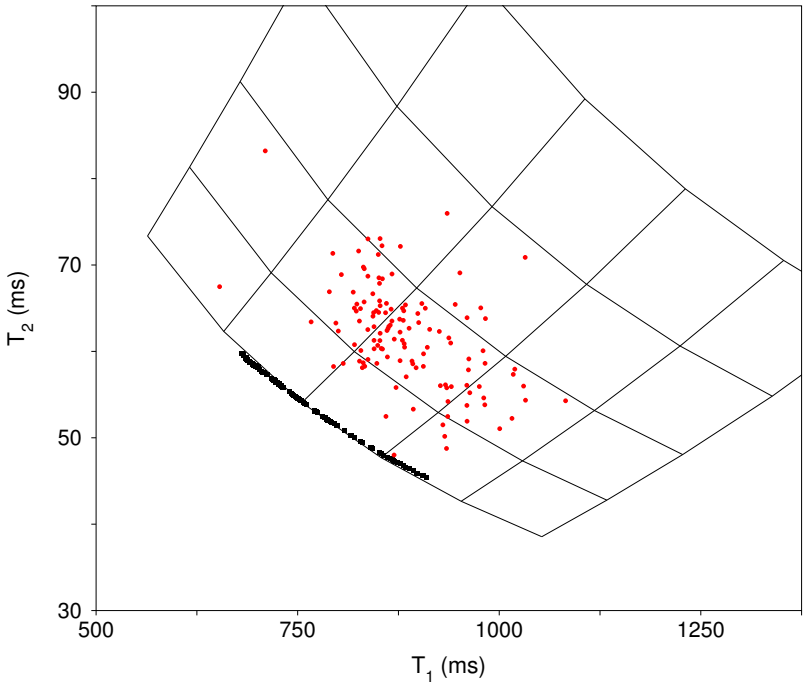

Figure 86: T_1 vs. T_2 plot for apo βPGM data recorded at 500 MHz (red circles), showing comparison of data predicted from HydroNMR (black squares).

As well as comparing the regions that the data occupies, it is also useful to assess how each T_1 and T_2 value corresponds to those predicted by HydroNMR. To partially remove variation due to flexibility on the ns–ps timescale, the T_1/T_2 ratio is compared. The plot of T_1/T_2 ratio with respect to residue number in Figure 88 shows that there is

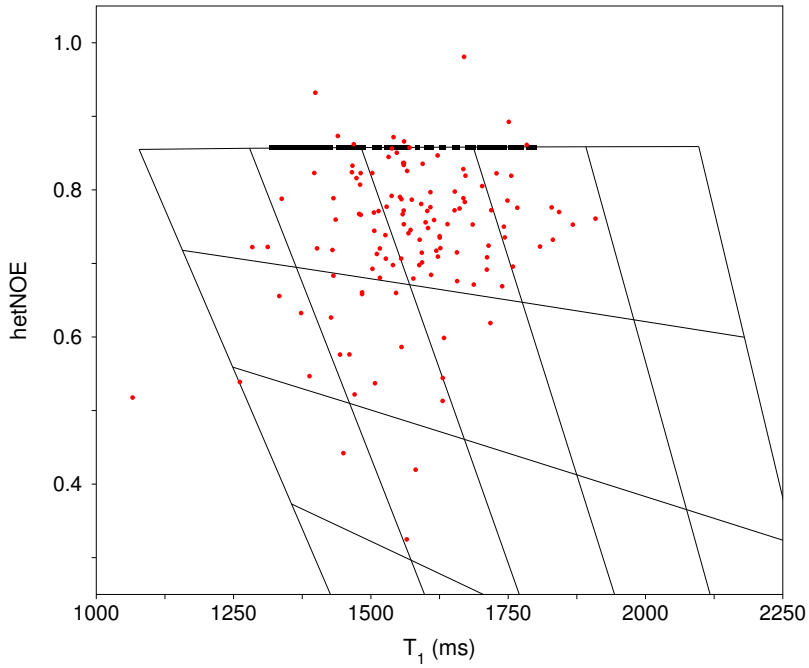

Figure 87: T_1 vs. hetNOE plot for apo βPGM data recorded at 800 MHz (red circles), showing comparison of data predicted from HydroNMR (black squares).

broad agreement between expected values, although in many cases the measured T_1/T_2 value is higher than that predicted; this could be due to widespread R_{ex} components which systematically reduce T_2. The section of residues from A17–F40 demonstrates that although the predicted and measured values do not always agree within error, the trends predicted based on N–H bond vectors in the Apo βPGM crystal structure are indeed present in the experimental relaxation data. This is further supported by Figure 89 which shows the T_1/T_2 ratio as a function of N–H bond vector. The experimental T_1/T_2 values clearly demonstrate the anisotropic diffusion of apo βPGM in solution, as expected by the predicted values.

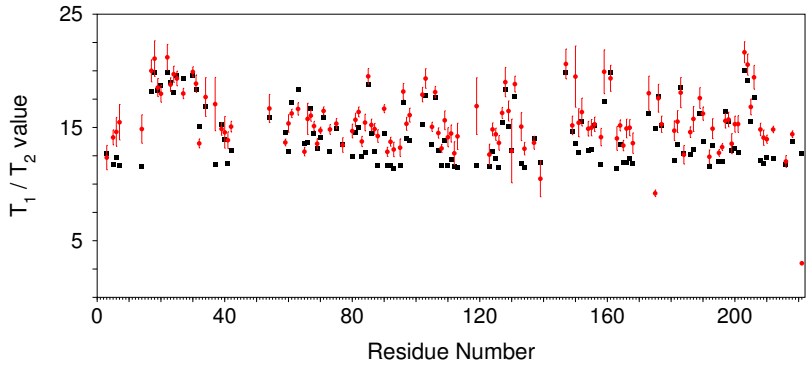

Figure 88: T_1/T_2 value plotted against Residue Number for apo βPGM data recorded at 500 MHz, showing broad agreement of predicted anisotropy with experimental results.

6.2.3 Apo βPGM Model Free Fitting and Validation

As the data were deemed to be of sufficient quality, they were then fit using the Relax program. After the fitting was complete, the fitted parameters were then used

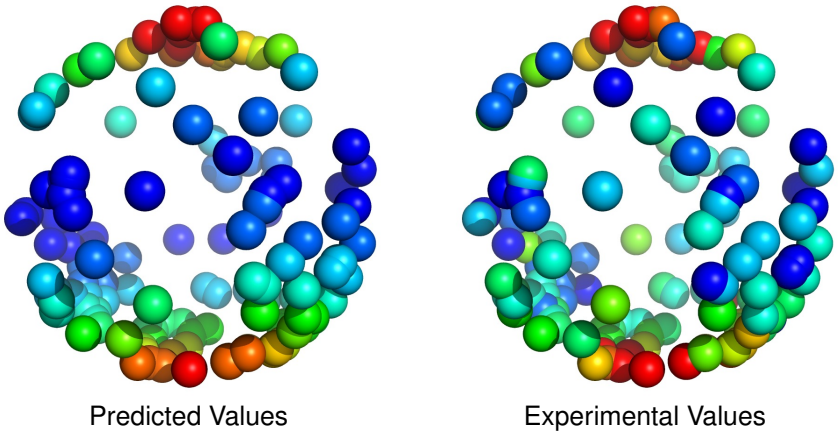

Figure 89: Anisotropic diffusion of apo βPGM demonstrated by T_1/T_2 ratios, both in predicted relaxation data from HydroNMR, and experimental relaxation data at 500 MHz. The positions of the spheres are determined by the N–H bond vector in the crystal structure and the colour is determined by the T_1/T_2 value.

to back-calculate the experimental values such that they could be compared to those input. The correlation coefficient between the experimental and back-calculated values for all the parameters was very high, although for the 500 MHz T_1 data the value is anomalously low at 0.877. The next lowest is the 500 MHz T_2 data at 0.961, and the lowest correlation coefficient for the 800 MHz data was 0.969 for the T_2 measurements. Table 17 shows the fitting statistics with which the fully anisotropic "ellipsoid" diffusion model (with three different diffusion tensors) was found to satisfy the data best, as determined by Akaike's Information Criterion (AIC). The principal axes of diffusion are plotted onto the crystal structure and displayed in Figure 90.

Diffusion Model	**Parameters**	**Data Points**	χ^2	**AIC**
Local τ_m	409	645	476.7	1294.7
Sphere	388	645	704.8	1480.8
Prolate Spheroid	357	645	523.6	1237.6
Oblate Spheroid	377	645	568.7	1322.7
Ellipsoid	362	645	503.1	1227.1

Table 17: Results of fitting apo βPGM data to different diffusion models. The models (in order of their ability to satisfy the data) are: ellipsoid, prolate spheroid, local τ_m, oblate spheroid, and sphere.

The global correlation time was fit as 14.5 ns which fits extremely well with the value inferred from the T_1 vs. T_2 plots shown in Figures 85 & 86, with a full rotation about the three principal axes of diffusion taking 67, 100, and 107 ns. The two longer sides of the molecule are found to be similar in length, and accordingly the "prolate spheroid" diffusion model satisfied the data almost as well. This matches the expectations from the crystal structure, where one side of the molecule is clearly much longer than the other, and the other two look approximately the same. The ratio of diffusion rates parallel and perpendicular to the principal axis of diffusion — $D_{\parallel}/D_{\perp}$, or D_{ratio} — was fit as 1.55, which is considerably lower than that predicted by HydroNMR (1.75) and infact closer to the value that HydroNMR predicts for the fully closed βPGM-MgF_3^--

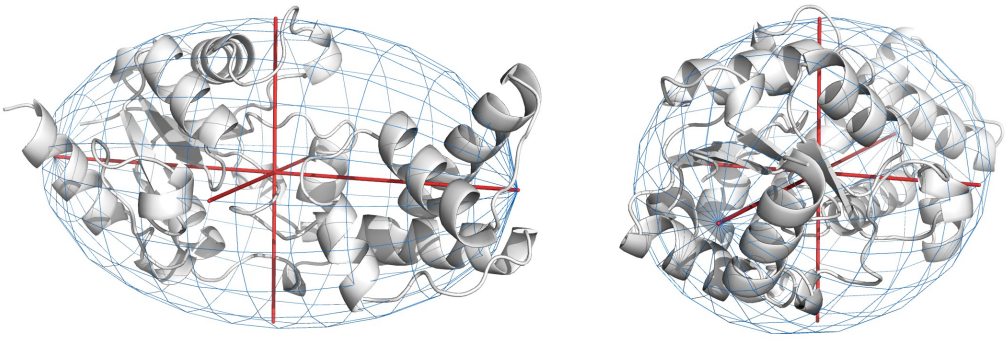

Figure 90: Anisotropic diffusion tensors from apo βPGM relaxation data (red) and their resulting ellipsoid (blue), plotted onto the crystal structure. The two images are rotated about a 90° angle from each other. It can be seen that although the ellipsoid generally agrees well with the crystal structure, the principal axis of diffusion is slightly shorter than would be expected from the crystal structure, with the end of the cap domain lying outside of the ellipsoid. This suggests that apo βPGM may transiently sample a closed conformation in solution.

βG6P TSA complex crystal structure, which is 1.43.

Although HydroNMR can perform poorly for proteins consisting of multiple domains connected by flexible linkers[233], the lower anisotropy may imply that apo βPGM transiently samples a closed-like state. As shown in Figure 90, the agreement between the diffusion axes and the crystal structure is good, although by fitting each domain separately it would be possible to assess if the principal axes as felt by ^{15}N nuclei in both domains are exactly co-linear.

Interestingly for the inappropriate diffusion models (sphere and oblate spheroid) the total number of parameters used was higher, demonstrating that more complicated Model-Free models were required to satisfy the data. It has been shown that there are two primary effects which result from an inaccurate diffusion model. Residues whose N–H bond vector points along a side of the molecule which has been modelled too short will contain falsely large R_{ex} components[234], and residues whose N–H bond vectors are such that their local correlation time have been overestimated may contain false internal correlation times on the ns timescale[235]. This information was used to critically assess the choice of diffusion model.

Both of these effects were immediately obvious upon plotting those residues which have ns τ_e values or R_{ex} components in the sphere model which were not present in the prolate spheroid or ellipsoid models. All the residues which had R_{ex} terms in the sphere model only had N–H bond vectors which were parallel to the principal axis of diffusion, and all the residues which had ns internal correlation times fit only in the sphere model were perpendicular. Importantly, residues which have R_{ex} terms fit in the prolate spheroid but not in the ellipsoid model were also orientated in a similar fashion, facing along what is presumably the longer of the two shorter sides. There is no correlation between N–H bond vectors of the remaining residues with R_{ex} terms or ns τ_e values in the ellipsoid diffusion model compared to the prolate spheroid model. These observations are shown in Figure 91.

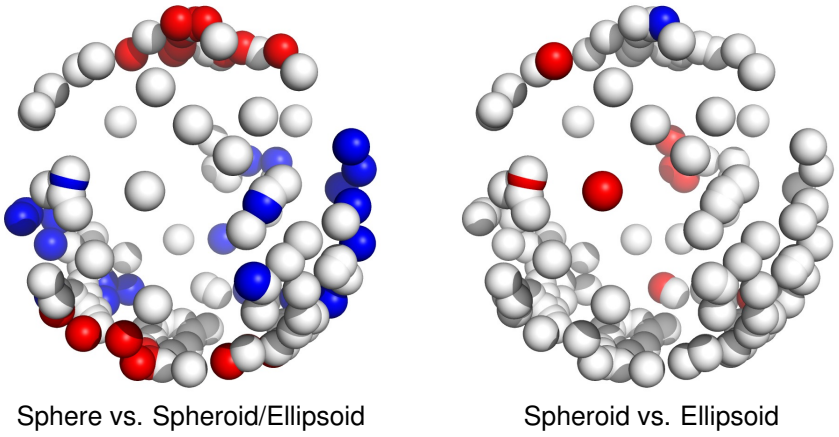

Figure 91: Dependence of R_{ex} and ns τ_e motion fitting on diffusion model. As before, the position of the spheres represents the N–H bond vector of a residue in apo βPGM. In both cases, R_{ex} terms present in the first model but not in the latter model(s) are coloured red, and ns τ_e motions present in the first model but not in the latter model(s) are blue. All other residues are coloured white. Note that the orientation of the spheres is the same as in Figure 89.

MF Model	**Parameters**	**Residues Fit**
m0	None	1
m1	S^2	12
m2	S^2, τ_e	32
m3	S^2, R_{ex}	3
m4	S^2, τ_e, R_{ex}	10
m5	S^2_s, S^2_f, τ_s	41
m6	S^2_s, S^2_f, τ_s, τ_f	0
m7	S^2_s, S^2_f, τ_s, R_{ex}	29
m8	S^2_s, S^2_f, τ_s, τ_f, R_{ex}	1
m9	R_{ex}	0

Table 18: Parameters fit in each Model Free model in Relax, and the number of residues which used each model in the analysis of apo βPGM data.

Table 18 demonstrates that there is no obvious bias in selecting models that have more parameters, i.e., the model selection is sufficiently harsh to avoid over-parameterisation. This was also tested by looking at the average fractional decreases in χ^2 upon the addition of a new parameter. For example, the χ^2 values for those residues which were fit using m4 (S^2, τ_e, and R_{ex}) were compared to the χ^2 values when those residues are fit using m2 ($S^2$and τ_e) to assess the impact of adding the R_{ex} parameter. In general, it was found that the decreases in χ^2 upon the addition of a new parameter were much greater than 2 which is required for AIC selection of the appropriate model. It was also commonly found that for the residues fit with an earlier model, the differences in χ^2 when another (unnecessary) parameter is added are very small. Although there are occasions when the improvement in χ^2 is close to 2 on the addition of a new parameter, for most residues the difference in χ^2 is either very large or much less than 2, making the model selection relatively straightforward. After this series of checks, it was clear that the fitting had been successful.

By doing the analysis of improvements in χ^2 on the addition of a new parameter for all models, it could be seen which are the best-defined parameters. By far the

largest improvement in χ^2 results from the addition of a first or second S^2 parameter. The differences on the addition of a second S^2 parameter are surprisingly large in the residues where it has been fitted, supporting that it is indeed possible to fit two independent internal motions to a single residue. The next most significant changes in χ^2 result from the addition of an internal correlation time parameter: τ_e, τ_s, or τ_f. The fitting of an R_{ex} component makes the least difference to χ^2. This data shows that S^2, even a second S^2, is very well defined by the data, the internal correlation time less so, and R_{ex} least of all.

As would be expected, these results are mirrored by the uncertainties on each of the fitted parameters, with S^2 having the smallest errors, and R_{ex} the largest. $\overline{\sigma S^2} = 2.8\%$ and accordingly, where appropriate the uncertainties on S_s^2 and S_f^2 are smaller. There are only three residues with σS^2 of >10%. The precision in the fit values is as good as can be expected for a system as large as βPGM. τ_e and τ_s have average errors of 33% and 41% respectively, with 13 residues having $\sigma\tau > 100\%$. It is curious that the addition of these parameters can significantly decrease χ^2, even if the value of the parameter is undefined. If the internal correlation time is not fitted, it is assumed to be 0 ps. Only one residue (N77) has a τ_f parameter fitted, with an uncertainty of 38%. Finally, the average error on the R_{ex} values is 43%.

Initially the large number of residues with R_{ex} terms fit (43 in total) was seen as a concern, however it was possible to rationalise the vast majority of these residues due to their proximity within the tertiary structure to those which are noticeably exchange broadened or unassignable in the spectra of apo βPGM. Furthermore, some of the residues with R_{ex} terms were those which have been shown to be involved in the transient unfolding motion at the C-terminus. With these also excluded, only 14 residues which cannot be immediately rationalised remain: 30, 32, 60, 61, 63, 98, 102, 105, 155, 196, 198, 205, 206, 212. Many of these are close to the C-terminus and therefore likely involved in the unfolding motion, and the remainder occur in clusters, supporting the assumption that these fitted R_{ex} terms report on something real. In addition, the fitted R_{ex} terms are very small, with the largest being 0.56 s^{-1} at 800 MHz for D37, and most being below 0.15 s^{-1}. Stated differently, the contribution of R_{ex} terms (where fitted) to the overall R_2 rate at 800 MHz is less than 1% except for D37 where it is 1.5%. It is only due to the highly sensitive R_2 measurements that so many R_{ex} terms could be fit by Relax.

6.2.4 Analysis of R_{ex} Terms

Because of the smaller D_{ratio} than expected from the Relax fitting, and the widespread exchange broadening in apo βPGM spectra, it was suspected that apo βPGM may transiently sample a closed conformation. Although the conformational changes between the open and closed conformation of βPGM are relatively small for most residues due to approximately rigid body motion of the domains relative to each other through a modest angle of 26°[130], if a significant population of βPGM molecules in

solution are in a closed-like conformation, the model free results may be biased due to the N–H bond vectors measured from the crystal structure of apo βPGM not reflecting the ensemble average.

TALOS[236] was used to predict the dihedral angles in apo βPGM from the measured δs of the ^{15}N and $^{1}H^{N}$ nuclei. These predicted ϕ & ψ angles were then compared to those calculated from the apo βPGM and βPGM-MgF_3^--βG6P TSA complex crystal structures to see which structure is better represented by apo βPGM in solution. As TALOS uses δs from adjacent residues, the results from any residues which were unassigned or adjacent to an unassigned residue were excluded. The comparison is shown in Figure 92: the selected residues are either those with large $\Delta\phi$ or $\Delta\psi$ between open and closed crystal structures, those known to be functionally important, or those which are exchange broadened in the apo βPGM spectra.

It is striking how few residues change dihedral angles between conformations. Only 6 residues could be reliably used to distinguish between the conformers: D15, N34, G35, A60, D61, and K62. Of these, D15 and G35 are not assigned in the apo βPGM spectra, and the prediction of ϕ & ψ angles from the chemical shifts of D61 match neither structure. The remaining residues, N34, A60, and K62 unambiguously show that in solution, apo βPGM is dominated by the open conformation seen in the crystal structure.

Of course this does not rule out that apo βPGM transiently samples a closed-like conformation in solution. The residues with the largest R_{ex} terms in the spectra were studied, comparing the interactions formed with other residues in the apo and MgF_3^- TSA complex. Much of the unassignable regions in the apo βPGM spectra correspond to active site loops which undoubtedly change their interactions with surrounding residues depending on the whether βPGM is open or closed. However, the large $\Delta\delta$s measured by relaxation dispersion on the βPGM-MgF_3^--βG6P TSA complex demonstrates that these loops are also capable of being flexible on the ms–μs timescale independent of hinge opening and closing. As such, it is difficult to comment on the cause of the exchange broadening of resonances from residues in the in the loop regions, as well as the cause of the R_{ex} terms fit to many of the residues which are adjacent to them.

D37 has the largest R_{ex} term and is adjacent to R38 which could not be assigned, but there are no loops or significant changes around this region on closure, with the apo βPGM and βPGM-MgF_3^--βG6P TSA complex crystal structures overlaying well around this region. The amides of both residues primarily form hydrogen bonds with surface water molecules in both crystal structures and so it is not apparent what could be the cause of such a significant perturbation. The next largest R_{ex} term is from residue N77, which is adjacent to two residues which, although assignable, have obvious linebroadening in the spectra of apo βPGM. This helical region does show some movement on closure of βPGM, although the hydrogen bonding partners (and

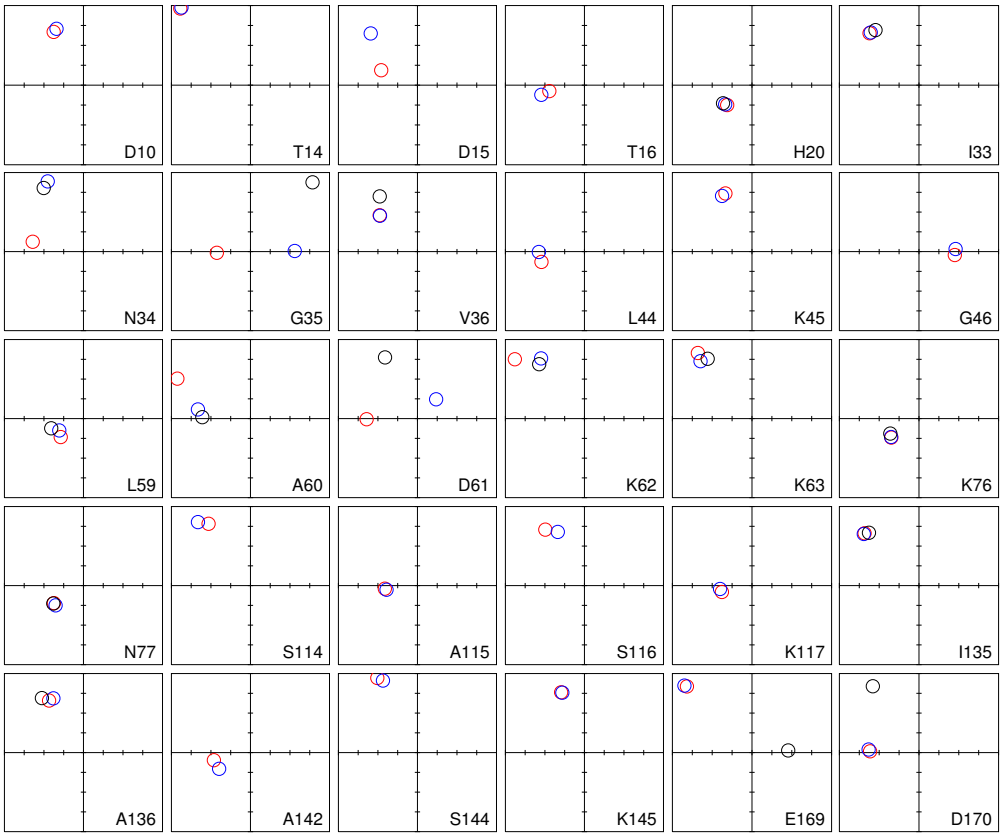

Figure 92: TALOS analysis of Apo βPGM δs showing it to be primarily open in solution. For each residue, an individual Ramachandran plot is shown with x & y axes (ϕ & ψ respectively) running $-180°$–$+180°$. The dihedral angles predicted from the Apo βPGM chemical shifts are shown as black circles, with circles corresponding to the dihedral angles calculated from the apo βPGM and βPGM-MgF_3^--βG6P TSA complex crystal structures shown in blue and red respectively.

even the bonding distances) do not change, these regions do not overlay well in the open and closed crystal structures. Their proximity to the outer hinge residues 85–87 suggests that the movement of the hinge on closure may affect the C-terminal end of this helix. However, that the exchange broadening is not worse at the very C-terminal end of the helix cannot easily be explained.

The average $\Delta\delta_{(\text{open}-\text{closed})}$ values for both ^{1}H and ^{15}N are significantly higher in the subset of residues with fitted R_{ex} terms (e.g., 0.72 ppm for ^{15}N) than in those without (e.g., 0.44 ppm for ^{15}N). Additionally, there is a very weak correlation between $\Delta\delta_{(\text{open}-\text{closed})}$ and R_{ex}, giving a correlation coefficient of 0.281 if D37 (an obvious outlier) is excluded. A weak correlation is expected as the $\Delta\delta_{(\text{open}-\text{closed})}$ figures will also include contributions from ligand binding. Finally, it is noted that the residues which display the largest changes in ϕ & ψ angles between open and closed structures are not necessarily those which are exchange broadened in solution. The conclusion from this is that a transiently closed substate of apo βPGM in solution cannot be ruled out, but it is insufficient to explain all the exchange broadening seen by NMR.

As the presence of a small population of closed enzyme could not be ruled out, the dependence of the model free results on the input crystal data was tested by fitting the

same relaxation data using the βPGM-MgF_3^--βG6P TSA complex coordinates. The results from the ellipsoid fully anisotropic diffusion model were strikingly similar to those acquired from the use of the apo coordinates, with the τ_m and D_{ratio} coming out respectively as 14.7 ns and 1.49. This suggests that the diffusion model is not strongly biased by the crystal structure. Secondly, the fitted S^2 values were surprisingly similar, with an average difference of 1.7%, which is less than the 2.8% uncertainty on the values fit using the apo crystal structure. There was no systematic bias in these differences: the S^2 value for some residues increased when fit to the βPGM-MgF_3^--βG6P TSA complex coordinates, for others it decreased. Lastly, in the parameters which have been previously shown to depend strongly on the diffusion model — R_{ex} terms and ns internal correlation times — again the results between the two data sets were very similar.

6.2.5 Analysis of S^2 Values

Roughly agreeing with the estimated average value of S^2 from the T_1 vs. T_2 plots shown in Figures 85 & 86 on Page 126, the average S^2 values is 0.830. The distribution of values shown in Figure 93 shows that despite the widespread exchange broadening in apo βPGM spectra, most residues are relatively static on the ns–ps timescale, with S^2 values of 0.7–0.9. Despite this small range of fit values, it is seen that most residues have S^2 values which are outside of error of their neighbours indicating that fast timescale dynamics in apo βPGM are highly heterogeneous. Interestingly, in most cases the residues which are adjacent to those which are exchange broadening and/or absent from the assignment do not show significantly more flexibility than the rest of the protein, which is perhaps surprising given the expectation for regions with large ms–μs dynamics to also be dynamic on faster timescales.

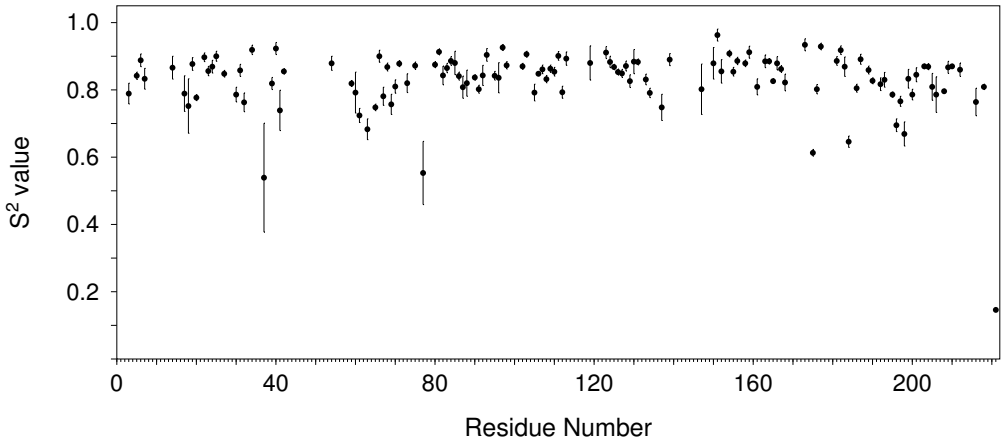

Figure 93: S^2 value plotted against Residue Number for apo βPGM, showing that most residues have values in the range of 0.7–0.9, displaying relatively stiff ns–ps timescale dynamics.

Given the variation within a relatively narrow range of S^2 values, it is difficult to assess regions that are particularly flexible. The two obvious exceptions to this are

the regions around residue 63 and residue 198. Both of these regions correspond to loops between secondary structural elements which are exposed to solvent, and as such their flexibility is not remarkable. Figure 93 also shows the presence of isolated flexible residues. Of these, the C-terminal residues K221 is unremarkable in its flexibility. The other outlying residues are D37, N77, I175, and L184. As noted earlier, D37 has the largest R_{ex} term fitted by Relax. As such it yielded relatively large uncertainties on the fitted parameters, giving an S^2 value of 0.54 ± 0.16, the upper boundary of which is in line with much of the rest of the protein. This residue may be genuinely flexible on the ns–ps timescale, but it seems probable that the quality of the data fitted to it is poor. N77 is interesting as it is in the part of the long helix which ends in the outer hinge region which also shows ms–μs dynamics, as mentioned earlier. Although the S^2 value is fairly poorly defined compared to the rest of the data, clearly it is more flexible than its surrounding residues. The functional significance of this is not clear. Lastly, I175 and L184 (the very well-defined outliers) both have hydrophobic side-chains which contact the flexible regions around residue 195 and the C-terminus respectively. Given their proximity to known flexible regions, the only point of interest here is that their neighbours are not flexible.

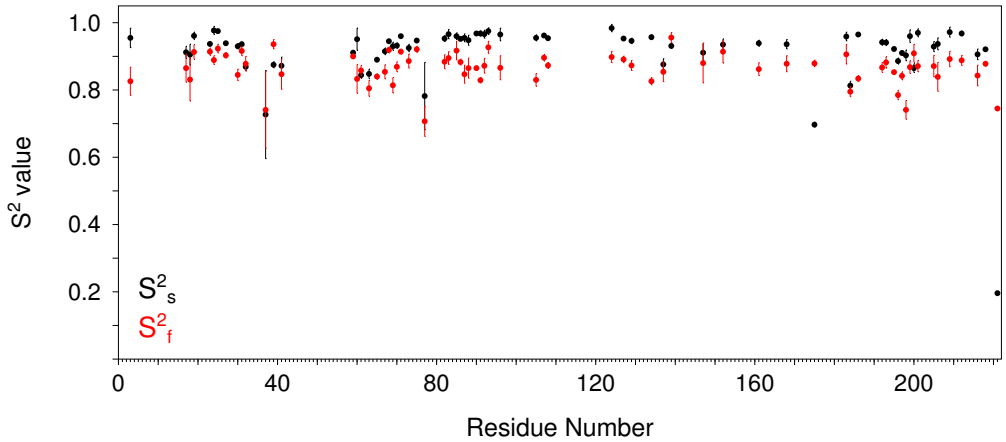

Figure 94: S^2_s (black) and S^2_f (red) values plotted against Residue Number for apo βPGM, showing that most residues have faster motions which are larger in amplitude than the slower motions.

Given that most residues infact have two S^2 values (S^2_s and S^2_f) fit, it is worth comparing the two sets of values, as shown in Figure 94. For the vast majority of residues, $S^2_s > S^2_f$, showing that slower motions tend to be more restricted in amplitude. Due to the higher average values of S^2_s, there is greater variation in this value, and the S^2_s values give a better impression of the regions of flexibility in apo βPGM: the C-terminal residue K221, the far ends of the cap domain residues 32–39 and 61–65, and also the region of residues 195–198. The S^2_f values are also lower around these regions, but the largest changes are in S^2_s. Interestingly, the uncertainties on the S^2_s values are lower which suggests that the spectral density in the high frequency regions where the fast internal correlation times lie is inadequately sampled, perhaps due to

only one measurement of the hetNOE values.

6.2.6 Analysis of τ_e, τ_s, and τ_f Values

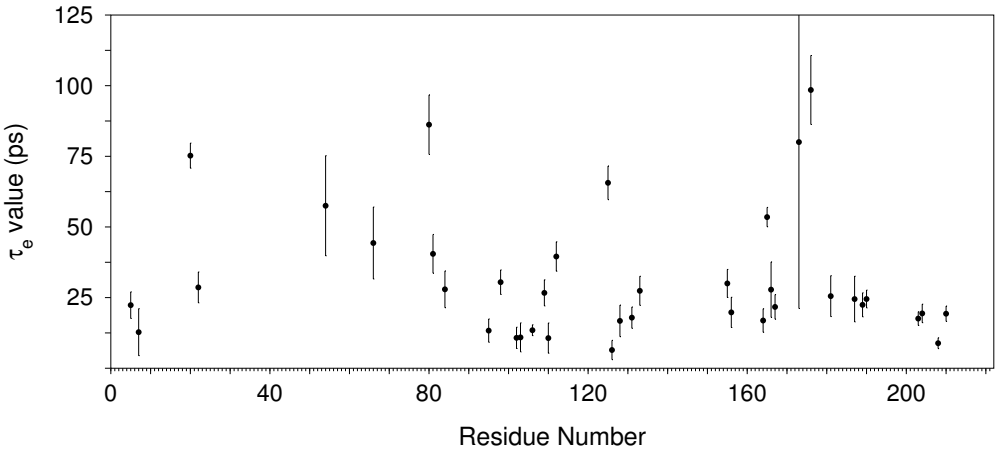

Figure 95: τ_e value plotted against Residue Number for apo βPGM, showing that most residues have values of below 100 ps, and that despite some very poorly defined values, generally the quality of the data is much higher than the average error of 33% would imply. Any residues whose uncertainty on the τ_e value was larger than the value itself have not been shown, nor have two very large values of τ_e which were very poorly determined.

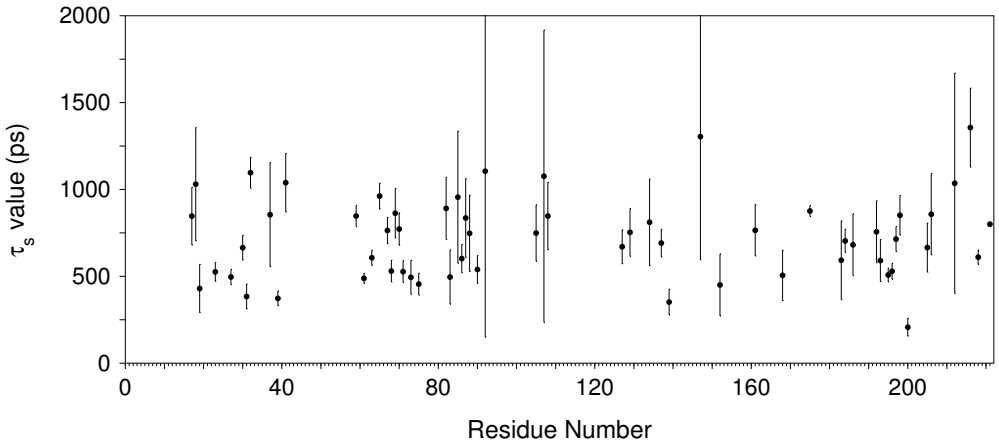

Figure 96: τ_s value plotted against Residue Number for apo βPGM, showing that all values are below 1500 ps, except for some very poorly determined large values, and that from residue N77, which has a τ_s value of 8 ± 4 ns. Any residues whose uncertainty on the τ_s value was larger than the value itself have not been shown.

The two plots of τ_e and τ_s as a function of residue number shown in Figures 95 & 96 show that there is little evidence for adjacent residues which have ns–ps motions induced by the same dynamic mode; few adjacent values of τ_e or τ_s are within error. It is seen that slower internal motion on the ns timescale is not rare, and most residues have τ_s values fit to them, at the expense of τ_e values. Interestingly, unlike in the case of the S^2 plots, the size of the uncertainties on the fit values varies greatly between residues, in a manner which cannot simply be explained by experimental error

on relaxation data. Although there are some extremely poorly determined internal correlation time values, around half of them have uncertainties of below 20%.

It can be seen that the values fitted for τ_e and τ_s do not overlap, the only exceptions being residues with very poorly defined internal correlation times, which are not shown in Figures 95 or 96. Given that only one residue (N77) has both a τ_s and τ_f value fitted, the implication is that residues can either have internal motion on the 1–100 ps timescale, or on the 200–1500 ps timescale, and not both. However it is important to remember that if an internal correlation time is not fit, it is assumed to be 0 ps, i.e., infinitely fast motion. Also, the F-tests used in the fitting of relaxation data by Relax will mean that only parameters which make a statistically significant improvement to the fit of the data will be included.

A more likely explanation of the seemingly mutually exclusive fitting of fast and slow internal correlation times to the data is that the fitting of a τ_s value to a residue prevents the inclusion of a τ_f parameter for the majority of the residues, as the τ_f value needs to be significantly different from the τ_s value, which may be as low as 200 ps, and also significantly different from the assumed value of 0 ps. Given that the average S^2 of 0.79 for those residues which have a τ_s value defined is lower than those residues which only have 1 internal motion defined, 0.86, it is tempting to speculate that many if not all of the residues have internal motions on both timescales, but only those residues which are sufficiently flexible on the 200–2000 ps timescale can define the τ_s value, which then may prevent the fitting of a τ_f value due to the fitting statistics used to avoid over-parameterisation.

6.3 βPGM-MgF$_3^-$-βG6P TSA Complex

6.3.1 Coverage

Unlike in the case of apo βPGM, the only exchange broadening present in the NMR spectra of the βPGM-MgF$_3^-$-βG6P TSA complex is slight, and did not prevent the assignment of all the non-proline resonances. The only exchange broadening which is seen in the NMR spectra results from the 2 different regimes of ms–μs dynamics discussed in Section 4. Only residue I166 has exchange broadening which proved sufficient to prevent the acquisition of sufficiently sensitive relaxation data for model-free analysis. D10 and A115 also had to be excluded as the ^{1}H–^{15}N TROSY peaks for both are split by $^1J_{FH}$ and $^1J_{FN}$ coupling[138], reducing the signal:noise on peak height measurements. Acquisition of the data using TROSY-detection on ^{2}H,^{15}N–labelled βPGM also significantly reduced the linewidths in both ^{1}H and ^{15}N dimensions and thus reduced the spectral overlap, thereby increasing the number of resonances which could be studied. In total, 175 out of the 211 non-proline residues (83%) could be studied which is excellent for a 24 kDa protein, and significantly better than the coverage of apo βPGM dynamics, where only 129 residues (61%) could be studied. The analogous figure to Figure 83 shown on Page 124 for the case of the βPGM-MgF$_3^-$-βG6P TSA complex coverage is shown in Figure 97.

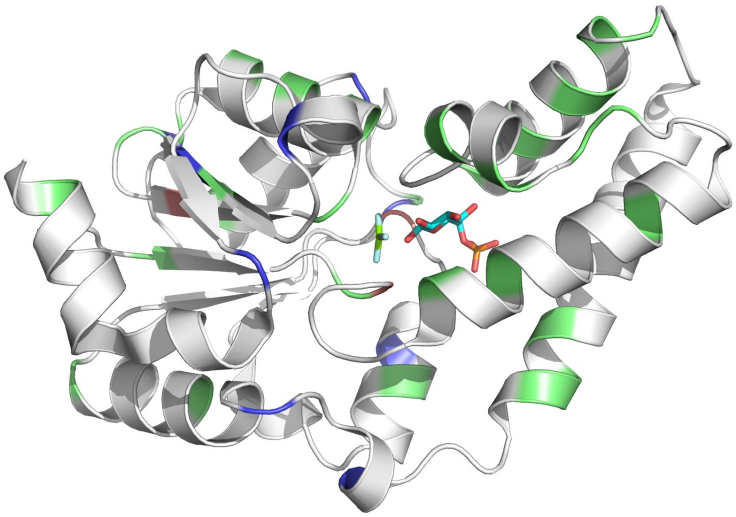

Figure 97: βPGM-MgF_3^--βG6P TSA complex co-ordinates represented as a cartoon of the backbone atoms and the MgF_3^- moiety and βG6P molecule are shown as sticks. Residues which were excluded from relaxation analysis due to exchange broadening or ^{19}F J-coupling are shown in pink. Unassignable (i.e., proline residues and the N-terminal methionine) residues are shown in blue, and those residues whose relaxation could not be studied due to signal overlap are green. It can be seen that the coverage is significantly better than in the apo βPGM case shown in Figure 83 on Page 124.

6.3.2 Relaxation Data

R_1, R_2, and hetNOE measurements were carried out at 600 MHz and 800 MHz as described in Section 2. The average percentage errors for the fitted T_1, T_2, and hetNOE values from each dataset are shown in Table 19. To counteract the lower intrinsic sensitivity of the RT probe fitted on the 800 MHz spectrometer compared to the 600 MHz spectrometer which is fitted with a cryoprobe, the 800 MHz data was carried out with twice the number of scans per increment than the 600 MHz data. The 600 MHz data still proved to have better signal:noise due to the sensitivity benefit of the cryoprobe, and also more stable shim coils. This latter effect is particularly apparent in the T_1 measurements made at 800 MHz, the decays from which are shown in Figure 98. The variation in shimming has led to an increase in random noise, and the measurements done after a relaxation delay of 0.36 s were particularly badly affected and therefore had to be excluded completely from the analysis.

As with the apo βPGM data, these relaxation data were validated using the grid plots and comparing to the expected data as predicted from the crystal structure by HydroNMR. The grid plots showing T_1 vs. T_2 at 800 MHz and 600 MHz are shown

Field Strength	Measurement	Mean Value	Mean Error
600 MHz	T_1	1110 ms	0.9 %
	T_2	62.3 ms	1.1 %
	hetNOE	0.785	2.1 %
800 MHz	T_1	1610 ms	7.5 %
	T_2	50.5 ms	2.9 %
	hetNOE	0.802	0.9 %

Table 19: Mean values and errors for relaxation measurements on βPGM-MgF_3^--βG6P TSA complex.

in Figures 99 & 100 respectively, and the plot showing T_1 vs. hetNOE at 800 MHz is shown in Figure 101. The T_1 vs. T_2 plots indicate that HydroNMR predicts a smaller spread in local τ_m values, as expected as the βPGM-MgF_3^--βG6P TSA complex has a more compact, less anisotropic structure than open apo βPGM. They also indicate that at both fields — even despite the shimming problems with the 800 MHz T_1 data — the spread in recorded T_1 and T_2 values fits exactly into that predicted by HydroNMR. As the relaxation measurements are done with the delay times in random order, problems with shimming are expected to be a source of random, rather than systematic, noise. This is clearly the case.

Additionally, the plots demonstrate that whilst a few residues deviate from the grid due to either R_{ex} terms or high flexibility on the ns–ps timescale, these deviations are smaller than are the case in the apo βPGM data. The T_1 vs. hetNOE plot shows that although the average hetNOE value is higher, the number of residues with impossibly high values is comparable to the apo βPGM data. The quality of the 600 MHz data is apparent when the T_1/T_2 values at each spin site are compared to those predicted from HydroNMR; the agreement between values is much better than for the apo βPGM data recorded at 500 MHz. This is shown in Figure 102. Accordingly, plotting the T_1/T_2 values as spheres representing the N–H bond vectors produces the expected pattern from HydroNMR as shown in Figure 89 on Page 128 for the apo βPGM data, though this is not shown for the βPGM-MgF_3^--βG6P TSA complex data.

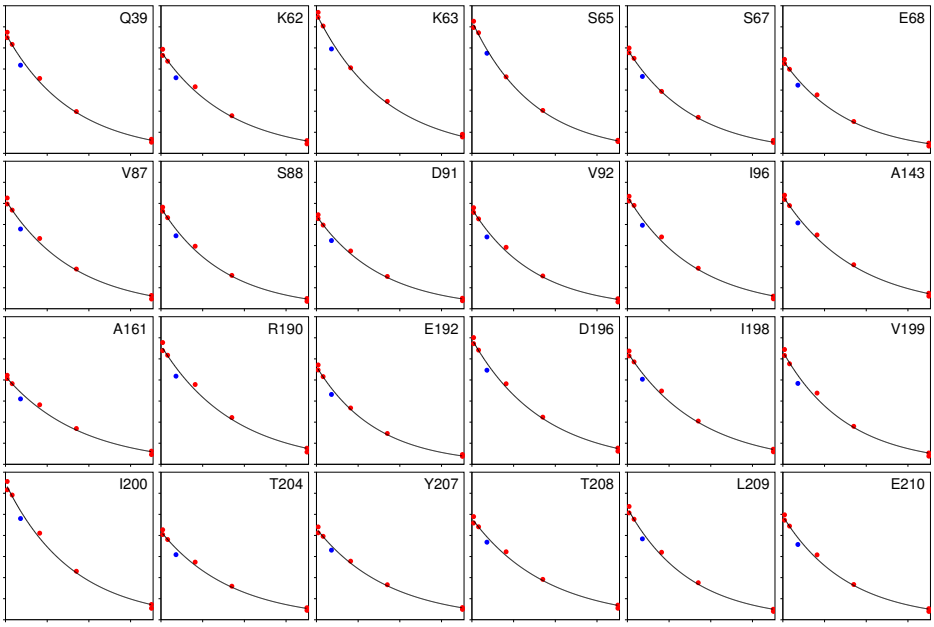

Figure 98: Selected ^{15}N R_1 decay curves recorded at 800 MHz on the βPGM-MgF_3^--βG6P TSA complex. The filled circles are experimental measurements, and the black lines correspond to fitted exponential decays for each residue. It can be seen that the drift in shimming has resulted in another source of random noise on the data, which is particularly obvious in the peak height measurements after a relaxation delay of 0.36 s, shown in blue: all these points lie significantly below the fit line due to B_0 field inhomogeneity causing an increase in peak line widths.

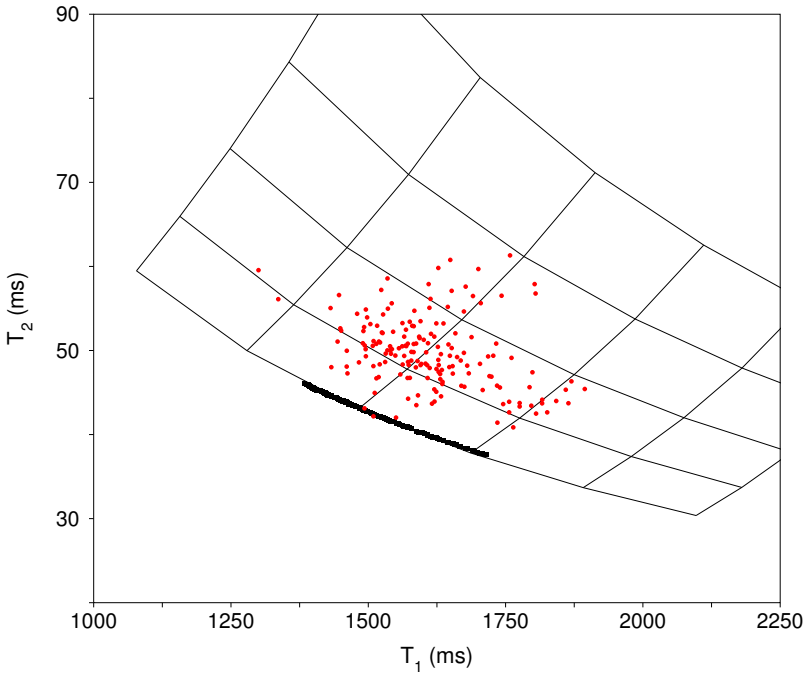

Figure 99: T_1 vs. T_2 plot for βPGM-MgF_3^--βG6P TSA complex data recorded at 800 MHz (red circles), showing comparison of data predicted from HydroNMR (black squares).

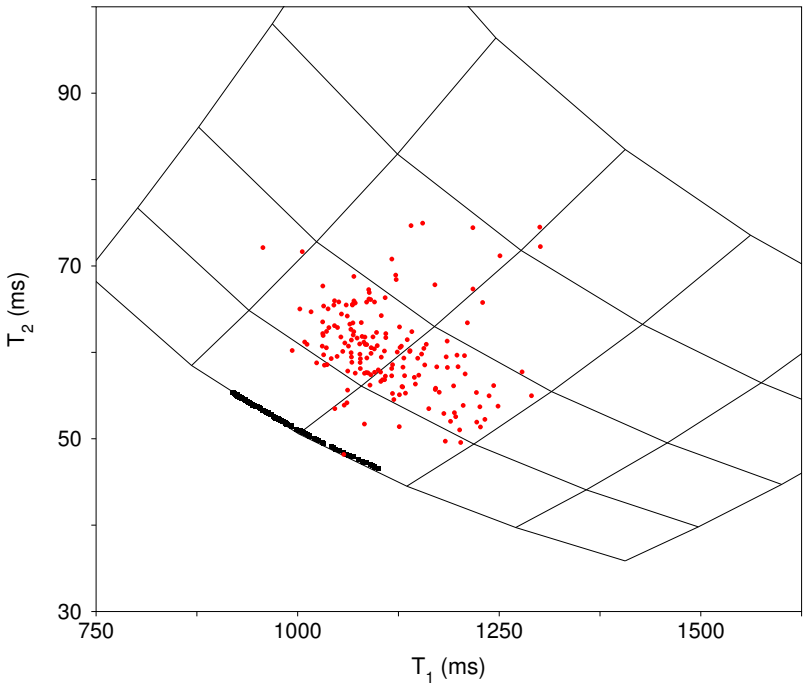

Figure 100: T_1 vs. T_2 plot for βPGM-MgF_3^--βG6P TSA complex data recorded at 600 MHz (red circles), showing comparison of data predicted from HydroNMR (black squares).

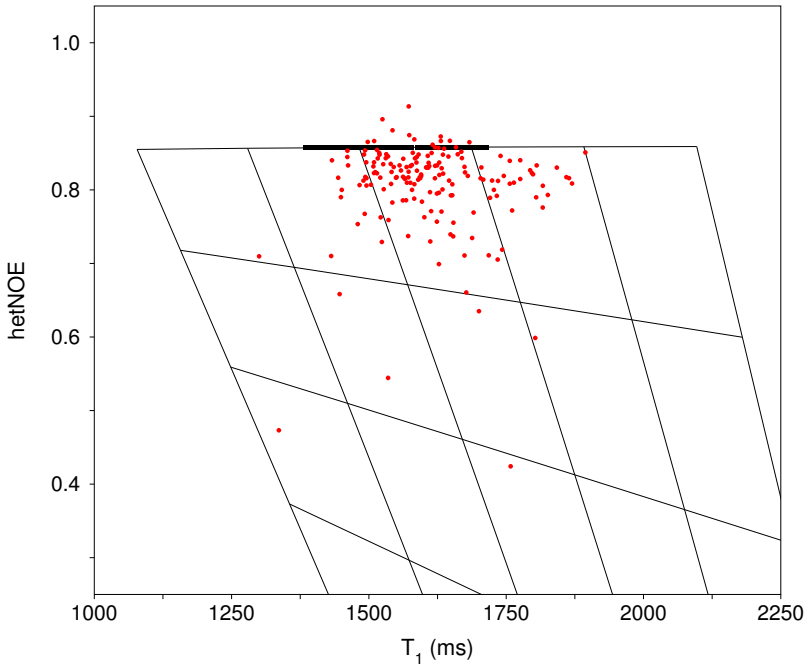

Figure 101: T_1 vs. hetNOE plot for βPGM-MgF_3^--βG6P TSA complex data recorded at 800 MHz (red circles), showing comparison of data predicted from HydroNMR (black squares).

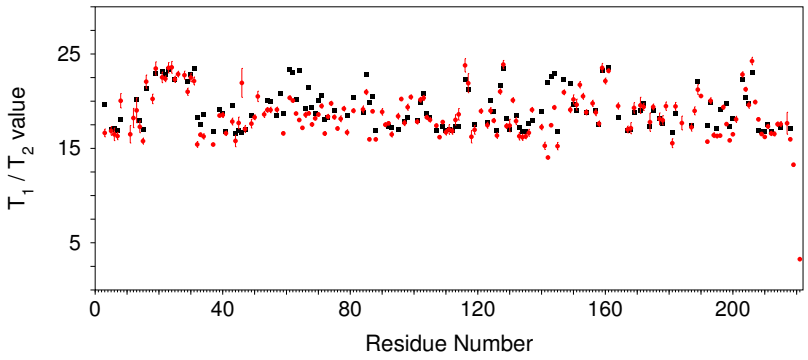

Figure 102: T_1/T_2 value plotted against Residue Number for βPGM-MgF_3^--βG6P TSA complex data recorded at 600 MHz, showing broad agreement of predicted anisotropy with experimental results.

6.3.3 βPGM-MgF_3^--βG6P TSA complex Model Free Fitting and Validation

Similar methods to those used for validation of the fitting results from the apo βPGM data were applied to the βPGM-MgF_3^--βG6P TSA complex data. Correlation coefficients between input T_1, T_2, and hetNOE values and those back-calculated from the fitted model free parameters were found to generally be higher in this case. Unsurprisingly, the T_1 values measured at 800 MHz score the lowest, with $r = 0.961$; all other values are above 0.970 indicating that the relaxation values supplied into the fitting procedure are extremely well accommodated by the fitted parameters. Similarly

to the apo βPGM data, of the global diffusion models, the fully anisotropic ellipsoid diffusion model was found to satisfy the data best.

Diffusion Model	**Parameters**	**Data Points**	χ^2	**AIC**
Local τ_m	578	1050	377.9	1533.9
Sphere	529	1050	1153.8	2211.8
Prolate Spheroid	530	1050	699.4	1759.4
Oblate Spheroid	528	1050	1054.3	2210.3
Ellipsoid	518	1050	635.0	1671.0

Table 20: Results of fitting βPGM-MgF_3^--βG6P TSA complex data to different diffusion models. The models (in order of their ability to satisfy the data) are: local τ_m, ellipsoid, prolate spheroid, oblate spheroid, and sphere.

As Table 20 demonstrates, the local τ_m model, i.e., that without any global diffusion parameters, actually performs the best, although the ellipsoid model is chosen by Relax as being the best model. Relax must therefore have another means by which to decide if it is reasonable to model global diffusion other than by relying on the AIC values. Figure 103 shows a histogram for all the values of local τ_m which are fit by Relax, and it clearly shows that — with the exception of the very flexible K221 residue — all local τ_m values are approximately normally distributed around a centre of around 13.5 ns, which demonstrates that the diffusion parameters at each spin site can be explained adequately by a global diffusion model.

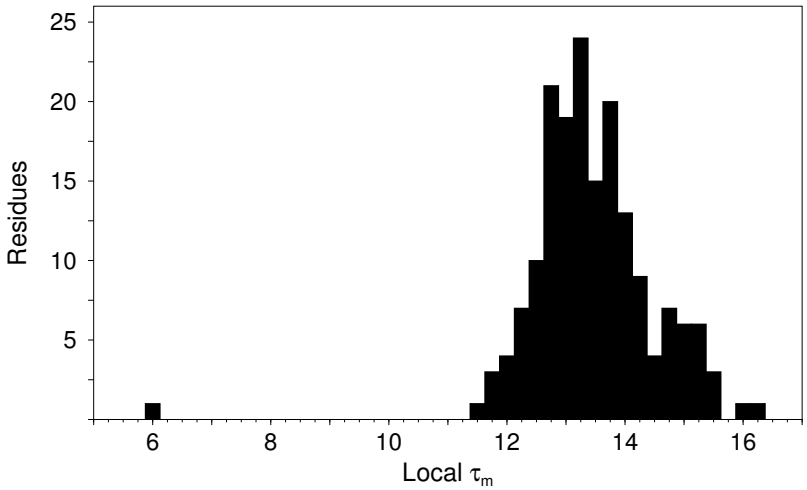

Figure 103: Distribution of local τ_m values fit to βPGM-MgF_3^--βG6P TSA complex relaxation data, demonstrating that the diffusion of the βPGM-MgF_3^--βG6P TSA complex can be satisfied by a global diffusion model.

The diffusion parameters resulting from the ellipsoid diffusion model give a global τ_m of 13.1 ns, which closely matches the average of the local τ_m values shown in Figure 103, and the evaluated average τ_m value from the T_1 vs. T_2 plots shown in Figures 99 & 100. The individual times for a full rotation along the three axes are 62.9, 85.4, and 95.9 ns, which shows that the differences between the two shorter sides of the βPGM-MgF_3^--βG6P TSA complex are more significant than in the apo βPGM

case, explaining why the AIC values for the prolate spheroid and ellipsoid models are more different in the case of the βPGM-MgF_3^--βG6P TSA complex. The D_{ratio} is 1.44, which almost exactly matches that predicted by HydroNMR, which was 1.43. The agreement between the experimental and predicted D_{ratio} values is reflected in plotting the diffusion tensors against the crystal structure as is shown in Figure 104. The N–H vector dependence of R_{ex} components or τ_e motions on the ns timescale which are present only in the isotropic diffusion model mirrors that shown for the apo βPGM data in Figure 91. However, because the prolate spheroid and the ellipsoid model have produced such different results, the differences between these models are not only demonstrated by residues which point along the relevant axes of the molecule, but are more global.

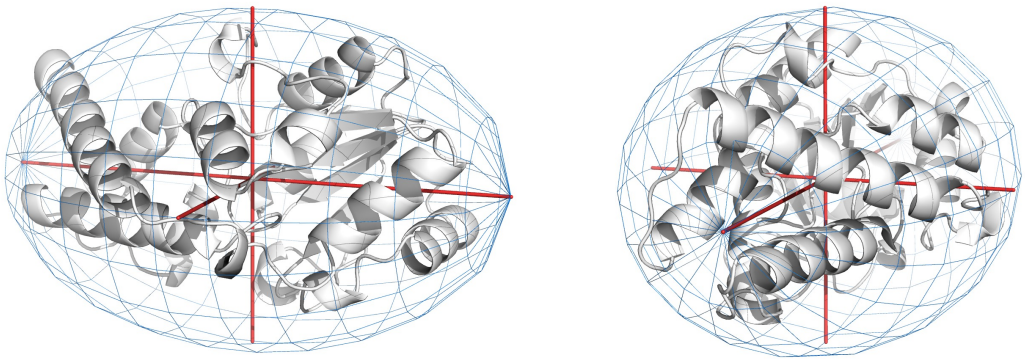

Figure 104: Anisotropic diffusion tensors from βPGM-MgF_3^--βG6P TSA complex relaxation data (red) and their resulting ellipsoid (blue), plotted onto the crystal structure. The two images are rotated about a 90° angle from each other. It is seen that the ellipsoid agrees very well with the crystal structure.

This data set has allowed the fitting of two separate internal correlation times for 18 residues, as compared to just one residue in the apo βPGM case. This is likely due to the inclusion of two measurements of the hetNOE value: being particularly sensitive to the spectral density at $J(\omega_H \pm \omega_N)$, the hetNOE measurement contains more information about faster timescale motions than T_1 or T_2 relaxation rates. The variation in hetNOE values between the two field strengths indicates that there is significant variation in spectral density even in the high-frequency $J(\omega_H \pm \omega_N)$ region. For example, K221 gives a hetNOE value of 0.056 at 600 MHz but 0.271 at 800 MHz.

Despite there being less evidence of R_{ex} terms from the T_1 vs. T_2 plots of the βPGM-MgF_3^--βG6P TSA complex data, 110 out of 175 residues have R_{ex} terms fitted by relax. In the case of the apo βPGM data, it seemed like the high number of R_{ex} terms fitted was merely a reflection of the sensitivity of the R_2 measurements. This being true, more residues with fitted R_{ex} terms in the βPGM-MgF_3^--βG6P TSA complex case is not surprising. As with the apo βPGM data, the largest value of R_{ex} — even at 800 MHz — is only small at 0.18 s^{-1}. This value of R_{ex} belongs to G46, which has been characterised by ^{15}N relaxation dispersion as having the largest R_{ex} value, i.e., it displays the largest $\Delta\delta$ value amongst the residues which are involved in the

major⇌minor dynamics (Section 4). The next largest R_{ex} value belongs to L217, which is known to have the largest $\Delta\delta$ of the residues involved in the "partial unfolding" dynamics (Section 4). Both of these observations agree with previous data and it is therefore believed that the fitted R_{ex} terms are likely real, and not an indication of a poor global diffusion model.

As is to be expected from more sensitive data, the error values on the fitted model free parameters compare favourably to the apo βPGM data. $\overline{\sigma S^2} = 1.5\%$, $\overline{\sigma\tau_e} = 20\%$, $\overline{\sigma\tau_s} = 41\%$, $\overline{\sigma\tau_f} = 18\%$, and finally, $\overline{\sigma R_{ex}} = 29\%$. As before, these parameters are discussed in turn, except for the R_{ex} components which have been characterised by more sensitive experiments and discussed in Section 4. The only comment concerns the residue I166 which had to be removed prior to analysis. Though having a relatively large $\Delta\delta$ for the residues involved in the "partial unfolding" dynamics, L217 and indeed G46 are expected to be worse. Why I166 yielded the worst data is not known, although it is possible that the relatively large R_{ex} term exacerbated what was already a short intrinsic R_2 due to a lack of flexibility on the ns–ps timescale and an N–H bond vector which points along the longer side of the molecule, though from the crystal structure the latter assertion does not seem to be true.

6.3.4 Analysis of S^2 Values

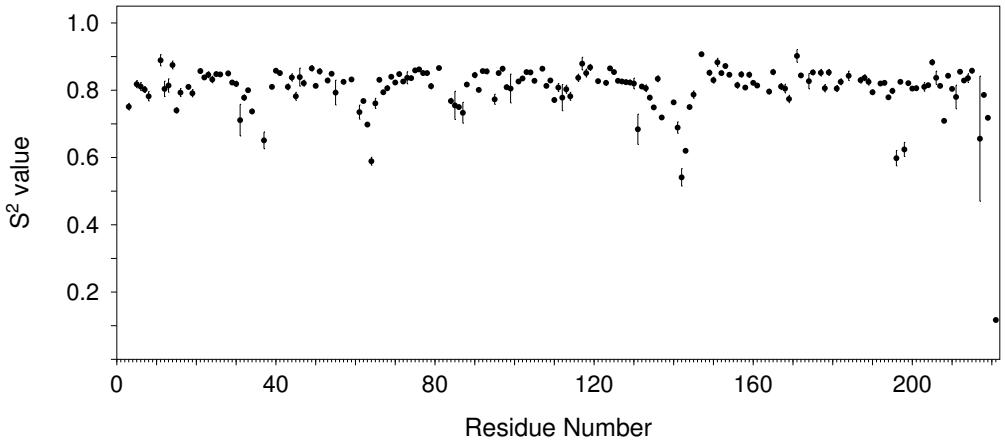

Figure 105: S^2 value plotted against Residue Number for the βPGM-MgF_3^--βG6P TSA complex, showing that most residues have values in the range of 0.7–0.9, as in the apo βPGM data. However, the general spread of S^2 values is lower and so flexible regions such as that between residues D133 and K145 can be more easily identified.

The plot of S^2 values against residue number shown in Figure 105 shows the improved quality of the data acquired on βPGM-MgF_3^--βG6P TSA complex when compared to apo βPGM data, shown in Figure 93 on Page 134. Most residues are relatively stiff on the ns–ps timescale, with the average S^2 values lying at 0.805, although this is slightly lower than that measured for apo βPGM. Although there are no residues which are particularly flexible on the ns–ps timescale, in general the distribution of S^2 values is much narrower than in the apo case, making it much easier to identify regions

which are less immobilised than the rest of the protein.

Mirroring the case for apo βPGM, the loops at the end of the cap domain, as well as the C-terminal residues, are relatively flexible on the ns–ps timescale. Interestingly, the most pronounced group of residues which are flexible is the loop which runs from residues D133 to K145 (K145 providing one of the positive charges in the active site). Other regions which show slight decreases in S^2 values compared to the rest of the protein are the residues close to the inner hinge region running from residues D15–Y19 (though these residues only show a very slight reduction in S^2 value), the outer hinge region running from residues I84–V87, and another active site loop region I110–S114. In addition, the same two residues with side-chain contacts to the C-terminus which show flexibility in the apo enzyme (D196 and I198) are also flexible in the TSA. Intriguingly, N77 which was flexible with an extraordinarily long internal correlation time in the apo data does not show any pronounced flexibility in the TSA data.

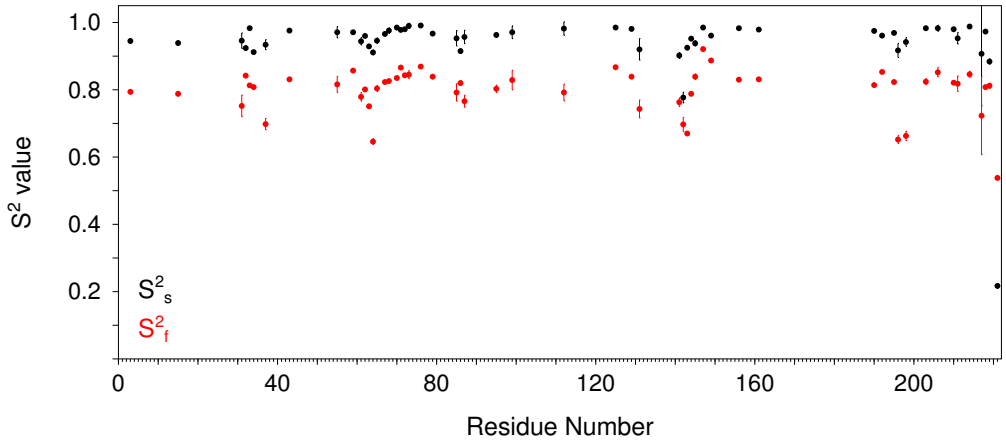

Figure 106: S^2_s and S^2_f values plotted against Residue Number for the βPGM-MgF_3^--βG6P TSA complex, showing that the two order parameters correlate strongly, and that as in the apo βPGM case, the S^2_f values are almost always lower.

The comparison between S^2_s and S^2_f in Figure 106 looks different to the apo case, whereby the order parameters are more strongly correlated, yielding a correlation coefficient of 0.667 (compared to 0.497 for the apo βPGM data). Any regions which are more flexible according to one order parameter are mirrored in the value of the other order parameter. This is interesting as it implies that any strong interactions with surrounding residues inhibit dynamics on both ps and ns timescales. This is implied in the apo βPGM data but is very clearly displayed in the more accurately determined values from the βPGM-MgF_3^--βG6P TSA complex data.

Further to the observation that the residues which surround those which are unassigned in the apo βPGM spectra do not have particularly low S^2 values, the S^2 values from the TSA complex were compared to the $\Delta\delta$ figures relating to the major$\rightleftharpoons$minor

exchange on the ms–μs timescale to study the interplay between dynamics on these very separate timescales. S^2 and $\Delta\delta$ figures give a correlation coefficient of 0.099, meaning they are not at all correlated. This is not surprising, as a more quantitative approach would point out that although many of the active site loops are implicated in the ms–μs exchange as well as showing slightly lower S^2 values, many of the regions distal to the active site which have lowered S^2 values — such as the outer hinge and the loops at the end of the cap — are not involved significantly in the ms–μs timescale dynamics. This again supports the impression from the apo βPGM data that dynamics on the different timescales are not coupled. Although the assertion in much of the literature that ms–μs fluctuations are the coincidental collective motions on faster timescales does not necessarily require that the very same residues are involved in dynamics on both timescales, it might be naively expected that some correlation would be seen.

6.3.5 Analysis of τ_e, τ_s, and τ_f Values

The fit values of the internal correlation times τ_e and, in the case where 2 internal correlation times are fitted, τ_f occupy similar ranges in values and are therefore plotted together in Figure 107. In general the values are smaller than in the apo βPGM case, meaning that although the amplitudes of the motion (determined by S^2) are similar in both states of βPGM, the motions themselves occur faster in the TSA complex. Interestingly, despite the much better quality data in the βPGM-MgF_3^--βG6P TSA complex case, the fraction of residues which have two order parameters fit is actually lower: 31% as compared to 55% in the apo βPGM case. Although the S^2 and τ_s values are correlated still ($r = 0.446$, if K221 and residues with $\tau_s > 3$ ns are excluded), this observation goes against the conclusion from the apo data that there are likely picosecond and nanosecond timescale motions occurring in every residue, and it was merely the fitting statistics that determined whether or not they were defined. As in the apo βPGM case, the plot of the τ_s values with the residue number for the βPGM-MgF_3^--βG6P TSA complex data as shown in Figure 108 is fairly unremarkable.

Given the much lower fitting uncertainties for the faster timescale motions in the βPGM-MgF_3^--βG6P TSA complex, it is possible to critically assess the principal assumption that is made during the conformational entropy calculations: that all ns–ps timescale motions are independent of each other. This can be done by observing if adjacent residues have similar internal correlation times, implying that it is the same motion that is causing fluctuations in the N–H bond vector for multiple residues.

This is clearly true for several groups of residues. The two residues which make up a tight turn in the inner hinge, V12 and I13, have almost identical S^2 values — 0.80 ± 0.02 and 0.81 ± 0.02 respectively — and similarly indistinguishable τ_e correlation times: 9 ± 2 and 8 ± 2 ps respectively. It is therefore very likely that the same type of motion is responsible for the internal dynamics felt by both ^{15}N–$^{1}H^N$ spin pairs. Another active site loop running from L44–V47 all have τ_e values within error, and also within error of the values for V12 and I13 which are on the adjacent loop. Likewise the region

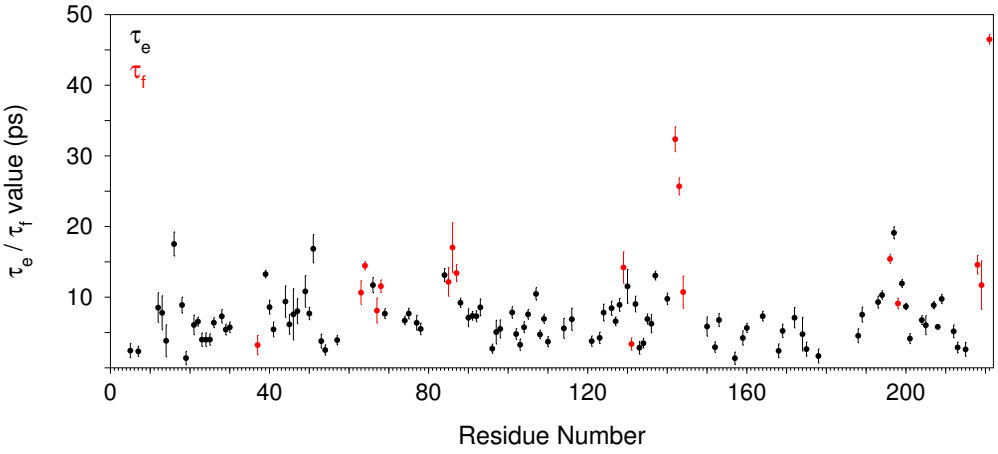

Figure 107: τ_e and τ_f values plotted against Residue Number for the βPGM-MgF_3^--βG6P TSA complex, showing that with the exception of the particularly long τ_f values, all τ_e and τ_f values are below 25 ps, and the vast majority below 10 ps, occupying a small range in values. It can be seen that there are blocks of residues with similar τ_e or τ_f values.

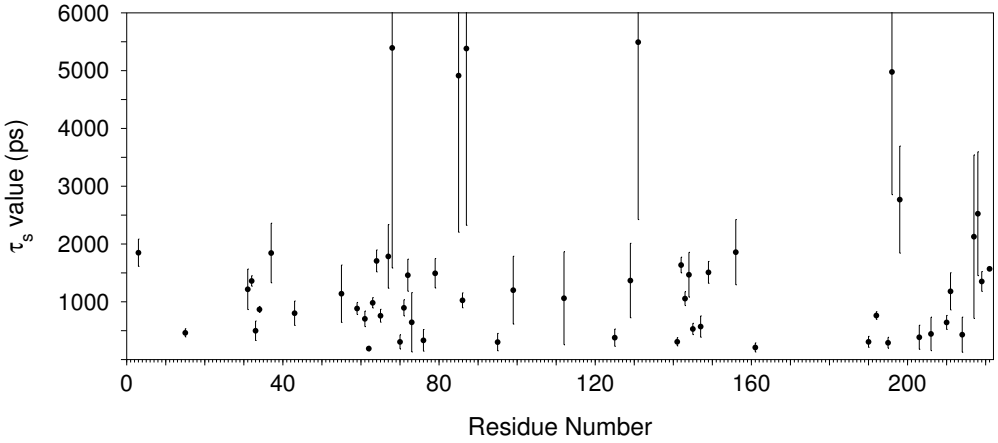

Figure 108: τ_s value plotted against Residue Number for the βPGM-MgF_3^--βG6P TSA complex, showing that the vast majority of residues fit τ_s as below 2500 ps, although there are some very poorly determined values in the range of 5000 ps.

immediately after the outer hinge (residues A90–Y93) which are also in contact the loop on which V12 and I13 lie also have τ_e values which match those for both regions. S114 and S116 which are on another loop in contact with many of these residues also show similar τ_e values: 6 ± 1 and 7 ± 2 ps respectively. Residues E169 and Q172 on yet another active site loop show similar, though not perfectly matching, τ_e values of 5 ± 1 and 7 ± 1 ps.

The group of residues in the long helix of the cap domain which display complicated behaviour in the apo βPGM case (R75–D78) appear to also have some interdependent motion, with τ_e values of 6 ± 1, 7 ± 1, 8 ± 1, and 6 ± 1 ps. Several residues from the spatially adjacent α-helix have τ_e values which also correspond to this value: F21,

R22, A26 and A28. Although there are outliers, many of the residues on both of these helices have either a τ_e or τ_f value defined within this range. Strikingly, these τ_e values also match those of the residues occupying active site loops as discussed in the above paragraph. Given that these α-helices — and therefore the N–H bond vectors — essentially run along the axis of the cap domain, this may suggest that the orientation of the cap domain with respect to the core domain changes on the ps timescale. This would almost certainly affect the active site loops as a consequence.

The outer hinge residues I84–V87 also all have τ_e or τ_f values which are within error of each other, though faster than the groups of residues discussed above. Although D86 has a completely different value for τ_s, it is striking that both Q85 and V87 have extraordinarily long and shared τ_s values of 5 ± 3 ns.

As a demonstration of contrast to much of the above, the region of residues from K101–I110 shows the kind of completely unrelated behaviour which was expected. Although many of the residues which have not been discussed are not as obviously independent as this region of residues, in general there are few groups of residues adjacent in the sequence with very similar τ_e or τ_f values as shown in Figure 107. This reinforces the significance of the observation that in the cases where there are two or more residues with indistinguishable internal correlation times, they are also indistinguishable from the correlation times of other groups of similarly clustered residues. There is evidence here that collective motions on the ns–ps timescale do occur, which would need to be taken into account during any subsequent conformational entropy analysis.

7 Discussion and Conclusions

7.1 βPGM Mutants

7.1.1 D15 and H20 Mutants

None of the βPGM mutants produced on the basis of the two near-attack complexes adopted by the WT βPGM-BeF_3^--βG6P GSA complex significantly altered the population of the "minor" conformer in the βPGM-MgF_3^--βG6P TSA complex. Despite this, ^{19}F NMR data demonstrates that the NAC I conformer of the βPGM-BeF_3^--βG6P GSA complex of D15A βPGM is destabilised, and that the mutation was therefore successful in removing the N-capping interaction which is only present in NAC I. This observation shows that the "minor" conformation of the βPGM-MgF_3^--βG6P TSA complex significantly differs from NAC I in the βPGM-BeF_3^--βG6P GSA complex. Given that this was the assumption which underpinned the choice of mutations made to βPGM, it is not surprising that this strategy was unsuccessful.

The D15A mutant may be useful in the future to study the BeF_3^- complexes of βPGM however: given that NAC I has been destabilised, it may provide a method for observing NAC II in solution. If the ratio of Mg^{2+}:Be^{2+} ions was decreased, it may be possible to populate this complex. Also, the fact that the mutant forms the βPGM-BeF_3^--βG1P complex (albeit at low concentrations) demonstrates that this mutant is catalytically active. No βG1P is added to the NMR samples, its presence is merely the result of incomplete inhibition of βPGM by BeF_3^-. This suggests that despite being one of the residues whose dihedral angles change the most between open and closed forms of βPGM (see Section 6.2.4), the sidechain of D15 is not important for catalysis.

The mutations of H20 (H20N and H20A) also failed to significantly perturb the major:minor conformer ratio in their respective βPGM-MgF_3^--βG6P TSA complexes. The similarity of the ^{19}F δs also implies that any interaction between D10 and H20 is likely to be very weak. Although it is suspected from the D15A mutant that the mutation strategy was flawed, D10 adopting its "out" configuration in the minor conformer of WT βPGM cannot be ruled out on the basis of the data collected on any of the mutants. Both mutants of H20 show greatly reduced saturation by substrate as compared to WT βPGM, which demonstrates a reduction in substrate affinity. This agrees with the apparent role of the H20 and Y80 side chains in "pushing" the substrate into the active site, as envisaged by the NAC complexes of the WT βPGM-BeF_3^--βG6P GSA complex[228].

The exchange broadening of the ^{19}F resonances in the H20 βPGM-MgF_3^--βG6P TSA complexes implies increased solvent exposure to the MgF_3^- moiety in the active site, and also that the F^- ions are able to leave the MgF_3^- moiety independently, with F_a being exchange broadened to a lesser extent than F_b or F_c. The observation that the minor conformer ^{19}F resonances do not suffer from this exchange broadening is intriguing, as it suggests that the minor conformer gives more protection to the MgF_3^-

TSA.

Finally, it is notable that — active site solvent exposure aside — the TSA complexes of the H20N and H20A mutants do not significantly differ from one another. This is despite the fact that the H20N mutation was designed to retain a hydrogen bond to the side chain amine group of K76, whereas the H20A mutant was designed to remove this interaction as well as any to the side-chain of D10. The implication here is that either the bond between H20 and K76 in WT βPGM is insignificant, or neither of the mutant complexes features a hydrogen bond between the side chains of N/A20 and K76, i.e., that the asparagine side-chain is actually too short to properly contact K76. Given that K76A mutants show reduced turnover rates despite K76 being well outside of the active site[229], the existence of an interaction between H20 and K76 seems likely. Instead it seems that H20N is too severe a mutation to accommodate this interaction. This is supported by kinetic measurements which show that H20Q βPGM is a much better functioning enzyme than H20N βPGM[229].

7.1.2 K219Δ Mutant

The K219Δ βPGM-MgF_3^--βG6P TSA complex gives identical ^{19}F spectra to that of the WT enzyme, showing that removal of the three C-terminal residues has had no effect on the major⇌minor equilibrium, and therefore that any dynamics around the C-terminal region are uncoupled from the slower dynamics around the active site. Instead, and in accordance with the hypothesis that the faster dynamics around the C-terminus may correspond to a transient local unfolding event, the mutant does display vastly reduced folding stability compared to the WT enzyme. The WT enzyme has a free energy of folding (ΔG_{fold}) value of 62 ± 1 kJ/mol, whereas the K219Δ mutant shows a loss of nearly half of the free energy of folding, with $\Delta G_{fold} = 32 \pm 1$ kJ/mol.

The scale of the differences between the WT and K219Δ mutant is surprising. It is noted that there does seem to be some small inaccuracy with the concentration of GuHCl used in the samples which correspond to the start of the unfolding transition of the K219Δ mutant, as is evidenced in Figure 80 on Page 121. However, the error estimates used during the fitting are reasonable and do deal with this apparent inconsistency between adjacent measurements. At best the differences in calculated ΔG_{fold} values between the WT and mutant βPGM may be very slightly overestimated.

The CD results unambiguously show that the K219Δ mutant has no reduction in the amount of secondary structure compared to the WT protein, whilst the ^{1}H NMR shows that its tertiary structure has changed. Given that non-native interactions have been shown to exist in folding intermediates[237], it could be suggested that the dynamics at the C-terminus involves the interchange between two equally-folded species, but that the lower-populated state is an intermediate in the unfolding of the enzyme. This idea could potentially explain the drastic loss of folding stability by the mutant, if it is that the equilibrium between the two conformers seen by ^{15}N relaxation dispersion has been shifted.

The population of "minor" state (defined by K221 only) is relatively high at 5%, and so only 15 kJ/mol would be required to completely reverse the equilibrium such that the mutant was 95% "minor" state. Given the lack of apparent interactions between the last 3 residues and any others, the expected shift in equilibrium would be far less than this, and even supposing that the equilibria before and after the mutation were as skewed as going from 1% to 99% "minor" conformation, the difference in free energy between the two conformers would only be 23 kJ/mol. Even when factoring in the loss of entropy due to removal of the flexible C-terminus, it is difficult to see how the loss in ΔG_{fold} would be as great as 30 kJ/mol.

The most intriguing difference between the two proteins is the differences between the m values. Although the interpretation of this value is ambiguous, it has been shown to correlate to some extent with the degree of burial of hydrophobic side chains in a protein[215]. The implication of this interpretation is that the removal of 3 predominantly disordered residues from the protein has exposed a third of the hydrophobic side chains in βPGM. Given that the change in structure — judging by both ^{1}H NMR and CD — is relatively small, it is impossible to conceive how this could be. Going further, the ^{19}F NMR spectra of the K219Δ βPGM-MgF_3^--βG6P TSA complex show that the structures of WT and mutant enzymes are absolutely identical around the active site. In conclusion, the titration which shows that the mutant is approximately half as stable as WT βPGM is difficult to square with the small structural changes displayed by CD and NMR, as well as with any expected perturbation in equilibrium between conformational substates due to the interactions destroyed by the mutation.

7.2 Fast Timescale Dynamics in βPGM

7.2.1 The Nature of Disorder in βPGM

Although the characterisation of fast timescale dynamics in βPGM is not yet complete — with further ^{15}N data to be collected on the βPGM-BeF_3^--βG6P GSA complex complex and analysis of the ^{2}H relaxation to be completed — it is still possible to make some comments from the data analysed and presented in Section 6. Firstly, and most apparently, with the exception of a few loop regions βPGM is fairly well ordered on the ns–ps timescale both as the apo enzyme and the MgF_3^- TSA complex, with the vast majority of S^2 values in the range 0.7–0.9. This is mirrored by several similar studies, and shows that although changes in conformational entropy between unligated and ligated proteins may be a significant driving force in the energetics of ligand binding, globular proteins are generally biased toward having favourable enthalpy rather than entropy[74]. Stated differently, the reason for the small amplitude of motions of backbone sites in βPGM and other proteins is likely to be that the interactions between backbone amide sites are relatively strong (of high enthalpy) and can therefore not be readily broken by thermal motions on the ns–ps timescale. It is notable that all the regions identified as having lower S^2 values belong to loops with no secondary structure, and therefore weak hydrogen-bonding interactions with other residues. In this context,

enthalpy and entropy are inherently inversely related as any strong interactions which give rise to favourable enthalpy require precise distance and/or angular relationships between interacting partners and therefore necessarily have poor entropy.

To return to the "energy landscape" view of protein dynamics, apo βPGM clearly has a landscape with several distinct "wells" — as displayed by ms–μs dynamics of the active site loops giving rise to exchange broadening in the NMR spectra — as does the βPGM-MgF_3^--βG6P TSA complex complex with its "minor" conformation (discussed later). However, within these wells, only a small number of conformational substates (the result of fluctuations of the ns–ps timescale) exist; the wells are relatively smooth. Given the stability of WT βPGM, and the relatively little conformational entropy as witnessed by ^{15}N relaxation, it seems that the protein has very well defined contacts between residues which give it favourable enthalpy. Stated another way, the wells in which βPGM resides are deep due to strong intramolecular interactions, which necessarily restricts the number of substates which can have similar energy. Stated differently, a deep well (favourable enthalpy) is necessarily smooth and narrow (less favourable entropy).

This idea of a deep energy well for the major conformer of the βPGM-MgF_3^--βG6P TSA complex explains the high $\Delta G^{\ddagger}$ of the transition between the major and minor conformers (discussed in more detail later). If one protein conformer is very stable in a deep energy well, it will take a large input of free energy in order for it to adopt another conformer, as this conformational exchange process likely involves the formation of an unstable conformational transition state. The more stable a protein ground-state conformation, the less likely the protein is to undergo conformational exchange. It is possible, on the other hand, that some conformational exchange processes may be slow not because the ground state is especially stable, but because the transition state is particularly unstable. This "high barrier" model differs from the "deep well" model because in the "high barrier" case, the ground state conformer may not have particularly favourable enthalpy, and may therefore have more favourable entropy as a result. Just as a deep well is necessarily slow and narrow, a shallow well can be wide and rugged.

The nature of any interplay between dynamics on fast and slow timescales has been much discussed, as outlined in Section 1.2.7. A particularly appealing model is that large conformational changes are slow because they are the result of the coincident individual motion of all the residues involved, an unlikely event. This model implies a strong link between dynamics on different timescales, and is much better suited to a "high barrier" model of conformational exchange, whereby the ground states possess significant dynamics on the faster timescale, which can — in the unlikely event that all necessary parts of the protein are moving in the same direction — give rise to larger, slower changes. In the "deep well" case, such fast timescale dynamics are unlikely to be sufficient to allow passage over the barrier.

As βPGM more closely resembles the "deep well" case, this model which suggests linkage between dynamics on different timescales may not be valid. Accordingly, there is no evidence of linkage between dynamics on the ns–ps and ms–μs timescales in βPGM from the relaxation data presented in Section 6. Residues either side of regions which have ms–μs conformational exchange in apo βPGM such as to be broadened beyond detection are no more dynamic on the faster timescale than any of the rest of the protein. And in the βPGM-MgF_3^--βG6P TSA complex data set, despite many of the active site loops which are involved in the major$\rightleftharpoons$minor exchange process showing relative flexibility on the faster timescale, there are many exceptions to this and accordingly the correlation between ^{15}N $\Delta\delta_{(\mathrm{major-minor})}$ and S^2 is very weak.

As a final comment to address the ultimate aim of comparing flexibility of βPGM in different states, the small differences in mean S^2 value from the apo and TSA data sets can most likely be accounted for by the fact that conformational exchange in apo βPGM has prevented the study of many of the residues which show elevated ns–ps dynamics in the βPGM-MgF_3^--βG6P TSA complex complex. There is no significant difference in the overall rigidity of the backbone of βPGM between apo and TSA states. This suggests no significant changes in conformational entropy as βPGM passes through its catalytic cycle, however insignificant changes in backbone flexibility are frequently found to be the case from backbone ^{15}N relaxation data measured on different ligation states, with differences in side-chain dynamics being much more significant[77]. It remains to be seen if the methyl ^{2}H relaxation data show any large differences in flexibility between the different states in the catalytic cycle.

7.2.2 Collective ns–ps Timescale Motion

In Section 6.3.5, it was shown that in the βPGM-MgF_3^--βG6P TSA complex, there are several groups of loop residues which are close in the tertiary structure but far from each other in the primary structure which all have motion on a timescale which is within error, or very close, according to their fitted τ_e values. This strongly implies that the same motion is affecting many of the active site loops. These residues are V12 and I13, F21 and A22, A26 and A28, L44–V47, R75–D78, A90–Y93, S114 and S116, and finally E169 and Q172. With a little imagination the residues D8–G11, A115, and D170–S171 which have no internal correlation times fit (or could not be studied due to spectral overlap or ^{19}F J-coupling) could be assumed to have internal correlation times which match those of their neighbours. If this assumption were true, this would represent almost all of the residues on the active site loops, and many of the residues which are involved in coordinating the TSA ligands. Whether or not these residues for which there is no information available are also involved, the matching of internal correlation times for such a large group of nearby and functionally-relevant residues would be an extraordinary coincidence if these motions are not coupled.

It has previously been thought that motion on this scale cannot take place on such a fast timescale, on the basis of the height of the energy barriers between dif-

ferent conformers. However, none of the S^2 values for any of the residues discussed thus far are particularly low, and so if this were true the orientations of the residues in question would not necessarily change enough to cause any energetically-costly reconfigurations of the interactions between them. It may be that the motion of the active site loops is such that none of the side chain interactions between the loops are significantly perturbed, i.e., all loops move together. This could avoid the transition state between conformers being unstable, and therefore avoid high barriers between them. In a similar case, an NMR relaxation study demonstrated the occurrence of fast timescale rigid-body motion of an entire α-helix, whereby none of the hydrogen bonding within the helix is broken during the transition[145]. Even if the active site loops are not moving as a whole, there are certainly groups of residues within it which show collective behaviour. The functional implication of this is not clear, but what is clear is that any conformational entropy calculations on the basis of this data set should be treated with caution, as the principal assumption that all ns–ps timescale motions are unrelated and independent does not seem to be true in this case.

7.3 Catalytic Role of Active Site ms–μs Dynamics

Assessing the structure of the minor conformer of the βPGM-MgF_3^--βG6P TSA complex as seen by ^{19}F and ^1H–^{15}N TROSY experiments has proved to be a difficult task, and as such it is necessary to include in this discussion data from other experiments which have shed light on this issue.

7.3.1 K145A Mutant and Minor Conformer Model

Although all the mutants of βPGM discussed in Section 5 failed to adopt the minor conformation, a K145A mutant produced by Dr. Joanna Griffin was more successful. The mutation was originally designed to test the effect of removing a positive charge from the active site: the TSA complex which is formed by the mutant responds by forming not an MgF_3^- complex, but an $MgF_2(H_2O)$ complex instead, maintaining the same overall charge in the active site[213]. The ^{19}F NMR spectrum of this complex in Figure 109 shows that the ^{19}F resonances have δs which closely resemble the minor conformation of the WT βPGM. Although there are minor resonances of this complex which resemble the major resonances of the WT βPGM-MgF_3^--βG6P TSA complex, the population of the major K145A βPGM resonances is more than 85%, which corresponds to a destabilisation of the ground-state WT conformation of over 9 kJ/mol, which is approximately the enthalpy of 1 hydrogen bond.

Crystallisation of the TSA complex of this mutant carried out by Erika Pellegrini showed that rather than the expected trigonal bipyramidal configuration, the Mg atom has octahedral geometry: its ligands are the 2 F^- ions, a water molecule which replaces F_b, and the backbone carbonyl from S114. This was unexpected, particularly as the amide bond between S114 and A115 must flip conformation in order for this interaction to occur. There are no other significant deviations of the protein conformation from that adopted by the WT βPGM-MgF_3^--βG6P TSA complex. In order to assess

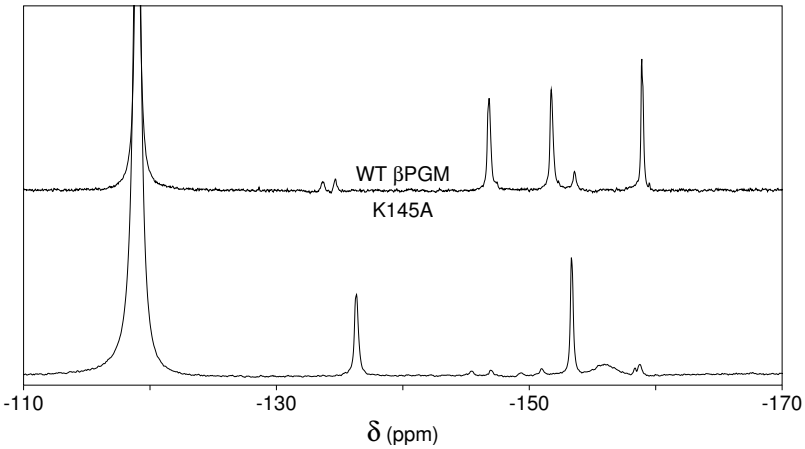

Figure 109: ^{19}F spectra of the K145A βPGM-$MgF_2(H_2O)$-βG6P TSA complex showing that to compensate for the loss of a positive charge in the active site, a negatively charged F^- ion has been replaced by a water molecule in the TSA. Comparison with the WT βPGM TSA complex shows that the 2 remaining ^{19}F nuclei have δs which closely resemble those of the minor conformer in the WT enzyme.

whether or not this resembles the minor conformation of the WT βPGM-MgF_3^--βG6P TSA complex, the structure of a crystal of the WT complex was solved to a resolution of 1.4 Å at room temperature by Dr. Matt Bowler. This data set shows that although the structure solved is identical to those solved from data recorded at cryogenic temperature, there is positive difference density, i.e., electron density present in the data but not accounted for by the model, which strongly suggests that the major⇌minor exchange does indeed involve this backbone amide flipping. Likewise with the K145A electron density, there are no other regions with significant alternative electron density. This is demonstrated in Figure 110. This amide flipping would also explain the large ^{19}F $\Delta\delta_{(\text{major-minor})}$ for F_b, which accepts a hydrogen bond from S115 in the "major" conformer.

The crystallographic data from WT βPGM and the K145A mutant are at odds with the ^{15}N relaxation dispersion study, which show that although clearly involved in the conformational exchange, the loop on which S114 and A115 lie does not show the most significant ^{15}N $\Delta\delta$s. The fact that there are larger changes elsewhere in the active site discourages the concern that the minor conformation is simply the product of the inorganic chemistry of magnesium, which is normally 6-coordinate. The residues with the largest chemical shift differences according to the dispersion study are D10 and G46. Neither of these residues show significant changes in the K145A crystal structure, nor significant difference density in the RT WT βPGM crystal structure. This is perhaps indicative that the conformational exchange event cannot occur within the constraints of the crystal packing, and only the backbone flipping at S114 and A115 is tolerated. The fact that the crystal data does not represent the full scope of the dynamics seen in solution is frustrating, as a structural picture of the minor conformer is still elusive.

G46 does not show any significant changes between crystal structures of any of the

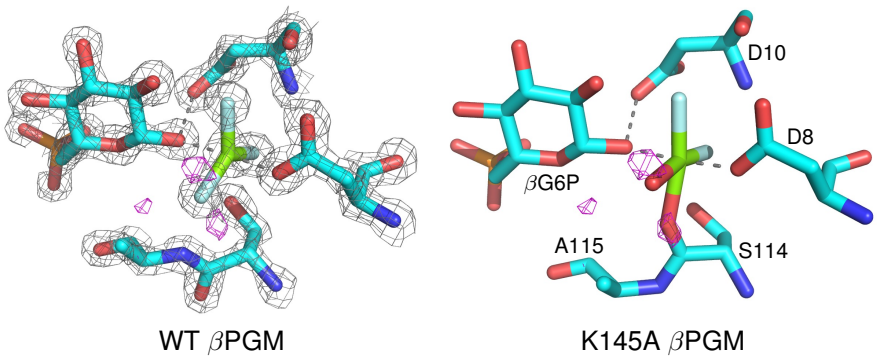

Figure 110: Electron density maps for WT and K145A βPGM-MgF_3^--βG6P TSA complexes, demonstrating that in the K145A mutant, the complex is octahedral at Mg^{2+} with the carbonyl of the S114–A115 amide bond providing the sixth ligand. Note that, in keeping with the way the MgF_3^- moiety in the WT PyMol images have been displayed throughout, all bonding to the Mg^{2+} in the mutant coordinates has been plotted as covalent linkages. The "amide flipping" is also visualised in the WT electron density recorded at room temperature, as visualised by positive difference density which closely resembles the K145A geometry. The difference density within 2.5 Å of the MgF_3^- moiety is contoured at 2.5σ and shown in purple. This difference density is also superimposed onto the K145A coordinates for illustrative purposes, and the $2F_o-F_c$ map of the WT RT density is also shown in grey.

complexes so far solved in βPGM, and as such it is extremely difficult to explain the very large 4 ± 2 ppm difference in ^{15}N δ between conformers. D10, on the other hand, alternates between "in" and "out" conformations as shown by the structures of the NACs formed by the βPGM-BeF_3^--βG6P GSA complex and discussed in Section 5.2. In its "in" conformation, it forms a hydrogen bond to the 1′ OH of the βG6P substrate, and in its "out" conformation, it forms a hydrogen bond to the 2′ hydroxyl instead. As well as mutating the enzyme to remove interactions, it is possible to mutate the substrate, and an MgF_3^- TSA complex using 2-deoxy-glucose-6-phosphate has been formed by Dr. Nicola Baxter. The ^{19}F NMR spectrum of this complex shown in Figure 111 reveals that there is no minor conformation; that it is destabilised by the removal of the 2′ hydroxyl on the sugar phosphate. This strongly suggests that, contrary to its position in the major form conformation, the D10 side chain is in its "out" position in the minor conformer.

Another deduction that can be made from the ^{15}N $\Delta\delta$s from the relaxation dispersion analysis is that given the relatively small changes in the regions which are most affected by opening/closing of the 2 βPGM domains, i.e., the inner and outer hinge regions, as well as the loops which lie at the end of the cap domain, the minor form of βPGM is (at least mostly) closed. The minor peak of K117 has been assigned in the ^1H–^{15}N TROSY spectra, and has ^{15}N and ^1H $\Delta\delta$s which are even further downfield than the major peak shifts, which indicates that the phosphate group is still closely associated with this residue in the minor conformer. The fitted ^{15}N $\Delta\delta$s from the relaxation dispersion data are only the magnitude, not the sign, of the changes, and so similar analyses for many other residues with large changes between conformers are not possible.

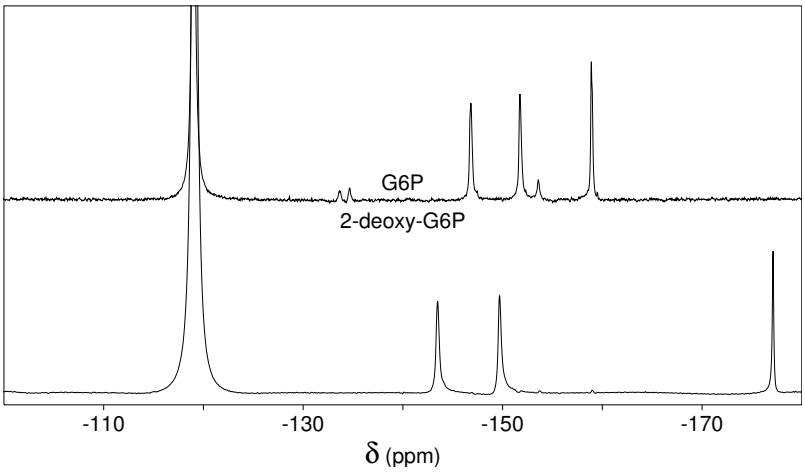

Figure 111: ^{19}F spectra of the βPGM-MgF_3^--β2deoxyG6P TSA complex. Comparison with the spectrum of the βPGM-MgF_3^--βG6P TSA complex shows that the only minor peaks visible in the 2-deoxy-G6P complex spectrum are from contaminating G6P molecules. The removal of the 2′ hydroxyl on the sugar phosphate has destabilised the minor conformer of the TSA complex and prevented its formation in solution. Removal of the 2′ sugar hydroxyl has also affected the active site electrostatics, particularly for F_c which is upfield shifted due to loss of its only hydrogen bond donor.

The following list are some of the known details about the structure of the minor conformer, given the data discussed in previous sections of this thesis and the above:

- It is a predominantly closed structure.
- The βG6P ligand is still bound.
- The TSA has octahedral geometry at Mg^{2+}.
- The backbone amide group which links residues S114 and A115 flips around such that the carbonyl of S114 interacts with the central Mg^{2+} ion.
- The side chain of the catalytic base D10 is in its "out" position, pointing away from the active site and forming an interaction with the 2′ OH of the βG6P molecule.

7.3.2 Role in Catalytic Turnover

Unlike the kinetic measurements presented in this thesis, enzyme kinetic studies on βPGM have not been carried out at a range of temperatures, and therefore $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ values for the βG16BP→βG6P dephosphorylation step during catalysis have not yet been measured. This prevents comparison of the temperature dependence of both conformational exchange and catalytic turnover rates, which would prove that the "minor" conformer is indeed of catalytic relevance. However at 298 K the rate at which the minor conformer of the βPGM-MgF_3^--βG6P TSA complex exchanges back into the major conformer — as measured both by ^{19}F SEXSY and ^{15}N relaxation dispersion — occurs on precisely the same timescale as the reaction step which it mimics providing a small error (0.5 K) in the temperature at which the enzyme kinetics were measured is allowed (see the end of Section 3.4). Further to this, numerical simulations by Dr. Matthew Cliff indicate that a conformational exchange event occurring at the rate of the minor→major exchange (as measured by ^{19}F SEXSY) immediately prior to the dephosphorylation step is fully consistent with the kinetic data presented in the

Goličnik et al, 2009 study[129].

It is clear even from the vague image of the structure of the minor conformer that it represents the enzyme in a catalytically-incompetent conformation: the flipping of a carbonyl into the active site would sterically prevent any phosphate transfer reactions, and the catalytic base D10 is pointing out of the active site and is therefore unable to activate the nucleophile for such chemistry to occur. The minor→major transition may represent a conformational exchange which takes place during catalysis, and with particular relevance to the catalytic step whose rate is matched, once the βG16BP bisphosphate intermediate binds to the enzyme. The minor→major exchange may consist of setting up the active site for catalysis once the substrate has bound. Conversely, the major→minor transition would then represent the enzyme withdrawing necessary catalytic groups and blocking the substrate from fully engaging with the active site.

Withdrawing catalytic groups and preventing subsequent catalytic steps may have a role in allowing βPGM to function as a mutase, rather than a phosphatase like many of its homologues. Phosphoserine Phosphatase (PSP) is a very close structural homologue of βPGM, and transfers the phosphate group from phosphoserine to the phosphate acceptor D11 (equivalent to D8). The serine molecule then dissociates, and a water molecule contacts the aspartyl-phosphate residue in its place. The catalytic base (D13 in PSP) then activates this water molecule, resulting in hydrolysis of the phosphate from D11. Ultimately this task is considerably easier than the task that βPGM has to carry out, for two related reasons. Firstly, βPGM must move the catalytic base away from the active site when the enzyme is phosphorylated, to prevent unwanted hydrolysis as a result of activating a water molecule to act as a nucleophile. This is why D10 is seen to alternate between two conformations when bound in complexes representing different stages of catalysis[130]. Secondly, and more subtly, βPGM must also protect against unproductive phosphorylation–dephosphorylation cycles when the bisphosphorylated sugar intermediate is formed. Assuming that βG1P is being converted to βG6P, βG1P is phosphorylated to form βG16BP as a product, but βG16BP (when reorientated) is also the substrate for the subsequent dephosphorylation step which forms βG6P as the final product. βPGM must phosphorylate βG1P, but then not dephosphorylate it again before it allows the formed βG16BP molecule time to dissociate and rebind such that the $1'$ phosphate is presented to the catalytic machinery for removal. Given that βPGM is capable of catalysing the conversion in either direction, this cannot be accomplished simply by differential binding constants for the βG16BP intermediate in the two conformations.

It can therefore be appreciated that the ability to remove the catalytic base and block the phosphate acceptor in the active site for long enough to allow release of the bisphosphate is catalytically advantageous. The hypothesis is that after forming the βG16BP intermediate, βPGM adopts the minor conformation, preventing hydrolysis and the return to the βG1P substrate. In a second conformational change, the enzyme opens to allow the bisphosphate to dissociate. It (or another free bisphosphate) would

then rebind in the alternative conformation, and the process is reversed: βPGM closes into the minor conformation, then repositions all catalytic residues to allow for the hydrolysis of the βG16BP intermediate into βG6P. It is this final step for which the catalytic kinetics and conformational exchange kinetics match. As such, mirroring many of the other perceived roles of enzyme dynamics on the ms–μs timescale, this conformational exchange does not "drive" catalysis, it deliberately hinders it, but in a way that is functionally useful and fine-tuned to the needs of the enzyme.

7.3.3 Further Work

The most obvious test of this hypothesis is to establish whether or not the same conformational exchange event occurs in a βPGM-βG16BP GSA complex, however this cannot be done as the bisphosphate would be immediately hydrolysed. It may only be the result of the inorganic chemistry of the central Mg^{2+} that the minor conformation is able to be significantly populated in the βPGM-MgF_3^--βG6P TSA complex. Although the βPGM-βG16BP GSA complex is not accessible, it may be possible to study the dynamics of a complex of βPGM and a non-hydrolysable bisphosphonate.

Regardless of the ability to define the structure of the minor conformer in high detail, the case for its involvement in turnover would be strengthened by having enzyme kinetic data at a variety of temperatures. Also, as this conformational exchange should be a feature of the conversion of the sugar phosphates in either direction, it would be advantageous to have enzyme turnover kinetic data available in the βG6P$\rightarrow\beta$G1P direction. This is difficult as there is no established spectroscopic method by which the production of βG1P can be followed, though in theory it is possible to pursue by ^{31}P NMR.

Finally, it would also be of interest to test the turnover kinetics of mutants which had different conformational exchange properties. Unfortunately, K145A is probably not a suitable candidate given that the electrostatics of the active site are so perturbed by this mutation, and given that most of the large structural changes (as measured by ^{15}N relaxation dispersion) are concentrated around the active site, it may be difficult to find a suitable candidate mutant for such a study.

7.4 Final Remarks

Although NMR is a powerful technique to measure fast (ns–ps) timescale dynamics, it is widely appreciated that dynamics on this timescale are not directly involved in bond forming or bond breaking steps in catalysis. Most of the interest in dynamics on this timescale comes from their contribution to thermodynamics through conformational entropy. Although this clearly has implications when comparing two ground states such as apo and ligand-bound proteins, the role of differential entropy between a ground state and a transition state — a necessity if conformational entropy has a catalytic role — is less easily justified. Thermodynamics arguments involve equilibrated systems, and although a TSA such as the βPGM-MgF_3^--βG6P TSA complex

is a thermally equilibrated system, the true transition state is not. Conformational entropy changes between substates along the catalytic pathway of an enzyme may be a crucial part of how the enzyme moves along the pathway, just as ms–μs dynamics are, but it is not easily envisaged how conformational entropy would reduce the size of the activation energy of a reaction.

As for ms–μs timescale events, the techniques for studying conformational exchange processes on a per-residue basis are still in their infancy. Given the limited data available, it seems a likely scenario that in any given enzyme, there are a great many conformational states which must be traversed in order to chaperone the substrate through the reaction and release the product at the end. In βPGM alone, there are a total of six conformational states which likely have relevance for catalysis: apo enzyme, the phospho-enzyme (as mimicked by the βPGM-BeF_3^- complex), the two NACs of the βPGM-BeF_3^--βG6P GSA complex, and the two conformers of the βPGM-MgF_3^--βG6P TSA complex. There may yet be more to be discovered. Enzymes which catalyse multi-step reactions involving multiple substrates are likely even more complex. Future studies of dynamics will allow for the discovery of more such non-ground state conformations, and a greater understanding of how enzymes function.

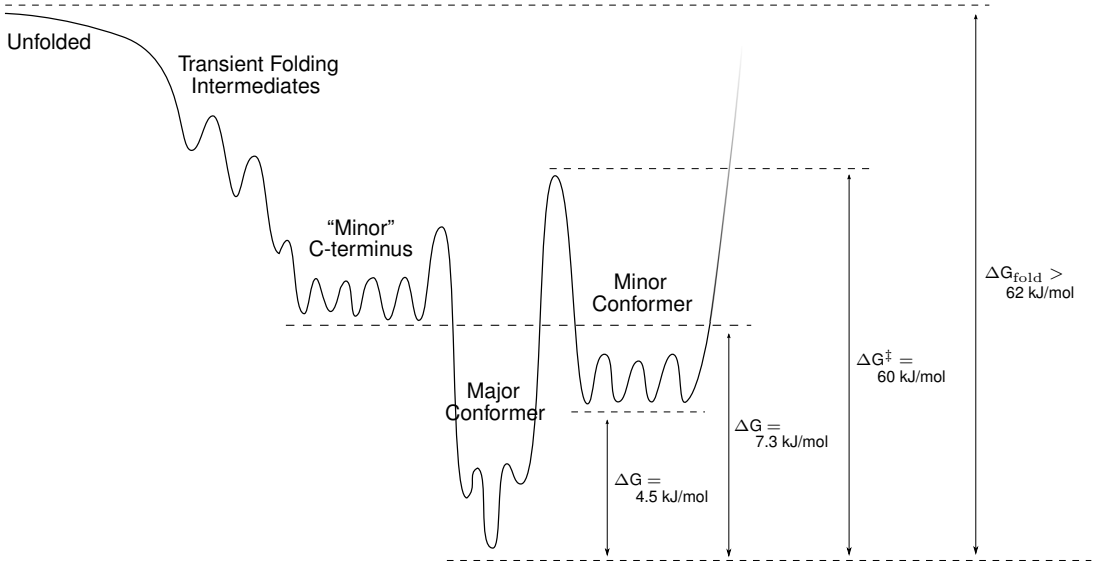

Figure 112: Semi-quantitative Energy Landscape for the βPGM-MgF_3^--βG6P TSA complex. Different wells and regions are labelled and represent entities which exchange on the ms–μs timescale. The substates within the wells represent dynamic excursions on a sub-μs timescale. In keeping with the idea that it is in a deep energy well with highly favourable enthalpy and little backbone conformational entropy, there are fewer substates in the major conformer well. Transient folding intermediates have not been studied in this enzyme, but are believed to be ubiquitous. It is seen that the well representing the "minor" conformer of the faster dynamics (labelled: "Minor" C-terminus) is along the way to unfolding, hence the lowered folding stability of a mutant which is believed to occupy this conformer at higher population. Because these dynamics were faster than those between the major and minor conformers of the active site, the barrier separating this well from that of the ground-state major conformer well is lower than that separating the major and minor conformers. The ΔG_{fold} value has been taken from the CD study of WT apo βPGM, although it is known from the stability of βPGM samples at higher temperatures that the βPGM-MgF_3^--βG6P TSA complex is substantially more thermally stable than apo βPGM. Even as a schematic, this energy landscape incorporates data and observations from all four results chapters of this thesis.

References

[1] E. Fischer. Einfluss der Configuration auf die Wirkung der Enzyme. *Ber Deut Chem Ges*, 27:2985–2993, 1894.

[2] L. A. Heppel and R. J. Hilmoe. Purification and Properties of 5-Nucleotidase. *J Biol Chem*, 188:665–676, 1951.

[3] D. E. Koshland. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc Natl Acad Sci U S A*, 44:98–104, 1958.

[4] D. Tobi and I. Bahar. Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc Natl Acad Sci U S A*, 102:18908–18913, 2005.

[5] O. F. Lange, N. Lakomek, C. Farès, G. F. Schröder, K. F. A. Walter, S. Becker, J. Meiler, H. Grubmüller, C. Griesinger, and B. L. de Groot. Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science*, 320:1471–1475, 2008.

[6] J. Gsponer, J. Christodoulou, A. Cavalli, J. M. Bui, B. Richter, C. M. Dobson, and M. Vendruscolo. A coupled equilibrium shift mechanism in calmodulin-mediated signal transduction. *Structure*, 16:736–746, 2008.

[7] L. Leder, C. Berger, S. Bornhauser, H. Wendt, F. Ackermann, I. Jelesarov, and H. R. Bosshard. Spectroscopic, calorimetric, and kinetic demonstration of conformational adaptation in peptide-antibody recognition. *Biochemistry*, 34:16509–16518, 1995.

[8] B. Ma, S. Kumar, C. J. Tsai, and R. Nussinov. Folding funnels and binding mechanisms. *Protein Eng*, 12:713–720, 1999.

[9] H. R. Bosshard. Molecular recognition by induced fit: how fit is the concept? *News Physiol Sci*, 16:171–173, 2001.

[10] T. Wlodarski and B. Zagrovic. Conformational selection and induced fit mechanism underlie specificity in noncovalent interactions with ubiquitin. *Proc Natl Acad Sci U S A*, 106:19346–19351, 2009.

[11] L. Pauling. Nature of Forces between Large Molecules of Biological Interest. *Nature*, 161:707–709, 1948.

[12] W. P. Jencks. *Catalysis in Chemistry and Enzymology*. McGraw-Hill. New York., 1969.

[13] A. Tramontano, K. D. Janda, and R. A. Lerner. Catalytic antibodies. *Science*, 234:1566–1570, 1986.

[14] D. Hilvert. Critical analysis of antibody catalysis. *Annu Rev Biochem*, 69:751–793, 2000.

[15] S. J. Benkovic and S. Hammes-Schiffer. A perspective on enzyme catalysis. *Science*, 301:1196–1202, 2003.

[16] K. I. Okazaki and S. Takada. Dynamic energy landscape view of coupled binding and protein conformational change: Induced-fit versus population-shift mechanisms. *Proc Natl Acad Sci U S A*, 105:11182–11187, 2008.

[17] K. A. Dill and H. S. Chan. From Levinthal to pathways to funnels. *Nat Struct Biol*, 4:10–19, 1997.

[18] P. G. Wolynes, J. N. Onuchic, and D. Thirumalai. Navigating the folding routes. *Science*, 267:1619–1620, 1995.

[19] S. Kumar, B. Ma, C. J. Tsai, N. Sinha, and R. Nussinov. Folding and binding cascades: dynamic landscapes and population shifts. *Protein Sci*, 9:10–19, 2000.

[20] J. B. S. Haldane. *Enzymes*. Longmans, Green, London, 1930.

[21] W. P. Jencks. Binding-Energy, Specificity, and Enzymic Catalysis - Circe Effect. *Adv Enzymol Relat Areas Mol Biol*, 43:219–410, 1975.

[22] C. C. Blake, D. F. Koenig, G. A. Mair, A. C. North, D. C. Phillips, and V. R. Sarma. Structure of Hen Egg-White Lysozyme. A Three-dimensional Fourier Synthesis at 2 ÅResolution. *Nature*, 206:757–761, 1965.

[23] D. C. Phillips. The Hen Egg-White Lysozyme Molecule. *Proc Natl Acad Sci U S A*, 57:484–495, 1967.

[24] D. J. Vocadlo, G. J. Davies, R. Laine, and S. G. Withers. Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate. *Nature*, 412:835–838, 2001.

[25] A. Warshel and M. Levitt. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J Mol Biol*, 103:227–249, 1976.

[26] G. R. Stockwell and J. M. Thornton. Conformational diversity of ligands bound to proteins. *J Mol Biol*, 356:928–944, 2006.

[27] Y. Yoshioka, K. Hasegawa, Y. Matsuura, Y. Katsube, and M. Kubota. Crystal structures of a mutant maltotetraose-forming exo-amylase cocrystallized with maltopentaose. *J Mol Biol*, 271:619–628, 1997.

[28] N. A. Khanjin, J. P. Snyder, and F. M. Menger. Mechanism of Chorismate Mutase: Contribution of Conformational Restriction to Catalysis in the Claisen Rearrangement. *J Am Chem Soc*, 121:11831–11846, 1999.

[29] L Stryer. *Biochemistry; 4th Edition*. W. H. Freeman and Company, New York, 1995.

[30] M. F. Perutz. Concluding Remarks. *Proc R Soc Lond B Biol Sci*, 167:448, 1967.

[31] A. Warshel. Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites. *J Biol Chem*, 273:27035–27038, 1998.

[32] J. Crosby, R. Stone, and G. E. Lienhard. Mechanisms of thiamine-catalyzed reactions. Decarboxylation of 2-(1-carboxy-1-hydroxyethyl)-3,4-dimethylthiazolium chloride. *J Am Chem Soc*, 92:2891–2900, 1970.

[33] S. C. Tucker and D. G. Truhlar. Effect of Nonequilibrium Solvation on Chemical-Reaction Rates. Variational Transition-State-Theory Studies of the Microsolvated Reaction $Cl^-(H_2O)_n + CH_3Cl$. *J Am Chem Soc*, 112:3347–3361, 1990.

[34] W. R. Cannon and S. J. Benkovic. Solvation, reorganization energy, and biological catalysis. *J Biol Chem*, 273:26257–26260, 1998.

[35] A. Warshel, J. Aqvist, and S. Creighton. Enzymes work by solvation substitution rather than by desolvation. *Proc Natl Acad Sci U S A*, 86:5820–5824, 1989.

[36] P. Gilli, V. Bertolasi, V. Ferretti, and G. Gilli. Covalent Nature of the Strong Homonuclear Hydrogen Bond. Study of the O-H-O system by Crystal Structure Correlation Methods. *J Am Chem Soc*, 116:909–915, 1994.

[37] A. D. Frey, P. A. & Hegeman. *Enzymatic Reaction Mechanisms*. Oxford University Press, New York, 2007.

[38] P. A. Frey, S. A. Whitt, and J. B. Tobin. A low-barrier hydrogen bond in the catalytic triad of serine proteases. *Science*, 264:1927–1930, 1994.

[39] J. A. Gerlt, M. M. Kreevoy, W. W. Cleland, and P. A. Frey. Understanding enzymic catalysis: the importance of short, strong hydrogen bonds. *Chem Biol*, 4:259–267, 1997.

[40] W. W. Cleland, P. A. Frey, and J. A. Gerlt. The low barrier hydrogen bond in enzymatic catalysis. *J Biol Chem*, 273:25529–25532, 1998.

[41] M. H. M. Olsson, W. W. Parson, and A. Warshel. Dynamical contributions to enzyme catalysis: Critical tests of a popular hypothesis. *Chemical Reviews*, 106:1737–1756, 2006.

[42] W. P. Jencks. *Catalysis in Chemistry and Enzymology*. Dover Publications, Inc. New York., 1987.

[43] D. E. Koshland and K. E. Neet. Catalytic and Regulatory Properties of Enzymes. *Ann Rev Biochem*, 37:359–411, 1968.

[44] D. R. Storm and D. E. Koshland. A source for the special catalytic power of enzymes: orbital steering. *Proc Natl Acad Sci U S A*, 66:445–452, 1970.

[45] M. I. Page and W. P. Jencks. Entropic Contributions to Rate Accelerations in Enzymic and Intramolecular Reactions and Chelate Effect. *Proc Natl Acad Sci U S A*, 68:1678–1683, 1971.

[46] A. Warshel, F. Sussman, and J. K. Hwang. Evaluation of catalytic free energies in genetically modified proteins. *J Mol Biol*, 201:139–159, 1988.

[47] S. Hur and T. C. Bruice. Enzymes do what is expected (chalcone isomerase versus chorismate mutase). *J Am Chem Soc*, 125:1472–1473, 2003.

[48] K. E. Ranaghan and A. J. Mulholland. Conformational effects in enzyme catalysis: QM/MM free energy calculation of the 'NAC' contribution in chorismate mutase. *Chem Commun*, pages 1238–1239, 2004.

[49] M. Strajbl, A. Shurki, M. Kato, and A. Warshel. Apparent NAC effect in chorismate mutase reflects electrostatic transition state stabilization. *J Am Chem Soc*, 125:10228–10237, 2003.

[50] A. Dafforn and D. E. Koshland. Theoretical aspects of orbital steering. *Proc Natl Acad Sci U S A*, 68:2463–2467, 1971.

[51] A. R. Fersht. *Enzyme Structure and Mechanism: 2nd Edition*. W. H. Freeman and Company, New York, 1985.

[52] M. P. Williamson. *How Proteins Work*. Garland Science, New York, 2011.

[53] J. Villa and A. Warshel. Energetics and dynamics of enzymatic reactions. *J Phys Chem B*, 105:7887–7907, 2001.

[54] K. E. Neet. Enzyme catalytic power minireview series. *J Biol Chem*, 273:25527–25528, 1998.

[55] S. C. L. Kamerlin and A. Warshel. At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis? *Proteins*, 78:1339–1375, 2009.

[56] K. Henzler-Wildman and D. Kern. Dynamic personalities of proteins. *Nature*, 450:964–972, 2007.

[57] G. G. Hammes, S. J. Benkovic, and S. Hammes-Schiffer. Flexibility, diversity, and cooperativity: pillars of enzyme catalysis. *Biochemistry*, 50:10422–10430, 2011.

[58] S. Hammes-Schiffer and S. J. Benkovic. Relating protein motion to catalysis. *Ann Rev Biochem*, 75:519–541, 2006.

[59] A. J. Adamczyk, J. Cao, S. C. L. Kamerlin, and A. Warshel. Catalysis by dihydrofolate reductase and other enzymes arises from electrostatic preorganization, not conformational motions. *Proc Natl Acad Sci U S A*, 108:14115–14120, 2011.

[60] G. Bhabha, J. Lee, D. C. Ekiert, J. Gam, I. A. Wilson, H. J. Dyson, S. J. Benkovic, and P. E. Wright. A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. *Science*, 332:234–238, 2011.

[61] P. Vallurupalli, D. F. Hansen, and L. E. Kay. Structures of invisible, excited protein states by relaxation dispersion NMR spectroscopy. *Proc Natl Acad Sci U S A*, 105:11766–11771, 2008.

[62] D. Antoniou, J. Basner, S. Nunez, and S. D. Schwartz. Computational and theoretical methods to explore the relation between enzyme dynamics and catalysis. *Chem Rev*, 106:3170–3187, 2006.

[63] V. L. Schramm. Enzymatic transition states: thermodynamics, dynamics and analogue design. *Arch Biochem Biophys*, 433:13–26, 2005.

[64] A. Lewandowicz and V. L. Schramm. Transition state analysis for human and Plasmodium falciparum purine nucleoside phosphorylases. *Biochemistry*, 43:1458–1468, 2004.

[65] L. Li, M. Luo, M. Ghanem, E. A. Taylor, and V. L. Schramm. Second-sphere amino acids contribute to transition-state structure in bovine purine nucleoside phosphorylase. *Biochemistry*, 47:2577–2583, 2008.

[66] P. K. Agarwal. Role of protein dynamics in reaction rate enhancement by enzymes. *J Am Chem Soc*, 127:15248–15256, 2005.

[67] V. L. Schramm. Enzymatic transition states and transition state analogues. *Curr Opin Struct Biol*, 15:604–613, 2005.

[68] J. Kraut. How do enzymes work? *Science*, 242:533–540, 1988.

[69] M. J. Sutcliffe and N. S. Scrutton. Enzymology takes a quantum leap forward. *Philos Transact A Math Phys Eng Sci*, 358:367–386, 2000.

[70] S. Hay, C. R. Pudney, T. A. McGrory, J. Pang, M. J. Sutcliffe, and N. S. Scrutton. Barrier compression enhances an enzymatic hydrogen-transfer reaction. *Angew Chem Int Ed Engl*, 48:1452–1454, 2009.

[71] D. B. Northrop. Unusual origins of isotope effects in enzyme-catalysed reactions. *Philos Trans R Soc Lond B Biol Sci*, 361:1341–1349, 2006.

[72] D. Antoniou and S. D. Schwartz. Internal enzyme motions as a source of catalytic activity: Rate-promoting vibrations and hydrogen tunneling. *J Phys Chem B*, 105:5553–5558, 2001.

[73] Q. Zhao, C. Abeygunawardana, and A. S. Mildvan. ^{13}C NMR relaxation studies of backbone and side chain motion of the catalytic tyrosine residue in free and steroid-bound delta 5-3-ketosteroid isomerase. *Biochemistry*, 35:1525–1532, 1996.

[74] L. Zídek, Novotny M. V., and Stone M. J. Increased protein backbone conformational entropy upon hydrophobic ligand binding. *Nat Struct Biol*, 6:1118–1121, 1999.

[75] L. K. Nicholson, L. E. Kay, D. M. Baldisseri, J. Arango, P. E. Young, A. Bax, and D. A. Torchia. Dynamics of Methyl-Groups in Proteins as Studied by Proton-Detected ^{13}C NMR-Spectroscopy - Application to the Leucine Residues of Staphylococcal Nuclease. *Biochemistry*, 31:5253–5263, 1992.

[76] D. Fushman, R. Weisemann, H. Thuring, and H. Ruterjans. Backbone dynamics of ribonuclease T1 and its complex with 2'GMP studied by two-dimensional heteronuclear NMR spectroscopy. *J Biomol NMR*, 4:61–78, 1994.

[77] C. Diehl, O. Engström, T. Delaine, M. Håkansson, S. Genheden, K. Modig, H. Leffler, U. Ryde, U. J. Nilsson, and M. Akke. Protein flexibility and conformational entropy in ligand design targeting the carbohydrate recognition domain of galectin-3. *J Am Chem Soc*, 132:14577–14589, 2010.

[78] J. R. Schnell, H. J. Dyson, and P. E. Wright. Structure, dynamics, and catalytic function of dihydrofolate reductase. *Annu Rev Biophys Biomol Struct*, 33:119–140, 2004.

[79] A. V. Pisliakov, J. Cao, S. C. L. Kamerlin, and A. Warshel. Enzyme millisecond conformational dynamics do not catalyze the chemical step. *Proc Natl Acad Sci U S A*, 106:17359–17364, 2009.

[80] E. J. Loveridge, L. Tey, E. M. Behiry, W. M. Dawson, R. M. Evans, S. B. Whittaker, U. L. Günther, C. Williams, M. P. Crump, and R. K. Allemann. The role of large-scale motions in catalysis by dihydrofolate reductase. *J Am Chem Soc*, 133:20561–20570, 2011.

[81] V. C. Nashine, S. Hammes-Schiffer, and S. J. Benkovic. Coupled motions in enzyme catalysis. *Curr Opin Chem Biol*, 14:644–651, 2010.

[82] G. P. Miller and S. J. Benkovic. Stretching exercises–flexibility in dihydrofolate reductase catalysis. *Chem Biol*, 5:105–113, 1998.

[83] P. T. Rajagopalan, S. Lutz, and S. J. Benkovic. Coupling interactions of distal residues enhance dihydrofolate reductase catalysis: mutational effects on hydride transfer rates. *Biochemistry*, 41:12618–12628, 2002.

[84] E. D. Watt, H. Shimada, E. L. Kovrigin, and J. P. Loria. The mechanism of rate-limiting motions in enzyme function. *Proc Natl Acad Sci U S A*, 104:11981–11986, 2007.

[85] E. L. Kovrigin and J. P. Loria. Enzyme dynamics along the reaction coordinate: critical role of a conserved residue. *Biochemistry*, 45:2636–2647, 2006.

[86] C. R. Wagner, Z. Huang, S. F. Singleton, and S. J. Benkovic. Molecular basis for nonadditive mutational effects in Escherichia coli dihydrofolate reductase. *Biochemistry*, 34:15671–15680, 1995.

[87] P. K. Agarwal, S. R. Billeter, P. T. Rajagopalan, S. J. Benkovic, and S. Hammes-Schiffer. Network of coupled promoting motions in enzyme catalysis. *Proc Natl Acad Sci U S A*, 99:2794–2799, 2002.

[88] G. M. Suel, S. W. Lockless, M. A. Wall, and R. Ranganathan. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat Struct Biol*, 10:59–69, 2003.

[89] K. F. Wong, T. Selzer, S. J. Benkovic, and S. Hammes-Schiffer. Impact of distal mutations on the network of coupled motions correlated to hydride transfer in dihydrofolate reductase. *Proc Natl Acad Sci U S A*, 102:6807–6812, 2005.

[90] E. Z. Eisenmesser, O. Millet, W. Labeikovsky, D. M. Korzhnev, M. Wolf-Watz, D. A. Bosco, J. J. Skalicky, L. E. Kay, and D. Kern. Intrinsic dynamics of an enzyme underlies catalysis. *Nature*, 438:117–121, 2005.

[91] M. Wolf-Watz, V. Thai, K. Henzler-Wildman, G. Hadjipavlou, E. Z. Eisenmesser, and D. Kern. Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. *Nat Struct Mol Biol*, 11:945–949, 2004.

[92] D. D. Boehr, D. McElheny, H. J. Dyson, and P. E. Wright. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science*, 313:1638–1642, 2006.

[93] A. K. Gardino, J. Villali, A. Kivenson, M. Lei, C. F. Liu, P. Steindel, E. Z. Eisenmesser, W. Labeikovsky, M. Wolf-Watz, M. W. Clarkson, and D. Kern. Transient non-native hydrogen bonds promote activation of a signaling protein. *Cell*, 139:1109–1118, 2009.

[94] D. D. Boehr, D. McElheny, H. J. Dyson, and P. E. Wright. Millisecond timescale fluctuations in dihydrofolate reductase are exquisitely sensitive to the bound ligands. *Proc Natl Acad Sci U S A*, 107:1373–1378, 2010.

[95] J. Villali and D. Kern. Choreographing an enzyme's dance. *Curr Opin Chem Biol*, 14:636–643, 2010.

[96] G. Manning, G. D. Plowman, T. Hunter, and S. Sudarsanam. Evolution of protein kinase signaling from yeast to man. *Trends Biochem Sci*, 27:514–520, 2002.

[97] B. E. Turk. Understanding and exploiting substrate recognition by protein kinases. *Curr Opin Chem Biol*, 12:4–10, 2008.

[98] C. Lad, N. H. Williams, and R. Wolfenden. The rate of hydrolysis of phosphomonoester dianions and the exceptional catalytic proficiencies of protein and inositol phosphatases. *Proc Natl Acad Sci U S A*, 100:5607–5610, 2003.

[99] M. W. Bowler, M. J. Cliff, J. P. Waltho, and G. M. Blackburn. Why did Nature select phosphate for its dominant roles in biology? *New J Chem*, 34:784–794, 2010.

[100] L. W. Tari, A. Matte, H. Goldie, and L. T. Delbaere. Mg^{2+}–Mn^{2+} clusters in enzyme-catalyzed phosphoryl-transfer reactions. *Nat Struct Biol*, 4:990–994, 1997.

[101] N. Strater, W. N. Lipscomb, T. Klabunde, and B. Krebs. Two-metal ion catalysis in enzymatic acyl- and phosphoryl-transfer reactions. *Angew Chem Int Ed Engl*, 35:2024–2055, 1996.

[102] S. R. Jones, L. A. Kindman, and J. R. Knowles. Stereochemistry of phosphoryl group transfer using a chiral [^{16}O, ^{17}O, ^{18}O] stereochemical course of alkaline phosphatase. *Nature*, 275:564–565, 1978.

[103] W. A. Blättler and J. R. Knowles. Stereochemical course of phosphokinases. The use of adenosine [γ-(S)-^{16}O,^{17}O,^{18}O]triphosphate and the mechanistic consequences for the reactions catalyzed by glycerol kinase, hexokinase, pyruvate kinase, and acetate kinase. *Biochemistry*, 18:3927–3933, 1979.

[104] G. Lowe, P. M. Cullis, R. L. Jarvest, B. V. Potter, and B. S. Sproat. Stereochemistry of phosphoryl transfer. *Philos Trans R Soc Lond B Biol Sci*, 293:75–92, 1981.

[105] J. K. Lassila, J. G. Zalatan, and D. Herschlag. Biological phosphoryl-transfer reactions: understanding mechanism and catalysis. *Annu Rev Biochem*, 80:669–702, 2011.

[106] W. W. Cleland and A. C. Hengge. Enzymatic mechanisms of phosphate and sulfate transfer. *Chem Rev*, 106:3252–3278, 2006.

[107] J. Florián and A. Warshel. Phosphate Ester Hydrolysis in Aqueous Solution: Associative versus Dissociative Mechanisms. *J Phys Chem B*, 102:719–734, 1998.

[108] I. Berente, T. Beke, and G. Náray-Szabó. Quantum mechanical studies on the existence of a trigonal bipyramidal phosphorane intermediate in enzymatic phosphate ester hydrolysis. *Theor Chem Acc*, 118:129–134, 2007.

[109] W. B. Knight, P. M. Weiss, and W. W. Cleland. Determination of Equilibrium ^{18}O Isotope Effects on the Deprotonation of Phosphate and Phosphate Esters and the Anomeric Effect on Deprotonation of Glucose-6-Phosphate. *J Am Chem Soc*, 108:2759–2761, 1986.

[110] I. Schlichting and J. Reinstein. pH influences fluoride coordination number of the AlF_x phosphoryl transfer transition state analog. *Nat Struct Biol*, 6:721–723, 1999.

[111] H. A. Krebs and R. Hems. Phosphate-transfer reactions of adenosine and inosine nucleotides. *Biochem J*, 61:435–441, 1955.

[112] P. C. Sternweis, J. K. Northup, M. D. Smigel, and A. G. Gilman. The regulatory component of adenylate cyclase. Purification and properties. *J Biol Chem*, 256:11517–11526, 1981.

[113] P. C. Sternweis and A. G. Gilman. Aluminum: a requirement for activation of the regulatory component of adenylate cyclase by fluoride. *Proc Natl Acad Sci U S A*, 79:4888–4891, 1982.

[114] T. Higashijima, M. P. Graziano, H. Suga, M. Kainosho, and A. G. Gilman. ^{19}F and ^{31}P NMR spectroscopy of G protein α subunits. Mechanism of activation by Al^{3+} and F^-. *J Biol Chem*, 266:3396–3401, 1991.

[115] J. Sondek, D. G. Lambright, J. P. Noel, H. E. Hamm, and P. B. Sigler. GTPase mechanism of Gproteins from the 1.7-Åcrystal structure of transducin α-GDP-AIF_4^-. *Nature*, 372:276–279, 1994.

[116] R. B. Martin. The chemistry of aluminum as related to biology and medicine. *Clin Chem*, 32:1797–1806, 1986.

[117] T. Higashijima, K. M. Ferguson, P. C. Sternweis, E. M. Ross, M. D. Smigel, and A. G. Gilman. The effect of activating ligands on the intrinsic fluorescence of guanine nucleotide-binding regulatory proteins. *J Biol Chem*, 262:752–756, 1987.

[118] D. L. Graham, P. N. Lowe, G. W. Grime, M. Marsh, K. Rittinger, S. J. Smerdon, S. J. Gamblin, and J. F. Eccleston. MgF_3^- as a transition state analog of phosphoryl transfer. *Chem Biol*, 9:375–381, 2002.

[119] N. J. Baxter, L. F. Olguin, M. Goličnik, G. Feng, A. M. Hounslow, W. Bermel, G. M. Blackburn, F. Hollfelder, J. P. Waltho, and N. H. Williams. A Trojan horse transition state analogue generated by MgF_3^- formation in an enzyme active site. *Proc Natl Acad Sci U S A*, 103:14732–14737, 2006.

[120] M. J. Cliff, M. W. Bowler, A. Varga, J. P. Marston, J. Szabó, A. M. Hounslow, N. J. Baxter, G. M. Blackburn, M. Vas, and J. P. Waltho. Transition state analogue structures of human phosphoglycerate kinase establish the importance of charge balance in catalysis. *J Am Chem Soc*, 132:6507–6516, 2010.

[121] S. Park, K. Ajtai, and T. P. Burghardt. Inhibition of myosin ATPase by metal fluoride complexes. *Biochim Biophys Acta*, 1430:127–140, 1999.

[122] M. Chabre. Aluminofluoride and beryllofluoride complexes: a new phosphate analogs in enzymology. *Trends Biochem Sci*, 15:6–10, 1990.

[123] J. Guhaniyogi, V. L. Robinson, and A. M. Stock. Crystal structures of beryllium fluoride-free and beryllium fluoride-bound CheY in complex with the conserved C-terminal peptide of CheZ reveal dual binding modes specific to CheY conformation. *J Mol Biol*, 359:624–645, 2006.

[124] J. Bigay, P. Deterre, C. Pfister, and M. Chabre. Fluoride complexes of aluminium or beryllium act on G-proteins as reversibly bound analogues of the γ phosphate of GTP. *EMBO J*, 6:2907–2913, 1987.

[125] M. A. Danielson and J. J. Falke. Use of ^{19}F NMR to probe protein structure and conformational changes. *Annu Rev Biophys Biomol Struct*, 25:163–195, 1996.

[126] A. Ramos, I. C. Boels, W. M. de Vos, and H. Santos. Relationship between glycolysis and exopolysaccharide biosynthesis in Lactococcus lactis. *Appl Environ Microbiol*, 67:33–41, 2001.

[127] F. Levander, U. Andersson, and P. Rådström. Physiological role of β-phosphoglucomutase in Lactococcus lactis. *Appl Environ Microbiol*, 67:4546–4553, 2001.

[128] S. D. Lahiri, G. F. Zhang, D. Dunaway-Mariano, and K. N. Allen. Caught in the act: The structure of phosphorylated beta-phosphoglucomutase from Lactococcus lactis. *Biochemistry*, 41:8351–8359, 2002.

[129] M. Goličnik, L. F. Olguin, G. Feng, N. J. Baxter, J. P. Waltho, N. H. Williams, and F. Hollfelder. Kinetic Analysis of β-Phosphoglucomutase and Its Inhibition by Magnesium Fluoride. *J Am Chem Soc*, 131:1575–1588, 2009.

[130] G. Zhang, J. Dai, L. Wang, D. Dunaway-Mariano, L. W. Tremblay, and K. N. Allen. Catalytic cycling in β-phosphoglucomutase: a kinetic and structural analysis. *Biochemistry*, 44:9404–9416, 2005.

[131] S. D. Lahiri, G. Zhang, D. Dunaway-Mariano, and K. N. Allen. The pentacovalent phosphorus intermediate of a phosphoryl transfer reaction. *Science*, 299:2067–2071, 2003.

[132] J. Knowles. Chemistry: Seeing is believing. *Science*, 299:2002–2003, 2003.

[133] G. M. Blackburn, N. H. Williams, S. J. Gamblin, and S. J. Smerdon. Comment on "The pentacovalent phosphorus intermediate of a phosphoryl transfer reaction". *Science*, 301:1184, 2003.

[134] K. N. Allen and D. Dunaway-Mariano. Response to comment on "The pentacovalent phosphorus intermediate of a phosphoryl transfer reaction". *Science*, 301:1184, 2003.

[135] L. W. Tremblay, G. F. Zhang, J. Y. Dai, D. Dunaway-Mariano, and K. N. Allen. Chemical confirmation of a pentavalent phosphorane in complex with β-phosphoglucomutase. *J Am Chem Soc*, 127:5298–5299, 2005.

[136] Z. Lu, D. Dunaway-Mariano, and K. N. Allen. The catalytic scaffold of the haloalkanoic acid dehalogenase enzyme superfamily acts as a mold for the trigonal bipyramidal transition state. *Proc Natl Acad Sci U S A*, 105:5687–5692, 2008.

[137] C. E. Webster. High-energy intermediate or stable transition state analogue: theoretical perspective of the active site and mechanism of β-phosphoglucomutase. *J Am Chem Soc*, 126:6840–6841, 2004.

[138] N. J. Baxter, M. W. Bowler, T. Alizadeh, M. J. Cliff, A. M. Hounslow, B. Wu, D. B. Berkowitz, N. H. Williams, G. M. Blackburn, and J. P. Waltho. Atomic details of near-transition state conformers for enzyme phosphoryl transfer revealed by MgF_3^- rather than by phosphoranes. *Proc Natl Acad Sci U S A*, 107:4555–4560, 2010.

[139] E. Marcos, M. J. Field, and R. Crehuet. Pentacoordinated phosphorus revisited by high-level QM/MM calculations. *Proteins*, 78:2405–2411, 2010.

[140] A. Abragam. *The Principles of Nuclear Magnetism*. Oxford University Press, London, 1961.

[141] F. Bloch. Nuclear Induction. *Phys Rev*, 70:460–474, 1946.

[142] N. K. Goto and L. E. Kay. New developments in isotope labeling strategies for protein solution NMR spectroscopy. *Curr Opin Struct Biol*, 10:585–592, 2000.

[143] G. Bodenhausen and D. J. Ruben. Natural abundance ^{15}N NMR by enhanced heteronuclear spectroscopy. *Chem Phys Lett*, 69:185–189, 1980.

[144] L. E. Kay. Protein dynamics from NMR. *Nat Struct Biol*, 5:513–517, 1998.

[145] M. W. F. Fischer, L. Zeng, Y. X. Pang, W. D. Hu, A. Majumdar, and E. R. P. Zuiderweg. Experimental characterization of models for backbone picosecond dynamics in proteins. Quantification of NMR auto- and cross-correlation relaxation mechanisms involving different nuclei of the peptide plane. *J Am Chem Soc*, 119:12629–12642, 1997.

[146] D. R. Muhandiram, T. Yamazaki, B. D. Sykes, and L. E. Kay. Measurement of ^{2}H T_1 and $T_{1\rho}$ Relaxation Times in Uniformly ^{13}C-Labeled and Fractionally ^{2}H-Labeled Proteins in Solution. *J Am Chem Soc*, 117:11536–11544, 1995.

[147] O. Millet, D. R. Muhandiram, N. R. Skrynnikov, and L. E. Kay. Deuterium spin probes of side-chain dynamics in proteins. 1. Measurement of five relaxation rates per deuteron in ^{13}C-labeled and fractionally ^{2}H-enriched proteins in solution. *J Am Chem Soc*, 124:6439–6448, 2002.

[148] G. Lipari and A. Szabo. Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .1. Theory and Range of Validity. *J Am Chem Soc*, 104:4546–4559, 1982.

[149] G. Lipari and A. Szabo. Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .2. Analysis of Experimental Results. *J Am Chem Soc*, 104:4559–4570, 1982.

[150] M. G. Clore, A. Szabo, A. Bax, L. E. Kay, P. C. Driscoll, and A. M. Gronenborn. Deviations from the Simple Two-Parameter Model-Free Approach to the Interpretation of Nitrogen-15 Nuclear Magnetic Relaxation of Proteins. *J Am Chem Soc*, 112:4989–4991, 1990.

[151] M. Akke, R. Brüschweiler, and A. G. Palmer. NMR Order Parameters and Free Energy: An Analytical Approach and Its Application to Cooperative Ca^{2+} Binding by Calbindin D9k. *J Am Chem Soc*, 115:9832–9833, 1993.

[152] N. Lakomek, K. F. A. Walter, C. Farès, O. F. Lange, B. L. de Groot, H. Grubmüller, R. Brüschweiler, A. Munk, S. Becker, J. Meiler, and C. Griesinger. Self-consistent residual dipolar coupling based model-free analysis for the robust determination of nanosecond to microsecond protein dynamics. *J Biomol NMR*, 41:139–155, 2008.

[153] B. D. Hilton and C. K. Woodward. Nuclear magnetic resonance measurement of hydrogen exchange kinetics of single protons in basic pancreatic trypsin inhibitor. *Biochemistry*, 17:3325–3332, 1978.

[154] C. L. Perrin and R. E. Engler. Quantitative Assessment by 1D-EXSY NMR of Stereoelectronic Control in Acid-Catalyzed Exchange between Stereoisomeric 2-Methoxy-1,3-dioxanes and Methanol. *J Am Chem Soc*, 119:585–591, 1997.

[155] B. H. Meier and R. R. Ernst. Elucidation of Chemical Exchange Networks by Two-Dimensional NMR Spectroscopy: The Heptamethylbenzenonium Ion. *J Am Chem Soc*, 101:6441–6442, 1979.

[156] J. G. Kempf and J. P. Loria. Measurement of intermediate exchange phenomena. *Methods Mol Biol*, 278:185–231, 2004.

[157] Y. Li and A. G. Palmer. TROSY-selected ZZ-exchange experiment for characterizing slow chemical exchange in large proteins. *J Biomol NMR*, 45:357–360, 2009.

[158] J. P. Loria, M. Rance, and A. G. Palmer. A Relaxation-Compensated Carr-Purcell-Meiboom-Gill Sequence for Characterizing Chemical Exchange by NMR Spectroscopy. *J Am Chem Soc*, 121:2331–2332, 1999.

[159] M. Akke and A. G. Palmer. Monitoring macromolecular motions on microsecond to millisecond time scales by $R_{1\rho}-R_1$ constant relaxation time NMR spectroscopy. *J Am Chem Soc*, 118:911–912, 1996.

[160] A. Mittermaier and L. E. Kay. New tools provide new insights in NMR studies of protein dynamics. *Science*, 312:224–228, 2006.

[161] D. F. Hansen, P. Vallurupalli, P. Lundström, P. Neudecker, and L. E. Kay. Probing chemical shifts of invisible states of proteins with relaxation dispersion NMR spectroscopy: how well can we do? *J Am Chem Soc*, 130:2667–2675, 2008.

[162] A. Cavalli, X. Salvatella, C. M. Dobson, and M. Vendruscolo. Protein structure determination from NMR chemical shifts. *Proc Natl Acad Sci U S A*, 104:9615–9620, 2007.

[163] P. Vallurupalli, D. F. Hansen, E. Stollar, E. Meirovitch, and L. E. Kay. Measurement of bond vector orientations in invisible excited states of proteins. *Proc Natl Acad Sci U S A*, 104:18473–18477, 2007.

[164] P. Vallurupalli, D. F. Hansen, and L. E. Kay. Probing structure in invisible protein states with anisotropic NMR chemical shifts. *J Am Chem Soc*, 130:2734–2735, 2008.

[165] D. M. Korzhnev, X. Salvatella, M. Vendruscolo, A. A. Di Nardo, A. R. Davidson, C. M. Dobson, and L. E. Kay. Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature*, 430:586–590, 2004.

[166] J. M. Berrisford, A. M. Hounslow, J. Akerboom, W. R. Hagen, S. J. J. Brouns, J. Van der Oost, I. A. Murray, G. M. Blackburn, J. P. Waltho, D. W. Rice, and P. J. Baker. Evidence supporting a cis-enediol-based mechanism for Pyrococcus furiosus phosphoglucose isomerase. *J Mol Biol*, 358:1353–1366, 2006.

[167] M. Findeisen, T. Brand, and S. Berger. A ^{1}H-NMR thermometer suitable for cryoprobes. *Magn Reson Chem*, 45:175–178, 2007.

[168] S. F. Bellon, D. Chen, and E. R. Johnston. Quantitative 1D Exchange Spectroscopy. *J Magn Reson*, 73:168–173, 1987.

[169] C. L. Perrin and R. E. Engler. Weighted Linear-Least-Squares Analysis of EXSY Data from Multiple 1D Selective Inversion Experiments. *J Magn Reson*, 90:363–369, 1990.

[170] W. Massefski Jnr. and P. H. Bolton. Quantitative Analysis of Nuclear Overhauser Effects. *J Magn Reson*, 65:526–530, 1985.

[171] J. Jeener, B. H. Meier, P. Bachmann, and R. R. Ernst. Investigation of Exchange Processes by Two-dimensional NMR spectroscopy. *J Chem Phys*, 71:4546–4553, 1979.

[172] J.T. Gerig. Fluorine NMR of Proteins. *Prog Nucl Mag Res Sp*, 26:293–370, 1994.

[173] M. S. Silver, R. I. Joseph, and D. I. Hoult. Selective spin inversion in nuclear magnetic resonance and coherent optics through an exact solution of the Bloch-Riccati equation. *Phys Rev A*, 31:2753–2755, 1985.

[174] R. E Hurd and B. K. John. Gradient-Enhanced Proton-Detected Heteronuclear Multiple-Quantum Coherence Spectroscopy. *J Magn Reson*, 91:648–653, 1991.

[175] M. Piotto, V. Saudek, and V. Sklenář. Gradient-Tailored Excitation for Single-Quantum NMR-Spectroscopy of Aqueous-Solutions. *J Biomol NMR*, 2:661–665, 1992.

[176] G. N. Yip and E. R. Zuiderweg. Improvement of duty-cycle heating compensation in NMR spin relaxation experiments. *J Magn Reson*, 176:171–178, 2005.

[177] R. Ishima and D. A. Torchia. Error estimation and global fitting in transverse-relaxation dispersion experiments to determine chemical-exchange parameters. *J Biomol NMR*, 32:41–54, 2005.

[178] J. L. Markley, W. J. Horsley, and M. P. Klein. Spin-Lattice Relaxation Measurements in Slowly Relaxing Complex Spectra. *J Chem Phys*, 55:3604–3605, 1971.

[179] C. Renner, M. Schleicher, L. Moroder, and T. A. Holak. Practical aspects of the 2D ^{15}N–{^{1}H} NOE experiment. *J Biomol NMR*, 23:23–33, 2002.

[180] S. Grzesiek and A. Bax. The importance of not saturating water in protein NMR. Application to sensitivity enhancement and NOE measurements. *J Am Chem Soc*, 115:12593–12594, 1993.

[181] D. Idiyatullin, V. A. Daragan, and K. H. Mayo. Improved measurement of ^{15}N–{^{1}H} NOEs in the presence of H(N)-water proton chemical exchange. *J Magn Reson*, 153:138–143, 2001.

[182] Q. Gong and R. Ishima. ^{15}N–{^{1}H} NOE experiment at high magnetic field strengths. *J Biomol NMR*, 37:147–157, 2007.

[183] F. Ferrage, D. Cowburn, and R. Ghose. Accurate sampling of high-frequency motions in proteins by steady-state ^{15}N–{^{1}H} nuclear Overhauser effect measurements in the presence of cross-correlated relaxation. *J Am Chem Soc*, 131:6048–6049, 2009.

[184] F. Ferrage, A. Piserchio, D. Cowburn, and R. Ghose. On the measurement of ^{15}N–{^{1}H} nuclear Overhauser effects. *J Magn Reson*, 192:302–313, 2008.

[185] A. G. Palmer, C. D. Kroenke, and J. P. Loria. Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol*, 339:204–238, 2001.

[186] J. Long, T. P. Garner, M. J. Pandya, C. J. Craven, P. Chen, B. Shaw, M. P. Williamson, R. Layfield, and M. S. Searle. Dimerisation of the UBA domain of p62 inhibits ubiquitin binding and regulates NFκB signalling. *J Mol Biol*, 396:178–194, 2010.

[187] C. Wang, M. J. Grey, and A. G. Palmer. CPMG sequences with enhanced sensitivity to chemical exchange. *J Biomol NMR*, 21:361–366, 2001.

[188] D. F. Hansen, P. Vallurupalli, and L. E. Kay. An improved ^{15}N relaxation dispersion experiment for the measurement of millisecond time-scale dynamics in proteins. *J Phys Chem B*, 112:5898–5904, 2008.

[189] K Levenberg. A Method for the Solution of Certain Non-Linear Problems in Least Squares. *The Quarterly of Applied Mathematics*, 2:164–168, 1944.

[190] D. Marquardt. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *SIAM Journal on Applied Mathematics*, 11:431–441, 1963.

[191] E. J. d'Auvergne and P. R. Gooley. Optimisation of NMR dynamic models I. Minimisation algorithms and their performance within the model-free and Brownian rotational diffusion spaces. *J Biomol NMR*, 40:107–119, 2008.

[192] E. J. d'Auvergne and P. R. Gooley. Optimisation of NMR dynamic models II. A new methodology for the dual optimisation of the model-free parameters and the Brownian rotational diffusion tensor. *J Biomol NMR*, 40:121–133, 2008.

[193] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in Fortran 77: 2nd Edition. The art of Scientific Computing*. Cambridge University Press, New York, 1992.

[194] H. Akaike. Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki, editors, *Proceedings of the Second International Symposium on Information Theory*, pages 267–281. Budapest Akademiai Kiado, Budapest, 1973.

[195] K. Pearson. Correlation coefficient. In *Royal Society Proceedings*, volume 58, page 214, 1895.

[196] J. P. Carver and R. E. Richards. A General Two-Site Solution for the Chemical Exchange Produced Dependence of T_2 Upon the Carr-Purcell Pulse Separation. *J Magn Reson*, 6:89–105, 1972.

[197] D. G. Davis, M. E. Perlman, and R. E. London. Direct measurements of the dissociation-rate constant for inhibitor-enzyme complexes via the $T_{1\rho}$ and T2 (CPMG) methods. *J Magn Reson B*, 104:266–275, 1994.

[198] M. Tollinger, N. R. Skrynnikov, F. A. Mulder, J. D. Forman-Kay, and L. E. Kay. Slow dynamics in folded and unfolded states of an SH3 domain. *J Am Chem Soc*, 123:11341–11352, 2001.

[199] H. M. McConnell. Reaction Rates by Nuclear Magnetic Resonance. *J Chem Phys*, 28:430–431, 1958.

[200] J. G. de la Torre, M. L. Huertas, and B. Carrasco. HYDRONMR: Prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J Magn Reson*, 147:138–146, 2000.

[201] F. Cordier, M. Caffrey, B. Brutscher, M. A. Cusanovich, D. Marion, and M. Blackledge. Solution structure, rotational diffusion anisotropy and local backbone dynamics of Rhodobacter capsulatus cytochrome c2. *J Mol Biol*, 281:341–361, 1998.

[202] M. Blackledge, F. Cordier, P. Dosset, and D. Marion. Precision and Uncertainty in the Characterization of Anisotropic Rotational Diffusion by ^{15}N Relaxation. *J Am Chem Soc*, 120:4538–4539, 1998.

[203] A. G. Palmer, M. Rance, and P. E. Wright. Intramolecular motions of a zinc finger DNA-binding domain from Xfin characterized by proton-detected natural abundance carbon-13 heteronuclear NMR spectroscopy. *J Am Chem Soc*, 113:4371–4380, 1991.

[204] A. M. Mandel, M. Akke, and A. G. Palmer. Backbone dynamics of Escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme. *J Mol Biol*, 246:144–163, 1995.

[205] V. Y. Orekhov, D. E. Nolde, A. P. Golovanov, D. M. Korzhnev, and A. S. Arseniev. Processing of heteronuclear NMR relaxation data with the new software DASHA. *Appl Magn Reson*, 9:581–588, 1995.

[206] A. G. Palmer. NMR Probes of Molecular Dynamics: Overview and Comparison with Other Techniques. *Annu Rev Biophys Biomol Struct*, 30:129–155, 2001.

[207] D. Fushman, N. Tjandra, and D. Cowburn. An Approach to Direct Determination of Protein Dynamics from ^{15}N NMR Relaxation at Multiple Fields, Independent of Variable ^{15}N Chemical Shift Anisotropy and Chemical Exchange Contributions. *J Am Chem Soc*, 121:8577–8582, 1999.

[208] G. S. Harbison, L. W. Jelinski, R. E. Stark, D. A. Torchia, J. Herzfeld, and R. G. Griffin. ^{15}N Chemical Shift and ^{15}N-^{13}C Dipolar Tensors for the Peptide Bond in [1-^{13}C]Glycyl[^{15}N]Glycine Hydrochloride Monohydrate. *J Magn Reson*, 60:79–82, 1984.

[209] C. J. Hartzell, M. Whitfield, T. G. Oas, and G. P. Drobny. Determination of the ^{15}N and ^{13}C Chemical Shift Tensors of L-[^{13}C]Alanyl-L-[^{15}N]alanine from the Dipole-Coupled Powder Patterns. *J Am Chem Soc*, 109:5966–5969, 1987.

[210] C. D. Kroenke, M. Rance, and A. G. Palmer. Variability of the ^{15}N Chemical Shift Anisotropy in Escherichia coli Ribonuclease H in Solution. *J Am Chem Soc*, 121:10119–10125, 1999.

[211] D. Fushman, N. Tjandra, and D. Cowburn. Direct Measurement of ^{15}N Chemical Shift Anisotropy in Solution. *J Am Chem Soc*, 120:10947–10952, 1998.

[212] J. G. Sośnicki, M. Langaard, and P. E. Hansen. Long-range deuterium isotope effects on ^{13}C chemical shifts of intramolecularly hydrogen-bonded N-substituted 3-(cycloamine)thiopropionamides or amides: a case of electric field effects. *J Org Chem*, 72:4108–4116, 2007.

[213] J Griffin. *Investigations of the Metal Fluoride Transition State and Ground State Analogue Complexes of HAD Superfamily Proteins by Nuclear Magnetic Resonance Spectroscopy*. PhD thesis, University of Sheffield, 2011.

[214] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes. Funnels, Pathways, and the Energy Landscape of Protein Folding: A Synthesis. *Proteins*, 21:167–195, 1995.

[215] M. J. Parker, M. Lorch, R. B. Sessions, and A. R. Clarke. Thermodynamic properties of transient intermediates and transition states in the folding of two contrasting protein structures. *Biochemistry*, 37:2538–2545, 1998.

[216] P. Vallurupalli, G. Bouvignies, and L. E. Kay. Increasing the exchange time-scale that can be probed by CPMG relaxation dispersion NMR. *J Phys Chem B*, 115:14891–14900, 2011.

[217] P. Neudecker, P. Lundström, and L. E. Kay. Relaxation dispersion NMR spectroscopy as a tool for detailed studies of protein folding. *Biophys J*, 96:2045–2054, 2009.

[218] E. L. Kovrigin, J. G. Kempf, M. J. Grey, and J. P. Loria. Faithful estimation of dynamics parameters from CPMG relaxation dispersion measurements. *J Magn Reson*, 180:93–104, 2006.

[219] I. Ando, H. Saito, R. Tabeta, A. Shoki, and T. Ozaki. Conformation-Dependent ^{13}C NMR Chemical Shifts of Poly(L-alanine) in the Solid State: FPT INDO Calculation of N-Acetyl-N'-methyl-L-alanine Amide as a Model Compound of Poly(L-alanine). *Macromolecules*, 17:457–461, 1984.

[220] M. Iwadate, T. Asakura, and M. P. Williamson. C^{α} and C^{β} carbon-13 chemical shifts in proteins from an empirical database. *J Biomol NMR*, 13:199–211, 1999.

[221] N. Tjandra and A. Bax. Solution NMR Measurement of Amide Proton Chemical Shift Anisotropy in ^{15}N-Enriched Proteins. Correlation with Hydrogen Bond Length. *J Am Chem Soc*, 119:8076–8082, 1997.

[222] S. Kuroki, N. Asakawa, S. Ando, I. Ando, A. Shoji, and T. Ozaki. Hydrogen Bond Length and ^{15}N Chemical Shift of the Glycine Residue of Some Oligopeptides in the Solid State. *J Mol Struct*, 245:69–80, 1991.

[223] N. J. Baxter and M. P. Williamson. Temperature dependence of ^{1}H chemical shifts in proteins. *J Biomol NMR*, 9:359–369, 1997.

[224] R. Auer, D. F. Hansen, P. Neudecker, D. M. Korzhnev, D. R. Muhandiram, R. Konrat, and L. E. Kay. Measurement of signs of chemical shift differences between ground and excited protein states: a comparison between H(S/M)QC and $R_{1\rho}$ methods. *J Biomol NMR*, 46:205–216, 2010.

[225] A. R. Fersht. *Structure and Mechanism in Protein Science*. W. H. Freeman and Company, New York, 1998.

[226] A. K. Antonczak, J. Morris, and E. M. Tippmann. Advances in the mechanism and understanding of site-selective noncanonical amino acid incorporation. *Curr Opin Struct Biol*, 21:481–487, 2011.

[227] J. S. Fraser, M. W. Clarkson, S. C. Degnan, R. Erion, D. Kern, and T. Alber. Hidden alternative structures of proline isomerase essential for catalysis. *Nature*, 462:669–673, 2009.

[228] J. L. Griffin, M. W. Bowler, N. J. Baxter, K. N. Leigh, H. R. W. Dannatt, A. M. Hounslow, G. M. Blackburn, C. E. Webster, M. J. Cliff, and J. P. Waltho. Near attack conformers dominate β-phosphoglucomutase complexes where geometry and charge distribution reflect those of substrate. *Proc Natl Acad Sci U S A*, 109:6910–6915, 2012.

[229] J. Dai, L. Finci, C. Zhang, S. Lahiri, G. Zhang, E. Peisach, K. N. Allen, and D. Dunaway-Mariano. Analysis of the structural determinants underlying discrimination between substrate and solvent in β-phosphoglucomutase catalysis. *Biochemistry*, 48:1984–1995, 2009.

[230] A. Mukhopadhyay. Inclusion bodies and purification of proteins in biologically active forms. *Adv Biochem Eng Biotechnol*, 56:61–109, 1997.

[231] H. P. Sørensen and K. K. Mortensen. Advanced genetic strategies for recombinant protein expression in Escherichia coli. *J Biotechnol*, 115:113–128, 2005.

[232] J. G. Thomas, A. Ayling, and F. Baneyx. Molecular chaperones, folding catalysts, and the recovery of active recombinant proteins from E. coli. To fold or to refold. *Appl Biochem Biotechnol*, 66:197–238, 1997.

[233] P. Bernadó, J. G. de la Torre, and M. Pons. Interpretation of ^{15}N NMR relaxation data of globular proteins using hydrodynamic calculations with HYDRONMR. *J Biomol NMR*, 23:139–150, 2002.

[234] N. Tjandra, P. Wingfield, S. Stahl, and A. Bax. Anisotropic rotational diffusion of perdeuterated HIV protease from ^{15}N NMR relaxation measurements at two magnetic fields. *J Biomol NMR*, 8:273–284, 1996.

[235] J. M. Schurr, H. P. Babcock, and B. S. Fujimoto. A Test of the Model-Free Formulas - Effects of Anisotropic Rotational Diffusion and Dimerization. *J Magn Reson B*, 105:211–224, 1994.

[236] G. Cornilescu, F. Delaglio, and A. Bax. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR*, 13:289–302, 1999.

[237] P. Neudecker, P. Robustelli, A. Cavalli, P. Walsh, P. Lundström, A. Zarrine-Afsar, S. Sharpe, M. Vendruscolo, and L. E. Kay. Structure of an intermediate state in protein folding and aggregation. *Science*, 336:362–366, 2012.