Investigating Macromolecular Rate Theory

A thesis

submitted in fulfilment

of the requirements for the degree

of

Doctor of Philosophy in Biological Sciences

at

The University of Waikato

by

Erica Jean Prentice

KO TE TANGATA

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2017

Abstract

Curvature with temperature is a defining characteristic of enzyme catalysed rates. Historically, curvature has been described by Arrhenius behaviour up to a temperature optimum (T_{opt}), while decreases in rates above T_{opt} have been attributed to protein denaturation. However, various evidence indicates that denaturation is insufficient to explain the decreases in rates above T_{opt}. Macromolecular rate theory (MMRT) has been proposed as a description of the temperature dependent curvature of enzyme rates based on the expansion of established thermodynamic principals. MMRT postulates that it is the unusually large change in heat capacity associated with enzyme catalysis ($\Delta C^{\ddagger}_{\mathrm{p}}$) that results in curvature of rates with temperature. This description is commensurate with the classical description of enzyme catalysis which involves the tighter binding of the transition state (TS) to the enzyme.

Here, the molecular origins of the $\Delta C^{\ddagger}_{\mathrm{p}}$ associated with catalysis are explored through crystallography and molecular dynamics simulations. Methods are developed which calculate a $\Delta C^{\ddagger}_{\mathrm{p}}$ commensurate with experimental values from *in silico* data. Simulations reveal that global rigidification is responsible for reduced heat capacity at the TS, although some protein regions contribute more significantly to differences at the TS. Exploration of the relationship between enzyme mass and catalytic efficiency implicates the vibrational modes captured in dynamics simulations as a reservoir for energy, some of which is available to drive catalysis. This provides a theoretical framework for the large size of enzymes and extraordinary rate enhancements enzyme catalysts achieve.

Curvature with temperature is also a feature of organism growth rates and the fluxes through ecosystems. Here, an experimental validation of the applicability of MMRT to these multi-enzyme systems is presented based on the characterisation of an enzymatic pathway. The inherent temperature dependence of an enzyme pathway is found to be an average of the temperature dependence of the constituent enzymes. This relationship shows the factors influencing the rates of enzymes are relevant at the organism level, providing an intrinsic temperature response upon which regulatory responses act. This proof gives an experimental justification of the application of MMRT to organism and ecosystem temperature data.

Acknowledgements

First and foremost, I would like to thank my primary supervisor, Professor Vic Arcus for your support and the opportunities you have provided. Your guidance and expertise have been invaluable in this journey. The enthusiasm you have for this project (and outlandish ideas) has driven this research, and continues to do so. Also to Dr Jo Lane, thank you for your expertise and help as a secondary supervisor.

Thank you to all the members, past and present, in the Waikato university Proteins and Microbes lab; Dr Emma Andrews, Dr Judith Burrows, Dr Jo Abell, Dr Emma Summer, Dr Jo Hicks, Dr Tifany Oulavallickal, Abby Sharrock, Claire Mulholland, Brooke Dillon, Vikas Chonira, Kirsty Mayall, Dr Chelsea Vickers, Kirsty Kraakman, Heng Zhang, Mitchell Murray, and Konny Shim. Even when the science isn’t working, the friendliness, help and support within the group makes any problem a smaller hurdle. A special thank you to Judith for looking after everyone so well and keeping everything running smoothly. Emma S and Jo H, thank you for taking the time to read chapters of this thesis.

To Professor Adrian Mulholland and Dr Marc van der Kamp at Bristol university, thank you for giving me the opportunity to learn from you and your group, especially to Marc for teaching me from the ground up and your ongoing support. An integral part of the work here could not have been completed without your shared expertise.

Thank you to the University of Waikato for my doctoral scholarship and also the Waikato Graduate Women Educational Trust for providing the funding to support me through this research.

To friends and family, thank you for your continued support throughout my studies. To my parents, family, and friends, thank you for providing the work life balance and your encouragement. Finally, Hayden, thank you for your unwavering support throughout all my studies.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

SI (Systeme Internationale d'Unités) abbreviations are used throughout this thesis for units, elements and chemical equations. Further abbreviations are listed below.

$\Delta G^{\ddagger}_{(\text{N-U})}$	Gibbs free energy for unfolding
3D	three-dimensional
AADH	aromatic amine dehydrogenase
Abs	absorbance
ADP	adenosine-5' diphosphate
AMP	ampicillin
APS	ammonium persulfate
ATP	adenosine-5' triphosphate
BPG	1,3-bisphosphoglycerate
C	Heat capacity
C_p	Heat capacity at constant pressure
C-terminal	carboxy-terminus
Da	Daltons
DHAP	dihydroxyacetone phosphate
DNA	deoxyribonucleic acid
DONM	1-deoxynojiriycin
E_A	activation energy
EC	Enzyme commission number
*ec*DHFR	*E. coli* dihydrofolate reductase
EDTA	ethylene-diamine-tetraacetic acid (disodium salt)
F6P	fructose-6-phosphate
FBP	fructose bisphosphate aldose

FBPA	fructose bisphosphate aldose
G6P	glucose-6-phosphate
g	times the force of gravity
G	Gibbs free energy
GAFF	general amber force field
GAP	glyceraldehyde-3-phosphate
GAPD	glyceraldehyde phosphate dehydrogenase
GITC	guanidinium thiocyanate
GK	glucokinase
Glc	glucose
h	Hill coefficient
H	enthalpy
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His-tag	poly-histidine tag
HLADH	horse liver alcohol dehydrogenase
IMAC	immobilised metal affinity chromatography
IPMDH	isopropylmalate dehydrogenase
IPTG	isopropylthio-β-D-galactosidase
ITC	isothermal titration calorimetry
k	rate constant
$K_{½}$	substrate binding constant for cooperative binding
k_{B}	Boltzmann constant
Kb	kilobase
k_{cat}	enzyme rate constant
kDa	kilo Dalton
k_{H20}	rate in water

K_i	inhibition constant
KIE	kinetic isotope effect
K_M	substrate binding constant
k_non	non-catalysed rate
k_obs	observed rate
KSI	ketosteroid isomerase
k_u	rate of unfolding
LB	Luria Bertani
LeuB	isopropylmalate dehydrogenase
MADH	methylamine dehydrogenase
MalL	isomaltose
MD	molecular dynamics
MM	Michaelis Menten
MMRT	macromolecular rate theory
MR	morphinone reductase
MW	molecular weight
NAD^+	nicotinamide adenine dinucleotide (oxidised)
NADH	nicotinamide adenine dinucleotide (reduced)
$NADP^+$	nicotinamide adenine dinucleotide phosphate (oxidised)
NADPH	nicotinamide adenine dinucleotide phosphate (reduced)
N-terminal	amino terminus
OD	optical density
OD_{600}	optical density at 600 nm
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	protein data bank

PEG	polyethylene glycol
PETNR	pentaerythritol tetranitrate reductase
PFK	phosphofructokinase
PGI	phosphoglucose isomerase
pI	isoelectric point
P_i	phosphate
pKa	acid dissociation constant
Q_{10}	temperature coefficient
QMT	quantum mechanical tunnelling
R	universal gas constant (8.314 $kJ.mol^{-1}.K^{-1}$)
RMSD	root mean squared deviation
RMSF	root mean squared fluctuation
RNA	ribonucleic acid
RNase	catalytic RNA
rpm	revolutions per minute
s	second
S	entropy
SAP	shrimp alkaline phosphatase
SD	standard deviation
SDS	sodium dodecyl sulfate
SE	standard error
SEC	size exclusion chromatography
SLO-1	soy bean lipoxygenase
Sub	substrate
T	temperature
T_0	reference temperature ($T_{opt} – 4$)

TE	tris ethylene diamine tetraacetic acid
TEMED	N, N, N, N,-tetramethylethylenediamine
T_{inf}	temperature inflection point
T_{m}	temperature mid-point for unfolding
T_{opt}	temperature optimum for activity
TPI	triose phosphate isomerase
TS	transition state; also indicated with ‡
υ	wave number
V	volts
v/v	volume per volume
VMD	Visual Molecular Dynamics
w/v	weight per volume
WT	wild type

Introduction

The effect of temperature on chemical rates has been well understood for decades (Laidler & King, 1983), yet the complexities involved in enzymatic catalysis with temperature are only just beginning to be understood. Macromolecular rate theory (MMRT) has recently been the focus of several publications, providing an extension of the established kinetic understanding of chemical rates to enzymatic catalysis (Hobbs *et al.*, 2013; Schipper *et al.*, 2014; Arcus & Pudney, 2015; Arcus *et al.*, 2016). This extension is not only critical for properly describing the temperature rate relationship for enzyme-catalysed rates, but also leads to insights over a range of scales. Publications on MMRT have delved into ideas ranging from a mechanistic insight into how enzymes achieve catalytic rates many orders of magnitude faster than their chemical equivalents, to the modelling of ecosystem responses to temperature, and the implications of climate change at a global scale.

1.1 The temperature dependence of reaction rates

The first formalisation of the relationship between chemical reaction rates and temperature was proposed in the late 1800's by Svante Arrhenius. The Arrhenius equation **(1)** describes the increase in reaction rate constant (*k*) with temperature in Kelvin (T), where A is a pre-exponential factor, E_{A} is the activation energy for the reaction and R is the universal gas constant. Due to the exponential relationship between rate and temperature, linear increases in temperature lead to exponential increases in chemical reaction rate. Likewise, the E_{A} term is also included in the exponential term, so small decreases in energy at the activation barrier for a reaction result in large increases in rate.

$k = Ae^{-E_{\mathrm{A}}/RT}$ ………………………………………….. *Arrhenius equation* **(1)**

The Arrhenius equation was further developed in the 1930's to better define both the pre-exponential factor and E_{A} term based on advancements in transition state (TS) theory, resulting in the Eyring-Polanyi equation (Laidler & King, 1983). In the Eyring-Polanyi equation [**(2)** and **(3)**], reaction rate is defined with the additional terms of the Boltzmann constant (k_B), Planck's constant (h) and the difference in Gibbs free energy between the reactant state and transition state ($\Delta G^{\ddagger}$) as the activation energy equivalent. For simplicity, κ, the transmission coefficient describing the probability of the reaction coordinate proceeding from transition state to product, is assumed to be one; i.e. from the point of the transition state, the reaction is irreversible. This relationship accurately describes the temperature dependence of the rate for a wide range of chemical reactions and is still in use today.

$k = \left(\frac{k_B T}{h}\right) e^{-\Delta G^{\ddagger}/RT}$ ……………………………………………….. *Eyring equation* **(2)**

$\ln(k) = \ln\left(\frac{k_B T}{h}\right) - \frac{\Delta G^{\ddagger}}{RT}$………………………………...…………… *Eyring equation* **(3)**

It is generally assumed that $\Delta G^{\ddagger}$ of the Eyring-Polanyi equation is linearly dependent on temperature over small temperature ranges according to the Gibbs equation ($\Delta G^{\ddagger} = \Delta H^{\ddagger} - T\Delta S^{\ddagger}$, where $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ are temperature independent; equation **4**). In chemical systems over small temperatures ranges, this assumption holds true. However, does this assumption apply for all systems? First, the factors which affect $\Delta G^{\ddagger}$ will be considered and the effect of temperature on these properties. Gibbs free energy, $\Delta G^{\ddagger}$, is defined by two terms, the difference in entropy ($\Delta S^{\ddagger}$) and enthalpy ($\Delta H^{\ddagger}$) between the ground state and TS (equation **4**). The temperature dependence of $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ [equations **(5)** and **(6)** respectively; Figure 1.1 A] are both defined by $\Delta C^{\ddagger}_{p}$, the change in heat capacity between the reactant state and transition state at constant pressure, where T_0 is a suitable reference temperature. Larger values for $\Delta C^{\ddagger}_{p}$ result in a steeper dependence of $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ with temperature. This in turn causes greater curvature in $\Delta G^{\ddagger}$ (Figure 1.1 A). At low temperatures, decreasing values of $\Delta H^{\ddagger}$ drive $\Delta G^{\ddagger}$ to a minimum where $\Delta S^{\ddagger} = 0$; at higher temperatures, the increasing influence of rising $\Delta S^{\ddagger}$ values causes $\Delta G^{\ddagger}$ to increase (Figure 1.1 A).

Including the temperature dependence of $\Delta G^{\ddagger}$ into the Eyring equation **(3)**, by defining $\Delta G^{\ddagger}$ in terms of equations **(4)**, **(5)** and **(6)** gives equation **(7)**, the temperature dependence of the rate of a reaction with a non-zero $\Delta C^{\ddagger}_{p.}$ Here T_0 is an arbitrary reference temperature; 4 °C below the highest rate measured experimentally is used consistently in this work in agreement with the convention seen in Hobbs (2013) and Schipper (2014). Thus, fit values of $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ are quoted at T_0, and vary according to equations **(5)** and **(6)** respectively; $\Delta C^{\ddagger}_{\mathrm{p}}$ is independent of temperature.

$\Delta G^{\ddagger} = \Delta H^{\ddagger} - T\Delta S^{\ddagger}$ ………………………………………… *Gibbs free energy* **(4)**

$\Delta H^{\ddagger}_{T} = \Delta H^{\ddagger}_{T_0} + \Delta C^{\ddagger}_{\mathrm{p}}(T - T_0)$ …………………. *Temperature dependence of H* **(5)**

$\Delta S^{\ddagger}{}_{T} = \Delta S^{\ddagger}_{T_0} + \Delta C^{\ddagger}_{\mathrm{p}}(ln\frac{T}{T_0})$ …………………….. *Temperature dependence of S* **(6)**

$\ln(k) = \ln\left(\frac{k_{\mathrm{B}}T}{h}\right) - \frac{\left[\Delta H^{\ddagger}_{T_0} + \Delta C^{\ddagger}_{\mathrm{p}}(T - T_0)\right]}{RT} + \frac{\left[\Delta S^{\ddagger}_{T_0} + \Delta C^{\ddagger}_{\mathrm{p}}\ln(T/T_0)\right]}{R}$……………….. *MMRT* **(7)**

In equation **(7)**, the sign and magnitude of the $\Delta C^{\ddagger}_{\mathrm{p}}$ term define the direction and extent of curvature of reaction rates with temperature (Figure 1.1 B). Negative values of $\Delta C^{\ddagger}_{\mathrm{p}}$ give a negative curvature with temperature and an optimum temperature, while positive values of $\Delta C^{\ddagger}_{\mathrm{p}}$ results in positive curvature. Values for $\Delta C^{\ddagger}_{\mathrm{p}}$ further from zero give more pronounced curvature in k with temperature due to the steeper dependence of the $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ terms. When $\Delta C^{\ddagger}_{\mathrm{p}}$ is zero, equation **(7)** collapses back into a simple Eyring function (Figure 1.1 B).

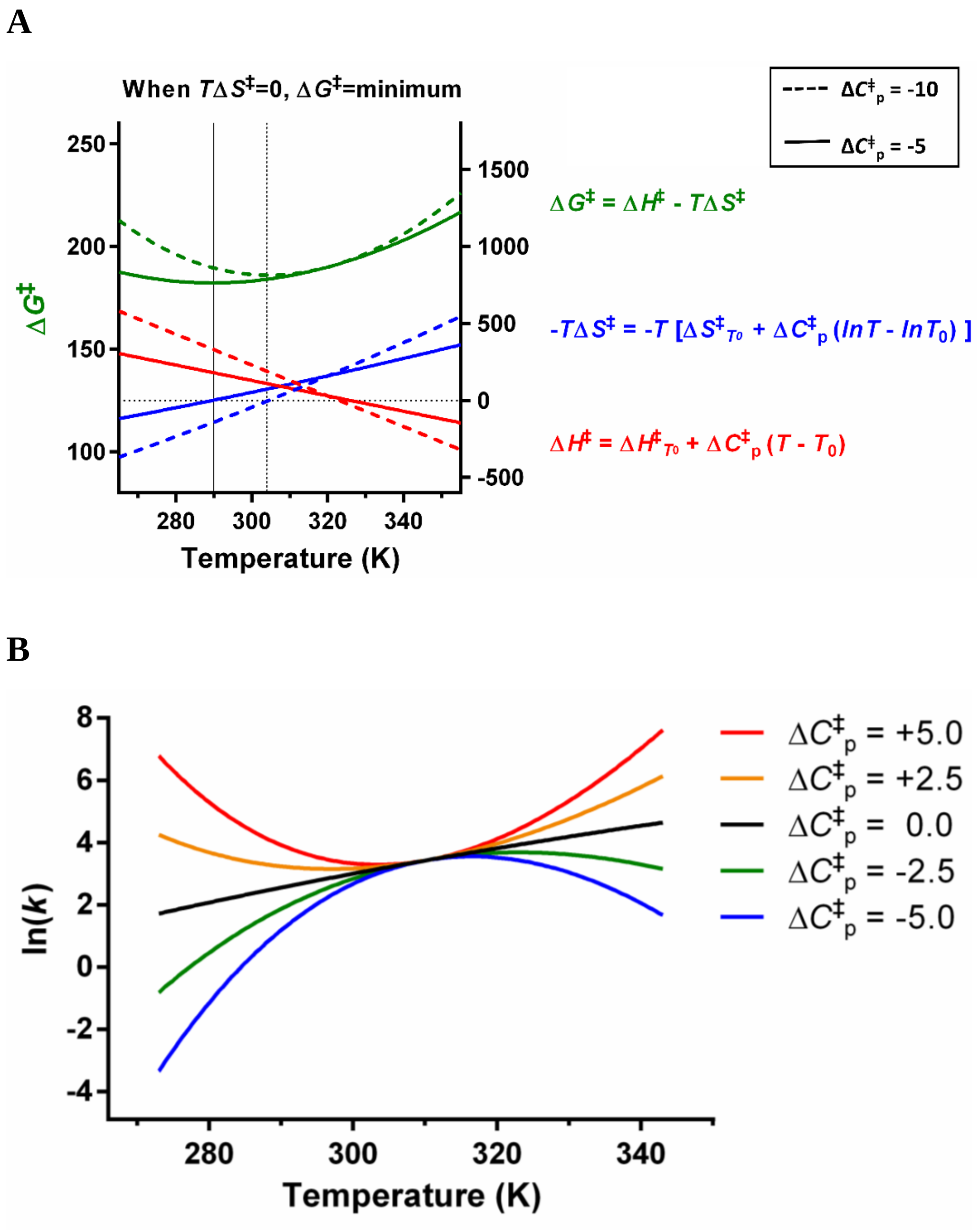


Figure 1.1: Effect of heat capacity on $\Delta G^{\ddagger}$, and the temperature dependence of reaction rates. ***(A)*** *The temperature dependence of $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ (right hand axis; kJ.mol^{-1}) due to variation in $\Delta C^{\ddagger}_{p}$, and the effect this has on $\Delta G^{\ddagger}$ (left hand axis; kJ.mol^{-1}). $\Delta H^{\ddagger}$ varies linearly with temperature, whereas $\Delta S^{\ddagger}$ varies by the natural log of the temperature. As $\Delta C^{\ddagger}_{p}$ becomes further from zero, $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ become more steeply temperature dependent, resulting in greater curvature in $\Delta G^{\ddagger}$.* ***(B)*** *The effect of the sign and magnitude of $\Delta C^{\ddagger}_{p}$ (kJ.mol^{-1}.K^{-1}) on the curvature of rates with temperature based on equation* ***(7)****. The direction and extent of curvature depends on the sign and magnitude of $\Delta C^{\ddagger}_{p}$ respectively; when $\Delta C^{\ddagger}_{p}$ is zero, equation* ***(7)*** *collapses back into simple Eyring function (black line).*

Negative curvature is a defining factor in the temperature dependence of enzymatic catalysis. Enzymes have a temperature optimum (T_{opt}) for activity, on either side of which activity steadily declines. In light of equation **(7)**, does enzyme catalysis also progress with the change in heat capacity? The classical model of enzymatic catalysis involves the progression of the weakly bound substrate (defined by K_{M}) through a transition state of high binding affinity (Pauling, 1946; Wolfenden, 1969). If binding to the transition state reduces conformational freedom in the protein (when compared to the enzyme substrate complex), there will be a negative heat capacity change associated with the catalytic process (Figure 1.2). This variable is not directly measurable over the course of enzymatic catalysis, however, with the use of transition state analogues, an approximation of the $\Delta C^{\ddagger}_{\mathrm{p}}$ associated with protein binding to the TS is experimentally accessible via isothermal titration calorimetry (ITC). For example, binding of a TS analogue to human 5′-methylthioadenosine phosphorylase enzyme has an associated ΔC_{p} of -2.5 kJ.mol^{-1}.K^{-1} (Guan *et al.*, 2011). The association of site-specific DNA binders has been characterised with ΔC_{p} values between -1.4 and -3.4 kJ.mol^{-1}.K^{-1} (Jen-Jacobson *et al.*, 2000; Bergqvist *et al.*, 2004). The binding of the iminosugar TS analogues 1-deoxynojirimycin (DONM) and isofagomine to the family GH-1 α-glucosidase of *Thermotoga maritima* is associated with a ΔC_{p} of -0.5 and -0.2 kJ.mol^{-1}.K^{-1} respectively (Zechel *et al.*, 2003). These negative changes in heat capacity are commensurate with the chemistry of the reaction where the TS binds with higher affinity in the enzyme active site than the substrate (K_{M}), and thus constrains degrees of freedom of the protein, decreasing the C_{p} of the protein-ligand complex along the reaction coordinate. From equation **(7)**, a negative $\Delta C^{\ddagger}_{\mathrm{p}}$ also corresponds to the negative curvature seen in the temperature dependence of enzyme activity. Thus, the changes in heat capacity associated with enzyme catalysis due to the strong interaction of the protein with the transition state and the effect this has on the temperature dependence of $\Delta G^{\ddagger}$ is sufficient to explain the curvature in rates with temperature in the absence of thermal denaturation (see Section 1.3.1 for comments on the influence of denaturation on enzymatic rates). This description for the temperature dependence of enzymatic rates is referred to as Macromolecular Rate Theory, MMRT, throughout this thesis (Schipper *et al.*, 2014).

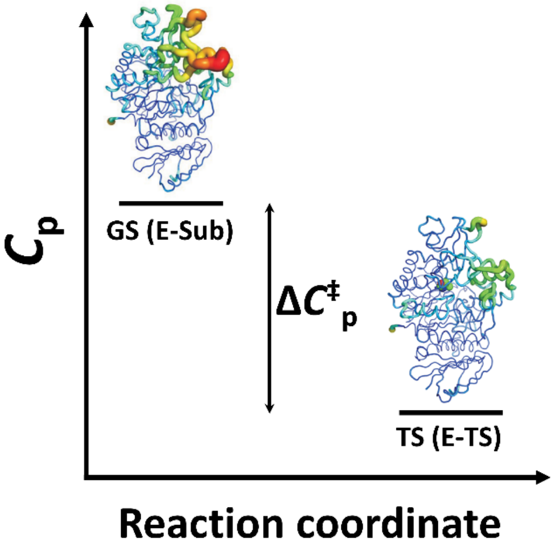


Figure 1.2: $\Delta C^{\ddagger}_{p}$ *change upon progression of an enzymatic reaction. Increasing affinity upon progression to the TS decreases* C_p *compared to the substrate bound state.*

1.2 Heat capacity of proteins

Heat capacity (C) defines the energy required to heat a compound by one degree, and is thus a measure of a systems ability to absorb energy in vibrational, rotational, translational, and electronic modes, increasing internal energy in the process. A high heat capacity means a large amount of energy is absorbed into these modes without contributing directly to temperature increases. In terms of proteins, only rotational, translational, and vibrational modes are accessible at biologically relevant temperatures. Electronic modes above the ground state are at too high an energy to be accessible at relevant temperatures, and thus are not a component of heat capacity in the conditions relevant to enzyme catalysis. Of the remaining modes, the number of vibrational modes (3N-6, where N is the number of atoms in the system) in a protein vastly outnumbers the translational and rotational modes (a total of six - three each in the x, y, and z coordinates).

Heat capacities of protein systems are consequently larger than chemical systems due to the greater number of vibrational modes present in a protein. Each amino acid adds an average of 0.25 $kJ.mol^{-1}.K^{-1}$ to the total C_p for a polypeptide chain [average calculated from the solid state $\Delta C^{\ddagger}_{p}$ values for each amino acid at 300 K (Makhatadze, 1998)]. Thus, a protein of 500 amino acids (roughly 55 kDa), has an estimated heat capacity of 125 $kJ.mol^{-1}.K^{-1}$. When considering the $\Delta C^{\ddagger}_{p}$ for a reaction catalysed by an enzyme, this parameter relates to the protein-ligand system and the changes in vibrational modes available/occupied between the starting state and TS bound form. Thus, values are expected to be, and are, larger than $\Delta C^{\ddagger}_{\mathrm{p}}$ values for a small chemical system alone. Although heat capacity changes over the course of a reaction have not been specifically considered previously for enzymes, the values presented for $\Delta C^{\ddagger}_{\mathrm{p}}$ in the literature (Hobbs *et al.*, 2013) and in this thesis are proportionate to measured values for ligand binding (including TS analogues as a proxy of TS binding during a catalysed reaction) available in the literature (Zechel *et al.*, 2003; Bergqvist *et al.*, 2004; Guan *et al.*, 2011).

The number and distribution of vibrational modes in a protein has become increasingly well understood (ben-Avraham, 1993; Na *et al.*, 2016). The total number of vibrational modes is given by 3N-6, where N is the number of atoms in the molecule. Various sources (ben-Avraham, 1993; Turton *et al.*, 2014; Na *et al.*, 2016) have shown that, at least for globular proteins, these modes are distributed in a “universal” pattern (Figure 1.3 A). Temperature dictates the probability of occupying a certain energy mode (via the Boltzmann distribution), and C_{p} in turn is determined by the energy taken to fill the available modes. When considering the $\Delta C^{\ddagger}_{\mathrm{p}}$ variable of MMRT, the changes in the vibrational mode distribution (i.e. protein dynamics) and occupancy, which are instigated by the binding of the TS, become important. Turton *et al.* (2014) showed shifts in vibrational modes (to higher energy modes) for modes in the < 200 cm^{-1} range upon binding of the non-cleavable inhibitor triacetylchitotriose to hen egg-white lysozyme (Figure 1.3 B). When bound, reduced conformational flexibility of the protein shifts vibrations to higher frequencies. With this shift, the contribution to heat capacity decreases compared to the unbound enzyme state (Figure 1.3 C).

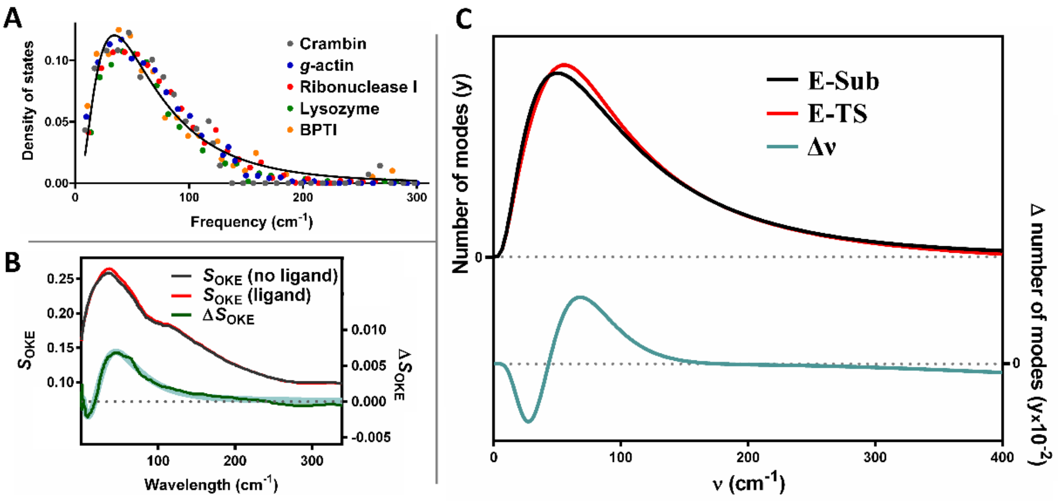


*Figure 1.3: Summary of the current literature regarding the distribution of vibrational states of a protein, and how this is effected by ligand binding, resulting in a negative $\Delta C^{\ddagger}_{p}$. **(A)** Vibrational mode distribution for a range of globular proteins (ben-Avraham, 1993). Over five globular proteins, a common distribution of modes is observed – this is modelled by a log-Gaussian function. This same universal vibrational mode distribution is also seen over 135 globular proteins in Na et al. (2016). **(B)** Effect of inhibitor binding on the distribution of states of lysozyme (Turton et al., 2014). Upon inhibitor (triacetylchitotriose) binding, vibrational modes shift to higher frequencies, especially over the 25-80 cm^{-1} range. The difference in the density of vibrational modes (green, right-hand axis) is fit with a sum-log-Gaussian function. **(C)** Proposed model for the effect of ligand binding on the distribution of available modes in a protein (personal communication, V. Arcus). The distribution of vibrational modes in a protein is roughly modelled by a log-Gaussian plot (black curve, left hand axis). The proposed model suggests that the vibrational modes shift to higher frequency upon restrictions caused by the tight binding to the TS (red line, left hand axis) – this shift occurs at the lower frequency modes of the protein, as a function of the sum-log-Gaussian curve (green, right hand axis). Curves depicted here are descriptive only, although have been based on the experimental evidence presented in panes **(A)** and **(B)**. An important caveat is that the shift in vibrational modes upon ligand binding must preserve the total number of modes at 3N-6 and thus, with this transformation, the area under the distribution of states curve must be constant.*

Enzyme hydration, specifically the interactions of the protein with closely associated solvation shells, is also a relevant factor, contributing an estimated 15 % of the *C* term for a globular protein. By comparison, vibrational modes are estimated to contribute 82 % of the heat capacity term, while the remaining three percent comprises effects from non-covalent interactions (Seewald *et al.*, 2000). It must also be noted that the relative contribution of these variables to the overall C_p changes (either for a ligand binding event, or for the tightening in binding occurring during a catalytic turn over) will vary between systems. For example, in the binding of transcription factors to promoter sequences, the change in solvation makes a significant contribution to the ΔC_p (Bergqvist *et al.*, 2004). Similarly, nitrogenase has been characterised with a significant contribution to ΔC_p from changes in solvation (Mayweather *et al.*, 2012), which contribute to non-Arrhenius behaviour. In the latter case, a domain change is associated with the reaction, in which case it is unsurprising that solvation has a more pronounced effect on ΔC_p, and may well be a general trend. Overall, given the close association of proteins with the solvent environment, it is reasonable to consider that the tightly bonded solvation shell has some currently unexplored relevance in MMRT.

1.3 Alternate theories for the temperature dependence of enzymatic catalysis

1.3.1 Protein denaturation

Various explanations have been put forward previously to explain the curvature in the temperature dependence of enzymatic rates. The most widely cited of these invokes denaturation at temperatures above the optimum. Here, increases in rate up to T_{opt} are attributed to Arrhenius behaviour (equation **1**), while increasing levels of denaturation at temperatures above T_{opt} are cited to explain the decreases in rate at high temperatures. This implies that there is no intrinsic T_{opt} to enzyme activity, as the observed phenomena is a function of multiple parameters (Daniel *et al.*, 2001). This explanation prevails despite numerous experimental discrepancies indicating denaturation is insufficient to account for the observed trend. For example, wild type barnase (EC 3.1.27.3) exhibits a T_{opt} of 312 K and a temperature mid-point of unfolding (T_m) of 326 K (Hobbs *et al.*, 2013). Thus, denaturation could conceivably be an influential factor at assay temperatures above T_{opt}. However, the A43C/S80C double mutant, with a stabilising disulfide bond, exhibits the same T_{opt} despite having a heightened T_m of 331 K. Therefore, denaturation is not a sufficient explanation of the temperature dependence of the stabilised double cysteine mutant. Denaturation has also been ruled out as the defining factor in rate decreases in phosphoglycerate kinase (Thomas & Scopes, 1998). In a comprehensive characterisation of temperature profiles and residual activity to determine denaturation rates with temperature, irreversible denaturation was found to not be sufficient to explain T_{opt} profiles.

Further discrepancies are seen in psychrophilic enzymes, which commonly have large differences between T_{opt} and T_{m} (Feller & Gerday, 2003). For example, α-amylase from the psychrophile *Pseudoalteromonas haloplanktis* exhibits a T_{opt} of 30 °C, significantly lower than the T_{m} at 45 °C (D'Amico *et al.*, 2003). In this instance, activity has already peaked and decreased down to almost half maximum rates at temperatures where denaturation begins to occur (Figure 1.4). This is in contrast to the example mesophilic and thermophilic α-amylases from pig pancreas and *Bacillus amyloliquefaciens* which exhibit closer consensus between T_{opt} and T_{m} temperatures [see D'Amico *et al.* (2003) Figure 4]. In order to account for this discrepancy, the concept of "localised flexibility" has been evoked to describe an active site which is more heat labile than the protein as a whole (D'Amico *et al.*, 2003)

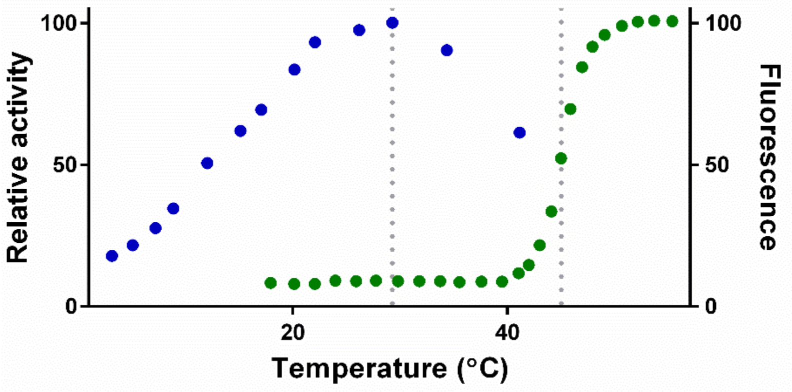


Figure 1.4: Discrepancy between activity (blue) and denaturation (green) for α-amylase from P. haloplanktis. T_{opt} and T_m are indicated with vertical dashed lines. Data from D'Amico et al. (2003).

It should be noted that denaturation is still important to consider for enzyme catalysis, especially at temperatures significantly above T_{opt} and at elevated temperatures over long time scales.

1.3.2 Equilibrium model

Following the conclusive data showing that denaturation was not sufficient to explain decreases in rates in variants of phosphoglycerate kinase (Thomas & Scopes, 1998), Daniel *et al.* (2001) published the equilibrium model that gave a denaturation independent explanation of enzymatic temperature behaviour. To describe the observed T_{opt}, the equilibrium model proposes a folded but inactive (or reduced activity) intermediate enzyme state, E_{inact}, in rapid equilibrium with the folded active enzymatic form (E_{act}) (Daniel *et al.*, 2001; Daniel & Danson, 2010). Irreversible unfolding to the observable thermally unfolded (X) state then proceeds from the E_{inact} form:

$$E_{act} \rightleftharpoons E_{inact} \rightarrow X$$

Under this model, catalytic rate is dependent on the concentration of E_{act} as well as k_{cat}. Temperature influences both of these parameters; k_{cat} as a function of the Arrhenius equation, and the concentration of E_{act} by altering the equilibrium constant (K_{eq}) between the active and inactive folded enzymatic forms. Decreases in rates with increased temperature come into effect as the equilibrium tends towards a greater proportion on E_{inact}.

Mathematical models from this idea fit experimental data from over 40 studied enzymes at time zero when denaturation has no effect on rates (Lee *et al.*, 2007), and as such is proposed as a universal phenomenon independent of quaternary structure variations and reaction type (Daniel *et al.*, 2009). Despite this fit to numerous data sets, the theory remains lacking in a formal description of the E_{inact} enzyme form. The $E_{\mathrm{act}} \rightleftharpoons E_{\mathrm{inact}}$ equilibrium is proposed to be rapid, up to two orders of magnitude faster than denaturation. From this, the reversibility of the conformational change and the low ΔH_{eq} attributed to the barrier between the two species, E_{inact} is proposed to more closely resemble the folded active state than the unfolded form (Peterson *et al.*, 2004; Daniel *et al.*, 2009). Proposed changes occur in the weak stabilising interactions, localised in the active site vicinity, although there is little experimental evidence to support this claim.

1.3.3 Multi state model

Recently an alternative model has been proposed which considers a conformational ensemble of protein forms over a free energy landscape (Elias *et al.*, 2014). Of these states, there are variations in partially active yet folded states. Temperature increases the breadth of the free energy landscape, increasing structural variability and diluting the highly active conformational states. Elias and colleagues estimate that for the average enzyme, fewer than 1 in 10000 enzyme-substrate encounters result in product formation, and this ratio is dependent on the spread of activities over conformational space, as well as potential sub-states for the substrate molecule (Bar-Even *et al.*, 2015).

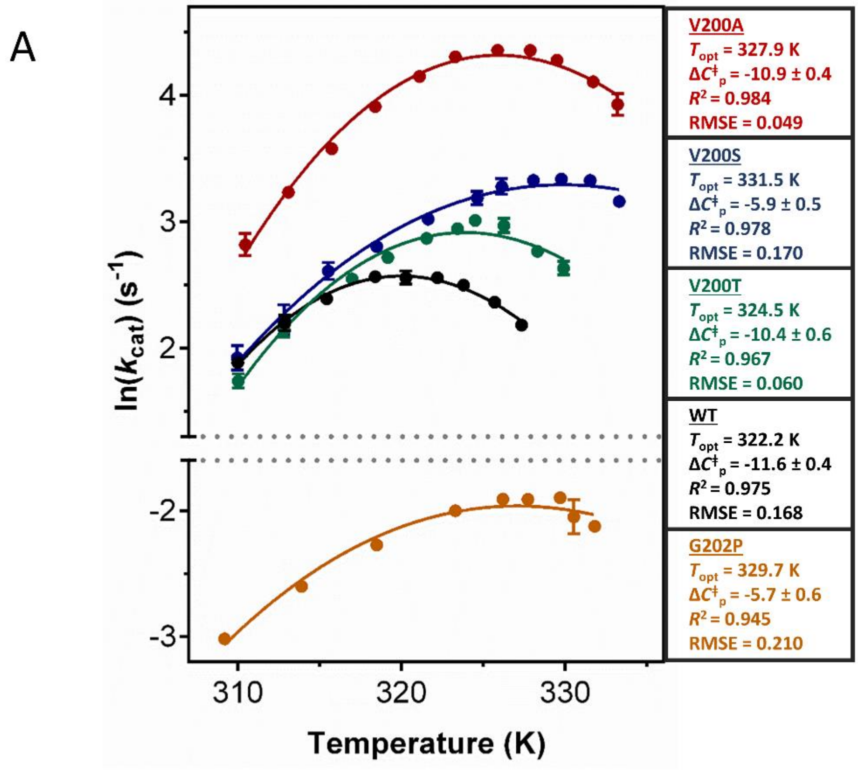
Again, the structural properties of the conformational ensemble are anticipated to include changes to backbone configurations, side chain states, bond lengths and angles, but is not backed up by experimental evidence. The major difference between MMRT and both the equilibrium model and multi state model is the $\Delta C^{\ddagger}_{p}$ term. Both alternate models assume $\Delta C^{\ddagger}_{p}$ = zero, however this parameter has been shown experimentally to be negative (Section 1.1), conducive with the treatment of temperature dependence under the MMRT model.

1.4 Experimental basis and implications of Macromolecular Rate Theory

The strength of any theory lies in its predictive power, through extrapolation to testable implications. From MMRT, a number of predictions have been projected, ranging from the mechanistic view of how enzymes work and enzymatic adaptation to temperature, to the temperature response of organisms and ecosystems. Some of these predictions, along with the testing of these and further implications to date are laid out below.

1.4.1 The relationship between $\Delta C^{\ddagger}_{p}$ and T_{opt}

Based on the potential application of the MMRT equation **(7)**, this model has been applied to experimental data. For the enzyme MalL (α-glucosidase/isomaltase; EC 3.2.1.10), four single amino acid mutants with increased T_{opt} were developed through site directed mutagenesis: V200A, V200S, V200T and G202P (Hobbs *et al.*, 2013). Variants V200A and V200T show slight increases in T_{opt} (ΔT_{opt} of 5.7 and 2.3 K respectively), and these are accompanied by small increases in $\Delta C^{\ddagger}_{p}$ compared to wild type (WT) MalL ($\Delta\Delta C^{\ddagger}_{p}$ of 0.7 and 1.2 kJ.mol^{-1}.K^{-1} respectively; Figure 1.5). The mutants V200S and G202P show larger increases in both T_{opt} and $\Delta C^{\ddagger}_{p}$ (ΔT_{opt} 9.3 and 7.3 K; $\Delta\Delta C^{\ddagger}_{p}$ of 5.7 and 5.9 kJ.mol^{-1}.K^{-1} respectively). Of these four variants, only G202P shows a significant reduction in k_{cat} (Figure 1.5). For the other variants, T_{opt} has been altered without effect on the efficiency of catalysis. Variants V200S and V200T show very similar k_{cat} values in the low temperature range compared to the WT enzyme (Figure 1.5 B).



B

MalL variant	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($10^3.s^{-1}.\mu M^{-1}$)
WT	205 ± 19	29.0 ± 0.6	141 ± 14
V200A	1560 ± 161	125 ± 4	80 ± 10
V200S	105 ± 11	36.0 ± 0.8	343 ± 40
V200T	122 ± 12	36.0 ± 0.7	295 ± 32
G202P	92 ± 13	0.100 ± 0.003	1.1 ± 0.2

Figure 1.5: Experimental data for MalL variants, adapted from (Hobbs et al., 2013). ***(A)*** *Temperature data fit with the MMRT equation* ***(7)****. Data shown is expressed as k_{cat} to account for enzyme concentration; values are the mean of three replicates, and error bars, where visible, are the SD of three replicates. T_{opt} values are taken as the temperature where the maximal activity was observed in the temperature data. $\Delta C^{\ddagger}_{p}$ values are in $kJ.mol^{-1}.K^{-1}$, and are quoted with the SE as determined by curve fit to the data.* ***(B)*** *Michaelis Menten parameters for the MalL variants, as calculated at 318 K. Standard error values from the fit to the data are given.*

The correlation between T_{opt} and $\Delta C^{\ddagger}_{\text{p}}$ observed in the MalL mutants agrees with theory. Equation **(8; a and b)** defines T_{opt} when the first derivative of equation **(7)** is equal to zero, given the $\Delta C^{\ddagger}_{\text{p}}$ is a non-zero value. The derivation of equation **(8)** is given in the box above the equation. Equation **(8b)** shows that changing $\Delta C^{\ddagger}_{\text{p}}$

alone is sufficient to alter T_{opt} without altering $\Delta H^{\ddagger}{}_{T0}$, hence having no impact on $\Delta G^{\ddagger}$ (activation energy) for the reaction. This relationship between T_{opt} and $\Delta C^{\ddagger}_{\mathrm{p}}$ extends to other enzymes over a range of temperatures, as shown by the relationship between temperature and mass corrected $\Delta C^{\ddagger}_{\mathrm{p}}$ values over various LeuB (EC 1.1.1.85), barnase and MalL enzyme variants [see (Hobbs *et al.*, 2013), Figure 3].

$$\ln(k) = \ln\left(\frac{k_{\mathrm{B}}T}{h}\right) - \frac{\left[\Delta H^{\ddagger}_{T_0} + \Delta C^{\ddagger}_{\mathrm{p}}(T - T_0)\right]}{RT} + \frac{\left[\Delta S^{\ddagger}_{T_0} + \Delta C^{\ddagger}_{\mathrm{p}}\ln(T/T_0)\right]}{R} \quad \ldots\ldots\ldots\ldots\ldots\ldots \mathbf{(7)}$$

Expanded

$$\ln(k) = \ln\left(\frac{k_{B}T}{h}\right) - \frac{[\Delta H^{\ddagger}_{T0}]}{RT} - \frac{[\Delta C^{\ddagger}_{\mathrm{p}}(T - T_0)]}{RT} + \frac{[\Delta S^{\ddagger}_{T0}]}{R} + \frac{[\Delta C^{\ddagger}_{\mathrm{p}}]}{R}\ln\left(\frac{T}{T_0}\right)$$

First derivative

$$\frac{d\ln(k)}{dT} = \left(\frac{1}{T}\right) + \frac{[\Delta H^{\ddagger}_{T0}]}{RT^2} - \frac{[\Delta C^{\ddagger}_{\mathrm{p}}T_0]}{RT^2} + \frac{[\Delta C^{\ddagger}_{\mathrm{p}}]}{RT}$$

Set the slope to zero to define T_{opt}

$$0 = \left(\frac{1}{T}\right) + \frac{[\Delta H^{\ddagger}_{T0}]}{RT^2} - \frac{[\Delta C^{\ddagger}_{\mathrm{p}}T_0]}{RT^2} + \frac{[\Delta C^{\ddagger}_{\mathrm{p}}]}{RT}$$

Rearranged

$$0 = \left(R + \Delta C^{\ddagger}_{p}\right)T_{opt} + \Delta H^{\ddagger}_{T_0} - \Delta C^{\ddagger}_{\mathrm{p}}T_0$$

Rearranged to make $\Delta C^{\ddagger}_{p}$, T_{opt} and $\Delta H^{\ddagger}$ the subjects to give equations ***(8a-c)***.

$$-\Delta C^{\ddagger}_{\mathrm{p}} = \frac{\Delta H^{\ddagger}_{T0} + RT_{opt}}{T_{opt} - T_0} \quad \ldots\ldots\ldots\ldots\ldots\ldots \mathbf{(8a)}$$

$$T_{opt} = \frac{-\Delta H^{\ddagger}_{T0} + \Delta C^{\ddagger}_{p}T_0}{R + \Delta C^{\ddagger}_{p}} \quad \ldots\ldots\ldots\ldots\ldots\ldots \mathbf{(8b)}$$

$$\Delta H^{\ddagger} = -RT_{opt} \quad \ldots\ldots\ldots\ldots\ldots\ldots \mathbf{(8c)}$$

1.4.1.1 Application to previous anomalies in the literature

Looking back at the barnase anomaly presented earlier (Section 1.3.1), the lack of difference in T_{opt} between the wild type and disulfide stabilised mutant can be rationalised by looking at the system in the context of MMRT and the relationship between T_{opt} and $\Delta C^{\ddagger}_{\text{p}}$. Both of the variants have similar $\Delta C^{\ddagger}_{\text{p}}$ values based on the fit of the data to equation **(7)** of -4.6 ± 0.8 and -5.3 ± 0.4 $\text{kJ.mol}^{-1}.\text{K}^{-1}$ respectively for wild type and stabilised barnase. As $\Delta C^{\ddagger}_{\text{p}}$ has not been greatly altered in the system, T_{opt} remains the same despite the alterations to the denaturation temperature of the mutant enzyme.

MMRT also disconnects the prevailing association between temperature profiles and denaturation, eliminating the issues surrounding the discrepancy between reduced rates and denaturation temperature in psychrophiles. This removes the necessity of the *ad hoc* explanation presented for this anomaly involving higher lability of the active site than the rest of the thermophilic enzyme’s structure (D'Amico *et al.*, 2003). Further implications of MMRT to psychrophilic enzymes are presented in Section 1.4.3.

1.4.2 Mechanisms of change for $\Delta C^{\ddagger}_{\text{p}}$

To instigate a change in $\Delta C^{\ddagger}_{\text{p}}$ for an enzymatic reaction, there needs to be changes in the C_{p} of the ground state (free enzyme ⇄ enzyme-substrate) and/or the enzyme-TS complex. To investigate this, the set of MalL mutants has been further used to explore the mechanism by which $\Delta C^{\ddagger}_{\text{p}}$ has been altered in V200S and G202P (Hobbs *et al.*, 2013). Crystallographic data sets of apo forms of MalL WT, V200S and G202P were coupled with molecular dynamics (MD) simulations to investigate the conformational dynamics changes caused by these mutations. The crystallographic structures of the three structures are very similar (RMSD values of 0.18 and 0.20 Å between WT - V200S and WT - G202P respectively), however, significant differences in the dynamics were observed. For the V200S variant, dynamics are greatly reduced (i.e. the range of thermally accessible conformational states is limited) compared to the apo WT simulations (Figure 1.6). This suggests that the change in $\Delta C^{\ddagger}_{\text{p}}$ is due to a rigidification of the ground state of V200S, decreasing C_{p}. This is consistent with the kinetic data for the variant which shows no change in k_{cat} associated with the mutation, indicating binding to the TS is unaltered. Analysis of the crystal structures shows the cause of this rigidification in the introduction of two new hydrogen bonds forming a network to a flexible loop neighbouring the active site [see Hobbs (2013), supplementary Figure 1; See also

Figure 5.7]. By comparison, G202P has very similar ground state dynamics to WT in the molecular dynamics. For this variant, it is proposed that the $\Delta C^{\ddagger}_{\mathrm{p}}$ has been altered by increasing the flexibility of the TS complex. This change has come at a cost to the k_{cat} of the G202P variant, with a rate constant significantly lower than that of the WT enzyme (Figure 1.5 B).

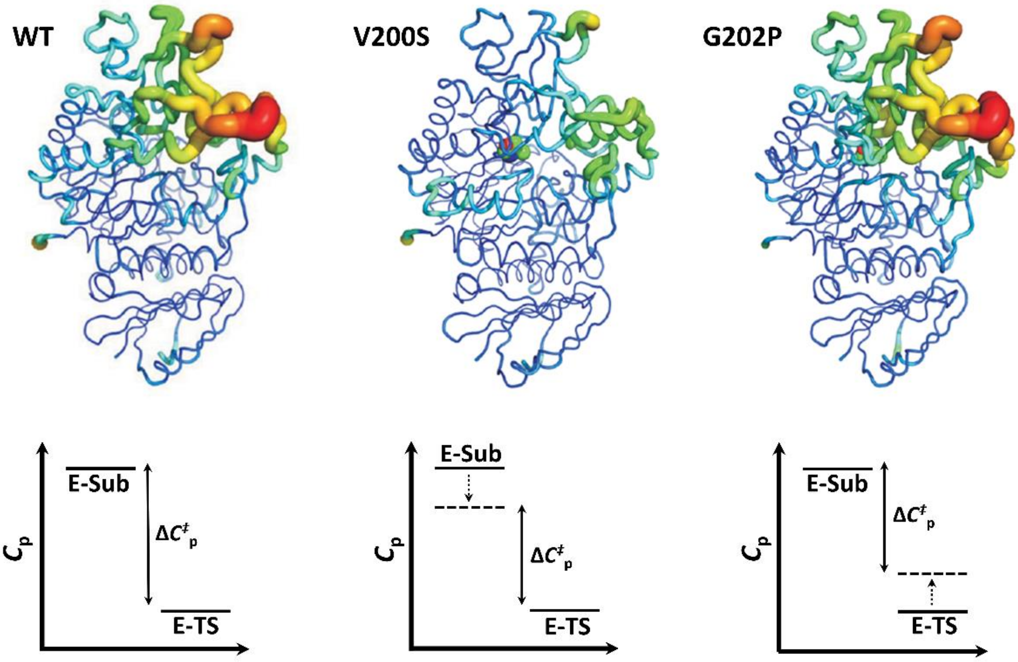
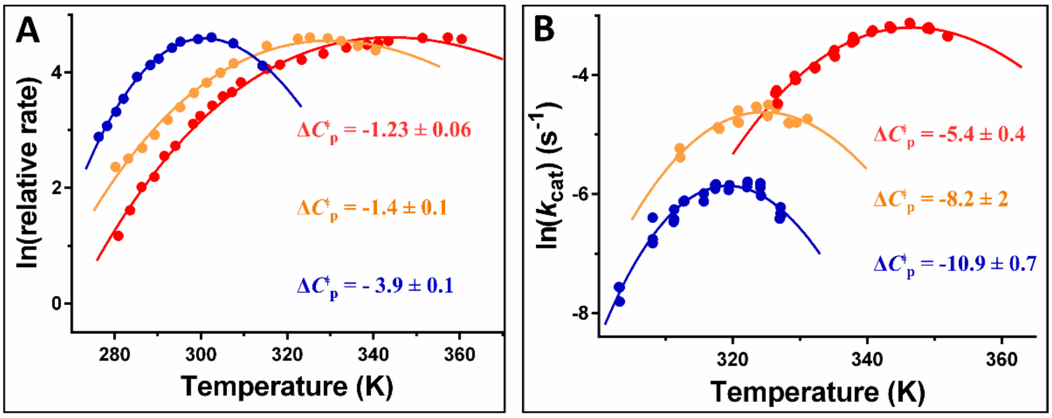


Figure 1.6: Representation of the dynamics of MalL variants WT, V200S and G202P. Structures depict the RMSD over 270 ns of simulation coloured red through blue from most to least flexible regions (Hobbs et al., 2013). The active site is indicated with red spheres (catalytic residues Asp199 and Glu255), and mutation positions with purple spheres. From the observed alterations at the ground state for V200S, the proposed mechanism for $\Delta\Delta C^{\ddagger}_{p}$ *is given in the lower panel. For G202P, the proposed mechanism of* $\Delta\Delta C^{\ddagger}_{p}$ *involves increases in flexibility of the TS complex due to similar flexibility as WT at the ground state. This is consistent with the compromised catalysis of G202P (Figure 1.5 B).*

1.4.3 Adaptation of psychrophilic and thermophilic enzymes

The relationship between $\Delta C^{\ddagger}_{\mathrm{p}}$ and T_{opt} (Section 1.4.1) has implications for the adaptation of enzymes to psychrophilic or thermophilic environments (Arcus *et al.*, 2016). Equation **(8b)** implies that $\Delta C^{\ddagger}_{\mathrm{p}}$ must alter to match T_{opt} to the environmental temperature. This trend is evident in the same enzyme from organisms with different temperature optima (Figure 1.7 A and Figure 1.7 B). In the case of the adaptation of psychrophilic enzymes, this relationship dictates that $\Delta C^{\ddagger}_{\mathrm{p}}$ must become large and negative to lower the T_{opt}, resulting in greater curvature in rates with temperature. This increased curvature results in a ‘psychrophilic trap’ at low temperatures as enzymes present greater sensitivity to small changes in environmental temperature. This factor may explain the rarity of true psychrophiles [organisms that predominantly grow at temperatures less than ~15 °C and are unable to develop at mesophilic temperatures (Feller & Gerday, 2003)] and the relative abundance of organisms adopting a cold tolerant approach to cooler temperatures (Feller *et al.*, 1996). If true psychrophiles are to be evolutionarily fit, stable environmental temperatures are essential.

In order to achieve large negative $\Delta C^{\ddagger}_{p}$ values, enzymes must either have increased C_p of the enzyme-substrate complex, or decrease the C_p of the enzyme-TS complex. The effect of both strategies is evident in psychrophilic enzymes. Compared to their mesophilic counterparts, psychrophilic enzymes generally have increased K_M, and faster k_{cat} values [Figure 1.7 C (Fields & Somero, 1998; D'Amico *et al.*, 2006)]. These measured kinetic changes are indicative of decreased binding strength of the substrate to increase C_p, and tighter binding of the TS to lower C_p respectively.



C

	Psychrophile		Mesophile	
Substrate	K_M (mg.L^{-1})	k_{cat} (s^{-1})	K_M (mg.L^{-1})	k_{cat} (s^{-1})
Starch	155	663	41	327
Amylopectin	258	636	53	222
Amylose	178	2148	36	700
Dextrin	586	716	61	311
Glycogen	1344	491	46	193

Figure 1.7: The effect of temperature adaptation on enzymes. ***(A)*** *Temperature dependence of activity for a psychrophilic Pseudoalteromonas haloplanktis (blue), mesophilic pig pancreatic (orange), and thermophilic Bacillus amyloliquefaciens (red) α-amylase (D'Amico et al., 2003).* ***(B)*** *The temperature dependence of isopropylmalate dehydrogenase (IPMDH) enzymes from psychrophilic (blue), mesophilic (orange) and thermophilic (red) Bacillus species (Hobbs et al., 2012; Arcus et al., 2016) Adaptation to colder temperatures steadily decreases $\Delta C^{\ddagger}_p$ in both amylase and IPMDH data.* ***(C)*** *Comparison of Michalis Menten parameters from a psychrophilic (Pseudoalteromonas haloplanktis) and mesophilic (pig) α-amylase, at 25 °C, on substrates of various lengths (D'Amico et al., 2006) illustrating the increased K_M and k_{cat} associated with the adaptation of psychrophilic enzymes to low temperature.*

In the instance of thermophilic enzymes, less negative $\Delta C^{\ddagger}_{\mathrm{p}}$ values are necessary to increase T_{opt}. This has the favourable consequence of resulting in less curvature with temperature, approaching Arrhenius behaviour, and thus less temperature sensitivity.

The molecular basis for $\Delta C^{\ddagger}_{\mathrm{p}}$ in terms of the adaptation of psychrophiles and mesophiles also fits with the observed relative flexibilities of these enzymes. The necessity of greater C_{p} in the enzyme-substrate complex for psychrophilic enzymes provides an explanation for the generally high flexibility of psychrophilic enzymes regardless of any reduced requirement for stability at low temperatures. For thermophiles, the opposite requirement for lesser C_{p} in the enzyme-substrate complex drives more rigid structures (Vieille & Zeikus, 2001). This is the same mechanism by which the MalL V200S mutant has gained a higher T_{opt} (Section

1.4.2). Additionally, the MalL mutant G202P illustrates the other potential mechanism for $\Delta C^{\ddagger}_{\mathrm{p}}$ change in high temperature adaptation; increasing the C_{p} of the TS complex. It is immediately clear why this strategy is highly improbable in evolutionary terms, given the significant decrease in G202P rates to gain a 7.3 degree T_{opt} increase. This factor may explain the apparent agreement in observed high stability in thermophilic enzymes (Vieille & Zeikus, 2001). Conversely for psychrophiles, decreasing C_{p} at the TS has the favourable consequence of increasing k_{cat}. This is evidently implemented to an extent (Figure 1.7 C), however, given that k_{cat} values are already highly optimised, one could imagine that alterations at the enzyme-substrate are also necessary/more easily accessible, despite the disfavourable increases in K_{M} affected by this. For psychrophiles, temperature adaptation is thus a balance of offsetting increases in K_{M} with improved turnover rates ($k_{\mathrm{cat}}/K_{\mathrm{M}}$).

1.4.4 Inclusion of a temperature dependent $\Delta C^{\ddagger}_{\mathrm{p}}$ parameter

Inclusion of $\Delta C^{\ddagger}_{\mathrm{p}}$ to the Eyring-Polanyi equation accounts for the temperature dependence of $\Delta S^{\ddagger}$ and $\Delta H^{\ddagger}$, however, $\Delta C^{\ddagger}_{\mathrm{p}}$ can also exhibit temperature dependence (Makhatadze & Privalov, 1990; Privalov & Makhatadze, 1990). Equation **(9)** describes the temperature dependence of $\Delta C^{\ddagger}_{\mathrm{p}}$ (Prabhu & Sharp, 2005); inclusion of this into MMRT (equation **7**) results in equation **(10)**.

$\Delta C^{\ddagger}_{p} = \Delta C^{\ddagger}_{p,T_0} + A(T - T_0)$............................ *Temperature dependence of* $\Delta C^{\ddagger}_{p}$ **(9)**

$$\ln(k) = \ln\left(\frac{k_{\mathrm{B}}T}{h}\right) - \frac{\Delta H^{\ddagger}_{T_0} + \left[\Delta C^{\ddagger}_{p,T_0} + A(T-T_0)\right](T-T_0)}{RT} + \frac{\Delta S^{\ddagger}_{T_0} + \left[\Delta C^{\ddagger}_{p,T_0} + A(T-T_0)\right]\ln\left(\frac{T}{T_0}\right)}{R}$$

…………………………...……. *MMRT including a temperature dependent* $\Delta C^{\ddagger}_{p}$ **(10)**

Under this model, $\Delta C^{\ddagger}_{\mathrm{p}}$ is linearly dependent on temperature, and, if A < 0, becomes more negative with increasing temperature (Figure 1.8 A). This imparts a curved response to temperature of $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$. The interplay between these two functions on $\Delta G^{\ddagger}$ gives a skewed curvature to enzymatic rates, with a tailed reduction in rates at low temperature. This fits the experimental data for MalL WT over a wider range of temperatures (Figure 1.8 B). Over the expanded range of temperatures presented in Figure 1.8 B (Kraakman, 2017), the temperature independent MMRT form fits poorly. It is only when temperature is sampled over a restricted range where $\Delta C^{\ddagger}_{\mathrm{p}}$

is essentially constant that this form fits well, as was the case in the original published data [(Hobbs *et al.*, 2013); Figure 1.5 A].

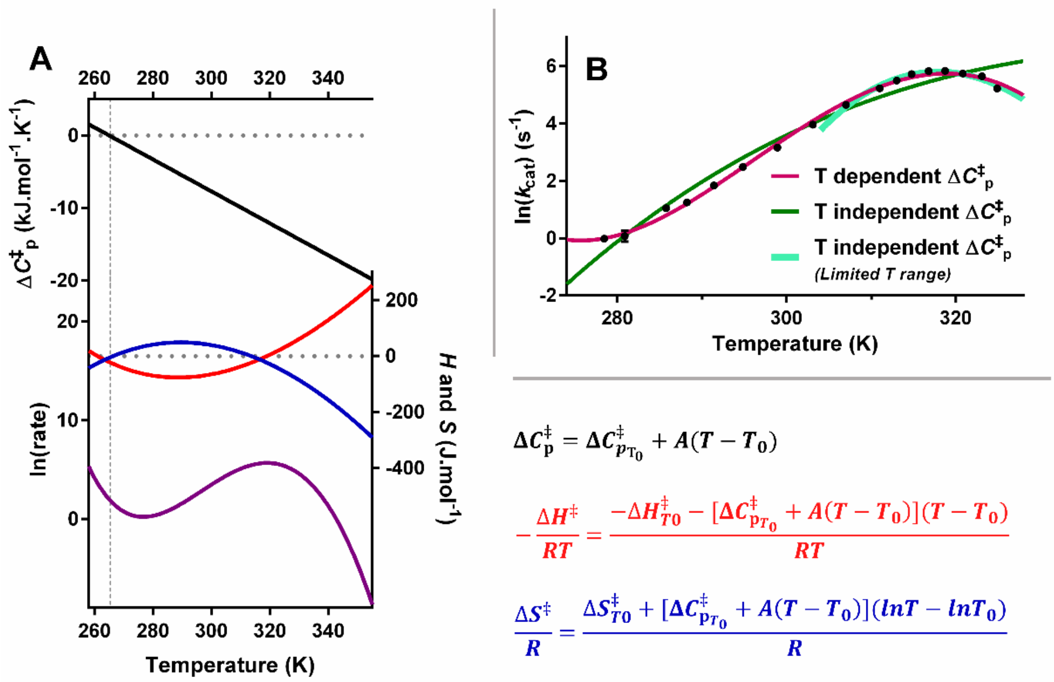


Figure 1.8: The effect of inclusion of a temperature dependent $\Delta C^{\ddagger}_{p}$ into MMRT. ***(A)*** *The temperature dependence of $\Delta C^{\ddagger}_{p}$ (black), and the effect this has on $-\Delta H^{\ddagger}/RT$ (red) and $\Delta S^{\ddagger}/R$ (blue), and thus overall rates (purple; equation* ***10****).* ***(B)*** *Experimental data for MalL WT over an extended temperature range illustrating improved fit to the temperature dependent $\Delta C^{\ddagger}_{p}$ equation (Kraakman, 2017). The fit of the temperature independent equation* ***(7)*** *is also shown over the whole temperature range and limited to the range presented in Hobbs et al. (2013). The original equation is suitable providing the temperature range is limited so that $\Delta C^{\ddagger}_{p}$ is close to constant/$-\Delta H^{\ddagger}/RT$ and $\Delta S^{\ddagger}/R$ are in the linear region of the curve illustrated in* ***(A)****.*

Equation **(10)** produces a rapid rise in rates with decreasing temperature below the familiar temperature curve (Figure 1.8 A below 280 K). In terms of enzyme catalysis, this represents an unrealistic scenario, as this region corresponds to $\Delta C^{\ddagger}_{p}$ approaching zero and becoming positive. Experimentally, rates are expected to approach zero at low temperatures and it appears that $\Delta C^{\ddagger}_{p}$ also approaches zero.

It is worth noting that the temperature profile resulting from a temperature dependent $\Delta C^{\ddagger}_{p}$ has been described in the literature previously under the Equilibrium model (Daniel *et al.*, 2009). From the Equilibrium model, this phenomenon is described as arising from the magnitude of ΔH_{eq}, the enthalpy change associated with the transition to the reversibly unfolded inactive state (Section 1.3.2).

1.4.5 $\Delta C^{\ddagger}_{P}$ as a predictor of vibrational coupling in hydrogen transfer reactions

The role of enzyme dynamics in enzyme catalysis, especially in terms of vibrational coupling between motions of the protein/substrate and the reaction coordinate is highly contentious in the literature (Schwartz & Schramm, 2009; Kamerlin & Warshel, 2010). Proteins exhibit motion over a wide range of time scales and degrees of movement, from domain shifts on the scale of milliseconds, to pico-nanosecond bond fluctuations (Henzler-Wildman *et al.*, 2007). It is well established that a range of enzymes use significant intra-molecular movements as part of the reaction mechanism. For example, the catalytic activity of adenylate kinase is accompanied by a lid opening-closing motion (Henzler-Wildman *et al.*, 2007), while 4-chlorobenzoate:CoA ligase undergoes a 140 ° domain rotation to switch between reactive states for each stage of the two step reaction (Reger *et al.*, 2008). However, the extent of dynamic contribution at the moment of the catalytic turnover, and specifically contributing to temporarily lowering the energetic barrier to reaction (Kandathil *et al.*, 2014), as opposed to active site preorganisation alone, is not well understood.

Vibrational coupling has been most extensively researched in the context of hydrogen transfer reactions (i.e. quantum mechanical tunnelling, QMT, reactions involving the transfer of H^-, $H^\bullet$ or H^+). Several QMT reactions have been investigated with respect for vibrational coupling; pentaerythritol tetranitrate reductase [PETNR, with NADH and NADPH cofactor (Pudney *et al.*, 2009; Pudney *et al.*, 2013)], morphinone reductase [MR, with NADH (Hay *et al.*, 2007; Pang *et al.*, 2008; Pudney *et al.*, 2010)], soybean lipoxygenase [SLO-1, with linoleic acid (Knapp *et al.*, 2002; Mincer & Schwartz, 2004; Tejero *et al.*, 2006; Meyer *et al.*, 2008)], *E. coli* dihydrofolate reductase [*ec*DHFR, with NADPH at pH 7.0 and 9.5 (Watney *et al.*, 2003; Loveridge *et al.*, 2011; Dametto *et al.*, 2012; Luk *et al.*, 2013) and (Swanwick *et al.*, 2006; Luk *et al.*, 2013) respectively], methylamine dehydrogenase [MADH, with methylamine (Basran *et al.*, 1999)], aromatic amine dehydrogenase [AADH, with tryptamine (Masgrau *et al.*, 2006; Johannissen *et al.*, 2007)] and horse liver alcohol dehydrogenase [HLADH, with benzyl alcohol (Rubach *et al.*, 2001; Caratzoulas *et al.*, 2002)]. Of these, the catalytic mechanisms of PETNR$_{\mathrm{NADPH}}$, MR, SLO-1, *ec*DHFR, AADH, and HLAD are largely acknowledged to involve vibrational coupling (Table 1.1), while PETNR$_{\mathrm{NADH}}$ and MADH have no evidence for vibrational coupling (Arcus & Pudney, 2015).

A widely-used tool for assessing the presence of coupling in an enzymatic reaction is the temperature dependence of the kinetic isotope effect (KIE), $\Delta\Delta H^{\ddagger}$, measuring the effect on $\Delta H^{\ddagger}$ when a heavy atom is introduced into the system to independently alter vibrational modes. A $\Delta\Delta H^{\ddagger}$ larger than ~5 kJ.mol^{-1}, greater than the additional contribution of zero-point energy alone, indicates a temperature dependent KIE, and is taken to reflect a contribution of vibrational coupling to the reaction. Conversely, a temperature independent KIE ($\Delta\Delta H^{\ddagger}$ ~ 0 kJ.mol^{-1}) suggests the low involvement, or absence, of vibrational coupling during the reaction. For the listed enzyme examples above however, the $\Delta\Delta H^{\ddagger}$ metric as a predictor of vibrational coupling has been shown to fail for *ec*DHFR$_{(\mathrm{pH}\ 9.5)}$, AADH and HLADH (Table 1.1). In these systems, $\Delta\Delta H^{\ddagger}$ is small/negative, indicating a lack of vibrational coupling. However, for these systems vibrational coupling has been shown through additional methods. For all the other systems with vibrational coupling marked in Table 1.1, there is experimental evidence beyond $\Delta\Delta H^{\ddagger}$ alone for vibrational coupling.

1.4.5.1 Application of Macromolecular Rate Theory to reactions involving vibrational coupling

Arcus and Pudney (2015) first hypothesised that MMRT would be a useful predictor of vibrational coupling of either the protein or substrate to the reaction coordinate. This prediction arises due to the expectation that, when vibrational coupling at the reactant state is present, this should reduce/disappear at the TS (i.e. the tunnelling ready conformation in QMT reactions), giving a difference in C_{p} between the ground state and TS. As a predictive tool, this is a useful approach, as vibrational coupling can be assessed from curvature (or lack of) in the temperature dependence of the hydrogen transfer reaction alone.

The MMRT equation **(7)** has been applied to the temperature data for the nine enzyme systems listed above (Arcus & Pudney, 2015). Overall, $\Delta C^{\ddagger}_{p}$ for the hydrogen transfer reaction proved to be a consistent predictor of vibrational coupling. For $\text{PETNR}_{\text{NADPH}}$, MR, SLO-1, *ec*$\text{DHFR}_{(\text{pH } 7.0)}$, which have a $\Delta\Delta H^{\ddagger}$ commensurate with vibrational coupling, the MMRT equation reported all negative $\Delta C^{\ddagger}_{p}$ values as expected (Table 1.1). The three enzymes for which the $\Delta\Delta H^{\ddagger}$ metric is not a good metric (*ec*$\text{DHFR}_{(\text{pH } 9.5)}$, AADH and HLADH), $\Delta C^{\ddagger}_{p}$ values correctly indicated the presence of vibrational coupling. In these cases, except HLADH, the $\Delta C^{\ddagger}_{p}$ metric is conclusively negative within error bounds. The $\Delta C^{\ddagger}_{p}$ metric also correctly predicted the two cases where there is no evidence of vibrational coupling, MADH and $\text{PETNR}_{\text{NADH}}$, giving $\Delta C^{\ddagger}_{p}$ values of 0.03 ± 0.14 and 0.23 ± 0.1 $\text{kJ.mol}^{-1}.\text{K}^{-1}$ respectively (Arcus & Pudney, 2015).

Table 1.1: Data for $\Delta\Delta H^{\ddagger}$, $\Delta C^{\ddagger}_{p}$, and $\Delta\Delta C^{\ddagger}_{p}$ as evidence for vibrational coupling. $\Delta\Delta H^{\ddagger}$, commonly used as an indicator of the presence of vibrational coupling (when ≥ 5 kJ.mol^{-1}) fails in three cases (bold). $\Delta C^{\ddagger}_{p}$ however accurately accounts for the known cases of vibrational coupling (negative $\Delta C^{\ddagger}_{p}$) and the absence of coupling ($\Delta C^{\ddagger}_{p}$ ≥ zero). The $\Delta\Delta C^{\ddagger}_{p}$ upon isotopic labelling also consistently accounts for the presence (negative) or absence (zero) of vibrational coupling, but with less certainty due to the associated errors.

Enzyme system	**$\Delta\Delta H^{\ddagger}$ (kJ.mol^{-1})**	**$\Delta C^{\ddagger}_{p}$ ($\text{kJ.mol}^{-1}.\text{K}^{-1}$)**	**$\Delta\Delta C^{\ddagger}_{p}$ ($\text{kJ.mol}^{-1}.\text{K}^{-1}$)**
Evidence for vibrational coupling			
$\text{PETNR}_{\text{NADPH}}$	6.5 ± 2.7	- 0.53 ± 0.03	- 0.36 ± 0.08
MR	7.4 ± 1.5	-0.22 ± 0.10	- 0.25 ± 0.13
SLO-1	4.2	-0.42 ± 0.15	- 0.63 ± 0.53
$\text{ecDHFR}_{(\text{pH } 7.0)}$	7.9 ± 0.9	-1.11 ± 0.23	- 0.22 ± 0.46
$\text{ecDHFR}_{(\text{pH } 9.5)}$	**-1.11 ± 0.79**	-1.40 ± 0.62	- 0.53 ± 0.49
AADH	**-3.0 ± 0.2**	-1.71 ± 0.89	1.89 ± 2.74
HLADH	**3.34 ± 4.2**	-0.12 ± 0.41	- 0.26 ± 0.86
No evidence for vibrational coupling			
MADH	0.4 ± 0.01	0.03 ± 0.14	- 0.04 ± 0.27
$\text{PETNR}_{\text{NADH}}$	- 0.8 ± 1.3	0.23 ± 0.1	- 0.07 ± 0.26

1.4.5.1.1 *Further evidence from deuterated substrate reactions*

All the QMT reactions listed above have literature data for rates with deuterated substrate, as required for calculation of the $\Delta\Delta H^{\ddagger}$ metric. These data have been analysed in the context of MMRT (Arcus & Pudney, 2015). If vibrational coupling is present, and the vibrational system extends to include the deuterated bond, a greater difference in vibrational energy would be expected over the course of the reaction, resulting in more curvature in the temperature data.

For all reaction systems, this is indeed the case (Table 1.1). For PETNR$_{\mathrm{NAPDH}}$, MR, SLO-1 and *ec*DHFR$_{\mathrm{(pH\ 9.5)}}$, the data is conclusive in that the $\Delta C^{\ddagger}_{\mathrm{p}}$ value is more negative for the deuterated reaction outside of error limits. For these cases, this suggests there is a contribution to catalysis from vibrational modes associated with the deuterated substrate bond/s. This coupling conceivably extends to vibrations present in the protein, which localise to within the active site. Data for the enzymes MADH and PETNR$_{\mathrm{NADH}}$ are zero within error limits, consistent with the absence of vibrational coupling in these systems.

1.4.6 Sensitivity of $\Delta C^{\ddagger}_{\mathrm{p}}$ to changes in the frequency of vibrational modes

Given the link established between $\Delta C^{\ddagger}_{\mathrm{p}}$ and vibrational coupling (Section 1.4.5), and the perturbations seen in this through the use of deuterated substrates (Section 1.4.5.1.1), other perturbations to vibrational modes are expected to prompt changes in $\Delta C^{\ddagger}_{\mathrm{p}}$ [as explored in (Arcus & Pudney, 2015)].

In addition to isotopically labelled substrates, whole proteins can be made with ^{13}C, ^{15}N and deuterium, decreasing the frequency of vibrational modes throughout the system. The effect this has on temperature dependence has been investigated for PETNR with NADH and NADPH (Pudney *et al.*, 2013), *ec*DHFR (Luk *et al.*, 2013), KSI and MalL (Kraakman, 2017). In these cases, isotopically labelled heavy enzymes show altered curvature ($\Delta C^{\ddagger}_{\mathrm{p}}$) when vibrational modes are altered, in most cases outside of error limits. However, there is no consistent direction of the change in $\Delta C^{\ddagger}_{\mathrm{p}}$, a detail which warrants further investigation into the possible effects of global versus local vibrational modes.

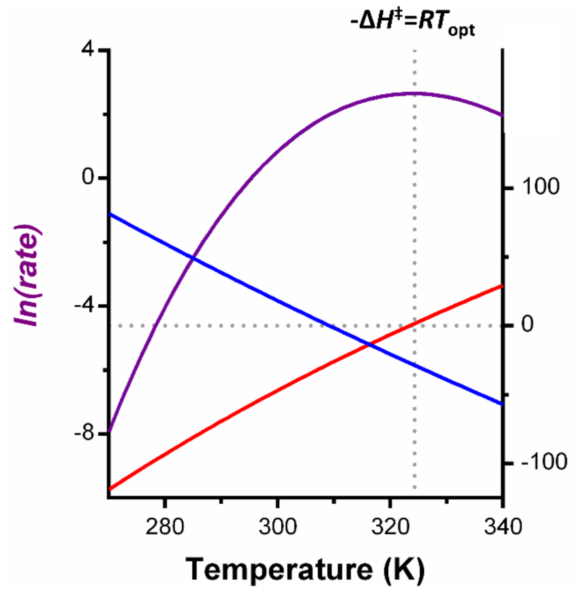
Arcus and Pudney (2015) also cite the effect of pressure of $\Delta C^{\ddagger}_{\mathrm{p}}$, although there are limited examples of this in the literature [MR (Hay *et al.*, 2007) and AADH (Hay *et al.*, 2012)]. The factors involved here are complex and require further deconstruction, however increasing pressure from one to 2000 bar results in a less negative $\Delta C^{\ddagger}_{p}$ in both cases, indicating a change in vibrational modes over the system.

The data presented here and Section 1.4.5, although controversial overall in terms of the impact of vibrational modes on catalysis, shows a consistent correlation between the predicted effects on $\Delta C^{\ddagger}_{p}$ through vibrational coupling and perturbations to vibrational frequencies.

1.4.7 Relative contributions of entropy and enthalpy to catalysis

Our established understanding of the influence of $\Delta C^{\ddagger}_{p}$ has implications for the discussion on the relative contribution to $\Delta S^{\ddagger}$ and $\Delta H^{\ddagger}$ to catalysis, about which there is a lack of consensus in the literature (Arcus *et al.*, 2016). For example, Wolfenden (2014) concludes that enzymatic catalysis is generally driven by favourable $\Delta H^{\ddagger}$, rather than a more positive $T\Delta S^{\ddagger}$, although acknowledges exceptions to this rule. Entropy changes as the driving force of catalysis are cited in other proposals (Page & Jencks, 1971), but the extent of this effect is questioned (Villà *et al.*, 2000), while Bruice and Benkovic (2000) expect a more prominent entropic effect in reactions with multiple substrates. Overall from the above evidence in the literature, it seems that different enzymes work by either optimising $\Delta H^{\ddagger}$ (less positive) or $\Delta S^{\ddagger}$ (more positive) to optimise reaction rates, and there is no general mechanism by which enzymes achieve lowered activation barriers ($\Delta G^{\ddagger}$).

From MMRT, the large $\Delta C^{\ddagger}_{p}$ associated with enzymatic reactions results in the steep temperature dependence of $-\Delta H^{\ddagger}/RT$ and $\Delta S^{\ddagger}/R$ (Figure 1.9). These two parameters vary in an opposing fashion over temperature; the contribution of the enthalpy term becomes more positive with increasing temperature, while the value of the entropic term decreases. Figure 1.9 shows that the relative contribution of the two factors will be strongly influenced by the temperature at which measurements are taken. Below T_{opt}, increasing rates are driven by the enthalpic term ($-\Delta H^{\ddagger}/RT$); however, as temperature rises, the entropic term ($\Delta S^{\ddagger}/R$) becomes more influential, leading to curvature and a decrease in rates (Arcus *et al.*, 2016).



$$-\frac{\Delta H^{\ddagger}}{RT} = \frac{-\Delta H^{\ddagger}_{T0} - \Delta C^{\ddagger}_{p}(T - T_{0})}{RT}$$

$$\frac{\Delta S^{\ddagger}}{R} = \frac{\Delta S^{\ddagger}_{T0} + \Delta C^{\ddagger}_{p}(lnT - lnT_{0})}{R}$$

Figure 1.9: The temperature dependence of the relative contribution of H and S to catalysis. Reaction rates with temperature (purple; equation **7***) are driven at low temperatures by the enthalpic term (red); above* T_{opt} *at* $-\Delta H^{\ddagger}=RT_{opt}$*, entropic effects become more influential, decreasing rates.*

MMRT however does make predictions relating to thermodynamic parameters which can be generalised over all enzymes. Providing $\Delta C^{\ddagger}_{p}$ is non-zero, T_{opt} is defined when the first derivative is equal to zero (equation **8c**; see page 14 for derivation).

$\Delta H^{\ddagger} = -RT_{opt}$ ……………………………………..……*Fixing of H at* T_{opt} **(8c)**

Equation **(8c)** fixes $\Delta H^{\ddagger}$ at T_{opt}, meaning $\Delta S^{\ddagger}$ is the only flexible term at optimum temperatures. Given that enzyme T_{opt} values vary between ~15 to 100 °C typically, this constraint fixes $\Delta H^{\ddagger}$ values between -2.4 and -3.1 kJ.mol^{-1}. Consequently, $\Delta S^{\ddagger}$ is the dominating term at T_{opt}, and the parameter with the most flexibility in temperature adaptation (Hobbs *et al.*, 2013; Arcus *et al.*, 2016).

1.5 Scaling Macromolecular Rate Theory up to multi-enzyme systems

Below is a review of the published understanding of how MMRT fits to multi-enzyme systems such as organism growth rates and ecosystem fluxes. In particular, it has long been acknowledged that both microorganism growth rates and ecosystem fluxes exhibit curvature in their temperature-rate profiles (Lloyd & Taylor, 1994). This prompted Schipper and Arcus (2014) to apply MMRT to investigate the temperature dependence of these larger scale processes.

1.5.1 Organism growth rate responses to temperature

The potential applications of MMRT extend further than single enzyme systems alone. Organisms' growth rates are defined by the rates of biochemical pathways driven by multiple enzyme catalysed steps. Each of these enzymes is expected to follow MMRT curvature, with T_{opt} values generally correlated to the environmental temperature (Hobbs *et al.*, 2012). As such, it is expected that the signature of MMRT will also be evident at the organism level, as an intrinsic response to temperature. Additional context-specific regulatory responses to temperature and related parameters (water availability, solubility) will then layer on top of this. The potential applicability of MMRT to organism growth rates was first shown in Hobbs (2013) through the successful fit of equation **(7)** to a range of growth rate data (Figure 1.10). MMRT fits the data well over the full range of temperatures, fitting data at high and low temperatures equally well (Figure 1.10). MMRT is shown to fit data from bacterial species from psychrophilic, mesophilic, and thermophilic environments as well as representatives from the fungal and archaeal domains of life [(Hobbs *et al.*, 2013); Figure 1.10].

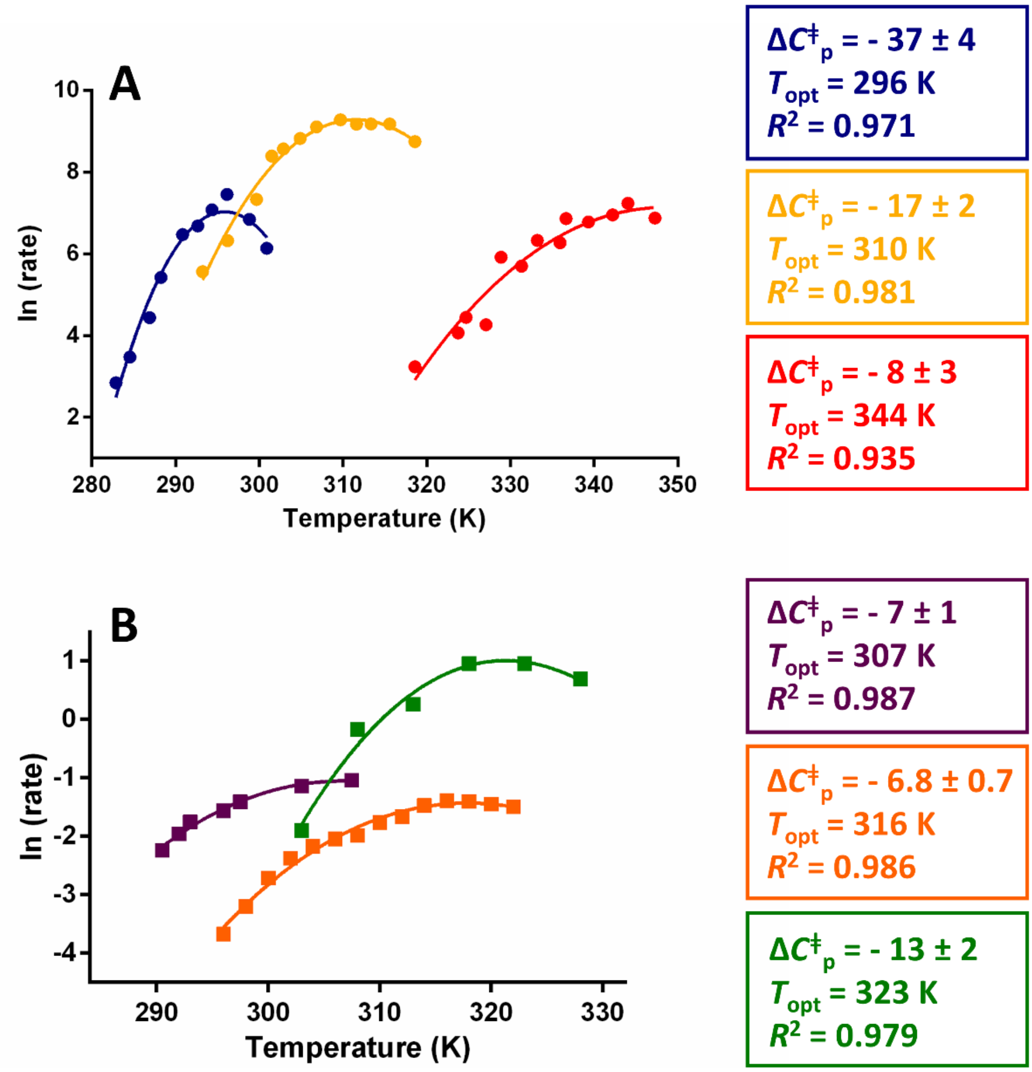


Figure 1.10: Growth rates fitted to MMRT for a range of organisms from **(A)** *different temperature environments - ● Vibrio marinus ATCC 15382 [psychrophile; (Mohr & Krawiec, 1980)], ● Bacillus subtilis [mesophile; (Mohr & Krawiec, 1980)], and ● Thermus aquaticus [thermophile; (Mohr & Krawiec, 1980)];* **(B)** *fungal and archaeal species – ■ Saccharomyces cerevisiae [fungal yeast; (Jagadish & Carter, 1978)], ■ Haloarcula vallismortis [archaeon;(Robinson et al., 2005)], and ■ Chaetomium thermophile* var. *dissitum 1A [fungal mould; (Tansey, 1972)].* $\Delta C^{\ddagger}_{p}$ *values are quoted as kJ.mol*$^{-1}$*. K*$^{-1}$ ± *SE. Figure adapted from Hobbs et al. (2013).*

The application of MMRT to organism growth rates is however only postulated, and remains to be experimentally demonstrated (Schulte, 2015).

1.5.2 Modelling simple ecosystem responses to temperature

Schipper et al. (2014) extended this idea further into the intrinsic temperature dependence of soil ecosystems on the basis that these populations comprise a consortium of species, each with a different T_{opt} for growth and curvature ($\Delta C^{\ddagger}_{p}$). For a consortium of individual species, the total rate for the ecosystem is the summation of individual species rates at each temperature. Total rates for models of this simple theoretical ecosystem fit well to MMRT (Figure 1.11). The application of MMRT to soil processes (respiration, enzyme activities, carbon and nitrogen cycling for example) is a critical area of interest considering potential changes in environmental temperature. For investigating the potential impacts of increased soil temperatures, models which accurately predict responses in biological turnover are essential.

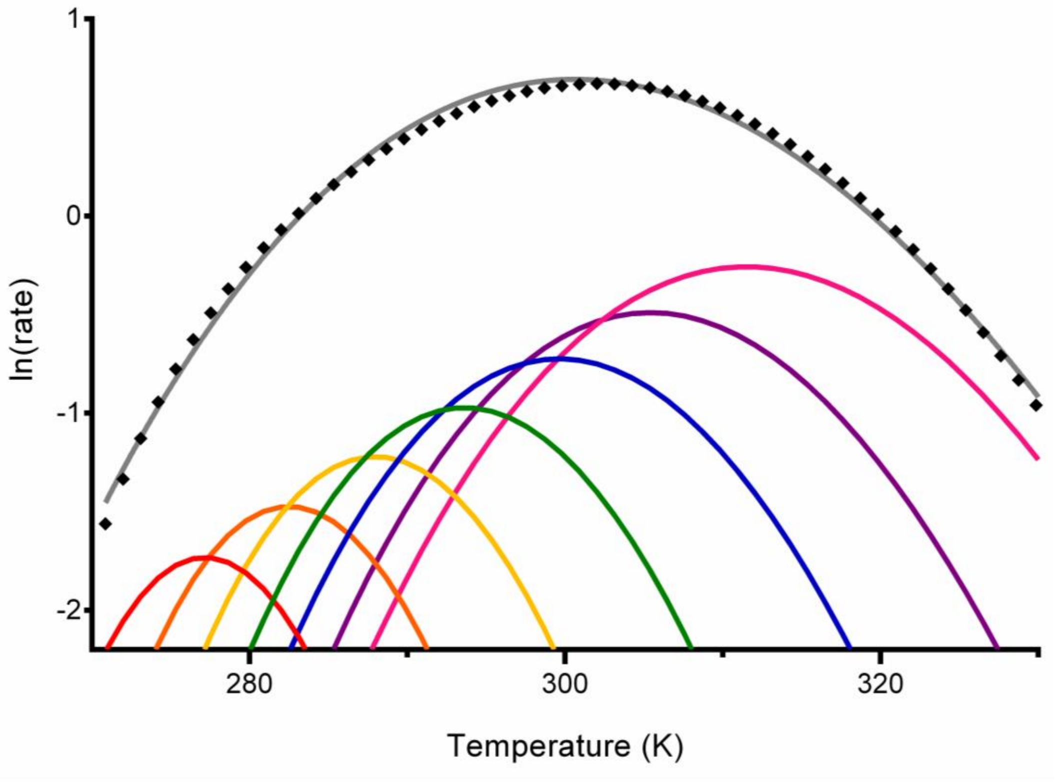


Figure 1.11: A consortium model (Schipper et al., 2014) of a simple seven species ecosystem (red through pink lines), with the overall ecosystem rates (♦) fit to MMRT (grey line). Individual curves range from $\Delta C^{\ddagger}_{p}$ *-15 to -5 kJ.mol*$^{-1}$*. K*$^{-1}$*, with corresponding* T_{opt} *values from 275 to 305 K. Adapted from (Schipper et al., 2014).*

Despite this, data are commonly fit with models based on Arrhenius equations which fail to encompass an optimum temperature (community T_{opt}) that is evident in laboratory and field data (Tuomi *et al.*, 2008). These functions used in the literature show increases in rates continually with increasing temperature, in contrast to the MMRT equation which predicts a clear temperature optimum (Figure 1.12 A and B). To explain these differences between the model and data, enzyme denaturation is routinely cited as the limiting factor at high temperature in the laboratory, while the lowered availability of water and water soluble components is often invoked to explain field data (Corkrey *et al.*, 2012).

The most convincing argument for the application of MMRT to soil data comes from the fit of equation **(7)** to experimental data [a representative range is given in Figure 1.12 D; a greater set of examples is available in Schipper *et al.* (2014), Figure 3]. Data can only be fit with the Arrhenius equation up to T_{opt}. By comparison, MMRT can be fit to data above T_{opt} also, and fits to the curvature seen over the full range of temperatures over a range of different measures of microbial activity both in the laboratory and field. Community $\Delta C^{\ddagger}_{\mathrm{p}}$ values gained from this fitting ranged from -1 to -12 $\mathrm{kJ.mol^{-1}.K^{-1}}$, consistent with the range of values gained from single enzymes (Hobbs *et al.*, 2013).

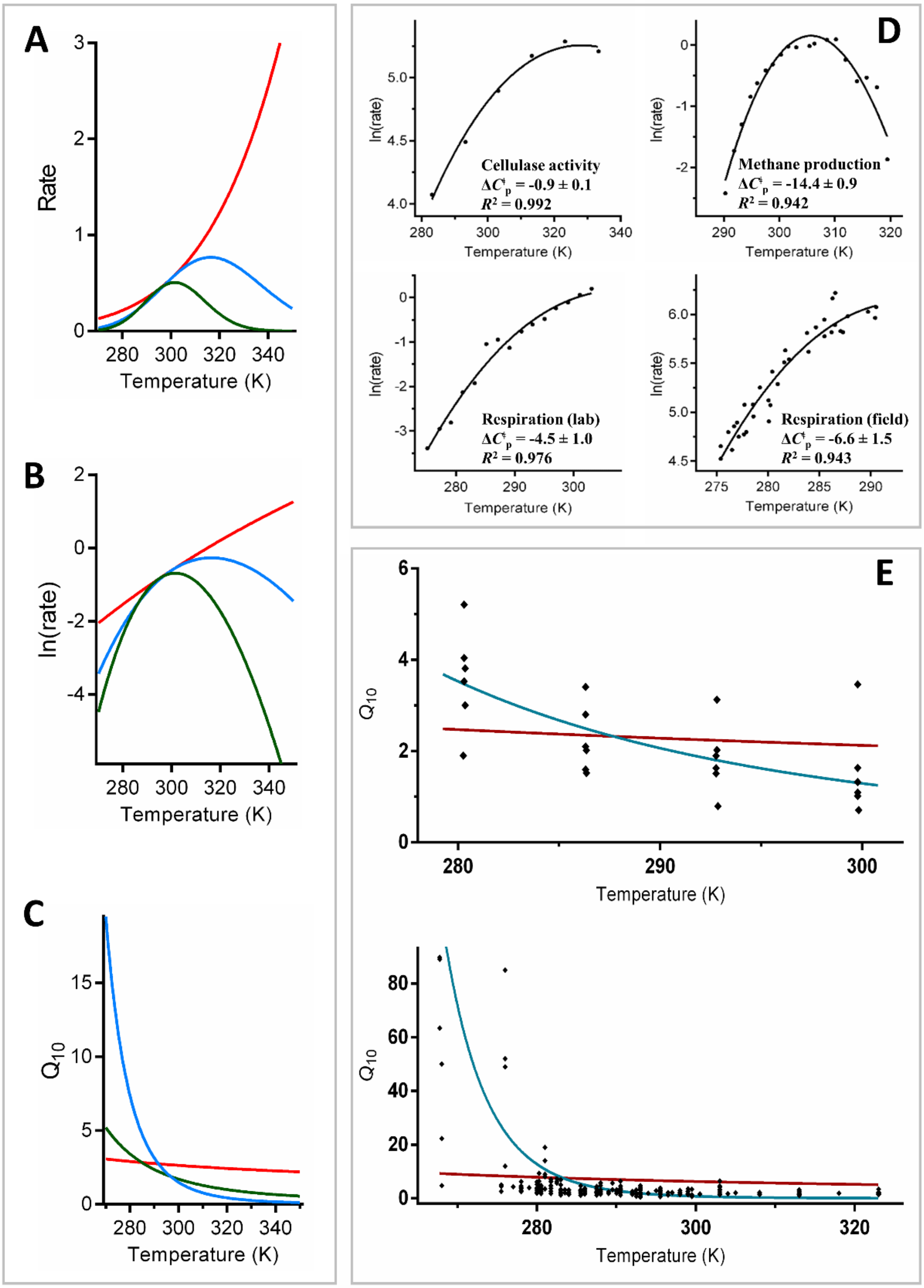


Figure 1.12: Comparison of MMRT to Arrhenius functions for modelling ecosystem data and fits to experimental data. ***(A)*** *Rates modelled by an Arrhenius function (red) and MMRT with $\Delta C^{\ddagger}_{p}$ values of –5 (green) and –2 kJ.mol^{-1}.K^{-1} (blue).* ***(B)*** *Natural log of rates modelled by an Arrhenius function and MMRT, and* ***(D)*** *the fit of the MMRT model to various ecosystem data (Schipper et al., 2014).* ***(C)*** *Trends in Q_{10} values with temperature based on the Arrhenius function, and MMRT at different values of $\Delta C^{\ddagger}_{p}$, and* ***(E)*** *the fit of Arrhenius (maroon) and MMRT (turquoise) models for Q_{10} to experimental data (Schipper et al., 2014).*

1.5.2.1 Implications for ecosystem temperature sensitivity (Q_{10})

Understanding the response of biological systems to increasing temperature is vital to accurate ecosystem modelling. The Q_{10} metric is widely used to describe relative temperature sensitivity of systems; this is the ratio of rates determined at a 10 °C interval (Sierra, 2012). Empirical studies have shown Q_{10} values to vary with temperature. Values tend to be higher at low temperatures, then decrease as temperature rises. Hamdi *et al.* (2013), reporting on a range of data from the literature, reported Q_{10} values ranging from 0.5 to over 300. Despite the wide agreement in this trend of Q_{10} values (Lloyd & Taylor, 1994; Chen & Tian, 2005; Sierra, 2012; Hamdi *et al.*, 2013), some global climate models fix Q_{10} at 1.5-2.0, or assign set values to different carbon pools (Chen & Tian, 2005; Friedlingstein *et al.*, 2006).

To illustrate the pitfalls of fixing Q_{10}, let us consider the Arctic tundra. The permafrost soils of the Northern tundra regions are of great concern under global climate change due to the large quantities of organic material held in the soils, the release of which is normally limited by low temperatures and water availability through freeze-thaw cycles (Schimel & Mikan, 2005). Since the late 1970's, measured Arctic temperatures have been rising by 0.5 °C per decade, and have been projected to rise in the range of 2-8 °C in the next 100 years (IPCC, 2013), warming 11.5 million km^2 of permafrost soils. Estimating increases in released carbon and other biogenic volatiles over such a large area is going to be highly sensitive to our estimate of the temperature sensitivity of the system. The generalised Q_{10} quoted earlier of 1.5-2 will vastly underestimate changes given true sensitivities may be significantly higher; the most extreme example of this is evident in the meta-analysis in Hamdi *et al.* (2013) with a reported Q_{10} value of 340 at zero degrees Celsius for polar desert soils.

By comparison, MMRT does predict higher Q_{10} values with decreasing temperature (Figure 1.12 C). Whereas, Arrhenius based models do not predict significantly increased relative temperature sensitivities at low temperatures, which do not fit data as well as the MMRT model [Figure 1.12 E; (Schipper *et al.*, 2014)]. The extent of Q_{10} values is dependent of $\Delta C^{\ddagger}_{p}$; the amount of curvature in the raw rates of the ecosystem with temperature (Figure 1.12 A). Given normal ranges of $\Delta C^{\ddagger}_{p}$

values observed, a MMRT approach predicts Q_{10} values at zero degrees Celsius between 20 and 100 (Schipper *et al.*, 2014), consistent with the values reported in Hamdi *et al.* (2013). As a predictive tool, this is more realistic than using Arrhenius based approaches as it accounts for the changes in Q_{10} values with temperature. However, to have a fully functioning predictive tool, community $\Delta C^{\ddagger}_{\mathrm{p}}$ will need to be accurately determined for the systems of interest to better define the range of Q_{10} values applicable.

1.6 Research objectives

To further our understanding of the temperature dependence of enzymes and biological rates at larger scales (cells, organisms, ecosystems), the following objectives were explored in my doctoral research:

Objective one:
Account for the effect of denaturation on MalL variant reaction rates, and correct published rate values for this effect, conclusively ruling out the effect of denaturation in MalL data.

Objective two:
Explore $\Delta C^{\ddagger}_{\mathrm{p}}$ as a measure of a ‘catalytic reservoir’ through the relationship between enzymatic mass and catalytic efficiency.

Objective three:
Test the scalability of MMRT for applicability to organism growth rates, in which enzymes with varying temperature profiles are operating in series and parallel via metabolic pathways.

Objective four:
Determine crystal structures of MalL variants in complex with TS analogues, and analyse these in comparison to previously determined apo structures for differences relevant to C_{p} changes through the reaction coordinate.

Objective five:
Test and apply *in silico* methods (molecular dynamics) in an effort to calculate

$\Delta C^{\ddagger}_{p}$ and compare with experimentally determined values, providing access to atomistic details pertaining to MMRT.

Materials and Methods

2.1 General methods

2.1.1 Purification of soluble protein

All proteins in this study were purified by the same two step immobilised metal affinity chromatography (IMAC) and size exclusion chromatography (SEC) protocol for isolating soluble protein from the cytoplasmic cellular fraction. Protein specific details such as buffers can be found in the methods sections pertaining to each protein (MalL 2.2.1.2; Glycolysis 2.3.2.3).

2.1.1.1 Cell lysis

Defrosted cell pellets were resuspended in ~20 ml of lysis buffer. Cells were lysed in 50 ml Falcon tubes on ice by sonication in six bursts of 15 seconds, with 30 second periods of cooling between each sonication. Cell debris was removed from the supernatant by centrifugation at 20,000 g for 20 minutes at 4 °C to separate the cellular supernatant from the insoluble debris.

2.1.1.2 Immobilised Metal Affinity Chromatography Purification

An initial protein purification step was performed via Immobilised Metal Affinity Chromatography (IMAC) purification based on the hexa-histidine tag. Columns with immobilised nickel ions (5 ml HiTrap HP/FF; GE Healthcare, UK) were prepared by stripping with two column volumes of 100 mM EDTA (pH 8.0) to remove Ni^{2+} from the column. Columns were recharged with one column volume of 100 mM $NiCl_2$. Washing steps with two column volumes of MQ H_2O followed each of these steps. Prepared columns were equilibrated in lysis buffer.

Cellular supernatant was filtered successively through 1.2 µm, 0.45 µm, and 0.2 µm Minisart syringe filters (Sartorius AG, Germany) prior to loading onto the equilibrated IMAC column.

Protein elution from the column was performed with an ÄKTA Basic, Prime, or Purifier system (GE Healthcare, Sweden). Weakly bound non-target proteins were eluted from the column with a solution comprising of 4 % elution buffer: 96 % lysis buffer at a flow rate of 1 ml.min^{-1}. Washing was continued until the absorbance at 280 nm had plateaued. Target protein was eluted from the column over a gradient elution from 0-100 % elution buffer over 50 ml at a flow rate of 1 ml.min^{-1}, and collected in 2 ml aliquots.

Fractions with target protein were identified via the 280 nm absorbance trace, and/or SDS-PAGE (Section 2.1.2.1).

2.1.1.3 Size exclusion chromatography

Protein containing fractions from the IMAC purification were pooled and concentrated (Section 2.1.3) to ≤ 20 mg.ml^{-1} for SEC. Protein was filtered through a 0.2 µm filter and 0.5 ml loaded onto a Superdex 200 10/300 GL column (GE Healthcare Life Science, UK) equilibrated with size exclusion buffer. Protein was separated and eluted with size exclusion buffer at a flow rate of 0.5 ml.min^{-1}, and collected in 0.5 ml aliquots.

Fractions with target protein were identified via the 280 nm absorbance trace, and/or SDS-PAGE (2.1.2.1).

Protein was concentrated (Section 2.1.3) to the required concentration for further analysis.

2.1.1.3.1 Molecular weight determination from size exclusion chromatography

A S200 10/300 analytical size exclusion column (GE Healthcare Life Science, UK) was previously calibrated using eight proteins standards from high and low molecular weight gel filtration calibration kits (GE Healthcare Life Science, UK) as per kit instructions. The void volume (V_0) of the column was determined with 1 mg.ml^{-1} Blue Dextran 2000.

Retention volumes of the protein standards were used to construct a standard curve of K_{av} (gel phase distribution coefficient; equation **11**) versus molecular weight, where V_e = elution volume and V_c = geometric column volume (Appendix D.4). Proteins of unknown molecular weight were interpolated from this standard curve.

$$K_{av} = \frac{V_e - V_0}{V_c - V_0} \ldots\ldots\ldots\ldots \textit{gel phase distribution coefficient} \quad \textbf{(11)}$$

2.1.2 Polyacrylamide gel electrophoresis

Polyacrylamide gel electrophoresis (PAGE) gels were made in the Hoefer gel casting system (Hoefer, US). The resolving gel (10, 12, 15 or 18 % acrylamide) was overlaid with a 5 % acrylamide stacker layer. All gels were made with 30 % acrylamide 37:5:1 (Bio-Rad Laboratories, USA). Both the resolving and stacking layer were polymerised with the addition of 0.05 % (w/v) ammonium persulfate (APS) and 0.05 % (v/v) TEMED. SDS-PAGE gels contained an additional 0.1 % (w/v) SDS in both the resolving and stacking layers.

2.1.2.1 SDS-PAGE

Protein samples were denatured by the addition of 4 x SDS loading buffer [250 mM TRIS-HCl, pH 6.8, 20 % (v/v) glycerol, 4 % (w/v) SDS, 10 % (v/v) β-mercaptoethanol, 0.025 % (w/v) bromophenol blue] and heated at 95 °C for five minutes. Aliquots of 5-15 µl of denatured protein were loaded into the wells of an appropriate polyacrylamide percentage SDS-PAGE gel (Section 2.1.2), alongside 10 µl of Precision Plus Protein Unstained Ladder (Bio-Rad Laboratories, USA). Gels were run in 1 x SDS-PAGE running buffer [25 mM Tris, 250 mM glycine, 0.1 % (w/v) SDS] at 15 mA as the protein passed through the stacker layer, and at 20 mA through the remainder of the gel until the dye front reached the base of the gel.

2.1.2.1.1 Gel staining and visualisation

Gels were stained with Coomassie stain [0.05 % (w/v) Coomassie blue R-250, 25 % (v/v) isopropanol, 10 % (v/v) acetic acid]. Gels were immersed in the stain, and microwave heated for 30 seconds and left in to stain at room temperature with shaking for an additional 15 minutes. Gels were removed from the staining solution, washed, and immersed in destain solution [10 % (v/v) acetic acid], microwave heated for 30 seconds, and left shaking at room temperature for half an hour. The destaining procedure was repeated till protein bands were well defined against the background.

2.1.2.2 Native PAGE

Native PAGE gels samples were prepared by the addition of 5 x native loading dye [300 mM TRIS-HCL, pH 6.8, 50 % (v/v) glycerol, 0.05 % (w/v) bromophenol blue] prior to loading into the wells of an appropriate polyacrylamide percentage native PAGE gel (Section 2.1.2), along with a lane of 10 µl of Precision Plus Protein Unstained Ladder (Bio-Rad Laboratories, USA).

Native gels were run, stained and visualised by the same methods as described for SDS-PAGE gels (Section 2.1.2.1.1), except without the inclusion of 0.1 % (w/v) SDS in the SDS-running buffer.

2.1.3 Protein concentration

Proteins were concentrated by centrifugation in 0.5 ml, 2 ml, or 20 ml Vivaspin concentrators with a 10 kDa molecular weight cut off (Sartorius AG, Germany) at 2600 g at 4 °C.

2.1.4 Determination of protein concentrations by A_{280}

Protein concentrations were determined via the Nanodrop 2000 UV-vis spectrophotometer (Thermo scientific, USA) using absorbance at 280 nm and corrected with the extinction coefficient as calculated by ProtParam (Gasteiger *et al.*, 2005) based on the protein sequence.

2.1.5 Storage of cell strains

Glycerol stocks of cell strains were made from an overnight culture in standard growth conditions for the cell strain with an additional 8 % (v/v) glycerol. Glycerol stocks were frozen and stored at -80 °C.

2.1.6 Agarose gel electrophoresis

Gels for DNA electrophoresis were made up between 1-2 % (w/v) agarose in TAE buffer (40 mM TRIS-acetate, 20 mM EDTA). Solutions were microwaved to dissolve all agarose. SYBR® safe (1000x; Invitrogen, USA) was added to gels prior to pouring to allow later visualisation of DNA bands. Samples were mixed with 5 x DNA loading dye [25 % glycerol (v/v), 0.2 % bromophenol blue (w/v)]. Gels were run at 100 V for 40 minutes submerged in TAE buffer. DNA bands were visualised under blue light (blue light box, Invitrogen USA), and size determined by comparison to 1kb-Plus DNA ladder (Invitrogen USA).

2.1.7 Processing of rate vs temperature data

Temperature-rate data were processed in GraphPad Prism (version 6.07 for Windows, GraphPad Software, La Jolla California USA, www.graphpad.com) by fitting equation **(7)** to the natural log of rate data. T_0 was set at four degrees lower than the temperature at which the highest rate was observed in accordance with previously published methods (Hobbs *et al.*, 2013; Schipper *et al.*, 2014).

Other equations were also fit to data where appropriate via the GraphPad software.

2.2 Methods pertaining to MalL variants

2.2.1 MalL protein expression and purification

2.2.1.1 Protein expression

A 10 mL starter culture inoculated from frozen glycerol stocks (Section 2.1.5) of the relevant strain was grown overnight at 37 °C with shaking in LB (Appendix A) containing 100 $\mu g.ml^{-1}$ ampicillin (AMP) in a 50 ml Falcon tube. An expression culture (1 L LB in a 2 L baffled flask) containing 100 $\mu g.ml^{-1}$ AMP was inoculated with the starter culture, and grown at 37 °C till mid log phase (OD_{600} between 0.5-0.7). Protein expression was induced in the mid log phase by the introduction of 1 mM IPTG. Protein expression was carried out at 18 °C with shaking for 20 hours, at which stage cells were harvested by centrifugation at 4500 g for 20 minutes at 4 °C. Cell pellets were frozen at -80 °C until immediately prior to protein purification steps.

2.2.1.2 Protein purification

Protein was purified as described in Section 2.1.1. Buffer compositions for this process are listed (Table 2.1).

Table 2.1: List of buffer compositions for IMAC and SEC purification of MalL variants

Buffer	**Composition**
IMAC lysis buffer	10 mM KCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 25 mM imidazole, pH 7.0 (Schönert *et al.*, 1998)
IMAC elution buffer	10 mM KCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 1 M imidazole, pH 7.0
SEC buffer	20 mM HEPES, pH 7.0

2.2.2 General MalL assay protocol

MalL was characterised via the cleavage of *p*-nitrophenyl-α-ᴅ-glucopyranoside to glucose and *p*-nitrophenol (Schönert *et al.*, 1998). Substrate and assay buffer (40 mM sodium-phosphate, pH 6.5, 150 mM NaCl) were mixed to give the appropriate substrate concentration in a final volume of 400 µl in a quartz cuvette (Starna Scientific Ltd, UK). Where necessary, reaction temperature was equilibrated and held with a single cell peltier (Thermo scientific, USA) coupled to a circulating water bath, with accurate temperatures measurements made via thermocouple. Reactions were started with the addition of 5 µl of enzyme (kept on ice), mixed quickly, and the reaction progression followed at 405 nm with a Heλios γ spectrophotometer (ThermoFisher, USA), taking 0.125 second readings in Vision32 software (version 1.25).

Rate data were analysed via linear regression over the initial 10 seconds of reaction time. Data with deviations from linearity were discarded, and where possible, triplicate data were analysed.

2.2.3 Crystallography of MalL

2.2.3.1 Protein crystallisation

2.2.3.1.1 Protein preparation

Protein was prepared as described in Section 2.2.1 and concentrated (Section 2.1.3). MalL WT was concentrated to 11 mg.ml^{-1}, and V200S was to 7 mg.ml^{-1} for crystallisation. 1-deoxynojirimycin (DONM; Sigma-Aldrich, USA) was added to a final concentration of 2 mM in the protein solution.

2.2.3.1.2 *Initial crystallisation condition determination*

General crystallisation conditions were determined by the sitting drop method. Pre-prepared screens (PEGRx HT - HR2-086, Crystal Screen HT - HR2-130, Index HT - HR2-134, and SaltRx HT - HR2-136) from Hampton Research (USA) were used as the precipitant solutions. Aliquots of 100 µl of precipitant solution were pipetted into 96-2 low profile Intelli-Plate protein crystallisation plates (Hampton Research, USA). Crystallisation drops of 100 nl protein to 100 nl precipitant solution were dispensed with a Mosquito crystallisation robot (TTP LabTech Ltd., USA). Trays were sealed with ClearSeal film (Hampton Research, USA) and left at 18 °C on shock proof shelving. Drops were checked regularly for crystal formation.

2.2.3.1.3 *Refinement of crystallisation conditions*

Crystallisation conditions were refined with the hanging drop vapour diffusion method from potential conditions identified in the Hampton Research crystallography screens. Conditions were varied from the initial condition by varying pH and precipitant concentrations. For this method, well tops of 24-well VDX plates (Hampton Research, USA) were greased with Glisseal N grease (Borer Chemie, Switzerland) and filled with 500 µl of precipitant solution. Protein was mixed with precipitant solution (1:1 ratio, 2 µl total) on a 22 mm siliconised cover slip, inverted and sealed over the mother liquor reservoir. Trays were left at 18 °C on shock proof shelving.

2.2.3.1.4 *Final crystallisation conditions*

2.2.3.1.4.1 *MalL WT with 1-deoxynojirimycin*

Final crystals were grown in precipitant solution containing 10 % (v/v) tacsimate (Hampton research, USA) pH 6.0, 100 mM MES monohydrate, pH 6.4, and 25 % (w/v) polyethylene glycol 4000.

2.2.3.1.4.2 *MalL V200S with 1-deoxynojirimycin*

Final crystals were grown in precipitant solution containing 100 mM BIS TRIS, pH 5.3, and 23 % (w/v) polyethylene glycol 3350.

2.2.3.2 Data collection

2.2.3.2.1 Crystal preparation

Crystals were removed from the crystallisation drops with a cryo-loop (Hampton Research, USA) of appropriate size for the crystal. Crystals were rapidly submerged into a cryo-protectant solution containing the same composition as the original precipitant solution as well as 20 % (v/v) glycerol. Crystals were rapidly frozen in either a gaseous nitrogen stream at 100 K (for testing on the SuperNova X-ray source; see Section 2.2.3.2.2) or in liquid nitrogen (for data collection at the Australian synchrotron; see Section 2.2.3.2.3).

2.2.3.2.2 Initial diffraction testing

Crystals were initially tested for diffraction quality on a SuperNova X-ray diffractometer (Agilent, USA). Crystals were prepared as described in Section 2.2.3.2.1 and frozen by a gaseous stream of liquid nitrogen at 100 K. Diffracting crystals were transferred to liquid nitrogen for storage and full data collection at the Australian synchrotron, as described in Section 2.2.3.2.3.

2.2.3.2.3 Data collection

X-ray diffraction data were collected at the Australian Synchrotron, Melbourne, using the MX1 beam line, with an ADSC Quantum 210r detector (Area Detector Systems Corp., USA). The MOSFLM (Leslie & Powell, 2007) strategy function was used in conjunction with the collection process to assist optimal data collection.

In both MalL WT and MalL V200S, both high and low data sets were collected to optimise data quality in the inner and outer shells.

2.2.3.3 Data processing

2.2.3.3.1 Indexing and integration

Diffraction data was visualised, scaled and integrated in MOSFLM (Leslie & Powell, 2007). Unit cell parameters were determined through the auto indexing function, and data was integrated in the appropriate space group.

2.2.3.3.2 Scaling

High and low resolution data sets were renumbered and combined in Sortmtz (P. J. Daly, Daresbury) in the CCP4 data suite (Winn *et al.*, 2011).

Combined data sets were scaled in SCALA (Evans, 2006) within the CCP4 program suite (Winn *et al.*, 2011). SCALA outputs were examined, and the data rescaled at decreasing resolution to optimise the R_{merge} value to below 0.8 in the outer resolution shell of the data.

2.2.3.3.3 Matthews coefficient

The number of monomers in the asymmetric unit was determined based on calculated solvent occupancy of the unit cell as calculated by Matthews coefficient (Matthews, 1968) within the CCP4 program suite (Winn *et al.*, 2011).

2.2.3.3.4 Molecular replacement

Molecular replacement of both MalL WT and MalL V200S with DONM bound was performed with the apo MalL WT structure (PDB 4M56) as the model. Molecular replacement was performed using PHASER (McCoy *et al.*, 2007) within the PHENIX software (Adams *et al.*, 2010).

2.2.3.3.5 Model refinement

The structure model was initially refined automatically using Autobuild (Terwilliger *et al.*, 2008) with simulated annealing in the PHENIX software (Adams *et al.*, 2010).

Manual refinement of the model was performed in COOT (Emsley & Cowtan, 2004) with the 2Fo-Fc electron density map contoured to one σ. Ligand parameters for 1-deoxynojirimycin were produced through ReadySet (Nigel W. Moriarty) within the PHENIX program suite, based off the NOJ ligand from the COOT library. Model refinement iterations were performed using phenix.refine (Afonine *et al.*, 2012) in the PHENIX software (Adams *et al.*, 2010).

2.2.3.4 Structure analysis

Final Ramachandran analysis was performed in Rampage (Lovell *et al.*, 2003) in the CCP4 suite (Winn *et al.*, 2011). B-factors were calculated in Baverage (Eleanor Dodson: York 1991) within CCP4. Structure analysis and image generation were performed in PyMOL (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC). Topology maps were generated in PDBsum (Laskowski *et al.*, 1997).

2.2.4 Correction of MalL rates for denaturation

2.2.4.1 Urea denaturation

Unfolding rates for enzymes were determined via fluorometry in a Hitachi F-7000 (Hitachi, Japan), with Peltier temperature control.

2.2.4.1.1 Excitation and emission wavelength determination

Excitation and emission wavelengths for each protein were determined with a 3D scan between excitation wavelengths of 200 nm to 350 nm, and emission wavelengths of 300 nm to 400 nm. Readings were taken at 10 nm windows, with a 5 nm wavelength window in both the excitation and emission domains. Optimal excitation and emission wavelengths as determined by peak intensity in the 3D scan were used for subsequent testing.

2.2.4.1.2 Determination of saturating isomaltose concentrations

Due to the turnover of isomaltose during the denaturation protocol, isomaltose concentrations were calculated to stay above 2x K_M [K_M of 0.205 mM; (Hobbs *et al.*, 2013)] for 45 seconds in the absence of denaturation (i.e. assuming a linear substrate consumption at maximum k_{cat} at T_{opt}, and the known concentration of enzyme in the reaction). Calculations were performed based on the maximal rate of 29 s^{-1} reported for MalL WT at T_{opt} (Hobbs *et al.*, 2013).

2.2.4.1.3 Unfolding assay

An 8 M stock solution of urea was made up in assay buffer (Section 2.2.2) in volumetric flasks at the assay temperatures (305, 310 and 315 K) to account for changes in volume with temperature. This stock solution was then diluted with assay buffer to get urea concentrations ranging from 3-7 M.

Denaturation assays were carried out in 500 µl of urea solution containing 2 mM isomaltose, as determined by methods described in Section 2.2.4.1.2. Urea-isomaltose solutions were brought up to temperature before the addition of 50 µg of protein. Solutions were mixed quickly, and the fluorescence measured until denaturation was complete. Fluorescence was followed with a Hitachi F-7000 fluorimeter (Hitachi, Japan), at the excitation and emission wavelengths of 280 and 330 nm respectively (Section 2.2.4.1.1).

2.2.4.2 Data processing

Denaturation data from three temperatures and various urea concentrations were fit to a single exponential decay curve in GraphPad Prism (version 6.07 for Windows, GraphPad Software, La Jolla California USA, www.graphpad.com) to determine a rate constant, k, for the unfolding process.

From the linear regression analysis of k versus urea concentration at each temperature, the k value in the absence of urea (k_{H2O}) was extrapolated. Data at and below urea concentrations where protein refolding was evident in the divergence of the data from a linear trend were excluded. The k_{H2O} value was used to calculate the Gibbs free energy change for unfolding ($\Delta G^{‡}_{N\text{-}U}$) via the Eyring equation **(12)** as follows, where k_B is the Boltzmann constant, h is Planck's constant, and R is the ideal gas constant.

$$\Delta G^{‡}_{(\mathrm{N-U})} = -\ln\left(\frac{k_{\mathrm{H_2O}}}{k_{\mathrm{B}}\mathrm{T}/h}\right) RT \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \textit{Eyring equation } \mathbf{(12)}$$

2.2.4.3 Correction of enzymatic rates

From the linear regression of $\Delta G^{‡}_{N\text{-}U}$ at three temperatures, $\Delta G^{‡}_{N\text{-}U}$ at the various activity assay temperatures was extrapolated. This was converted to a rate constant for denaturation, k_{H2O}, at each assay temperature via the reverse of the process described in Section 2.2.4.2. k_{cat} rates were then corrected via the following relationship:

$$k_{\mathrm{cat}} = k_{\mathrm{observed}} + k_{\mathrm{H_2O}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \textit{Correction of enzymatic rates } \mathbf{(13)}$$

2.3 Characterisation of glycolysis

2.3.1 Cloning

Gene sequences were obtained from the genome of *Escherichia coli* BL21 Gold (DE3) (NC_012947.1) for the enzymes glucokinase (GK; EC 2.7.1.2, ACT28349.1), phosphoglucose isomerase (PGI; EC 5.3.1.9, ACT30997.1), phosphofructokinase I (PFK; EC 2.7.1.11, ACT30997.1), fructose-bisphosphate aldolase class II (FBPA; EC 4.1.2.13, ACT27881.1), triosephosphate isomerase (TPI; EC 5.3.1.1, ACT31088.1) and glyceraldehyde phosphate dehydrogenase (GAPD; EC 1.2.1.12, ACT28912.1). Primers were designed based on these sequences. All primers were comprised of 20 base pairs at either end of the genes with 5’ or 3’ flanking restriction enzyme sites that included an extra 2-4 base pairs to ensure efficient cleavage (where the length of the overhang was optimised for the specific restriction enzyme). Primers were synthesised by Integrated DNA Technologies (USA). Primer sequences and utilised restriction sites are given in Table 2.2. Freeze dried primers were made up in TE buffer (10 mM Tris HCl pH 8, 1 mM EDTA) to a final concentration of 100 µM.

Table 2.2: Primer sequences for cloning of the first six genes in the glycolysis pathway from E. coli. Restriction enzyme cut sites are indicated in bold.

Gene	**Forward primer (5'-3')**	**Reverse primer (5'-3')**	**Restriction sites**
GK	CGC**GGATCC**ATGACAAAGTATGCATTAGT	AAAA**CTGCAG**TTACAGAATGTGACCTAAGG	BamHI/pstI
PGI	CGC**GGATCC**ATGAAAAACATCAATCCAAC	AAAA**CTGCAG**TTAACCGCGCCACGCTTTAT	BamHI/pstI
PFK_I	CGC**GGATCC**ATGATTAAGAAAATCGGTGT	AAAA**CTGCAG**TTAATACAGTTTTTTCGCGC	BamHI/pstI
FBPA_II	CGC**GGATCC**ATGTCTAAGATTTTTGATTT	CCG**CTCGAG**TTACAGAACGTCGATCGCGT	BamHI/XhoI
TPI	CGC**GGATCC**ATGCGACATCCTTTAGTGAT	CCC**AAGCTT**TTAAGCCTGTTTAGCCGCTT	BamHI/HindIII
GAPD	CGC**GGATCC**ATGACTATCAAAGTAGGTAT	AAAA**CTGCAG**TTATTTGGAGATGTGAGCGA	BamHI/pstI

2.3.1.1 *E. coli* BL21 genomic DNA extraction

E. coli BL21 (DE3) cells (Appendix B.1) were cultured overnight in LB media with 100 µg.ml^{-1} AMP at 37 °C with shaking. An aliquot of cells was added to 1/10th the volume of 5 M guanidinium thiocyanate (GITC) pH 7.0 in a 2 ml screw cap tube containing approximately 0.3 g of 0.1 mm and 2.5 mm zirconia beads. Cells were ruptured with three 30 second rounds of bead beating at setting six in a FastPrep cell disruptor (FP120 Thermo Savant). The lysed cell mixture was deproteinised by the addition of 50 µl 2 M sodium acetate, pH 4.0, and 500 µl phenol-chloroform (50:50 ratio), and gently mixed for 10 minutes. The liquid phases were separated by centrifugation at 11,000 g for one minute, and the top layer of liquid removed to a new tube. An equal volume of isopropyl alcohol was added to the removed top layer, and left at room temperature for 20 minutes. DNA was pelleted by centrifugation at 11,000 g, washed with successive steps of 70 % and 100 % ethanol, and air dried. DNA was resuspended in 50 µl of TE buffer (10 mM Tris HCl, pH 8, 1 mM EDTA) and stored at – 20 °C.

2.3.1.2 Gene insert preparation

2.3.1.2.1 PCR amplification of gene sequences

PCR solutions comprised of 1.5 µl Taq 10 x buffer (without $MgCl_2$), 0.45 µl 50 mM $MgCl_2$, 0.3 µl 10 mM dNTPs, 1.5 µl each of forward and reverse primers (10 pmol.µl^{-1}), 2.5 U *Taq* polymerase (Invitrogen, USA), and 0.15 µl BL21 genomic DNA (10 ng.µl^{-1}), made up to 15 µl with MQ H_2O.

PCR cycling parameters are given in Table 2.3.

Table 2.3: PCR protocol for amplification of glycolysis genes from E. coli BL21 genomic DNA. Steps for denaturation, annealing, and extension were cycled through 30 times.

Step	Temperature (°C)	Step length
Pre-denaturation	95	2 minutes
Denaturation	95	15 seconds
Annealing	55	30 seconds
Extension	72	45 seconds
Post-extension	72	5 minutes

A total of four PCR reactions for each gene were pooled. The purity of the PCR product was assessed by running a 5 µl sample on an agarose gel (Section 2.1.6).

For clean PCR products, DNA was purified from solution using the High Pure PCR Product Purification Kit (Roche Diagnostics, USA) following the manufacturer's instructions. If primer dimers or non-target products were present from the PCR, the whole PRC product was separated by agarose gel electrophoresis. The target band was excised from the gel, and DNA purified through gel extraction (QIAquick Gel Extraction; Qiagen, Germany) according to the manufacturer's instructions.

2.3.1.2.2 Restriction digests

Amplified genes were digested with the appropriate restriction enzyme in a double digest (Roche Diagnostics, USA) (see Section 2.3.1, Table 2.2). Restriction reactions comprised of 40 µl of vector DNA, 10 U per restriction enzymes, and 10 µl of the appropriate 10 x buffer for the double digest, made up to 100 µl with MQ water. All reactions were left at 37 °C for three hours, and purified from solution using the High Pure PCR Product Purification Kit (Roche Diagnostics, USA) as per the manufacturer's instructions.

2.3.1.3 Vector (pPROEX HTb) preparation

2.3.1.3.1 Vector isolation and purification

E. coli pPROEX vector was extracted from 5 ml of overnight culture using the QIAprep Spin Miniprep Kit (Qiagen, Netherlands) according to the manufacturer's instructions. DNA was eluted in 30 µl of elution buffer and stored at - 20 °C.

2.3.1.3.2 Restriction digest and plasmid dephosphorylation

Purified vector was digested with the appropriate restriction enzymes (Roche Diagnostics, USA; see Section 2.3.1, Table 2.2). Restriction digest reactions comprised of 30 µl of vector DNA, 10 U per restriction enzyme, and 10 µl of the appropriate 10 x buffer for the double digest, made up to 100 µl with MQ water. Reactions were left at 37 °C for three hours, and purified from solution using the High Pure PCR Product Purification Kit (Roche Diagnostics, USA) per the manufacturer's instructions.

Cut vector was dephosphorylated using shrimp alkaline phosphatase (SAP; Roche Diagnostics, USA) according to the manufacture's instructions. Dephosphorylated plasmid was purified from solution using the High Pure PCR Product Purification Kit (Roche Diagnostics, USA) per the manufacturer's instructions, and the final

DNA concentration determined using a Nanodrop 2000 UV-vis spectrophotometer (Thermo scientific, USA).

2.3.1.4 Ligation

Ligation reactions were performed using T4 DNA Ligase (Invitrogen, USA) in 20 µl reactions according to the manufacturer's instructions. A molar insert to vector ratio of 3:1 was used in all ligations. Ligations were left at 18 °C overnight.

2.3.1.5 DNA transformations

2.3.1.5.1 Preparation of electrocompetent E. coli

A glycerol stock of *E. coli* DH5α was streaked out on LB agar plate and grown overnight at 37 °C. A single colony was used to start a 5 ml LB culture, and grown overnight at 37 °C with shaking. This starter culture was used to seed a 500 ml LB culture, which was grown with shaking at 37 °C until an OD_{600} of 0.5-0.7 was reached. Cells were chilled on ice for at least 20 mins, then transferred to sterile pre-chilled centrifuge bottles for centrifugation at 3400 g at 4 °C for 20 mins. The supernatant was discarded, and cells resuspended in 500 ml ice cold sterile 10 % (v/v) glycerol. This pelleting and resuspension was repeated with volumes of 250 ml, 20 ml and 1-2 ml of ice cold sterile 10 % (v/v) glycerol. 50 µl aliquots of the final cell suspension in 1-2 ml 10 % (v/v) glycerol were flash frozen in liquid N_2, and stored at -80 °C until use.

2.3.1.5.2 Electroporation

The 20 µl of ligation reaction was added to a 50 µl aliquot of electrocompetent *E. coli* cells (defrosted on ice), along with 50 µl of sterile 10 % (v/v) glycerol. The mixture was placed in the bottom of a 2 mm electroporation cuvette (Bio-Rad Laboratories, USA), ensuring no bubbles were created, and electroporated in a Bio-Rad Gene Pulser^{TM} (Bio-Rad Laboratories, USA) at 2.5 kV, 25 µF capacitance and 200 Ω resistance. Media (1 ml) was added to the cells immediately after electroporation, and incubated at 37 °C with shaking (300 rpm) for one hour. An aliquot (~100 µl) was plated out on media plus 100 $\mu g.ml^{-1}$ AMP plates; the remainder of the cells was pelleted by centrifugation and pelleted out on a second plate. Colonies were grown overnight at 37 °C.

For the above electroporation method, the standard method using LB media throughout was used to clone FBPA, PFK, and TPI. For PGI (F433L) and GAPD, LB with 5 $g.L^{-1}$ glucose was used as the media in all steps. For GK, initial cloning was performed by the standard method with LB media, resulting in a mixed population of mutated strains. The WT sequence was isolated from this from sequencing of multiple strains.

2.3.1.6 Colony PCR

Single colonies from the electroporation plates were selected for PCR confirmation of insert integration. Single colonies were resuspended in 10 µl of H_2O with 100 $\mu g.ml^{-1}$ AMP (and 2 $g.L^{-1}$ glucose for PGI and GAPD enzymes). Resuspended colonies were used as template DNA (2 µl) in PCR reactions (Table 2.3) with general pPROEX primers (Appendix B.5). PCR product was run on an agarose gel to confirm insert presence (Section 2.1.6).

Colonies confirmed as containing an insert of the correct length were inoculated from the remaining suspension in H_2O into 5 ml of LB (with 2 $g.L^{-1}$ glucose for enzymes PGI and GAPD) containing 100 $\mu g.ml^{-1}$ AMP and grown overnight at 37 °C with shaking. Cultures were used to make glycerol stocks (Section 2.1.5) and for plasmid extraction (Section 2.3.1.3.1) for sequencing. Sequencing was done at Massey Genome Services (Massey University, Palmerston North, NZ) with the pPROEX primers (Appendix B.5) as per the standard protocols for the sequencing service.

2.3.2 Protein expression and purification

2.3.2.1 Protein expression trials

Protein expression was initially tested with 100 ml cultures in LB or LB glucose media grown at 37 °C with shaking. Cultures were inoculated with 1 ml of culture grown into lag phase overnight in LB at 37 °C. Cultures were grown to an OD_{600} between 0.5-0.7 before induction with 1 mM IPTG, then left overnight under the same growth conditions as described above. Cells were harvested by centrifugation at 4500 g, and resuspended in 5 ml 50 mM sodium phosphate buffer, pH 7.4, with 50 mM NaCl. Cells were lysed by sonication (Section 2.1.1.1), and centrifuged at

20,000 g, 4 °C. Protein contained in the soluble and insoluble divisions was analysed by SDS-PAGE as described in Section 2.1.1.

2.3.2.2 Large scale protein expression protocols

All six glycolytic enzymes were expressed by the same basic protocol as described below. Variations to this basic method are listed in Table 2.4.

Starter cultures were grown overnight at 37 °C with shaking. Starter cultures (10 ml) were used to inoculate 1 L of media in 2 L baffled flasks. Cultures were grown to an OD_{600} of 0.5-0.7 before induction with 1 mM IPTG. Cultures were grown at 37 °C with shaking over the expression period, then harvested by centrifugation (20,000 g).

Table 2.4: Specific details of the expression protocols for glycolytic enzymes. Expression details were determined based on tendency of the genes to mutate during the cloning processes. Glc = glucose.

Enzyme	Media	Starter volume	Expression period
GK	LB	40 ml	4 hours
PGI	LB + 2 $g.L^{-1}$ Glc	40 ml	4 hours
PFK	LB	10 ml	Overnight
FBPA	LB	10 ml	Overnight
TPI	LB	10 ml	Overnight
GAPD	LB + 2 $g.L^{-1}$ Glc	40 ml	4 hours

2.3.2.3 Protein purification

Protein was purified from the intracellular soluble fraction as described in Section 2.1.1. Buffer compositions for this process are listed in Table 2.5.

Table 2.5: Buffer composition for the IMACS and SEC purification of glycolytic enzymes.

Buffer	**Composition**
IMAC lysis buffer	50 mM HEPES, pH 7.4 150 mM NaCl, 25 mM imidazole
IMAC elution buffer	50 mM HEPES, pH 7.4 150 mM NaCl, 1 M imidazole
SEC buffer	50 mM HEPES, pH 7.4 150 mM NaCl, 10 mM $MgCl_2$

2.3.3 Single enzyme assays – Michaelis Menten and T_{opt} profiles

K_M values for the substrate(s) of each enzyme were determined at 37 °C and either 44 or 45 °C, with saturating concentrations of additional substrates in the reaction. Where substrate inhibition was observed, the substrate was used at the concentration which gained the greatest activity in further assays.

Saturating substrate concentration refers to at least 2 times the K_M concentration of the substrate. Ideally a saturating substrate concentration of 10 x K_M was used. Specific details of all substrate and enzyme concentrations in the Michaelis Menten and T_{opt} assays can be found in Appendix D.2.

Assay data from only the first ten seconds of reaction were analysed in all cases to eliminate any possibility of relevant levels of enzyme denaturation occurring, in consistency with MalL assay protocols (Hobbs *et al.*, 2013).

The same assay buffer was used for all enzymes. The assay buffer comprised of 50 mM HEPES, pH 7.4, 150 mM NaCl, 10 mM $MgCl_2$, and 2.5 mM $(NH_4)_2SO_4$. Assay components were also made up in this buffer.

In cases where coupled enzyme reactions are used, rates were checked to ensure the reaction of interest was the rate determining step. The enzyme of interest was added to the reaction at one and two times concentrations, and a doubling of rate was observed if the step was rate limiting.

T_{opt} assays were carried out at saturating concentration of all substrates, or at the concentration which gave the highest rate where substrate inhibition was observed.

In all cases, reaction mixtures comprising of all components were brought together in a quartz cuvette and heated in a Heλios γ spectrophotometer (ThermoFisher, USA) coupled to a single cell peltier (Thermo Scientific, USA). Assay temperatures were checked in the solution by way of a thermocouple, before reaction initiation with 5 μl of enzyme. Assay progression was followed continuously with readings every 0.125 seconds at the relevant wavelength via Vision32 software (version 1.25). In T_{opt} assays, temperature was also checked via thermocouple at the end of the measurement period.

2.3.3.1 Glucokinase

GK was assayed in a coupled reaction with the indicator enzyme glucose-6-phosphate dehydrogenase (G6PDH from *Leuconostoc mesenteroides*; Sigma Aldrich) to detect the formation of glucose-6-phosphate through the reduction of $NADP^+$ to NADPH at 340 nm (Miller & Raines, 2004). G6PDH was used at concentration high enough to ensure the first step (GK) was rate limiting. $NADP^+$ was used at a concentration of 1 mM (10 x K_{M}) in subsequent assays, based on determined K_{M} values (Section 4.3.1.1).

2.3.3.1.1 Characterisation of Glucose-6-phosphate dehydrogenase

The K_{M} of glucose-6-phosphate dehydrogenase (G6PDH; *Leuconostoc mesenteroides*; Sigma Aldrich, USA) for both glucose-6-phosphate and $NADP^+$ was determined at 37 °C and 44 °C through the detection of the product, NADPH, at 340 nm (Zhu *et al.*, 2011).

A rough temperature profile was carried out to check for sufficient enzyme activity over the range of temperatures necessary for the characterisation of GK and PGI (Appendix D.3).

2.3.3.2 Phosphoglucose isomerase

The reverse reaction of PGI, converting fructose-6-phosphate to glucose-6-phosphate, was assayed in a coupled reaction with the indicator enzyme glucose-6-phosphate dehydrogenase (G6PDH from *Leuconostoc mesenteroides*;

Sigma Aldrich, USA) to detect the formation of glucose-6-phosphate through the reduction of $NADP^+$ to NADPH at 340 nm (Hansen *et al.*, 2004). G6PDH was used at high concentration ensure the reaction of interest was rate limiting. $NADP^+$ was used at a concentration of 1 mM (10 x K_M), based on the determined K_M value.

2.3.3.3 Phosphofructokinase

PFK was assayed in a coupled reaction with FBPA and GAPD to follow the production of NADH at 340 nm by GAPD (Galzigna *et al.*, 1977). Enzyme ratios were determined to ensure the PFK step was rate limiting in the reaction. Additional components necessary for GAPD, P_i (neutralised orthophosphoric acid) and NAD^+, were added at concentrations of ≥ 80 mM and 10 mM respectively based on the K_M values determined in Section 4.3.1.1.

2.3.3.4 Fructose bisphosphate aldose

FBPA was assayed in a coupled reaction with GAPD (Nicholas, 1988). GAPD was used to detect the production of product (glyceraldehyde-3-phosphate: G3P) during the assay via the production of NADH followed at 340 nm. GAPD concentration was set to ensure this enzymatic step was not rate limiting. Additional components necessary for GAPD, P_i (orthophosphoric acid) and NAD^+, were added at concentrations of ≥ 80 mM and 10 mM respectively based of the K_M values determined in Section 4.3.1.1.

2.3.3.5 Triosephosphate isomerase

TPI was assayed in a coupled reaction with GAPD through the detection of product (glyceraldehyde-3-phosphate) coupled to the reduction of NADH followed at 340 nm (Sullivan *et al.*, 2011). Assays were set up so that TPI concentrations were not rate limiting. Additional components necessary for GAPD, P_i (as orthophosphoric acid) and NAD^+, were added at concentrations of ≥ 80 mM and 10 mM respectively based of the K_M values determined in Section 4.3.1.1, Figure 4.12.

2.3.3.6 Glyceraldehyde phosphate isomerase

GAPD was assayed in a coupled reaction with FBPA. FBPA was used to produce glyceraldehyde-6-phosphate prior to and/or during the assay of GAPD. Reaction mixtures also contained P_i (as orthophosphoric acid, neutralised with NaOH) and NAD^+ as substrates for GAPD. The reaction was followed through the production of NADH at 340 nm (Eyschen *et al.*, 1999a).

Prior to the addition of GAPD, 330 µl of FBP (variable concentration) was made up in assay buffer in a 1.5 ml Eppendorf tube. FBPA was added (2 µl, 9 mg.ml^{-1}), quickly mixed and warmed in a 37 °C water bath for two minutes. After two minutes of reaction, 320 µl of substrate was removed and added to 40 µl each of NAD^+ (in assay buffer) and phosphate solution (orthophosphoric acid, adjusted to pH 7.4 with NaOH) in a quartz cuvette. This mixture was brought up to temperature by a single cell Peltier (Thermo Scientific, USA) before the addition of 2 µl of GAPD to start the reaction.

Note that in this case, $K_{M\ (\text{fructose-1,6-bisphosphate})}$ does not refer to the affinity of GAPD for glyceraldehyde-6-phosphate, but instead the amount of fructose-1,6-bisphosphate required under the specific conditions described to produce glyceraldehyde-6-phosphate at a level required to half saturate GAPD active sites.

2.3.4 Calculation of T_{opt} and T_{inf} from Macromolecular Rate Theory fits

From the fit of equation **(7)**, precise T_{opt} and T_{inf} (the temperature of the lower inflection point) temperatures were calculated. Providing $\Delta C^{\ddagger}_{p} < 0$, T_{opt} and T_{inf} were calculated with equation **(14)** and **(15)** respectively.

$$T_{\text{opt}} = \frac{\Delta H^{\ddagger}_{T_0} - \Delta C^{\ddagger}_{p}.T_0}{-\Delta C^{\ddagger}_{p} - R} \text{ .... } T_{\text{opt}} \text{ calculation } \mathbf{(14)}$$

$$T_{\text{inf}} = \frac{\Delta H^{\ddagger}_{T_0} - \Delta C^{\ddagger}_{p}.T_0}{-\Delta C^{\ddagger}_{p} \pm \sqrt{-\Delta C^{\ddagger}_{p}.R}} \text{ .... } T_{\text{inf}} \text{ calculation } \mathbf{(15)}$$

2.3.5 Pathway characterisations – T_{opt} profile

Full pathway characterisation was run in assay buffer with additional $MgCl_2$ (15 mM final) to optimise formation of the active ATP-Mg complex. Substrate concentrations used were based on K_M values for the individual enzymes (Section

4.3.1.1), giving final concentrations of 10 mM glucose, 7 mM ATP, 10 mM NAD^+, and 100 mM P_i. Substrates were made up in assay buffer. Specific details of all substrate and enzyme concentrations in the assays can be found in Appendix D.2.

All components except the enzymes were brought up to temperature in a quartz cuvette in a heated Heλios γ spectrophotometer (ThermoFisher, USA) coupled to a single cell Peltier (Thermo Scientific, USA). Precise solution temperature was checked with a thermocouple. At a stable temperature, the reaction was started with 13 µl of enzyme mixture, mixed quickly, and the reaction followed via the production of NADH at 340 nm every 0.125 seconds for 60 seconds via Vision32 software (version 1.25). At the culmination of the assay period, temperature was taken again with a thermocouple to ensure stable temperatures over the assay course.

Data were analysed by best fit to the linear period past the 20 second lead in period at each temperature, after which stead state rates were achieved. Final temperature readings were taken at the reaction temperature for each assay.

2.3.6 Differential scanning fluorimetry T_m determination

Rough protein T_m values were determined by differential scanning fluorimetry using the SYPRO dye method (Lo *et al.*, 2004). SYPRO dye (Life Technologies, US; 5000x) was diluted down to 300x in buffer. SYPRO dye (7.5 µl, 300x) was added to glycolysis assay buffer (Section 2.3.5) and protein to a final volume of 25 µl with 0.1 $mg.ml^{-1}$ protein. Melts were performed from 25-99 °C with continuous monitoring at excitation and emission wavelengths of 490 and 530 nm respectively using the Roto-Gene™ real time PCR machine (Corbett Life Science, Australia). Blanks were performed under the same conditions without protein.

Two sets of melts were performed with different buffers used in the differential scanning fluorimetry preparation and protein dilution – assay buffer and assay buffer with ligands (NAD^+, P_i, and ATP) present to simulate the assay conditions as closely as possible. Ligand concentrations were used at the concentrations used in the pathway assays (10 mM NAD^+, 100 mM P_i, 7 mM ATP).

Data were analysed in GraphPad Prism (version 6.07 for Windows, GraphPad Software, La Jolla California USA, www.graphpad.com) by taking the first derivative of the raw fluorescence values. The T_m was taken from this plot as the highest dF/dT values (inflection point). Second derivatives were not taken as the process generated too much noise from the raw data.

2.3.7 Pathway modelling

Two modelling approaches were undertaken to generate an estimate of the overall pathway curvature from individual enzyme data.

2.3.7.1 Inverse rates

Experimentally determined temperature profiles were fit with MMRT equations (temperature dependent or independent as necessary); best fit lines were used to generate k_{cat} values at equal temperature intervals.

Overall pathway rates were calculated with equation **(16)** based on k_{cat} values at each temperature, and fit to MMRT.

$$\frac{1}{k_{tot}} = \frac{1}{k_1} + \frac{1}{k_2} + \cdots \frac{1}{k_n} \text{ .... } \textit{pathway rate } \textbf{(16)}$$

2.3.7.2 CellML modelling

More extensive modelling was performed in CellML software (Yu *et al.*, 2011) to consider changing component concentrations, and the effect of substrates below K_M through the pathway.

For the model, K_M values, and other relevant parameters, were taken from Michaelis Menten data at 37 °C. Rate values were taken from MMRT fit to the experimental temperature profiles at various temperatures. Equations to determine the rate of each enzymatic step are given in Table 2.6; the full script is given on the attached disk.

Due to the reversible and fast rates of TPI, instead of modelling the enzymatic turn over through TPI, the triose pool components (DHAP and GAP) were modelled by an equilibrium constant (K_{eq}; for the reaction DHAP → GAP). A K_{eq} value of 0.045 (favouring DHAP) was set based on literature values (Teusink *et al.*, 2000). All other enzymes in the pathway were modelled as irreversible, as the experimental set up highly favours forward progression of substrate through the pathway.

Table 2.6: Equations used to model enzymatic rates, based on experimentally determined parameters. Any assumptions and simplifications for each enzyme are also given.

Enzyme	Equation
GK	$v = \dfrac{V_{max}.[ATP]^{h_{ATP}}.[GLC]}{K_{hlf_{ATP}}^{h_{ATP}}.K_{M_{GLC}} + [ATP]^{h_{ATP}}.[GLC]}$
	Glucose inhibition removed from the equation; K_i values are too high (30-50 mM) to have a significant effect on pathway rates.
PGI	$v = \dfrac{V_{max}.[G6P]}{K_{M_{G6P}} + [G6P]}$
	Assumed that V_{max} in the forward direction is the same as the reverse (as measured) and $K_{M\ G6P} \approx K_{M\ F6P}$ as measured for the reverse reaction. These assumptions are justified based on the previous characterisation of *E. coli* PGI in forward and reverse directions, demonstrating similar K_M (0.28 and 0.147 mM respectively) and V_{max} (213 and 123 μM.min^{-1}.mg^{-1} respectively) (Gao *et al.*, 2005).
PFK	$v = \dfrac{V_{max}.[F6P]^{h_{F6P}}.[ATP]}{K_{hlf_{F6P}}^{h_{F6P}}.K_{M_{ATP}} + [F6P]^{h_{F6P}}.[ATP]}$
	ATP inhibition removed from the equation; K_i values are too high (30 mM) to have a relevant effect.
FBPA	$v = \dfrac{V_{max}.[FBP]}{K_{M_{FBP}} + \dfrac{[FBP].(1 + [FBP])}{K_{i_{FBP}}}}$
GAPD	$v = \dfrac{V_{max}.[GAP].[NAD^{+}].[P_i]}{K_{M_{GAP}}.K_{M_{NAD^{+}}}.K_{M_{P_i}} + [GAP].[NAD^{+}].[P_i]}$
	Model does not incorporate inhibition by GAP and NAD^{+}. The K_i value for NAD^{+} is high (90-170 mM); inhibition seen in GAP Michaelis Menten characterisations is not significant at low temperatures (K_i = 240), but may become significant at higher temperatures. It is also assumed that true $K_{M\ GAP}$ is slightly less than or equal to the experimentally determined $K_{M\ FBP}$ (Section 4.3.1.1).

2.4 Molecular dynamics of MalL

Molecular dynamics (MD) was performed within the AMBER 14 suite of programs (Case *et al.*, 2014). Trajectories were visualised in VMD (Humphrey *et al.*, 1996).

2.4.1 System set up

2.4.1.1 Structure preparation

Dynamics for MalL WT and V200S was based off crystallographic structures with the TS analogue DONM bound (Section 2.2.3). MalL WT was reduced to one molecule from the asymmetric unit; chain A was selected due to higher quality data supporting atomic positions compared to the second molecule in the asymmetric unit (Table 5.2). Structures were prepared by building in loops missing from the crystallographic structures in COOT (Emsley & Cowtan, 2004). This included loop Thr215-Arg219 in the WT structure, and two N-terminal residues in both structures. Optimisation of ideal bond parameters was relied on to build in loops in a realistic orientation; electron density too poor for absolute atomic placement was used to guide building if present. Bound molecules from the precipitant and cryo-protectant solutions were removed. For residues with two conformations, only the most well represented in the electron density was retained. For V200S, starting structures with Ser200 in both crystallographic orientations were prepared due to the importance of this residue (Section 5.3.3.2).

Substrate (isomaltose) was modelled into the MalL structures based on the PyMOL (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC) overlay with E277A MalL from *Saccharomyces cerevisiae* (Yamamoto *et al.*, 2011). Building of the disaccharide TS analogue was based from isomaltose; additional components were constructed using PyMOL builder. Crystallographic waters were retained in the structures unless clashes with atoms of the ligand positioned in the active site were present. Crystallographic waters were retained on the basis that these are in positions supported by experimental data.

2.4.1.2 Determination of system charges and protonation states

The protonation state of residues was determined with ligand included in the structure with PROPKA 3.1 (Olsson *et al.*, 2011; Søndergaard *et al.*, 2011). Suggested protonation states were manually checked in the structure. Hydrogen bonding networks and side chain tautomer alterations were checked using the WHAT IF optimal hydrogen bonding network tool (Vriend, 1990). No side chain orientations were found to require alteration to optimise hydrogen bonding networks in either of the MalL structures. Partial charges of the ligands were determined with the online R.E.D. Server (Vanquelef *et al.*, 2011).

2.4.1.3 Force fields used for simulations

The protein component of systems was parameterised with the General Amber Force Field [GAFF; (Wang *et al.*, 2004)]. For the ground state system, ligand [isomaltose; D-Glc-α(1-6)-D-Glc] was modelled with the GLYCAM-06 force field specifically designed for parameterisation of carbohydrates (Kirschner *et al.*, 2008). For consistency, the unmodified glucose unit of the TS analogue was modelled with GLYCAM. This was combined with GAFF parameters where limitations in the GLYCAM pre-set sugar units did not allow GLYCAM to be applied across the modified sugar region of the TS analogue. Further details including the structure and force fields used across the TS structure are presented in Section 6.1.1 and Figure 6.2.

2.4.1.3.1 Confirmation of GAFF as an accurate force field for parameterisation of sugars

In order to verify the comparison between systems modelled with GAFF and GLYCAM force fields, 60 ns of simulation (set up as described in Section 2.4) was analysed for MalL WT with glucose bound in the -1 binding pocket modelled with the two parameter sets. Dihedrals, angles and bond distances about the sugar ring were analysed with CPPTRAJ (Roe & Cheatham, 2013) and compared to confirm the similar parameterisation of a sugar unit with the two force fields.

2.4.1.4 Solvation and balancing of system charges

The system was solvated using tLEaP (Wei Zhang, Tingjun Hou, Christian Schafmeister, Wilson S. Ross, David A. Case) with TIP4P-Ew water molecules

(Horn *et al.*, 2004) in a truncated octahedral periodic box about the protein solute. The size of the periodic box was set so that no atom of the protein was less than 11 Å from its edge.

System charges were balanced with the random addition of Na^+ ions in tLEaP (Wei Zhang *et. al.*). Different starting structures (ten per state) were made by randomising Na^+ positions in the solvent with CPPTRAJ (Roe & Cheatham, 2013).

2.4.2 Minimisation

To relax unfavourable contacts, systems were minimised by gradually reducing constraints on the system. Solvent and hydrogens were initially minimised followed by heating from 50 to 300 K. The system was minimised and heated (25-320 K) with constraints on Cα positions of the protein. Restrictions at Cα positions were reduced from 5 to 1 $kcal.mol^{-1}.Å^2$ in four steps, followed by one ns of unrestrained simulation to equilibrate the system.

2.4.3 Dynamics simulations

Dynamics simulations were run for 500 ns in steps of 2×10^{-6} ns. The system state was written out at 0.01 ns intervals. Simulations were run at 320 K to match the active range of the MalL variants (Hobbs *et al.*, 2013). Initial simulation with substrate resulted in loss of ligand from the active site; final simulations included constraints between the substrate (centre of mass of the glucose unit in the -1 binding site) and active site residues (Cα of 255 and 332) in the binding pocket.

2.4.4 Analysis of trajectories

Trajectories were analysed within the AMBER 14 suite of programs (Case *et al.*, 2014) and visually in VMD (Humphrey *et al.*, 1996).

2.4.4.1 RMSD and RMSF analysis

RMSD over individual trajectories was calculated in CPPTRAJ (Roe & Cheatham, 2013) for Cα atoms. Individual frames were overlaid prior to RMSD calculation.

RMSF values were calculated for Cα atoms for each trajectory in CPPTRAJ (Roe & Cheatham, 2013) using a sliding average window of 60 ns. A window size of 60 ns was used to replicate the vibrational modes captured in successful $\Delta C^{\ddagger}_{\mathrm{p}}$ calculations (Section 6.3.2).

2.4.4.2 Outlier analysis

Cluster analysis within the CPPTRAJ program suite (Roe & Cheatham, 2013) was used to assess outlier conformations across multiple trajectories. Cluster was run with the N-terminal truncated (residues 7-561 included) to avoid large influences from the flexible terminal residues. Sieves were reduced through cluster iterations till all clusters had reasonable occupancy from at least one simulation.

Simulations with high occupancy of a conformation not highly populated in other simulations were further investigated with RMSD data (Section 2.4.4.1) and visualised in VMD (Humphrey *et al.*, 1996). Trajectories were only considered for removal if multiple data indicated an anomalous run, including identifiable conformations unamenable to the reaction coordinate.

2.4.4.3 Moving variance analysis to calculate $\Delta C^{\ddagger}_{\mathrm{p}}$

System energies without water were recalculated for $\Delta C^{\ddagger}_{\mathrm{p}}$ analysis. Waters were stripped using Closest within the CPPTRAJ software (Roe & Cheatham, 2013). Energies were calculated at 0.01 ns intervals to give 50,000 instances per 500 ns simulation. The first 50 ns of this was not considered further as the system was still equilibrating in this period.

A sliding window of varying time length (5-100 ns) was used to calculate mean and variance values for enthalpy across 50-500 ns of simulation time for each run. Variance values were averaged over the run to get the variance per trajectory. A total of ten runs per state were considered and averaged; differences in enthalpy variation ($\Delta{<}\delta H^2{>}$) between the substrate and TS analogue simulations for each window size were taken through to $\Delta C^{\ddagger}_{\mathrm{p}}$ calculations via equation **(17b)** (Prabhu & Sharp, 2005).

$$\Delta C_p = \frac{\Delta{<}\delta H^2{>}}{kT^2} \text{……………..………..} \textit{Heat capacity as a function of enthalpy} \quad \textbf{(17b)}$$

$\Delta C^{\ddagger}_{\mathrm{p}}$ was plotted against window size and a final *in silico* $\Delta C^{\ddagger}_{\mathrm{p}}$ values assessed from the plateau in $\Delta C^{\ddagger}_{\mathrm{p}}$ with increasing window size.

Proofs and implications of MMRT

3.1 Accounting for the effect of denaturation on MalL rates

Protein denaturation at temperatures above enzymatic T_{opt} remains a prevalent explanation for the declines in enzymatic activity at high temperatures, despite contradictory evidence that denaturation is insufficient to explain the data (Thomas & Scopes, 1998). Despite the short (ten second) assay time used in MalL assays to greatly reduce the influence of denaturation, further proof that denaturation is not the cause of decreasing rates at high temperatures is presented here, specific to the enzyme used as the central model in MMRT analysis.

The temperature dependence of Gibbs free energy of unfolding ($\Delta G^{\ddagger}_{(\mathrm{N\text{-}U})}$) and the unfolding rate (k_{u}) for both MalL WT and V200A were determined through urea denaturation (Figure 3.1 and Figure 3.2 A). Denaturation was followed via the intrinsic fluorescence of MalL in increasing concentrations of urea, allowing extrapolation to denaturation rates in the absence of urea (i.e. assay conditions; Figure 3.1). Denaturation was performed in assay buffer with substrate (isomaltose) present at an excess concentration to replicate assay conditions (Section 2.2.4.1.2). The relative stability of the two enzyme variants is consistent with previous data on the rigidification and additional hydrogen bonds of V200S compared to WT MalL, as summarised in Section 1.4.2 (Hobbs *et al.*, 2013). The linear change in $\Delta G^{\ddagger}_{(\mathrm{N\text{-}U})}$ with temperature allowed extrapolation of $\Delta G^{\ddagger}_{(\mathrm{N\text{-}U})}$ to assay temperatures, and the calculation of denaturation rates in these conditions (via $\Delta G = RT.\ln k_{\mathrm{u}}$, where k_{u} is the denaturation rate; Section 2.2.4.2). The extent of denaturation over a ten second assay extrapolated to assay temperatures is minimal (Figure 3.2 B), with a denaturation rate of 6.10×10^{-2} s^{-1} for the WT enzyme at the highest assay temperature. Thus, the observed assay rate ($k_{\mathrm{obs}} = 8.88\ \mathrm{s}^{-1}$) is only minimally altered in the correction, resulting in a true rate (k_{cat}) of 8.94 s^{-1}, an insignificant difference given the errors associated with the data ($\pm\ 0.1\ \mathrm{s}^{-1}$). With denaturation corrected for, rates with temperature for both enzyme variants are still distinctly curved (Figure 3.2 C). $\Delta C^{\ddagger}_{p}$ values from the fit to the data, or measures of fit of MMRT to the data (R^2), are not affected by the correction for denaturation compared to the original k_{obs} data.

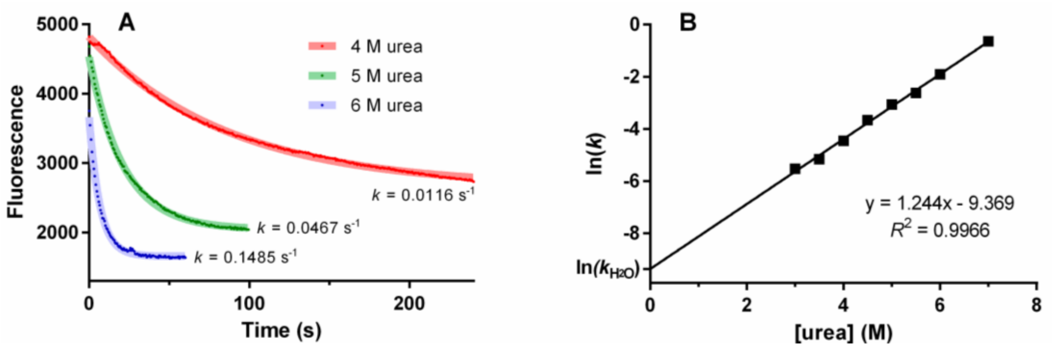
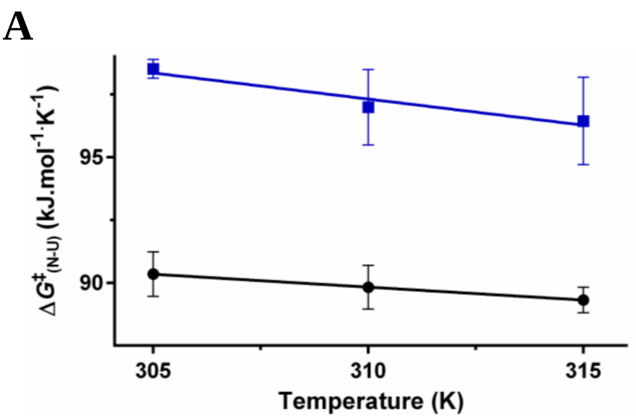


Figure 3.1: Urea denaturation data for MalL V200S at 305 K. ***(A)*** *Unfolding of V200S in various concentrations of urea. Data are shown in dark squares, fit to an exponential decay curve (pale line).* ***(B)*** *Linear regression of rates of denaturation in urea concentrations, extrapolated to* k_{H2O} *(rate of denaturation in the absence of urea).* $\Delta G^{\ddagger}_{(N\text{-}U)}$ *is calculated from* k_{H2O} *using equation* ***(12)****;* $\Delta G^{\ddagger} = -RT.\ln k_u$.



V200S ■
$\Delta G^{\ddagger} = -0.2078.T + 161.7$
$R^2 = 0.9318$

WT ●
$\Delta G^{\ddagger} = -0.1029.T + 121.7$
$R^2 = 0.9998$

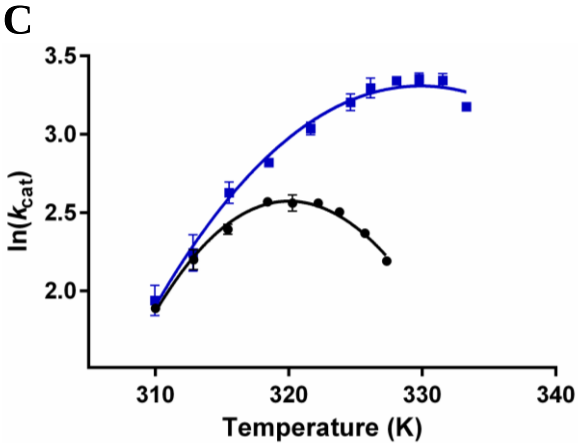
B

WT

T (K)	k_{obs} (s^{-1})	k_u (s^{-1})	k_{cat} (s^{-1})
310.0	6.60	4.74×10^{-3}	6.61
312.9	9.03	7.37×10^{-3}	9.04
315.4	10.95	1.09×10^{-2}	10.96
318.4	13.04	1.69×10^{-2}	13.05
320.3	12.94	2.23×10^{-2}	12.96
322.2	12.93	2.95×10^{-2}	12.96
323.8	12.19	3.70×10^{-2}	12.23
325.7	10.64	4.85×10^{-2}	10.68
327.3	8.88	6.10×10^{-2}	8.94

V200S

T (K)	k_{obs} (s^{-1})	k_u (s^{-1})	k_{cat} (s^{-1})
310.0	6.97	2.52×10^{-4}	6.97
312.8	9.46	4.51×10^{-4}	9.46
315.5	13.84	7.75×10^{-4}	13.84
318.5	16.78	1.39×10^{-3}	16.78
321.6	20.85	2.55×10^{-3}	20.85
324.6	24.66	4.49×10^{-3}	24.66
326.1	27.00	5.95×10^{-3}	27.01
328.1	28.28	8.50×10^{-3}	28.29
329.8	28.60	1.16×10^{-2}	28.61
331.5	28.32	1.60×10^{-2}	28.33
333.3	23.94	2.19×10^{-2}	23.96



V200S ■
$\Delta C^{\ddagger}_p = -5.9 \pm 0.5$ kJ.mol^{-1}.K^{-1}
$R^2 = 0.978$

WT ●
$\Delta C^{\ddagger}_p = -11.6 \pm 0.4$ kJ.mol^{-1}.K^{-1}
$R^2 = 0.975$

Figure 3.2: Correction of MalL WT and V200S rates (k_{cat}) for denaturation rates (k_u) through urea denaturation. ***(A)*** *The temperature dependence of $\Delta G^{\ddagger}_{(N-U)}$. Error bars represent the standard error of the y intercept from the linear regression performed in the determination of k_{H2O} (Section 2.2.4.2).* ***(B)*** *Measured rate (k_{obs}) corrections based on unfolding rates over a ten second assay, giving true k_{cat} rates.* ***(C)*** *Corrected k_{obs} rates for denaturation occurring over the ten second assay course to reflect true k_{cat}. Error bars where visible are the standard deviation of three assay replicates.*

As shown previously for other systems (Thomas & Scopes, 1998), denaturation is not a confounding parameter in MalL assay data over the temperature range covered in assays (Figure 3.2 C). The single exponential fit to data here simplifies the unfolding process which likely involves unfolding intermediates, however, this approach captures the major global unfolding events with a significant impact on enzyme rates. Rate data after these corrections for denaturation display a curved relationship with temperature, which is accounted for by the MMRT function.

3.2 Enzyme mass and efficiency

Protein synthesis is an energetically expensive process for cells, consuming an estimated 80 % of cellular ATP, requiring the cleavage of four phosphate bonds per amino acid addition (Cox & Cook, 2007). Given this large energetic cost of synthesis, and given the strong evolutionary pressures driving energy efficiency in cells, it is expected enzymes will be the minimum size necessary to function. Enzymes are large biomolecules, especially when considering the comparison to small inorganic catalysts which essentially perform the same basic function. Given these factors, the large mass of enzymes should logically be playing some essential role in the catalytic process. It is also interesting to note that enzymes achieve rate enhancements many orders of magnitude greater than small molecule based catalysts (Wolfenden & Snider, 2001), suggesting that the source of an enzymes catalytic power lies, in part, in the greater mass.

Previously this concept has been addressed in the literature with a wide range of explanations. Kell (1982) explored the idea of enzymes functioning as energy funnels, conveying translational energy from the solvent to the active site. Reyes *et al.* (2015) describes loops outside the active site acting to create a caged environment ideal for stabilising the TS, while TS stabilisation has also been invoked in Britt (1997), with a correlation found between TS stabilisation and the enzyme-substrate mass ratio. Alternative ideas place the extra mass of an enzyme in a structural role, concentrating substrates at the active site (Payens, 1983), providing binding surfaces for integration into metabolic complexes or cellular binding sites (Srere, 1984), or a general requirement for maintaining the active site under an active-site preorganisation model (Warshel *et al.*, 2006).

In MMRT, the proposed concept is that protein flexibility is partially lost in the shift from the substrate to the TS bound state as vibrational modes shift to higher frequencies. Here it is suggested that a portion of this energy difference represented by the $\Delta C^{\ddagger}_{\mathrm{p}}$ is used to drive the catalytic process. Linking $\Delta C^{\ddagger}_{\mathrm{p}}$ to enzymatic catalysis is consistent with the finding in Arcus and Pudney (2015) where the effect of $\Delta C^{\ddagger}_{\mathrm{p}}$ is evident in systems where vibrational modes have been linked to catalysis. From this hypothesis, an enzyme's heat capacity, as a measure of the internal energy partitioned over various vibrational modes, acts as an energy reservoir, some of which is available to facilitate catalysis (Arcus *et al.*, 2016). In such a model, larger enzymes, as greater reservoirs of C_{p}, have the capacity to be more effective catalysts. Assuming this is true, enzymes should increase in size to an extent related to their catalytic proficiency (k_{cat}/k_{non}, where k_{non} is the uncatalysed rate in the absence of enzyme). This is the trend observed in a range of enzymes catalysing unimolecular reactions (Figure 3.3). There is an equivalent way of considering this, given that uncatalysed rates vary over many orders of magnitude, whereas enzymatic rates tend to fall within a narrow range of rates (Wolfenden & Snider, 2001); the more difficult the chemistry (i.e. the slower the k_{non}), the larger an enzyme needs to be to bring the catalysed rate into this biologically relevant range of rates. This argument implies that dynamics over the whole protein system is important to catalysis, and may be one of the reasons that biomimetic catalysts, which bind to TS analogues tightly, do not achieve the rate enhancements of their enzymatic counterparts (Schwartz & Schramm, 2009).

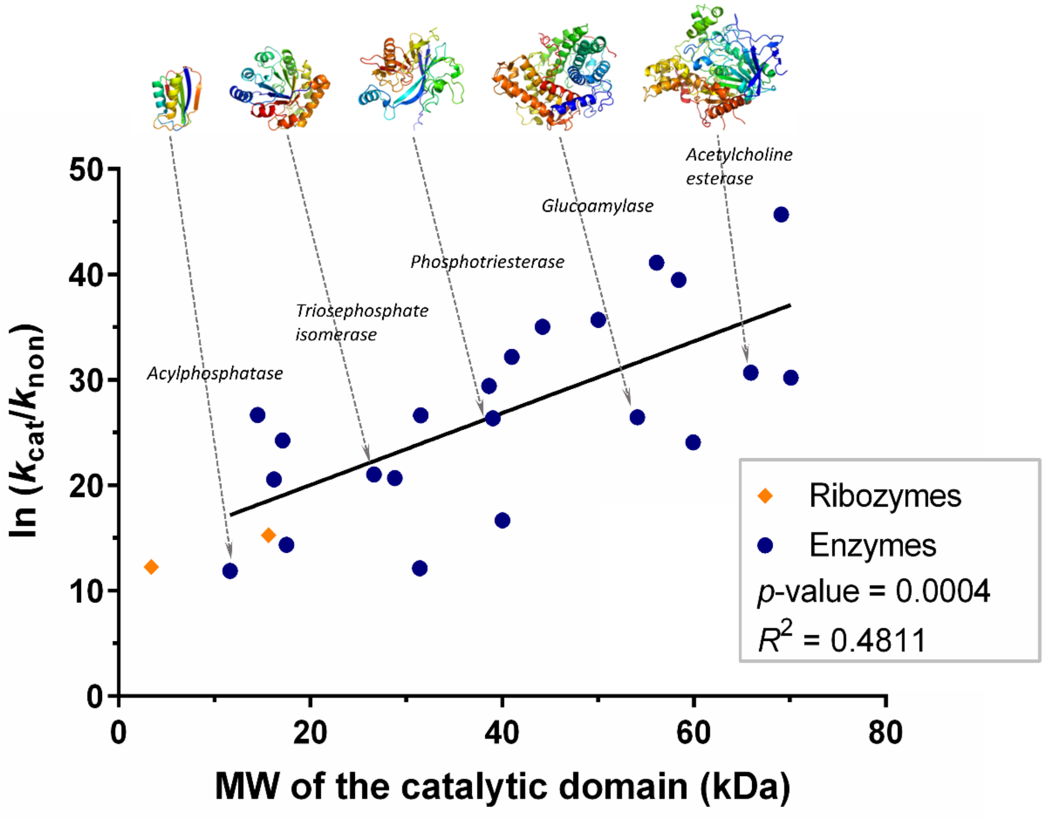


Figure 3.3: Relationship between catalytic proficiency (k_{cat}/k_{non}) and molecular weight of the catalytic domain (Arcus et al., 2016). Representative structures corresponding to data points are given across the top. Details, including data sources, is given in Appendix D.1.

To investigate the relationship between enzyme mass and catalysis, only unimolecular reactions have been considered (hydrolases, ligases, and isomerases), as these are not confounded by the additional factor of bringing multiple substrates together. For a process involving more than one substrate, preorganisation of the active site is likely to have a significant influence. Catalytic domains alone have been considered when auxiliary domains are not part of the main globular structure as these are unlikely to be part of the vibrational modes which contribute energy to the reservoir available for catalysis. Furthermore, only monomer weights have been considered for simplicity. The one exception to this is orotidine-5'-phosphate decarboxylase [mass 58.4 kDa and ln(k_{cat}/k_{non}) 39.5], where residues interact with the substrate between the two dimeric units (Harris *et al.*, 2000). Due to this high level of cooperation over the two subunits, the dimeric weight of this enzyme has been graphed (Figure 3.3).

On this graph, it is of interest to note that monomers and dimers are situated to the bottom left of the graph, representing low rate enhancements for weight, whereas enzymes with high rate enhancements are multimers situated to the top right of the trend (colour-coded in Figure 3.4 A). Consistent with this, catalytic efficiency plotted against MW of the catalytic domain multiplied by the oligomeric state maintains a linear relationship (Figure 3.4 B). This suggests there is potential for vibrational interaction across the oligomers, resulting in a shared reservoir for catalysis, although it is acknowledged that this is a highly controversial aspect of the data, and confounded by complexities such as assay condition effects on multimerization.

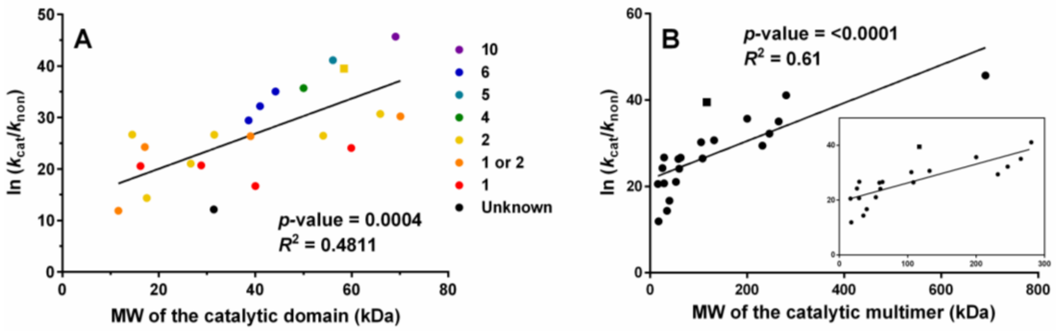


Figure 3.4: Incorporating the effect of oligomerisation into the relationship between molecular weight and enzyme efficiency. ***(A)*** *enzymes coloured by oligomeric state. Enzymes active in tetramers and above are consistently more efficient for their size than the overall trend. The exception to this is orotidine-5'-phosphate decarboxylase (square), which has an unusual mode of action (discussed in text).* ***(B)*** *Data replotted as MW of the catalytic domain multiplied by the oligomeric state. Data fits a linear equation with an improved R^2 compared to monomeric weight alone (Figure 3.3). Removal of the high value (arginine decarboxylase decamer) is illustrated; the linear relationship is maintained with no effect on R^2 or p-value. Where the literature has been ambiguous between a monomeric or dimeric state, both possibilities are labelled* ***(A)****, and the MW has been multiplied by 1.5 in* ***(B)*** *to encompass both possibilities.*

The relationship between enzyme efficiency and mass also has implications for thermophilic enzymes. As reactions at higher temperatures have faster k_{non} values, especially for slow reactions which respond more dramatically to temperature (Wolfenden, 2014), there is less requirement for enzyme input to bring catalysed rates into the bracket of biologically functional rates. Thus, thermophilic enzymes have the potential to be smaller (saving production costs), while still maintaining rate enhancements commensurate with sustaining life. This trend towards smaller thermophilic enzymes through loop deletions has been noted in a comparison between the complete genomes of 20 organisms (Thompson & Eisenberg, 1999). The less stringent requirements for efficiency on thermophilic enzymes has been proposed to alleviate the requirement for highly efficient (i.e. evolved) enzymes in primitive life forms, provided life originated in hot environments (Wolfenden, 2014).

The relationship between ribozymes and size has also been investigated. Few catalytic RNA molecules exist, of which most are small. The two small ribozymes investigated are consistent with the trends seen for the small enzymes (Figure 3.3). However, of special interest is the large RNAse P, which despite a molecular weight of 120 kDa, larger than the range of enzymes studied here, only achieves a slow rate of 0.02 s^{-1} ($\ln k_{cat}/k_{non} = 15$) (Wolfenden & Snider, 2001; Kikovska *et al.*, 2005). This trend may be consistent with the relative inflexibility of RNA compared to protein. Given that structured RNA is a less flexible molecule, additional molecular weight does not add significantly to the overall energy reservoir available to drive catalysis. Thus, much larger molecules are required to realise the same catalytic proficiency as achieved in proteins. It is also possible that this relationship is relevant to the RNA world hypothesis (Forterre, 2005). As ribozymes are poor catalysts of increasingly difficult chemistries, this may have been a driving factor in the switch to utilisation of proteins for this process (Arcus *et al.*, 2016).

Does MMRT scale from enzymes to ecosystems?

In previous publications on MMRT, the theory has been extended in possible applications to organism growth rates (Hobbs *et al.*, 2013) and ecosystem responses to temperature (Schipper *et al.,* 2014). Hobbs *et al.* (2013) initially proposed the application of the model to organisms based on the fact that enzymes are the primary driver of cellular metabolism and MMRT showed a good fit (R^2 consistently greater than 0.95) to growth rate data from various bacteria and yeast, with similar $\Delta C^{‡}_{p}$ values as observed for single enzymes (Section 1.5.1, Figure 1.10). In Schipper *et al.* (2014), MMRT was extended to ecosystems based on the fit to the summed rates of a consortium of species with different T_{opt} values and curvatures with temperature. Again, MMRT showed a good fit, with expected $\Delta C^{‡}_{p}$ values, to data collected both in the field and under laboratory controlled conditions [Section 1.5.2; Figure 1.12 and also Schipper *et al.* (2014) Figure 3 for additional experimental fits].

However, both applications rely on the assumption that MMRT, as developed for single enzymes, scales to organisms with pathways of enzymatic steps. MMRT was extended from the established understanding of chemical reactions to better describe the more complex thermodynamics of enzymatic systems and observed temperature dependence of rates. Application of the MMRT function to organism growth rates in Hobbs *et al.* (2013) has not been shown to be more than a fortuitous fit.

Here, to confirm that the principles of MMRT are affecting organism growth rates, a short metabolic pathway is characterised as a proxy for a simplified cell metabolism, to investigate the application of MMRT to organism growth rates, and how individual enzymes effect overall pathways.

4.1 Pathway selection: Glycolysis

To test the scalability of MMRT, the first six steps of the glycolytic pathway were selected as an experimental test system to characterise *in vitro*. Overall, an *in vitro* pathway is the ideal experimental system as it is easy to manipulate and eliminates many of the confounding factors at the cellular level [e.g. regulation at the expression and effector binding level, compartmentalisation of components, and macromolecular crowding (Teusink *et al.*, 2000)]. In a cellular system, these factors will be condition dependent and layer on top of MMRT to determine overall responses to temperature (Schipper *et al.*, 2014).

The glycolytic pathway was selected as it is a core metabolic process, essential to a wide range of organisms; the growth rate of *E. coli* is fuelled by glucose as the carbon source through the glycolytic pathway when available (Brückner & Titgemeyer, 2006). The first six enzymes were selected for characterisation as all have previously established, continuous assay protocols. The last of the six steps (glyceraldehyde phosphate dehydrogenase; GAPD) includes the production of NADH, making the full pathway readily measurable in a continuous assay. The first part of the glycolytic pathway (Figure 4.1) involves a range of classes of enzymes including isomerases (phosphoglucose isomerase; PGI and triosephosphate isomerase; TPI), kinases (glucokinase; GK and phosphofructokinase; PFK), an oxidoreductase (GAPD) and a lyase (fructose-bisphosphate aldolase; FBPA). The pathway has a branch point, where the dihydroxyacetone phosphate (DHAP) produced in the cleavage by FBPA is fed back into the main pathway via isomerisation catalysed by TPI. This feature starts to capture some of the complexity occurring in metabolic systems, where enzymes are not working exclusively in a linear series.

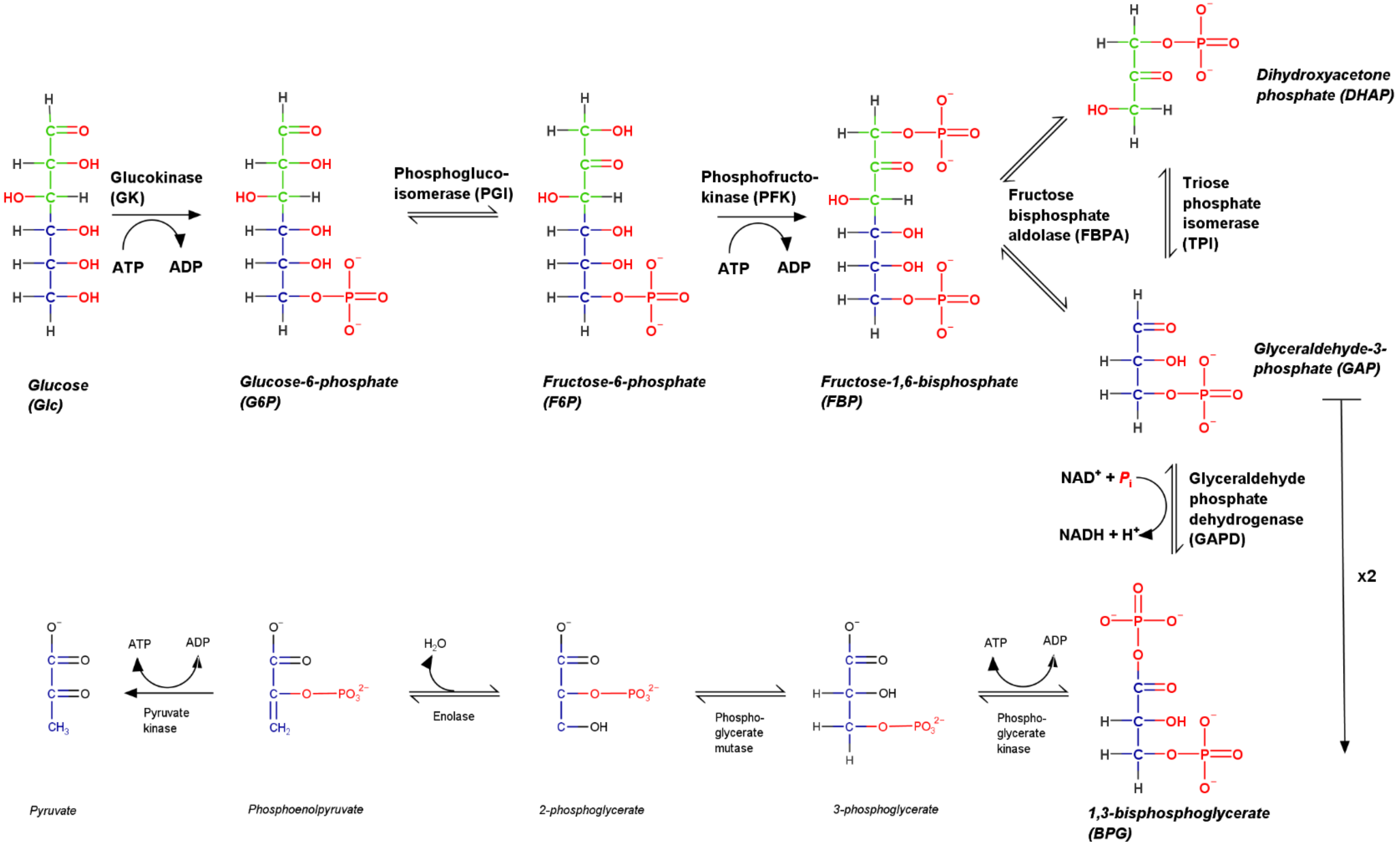


Figure 4.1: The glycolytic pathway from E. coli. The six steps studied here are in bold. The pathway involves the preparative phase of glycolysis, where ATP is consumed to break down six carbon sugars to three carbon units. The first steps involve the production of FBP through the phosphorylation of glucose (Glc; catalysed by GK), isomerisation to F6P (PGI) and phosphorylation at C_1 of the fructose unit (PFK). FBP is broken down to two three carbon units (DHAP and GAP) via the action of FBPA. GAPD passes GAP through the glycolytic pathway through oxidative phosphorylation to form BPG. The DHAP produced in the branch point of the pathway is fed back into the main chain through the action of TPI, isomerising DHAP to GAP. The final step in the studied cascade, GAPD, can be measured by the production of NADH at 340 nm.

4.2 Cloning, enzyme expression and purification

4.2.1 Optimisation of protein expression

Glycolytic enzymes were cloned from *E. coli* BL21 DE3 Gold (Appendix B.1) and ligated into pPROEX vector for expression in DH5α cells (Section 2.3.1). Standard cloning procedures (Section 2.3.1.5) were successful for PFK, FBPA and TPI. For GK, PGI, and GAPD, however, this method resulted in nonsynonymous point mutations in each gene over two cloning attempts.

Thus, the cloning for PGI and GAPD genes was carried out with the addition of two g.L^{-1} glucose in all plate and liquid media to down regulate the background expression from the TRC transcription start site in the pPROEX vector prior to induction (Eswari Pandaranayaka *et al.*, 2008). This method was successful for the GAPD gene. The final expression system for PGI was also gained via this method, but contains a single amino acid mutation (F433L). Several cloning approaches and attempts were unsuccessful at cloning the PGI WT sequence. Final use of a select single amino acid variant of PGI was rationalised since the same enzyme variant was to be used in both the single enzyme characterisation and the pathway characterisation, mutations were not in the proximity of the active site or the interior of the protein (Figure 4.2), and the F433L variant was highly active, with k_{cat} values approaching 2000 s^{-1} (Figure 4.11). The WT GK clone was successfully isolated after sequencing a number of clones from standard cloning processes (Section 2.3.1).

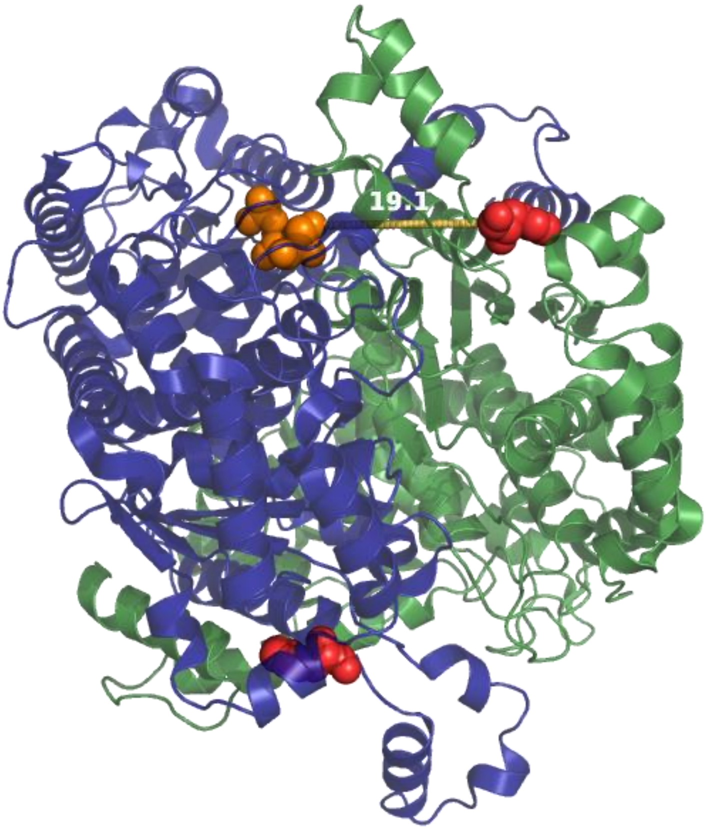


Figure 4.2: Position of the PGI mutation site (F433L) in the dimeric enzyme. Mutation positions are shown as red spheres in the E. coli PGI enzyme [PDB 3NBU (Totir et al., 2012)]. In the blue monomer, the substrate F6P is shown in the active site as orange spheres. F6P placement was achieved by overlay with the Francisella tularensis PGI structure with substrate bound (PDB 3M5P). The closest interaction between mutation position and substrate occurs between the two units of the dimer, at 19.1 Å.

Enzyme expression was tested under standard expression conditions (100 ml LB media, one mM IPTG induction at OD_{600} = 0.5-0.7, 37 °C expression overnight with shaking). Gel electrophoresis of sonication lysed cells pellet and supernatant fractions confirmed over expression of soluble protein for all six enzymes in pH 7.4 buffer (Figure 4.3).

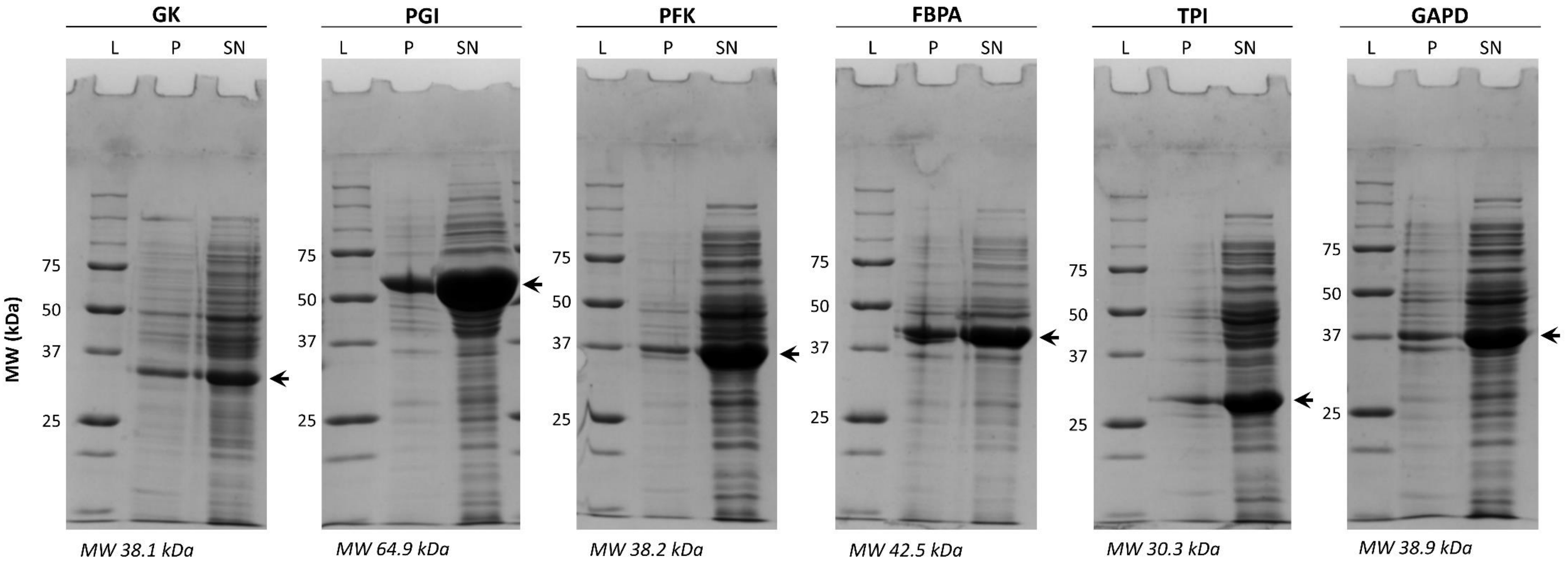


Figure 4.3: Expression trial of glycolytic enzymes. L = ladder; P = insoluble pellet; SN = soluble supernatant. Expression was carried out in 100 ml LB media overnight at 37 °C after IPTG induction. Cells were lysed by sonication in lysis buffer (50 mM sodium phosphate, pH 7.4, 50 mM NaCl) and spun down to pellet insoluble components prior to separation by SDS-PAGE. All proteins show overexpression in the soluble fraction (as indicated with arrows); molecular weights (MW) are listed below individual gels.

Final expression protocols (if different from the standard ON at 37 °C in LB media) for each enzyme (Section 2.3.2.2) were selected based on experience from the cloning procedure. Enzymes for which successful cloning was achieved with the addition of two g.L^{-1} glucose to media were also grown and expressed with LB glucose media to limit background expression from the vector prior to IPTG induction. These cultures were harvested four hours after induction to limit the time over which overexpression of the protein could select for mutations in the gene. Starter cultures for PGI and GAPD were increased to 40 ml to limit glucose exhaustion over the overnight growth period. GK was also harvested after a four hour induction period due to the high occurrence of mutations arising during the cloning procedure. Plasmids were extracted and sequenced to confirm no mutations were occurring over the expression protocol.

4.2.2 Enzyme purifications

All six glycolytic enzymes were purified from the soluble fraction by the same two step IMAC and gel filtration protocol (Section 2.1.1; Figure 4.4-Figure 4.9) in 50 mM HEPES, pH 7.4, 150 mM NaCl based buffers (Section 2.3.2.3 for full details)[1]. Protein presence and purity was confirmed during the purification by SDS-PAGE (Section 2.1.2.1; Figure 4.4-Figure 4.9).

The oligomeric state of each enzyme was calculated from estimated molecular weight from SEC (Section 2.1.1.3.1). A S200 10/300 analytical size exclusion column (GE Healthcare Life Science, UK) was previously calibrated with eight proteins standards from high and low molecular weight gel filtration calibration kits (GE Healthcare Life Science, UK) as per the kit instructions. The void volume (V_0) of the column was determined with 1 mg.ml^{-1} Blue Dextran 2000 dye. Based on elution volume (V_e) and column volume (V_c), K_{av} values (gel phase distribution coefficient; equation **11**) were calculated to construct a calibration curve. The calibration curve is given in Appendix D.4.

$$K_{av} = \frac{V_e - V_0}{V_c - V_0} \text{..................................... } \textit{gel phase distribution coefficient} \ \mathbf{(11)}$$

---

[1] Pertaining to gel filtration results, small peaks to the left of the primary peak are due to an imperfection in the gel bed rather than a property of the protein. Gels include this peak to show presence of target protein in the minor peak.

Elution volumes of each glycolytic enzyme were used to calculate molecular weight based on the calibration curve. This information, along with the oligomeric state, is given for each enzyme in Figure 4.4 - Figure 4.9.

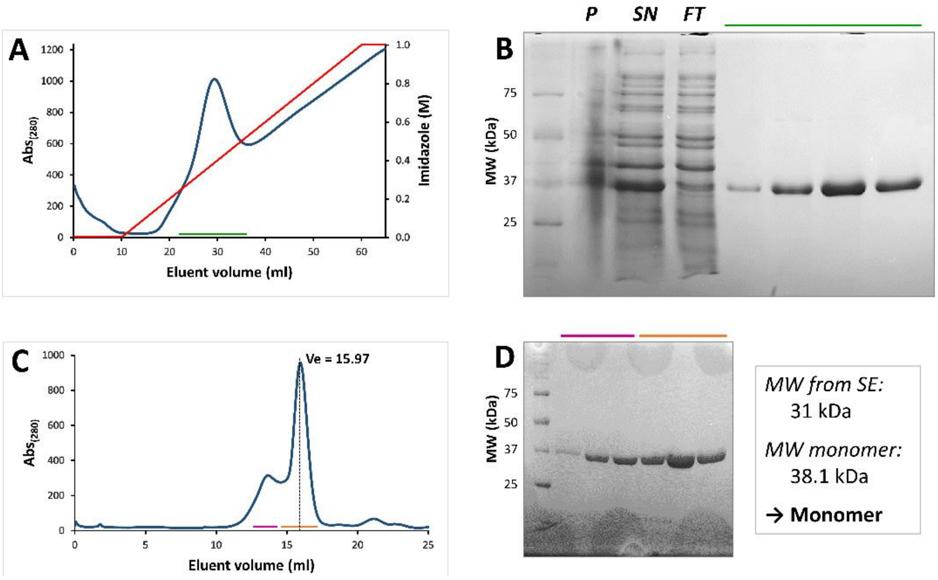


Figure 4.4: Glucokinase purification. ***(A)*** *IMAC purification of GK from zero to one M imidazole. Analysis of indicated fractions for the protocol are given in* ***(B)*** *SDS-PAGE of the IMAC purification process. P = insoluble pellet; SN = soluble protein in the supernatant; FT = unbound column flow through. Remaining samples correspond to the indicated section in* ***(A)****.* ***(C)*** *SEC trace of GK purification. Analysis of the indicated fractions for the protocol are as indicated in* ***(D)*** *SDS-PAGE of the SEC trace fractions. Molecular weight, as calculated from the SEC trace (Section 2.1.1.3.1) is also given.*

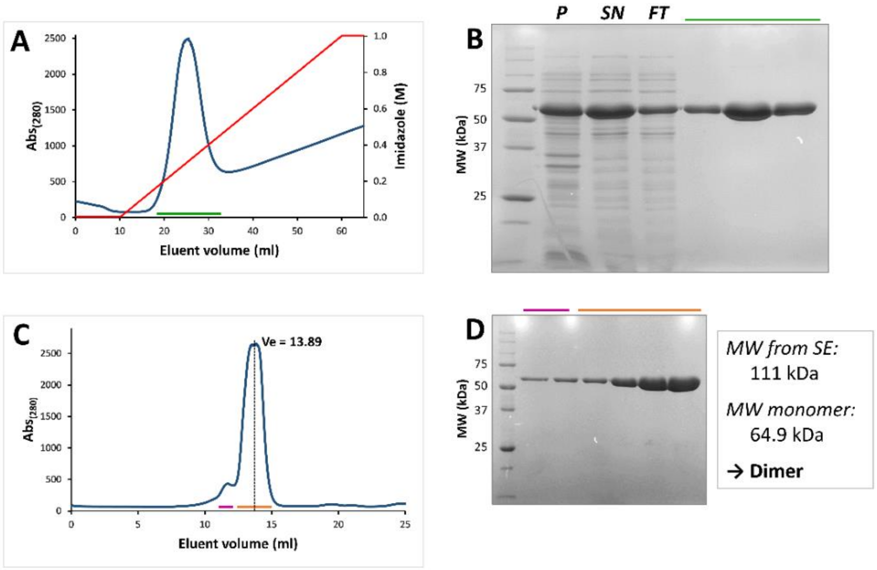


Figure 4.5: Phosphoglucose isomerase purification. ***(A)*** *IMAC purification of PGI from zero to one M imidazole. Analysis of indicated fractions for the protocol are given in* ***(B)*** *SDS-PAGE of the IMAC purification process. P = insoluble pellet; SN = soluble protein in the supernatant; FT = unbound column flow through. Remaining samples correspond to the indicated section in* ***(A)****.* ***(C)*** *SEC trace of PGI purification. Analysis of the indicated fractions for the protocol are as indicated in* ***(D)*** *SDS-PAGE of the SEC trace fractions. Molecular weight, as calculated from the SEC trace (Section 2.1.1.3.1) is also given.*

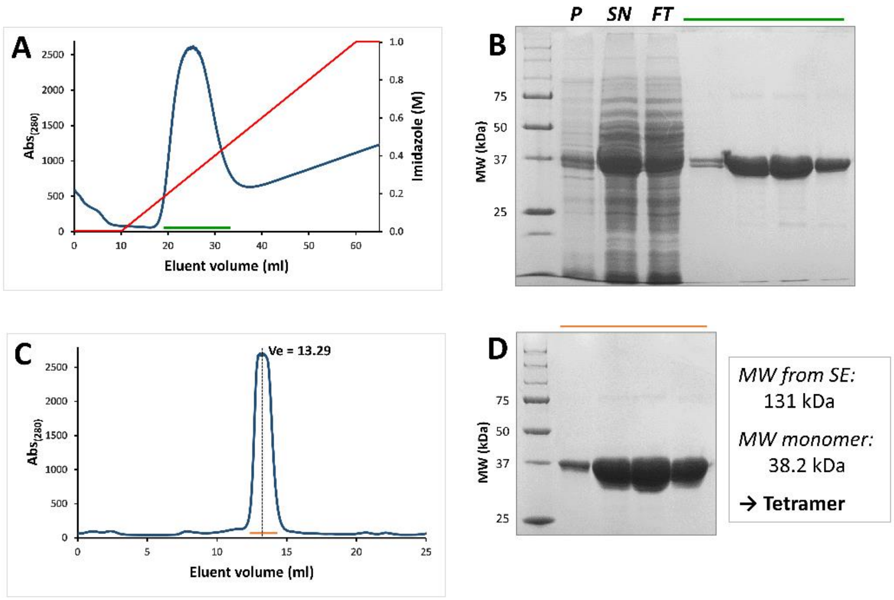


Figure 4.6: Phosphofructokinase purification. ***(A)*** *IMAC purification of PFK from zero to one M imidazole. Analysis of indicated fractions for the protocol are given in* ***(B)*** *SDS-PAGE of the IMAC purification process. P = insoluble pellet; SN = soluble protein in the supernatant; FT = unbound column flow through. Remaining samples correspond to the indicated section in* ***(A)****.* ***(C)*** *SEC trace of PFK purification. Analysis of the indicated fractions for the protocol are as indicated in* ***(D)*** *SDS-PAGE of the SEC trace fractions. Molecular weight, as calculated from the SEC trace (Section 2.1.1.3.1) was ambiguous between a trimer and tetramer, but was later confirmed as a tetramer by a Hill coefficient approaching four (Figure 4.11). A faint double banding is evident in one lane; this was not investigated further.*

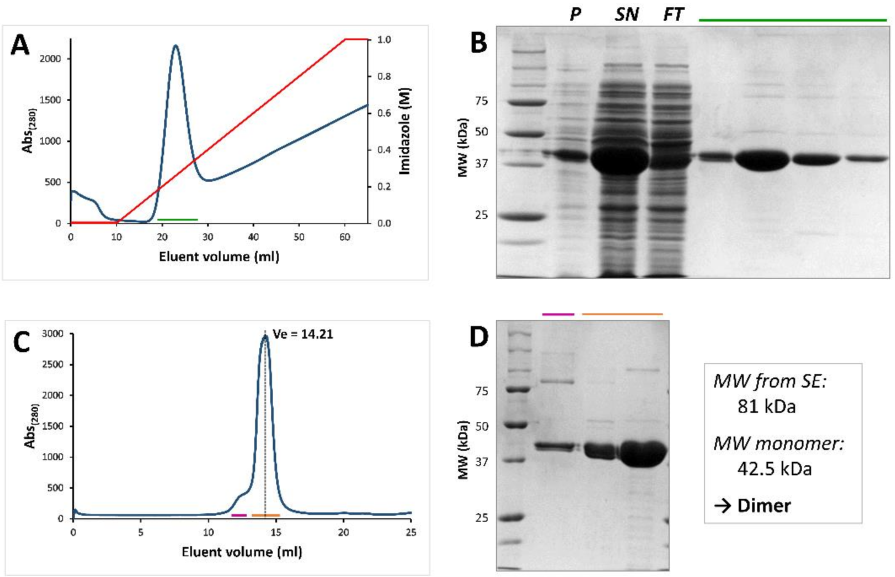


Figure 4.7: Fructose bisphosphate aldolase purification. ***(A)*** *IMAC purification of FBPA from zero to one M imidazole. Analysis of indicated fractions for the protocol are given in* ***(B)*** *SDS-PAGE of the IMAC purification process. P = insoluble pellet; SN = soluble protein in the supernatant; FT = unbound column flow through. Remaining samples correspond to the indicated section in* ***(A)****.* ***(C)*** *SEC trace of FBPA purification. Analysis of the indicated fractions for the protocol are as indicated in* ***(D)*** *SDS-PAGE of the SEC trace fractions. Molecular weight, as calculated from the SEC trace (Section 2.1.1.3.1) is also given.*

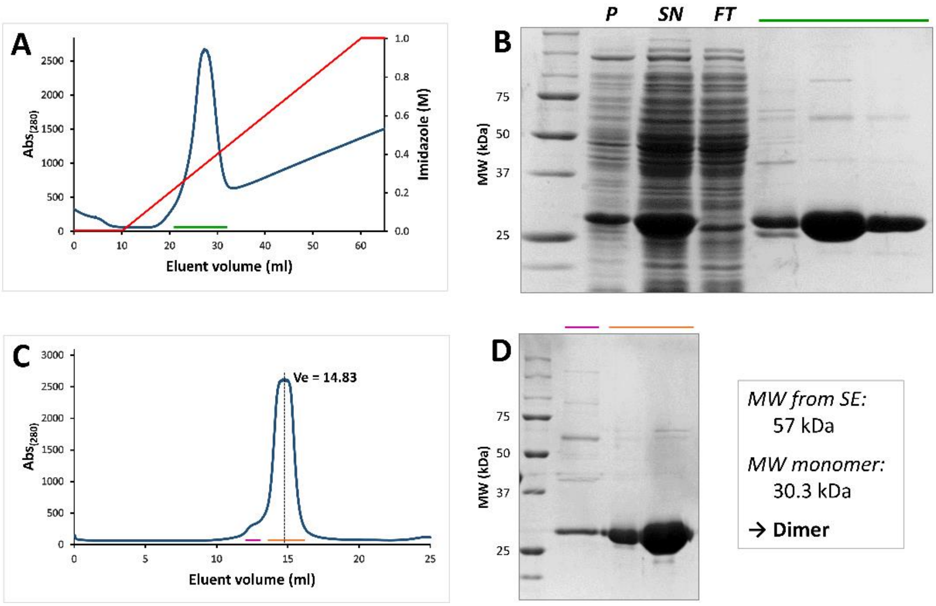


Figure 4.8: Triosephosphate isomerase purification. ***(A)*** *IMAC purification of TPI from zero to one M imidazole. Analysis of indicated fractions for the protocol are given in* ***(B)*** *SDS-PAGE of the IMAC purification process. P = insoluble pellet; SN = soluble protein in the supernatant; FT = unbound column flow through. Remaining samples correspond to the indicated section in* ***(A)****.* ***(C)*** *SEC trace of TPI purification. Analysis of the indicated fractions for the protocol are as indicated in* ***(D)*** *SDS-PAGE of the SEC trace fractions. Molecular weight, as calculated from the SEC trace (Section 2.1.1.3.1) is also given.*

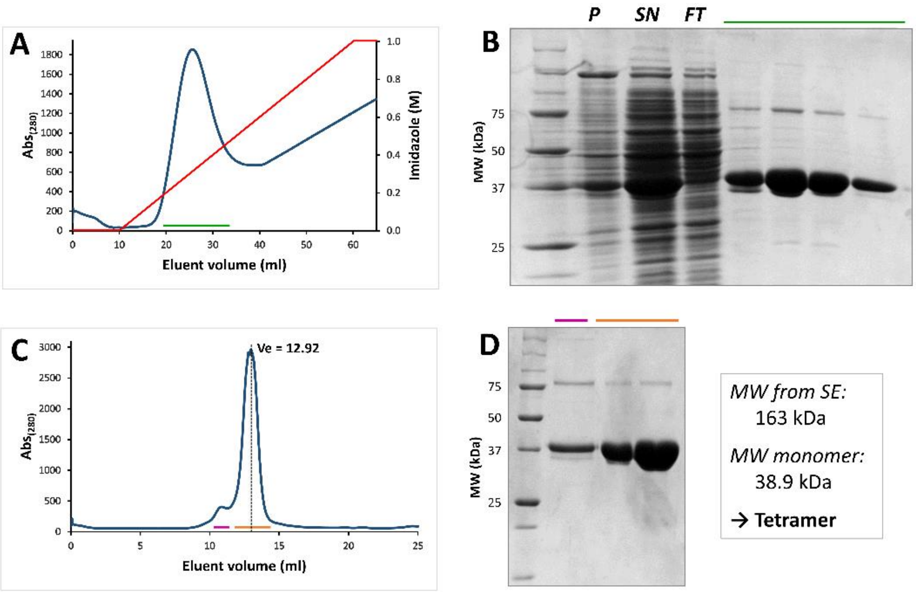


Figure 4.9: Glyceraldehyde-3-phosphate dehydrogenase purification. ***(A)*** *IMAC purification of GAPD from zero to one M imidazole. Analysis of indicated fractions for the protocol are given in* ***(B)*** *SDS-PAGE of the IMAC purification process. P = insoluble pellet; SN = soluble protein in the supernatant; FT = unbound column flow through. Remaining samples correspond to the indicated section in* ***(A)****.* ***(C)*** *SEC trace of GAPD purification. Analysis of the indicated fractions for the protocol are as indicated in* ***(D)*** *SDS-PAGE of the SEC trace fractions. Molecular weight, as calculated from the SEC trace (Section 2.1.1.3.1) is also given. A faint double banding is evident in some lanes; this was not investigated further.*

4.3 Glycolytic pathway characterisation

Enzymatic assays (Section 2.3.3) were developed for each enzyme based on published protocols. HEPES [4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid] was used for the purification and assay buffers for all enzymes and the full pathway due to the relatively low pH changes with temperature associated with this buffer (Baicu & Taylor, 2002). Phosphate buffer, the optimal system for maintaining pH with temperature, was not selected due to possible interference with the glycolytic pathway steps involving phosphate transfers, and the potential to bind to divalent metals such as Mg^{2+} (Van Wazer & Callis, 1958). This setup allowed inorganic phosphate (P_i) to be added independently and varied for Michaelis Menten characterisation of GAPD (Section 4.3.1.1). Assays were run at pH 7.4, the proposed *E. coli* intracellular pH (Wilks & Slonczewski, 2007). Additional salts were added to complex ATP to form the active ATP-Mg complex (10-15 mM $MgCl_2$), and to aid the activity of PFK [2.5 mM $(NH_4)_2SO_4$] (Kamemoto & Mansour, 1986).

4.3.1 Characterisation of individual enzymes

4.3.1.1 *Michaelis Menten characterisations*

Michaelis Menten plots (Section 2.3.3) at 37 °C and 44-45 °C were performed to determine K_M and k_{cat} values for the substrate/s over a relevant temperature range for all steps of the pathway (Figure 4.10 through Figure 4.13). Where more than one substrate is required, assays were performed with saturating concentrations of other substrates so that these did not become rate limiting. Where substrate inhibition was observed, substrates were used at the concentration which gave greatest turn over. Values gained are commensurate with values in the literature under similar conditions (Table 4.1). Specific details of all component concentrations in the Michaelis Menten assays can be found in Appendix D.2.

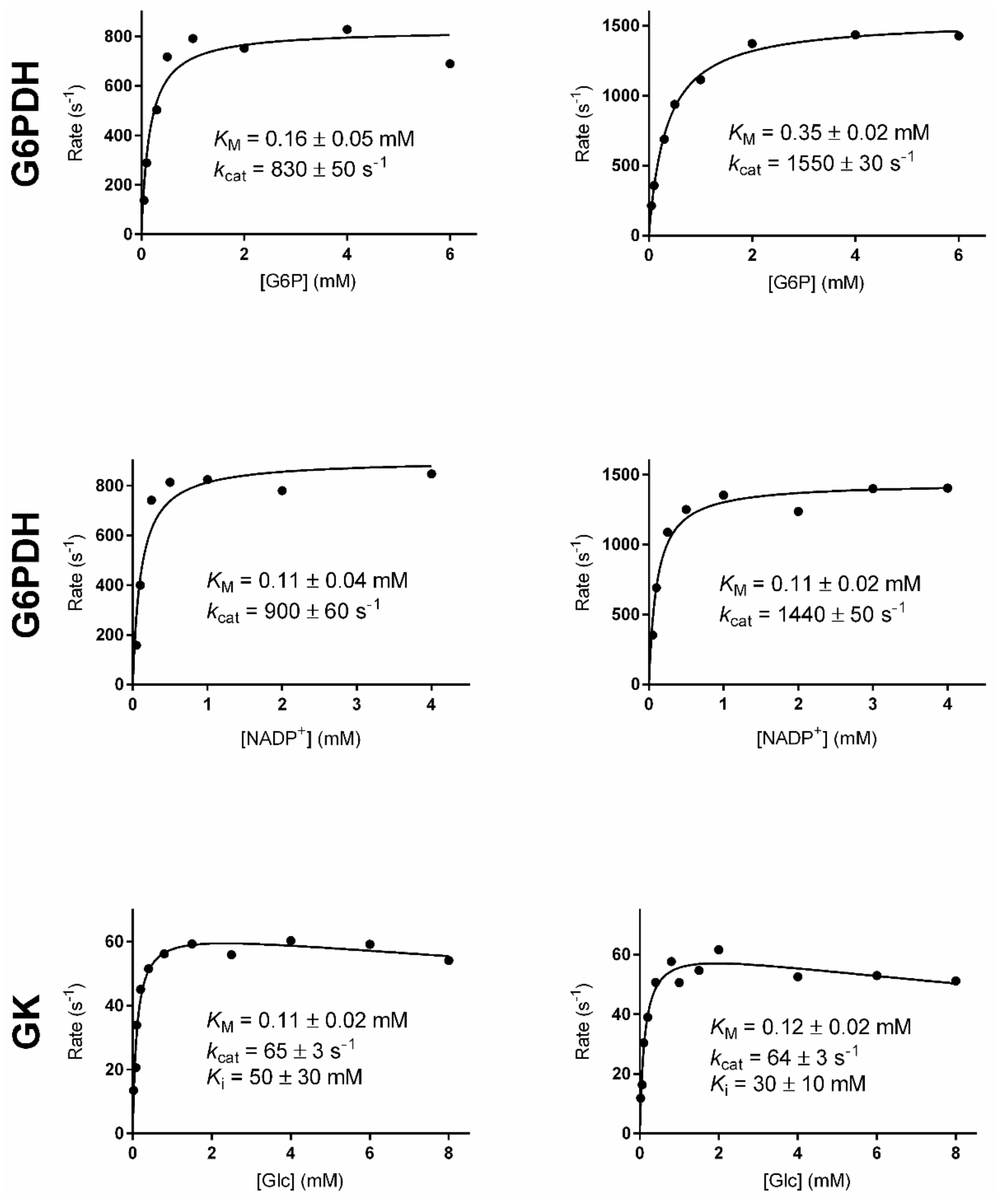


Figure 4.10: Michaelis Menten graphs for the indicator enzyme G6PDH, and GK (for glucose). Plots at 37 °C are given on the left, and 44 °C on the right. The substrate requirements of G6PDH were characterised to determine the amount of NADP$^+$ to add for subsequent GK and PGI assays where the enzyme was used to detect the production of G6P via the coupled reduction of NADP$^+$ at 340 nm.

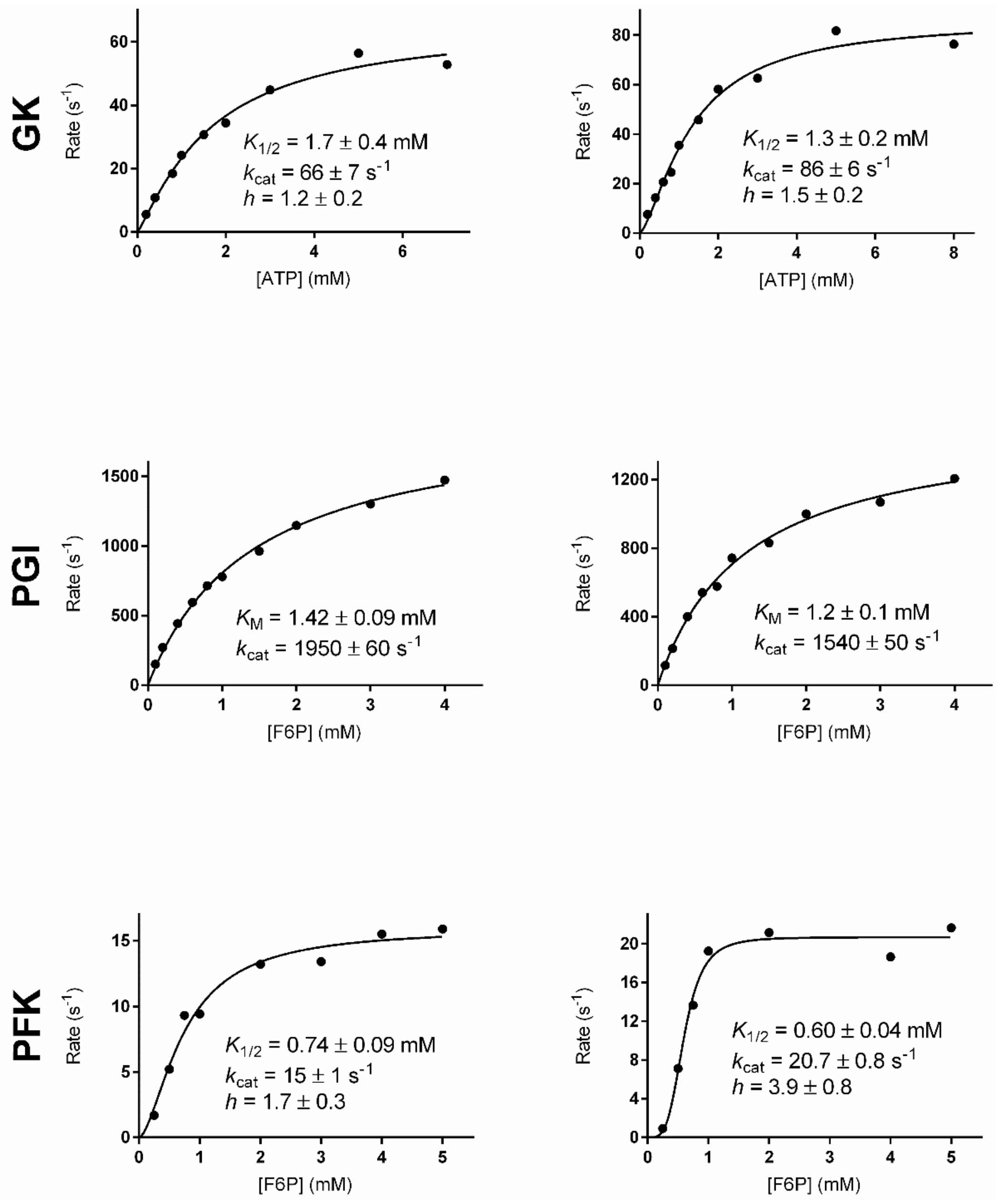


Figure 4.11: Michaelis Menten kinetics for GK (for ATP), PGI and PFK (for F6P). Plots at 37 °C are given on the left, and 44 °C on the right, except for PGI which was characterised at 45 °C (right).

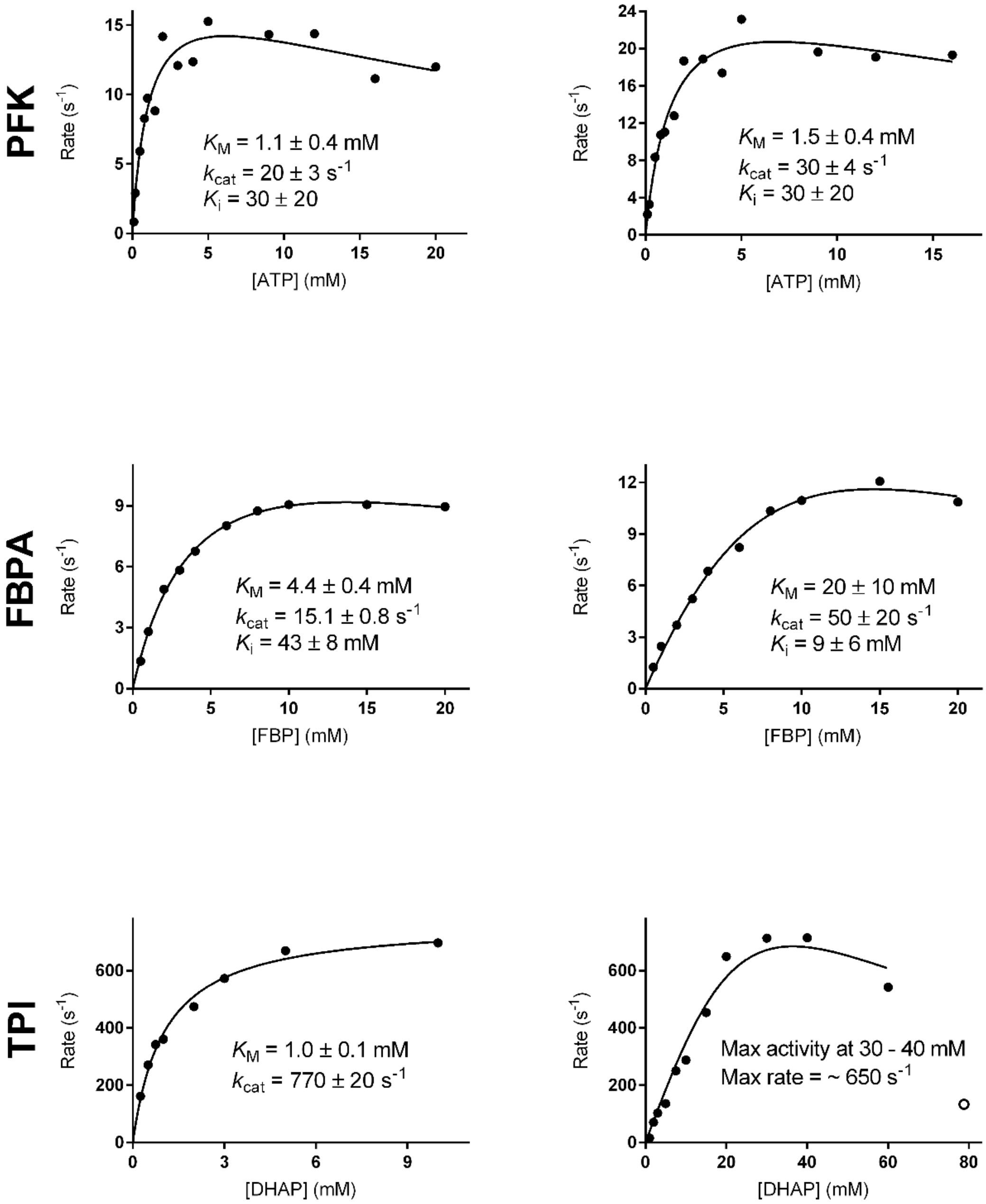


Figure 4.12: Michaelis Menten kinetics for PFK (for ATP), FBPA and TPI. Plots at 37 °C are given on the left, and 44 °C on the right. TPI at 44 °C showed strong substrate inhibition, but did not fit an inhibition model well at high concentration (open circle), so data were only fit up to 60 mM.

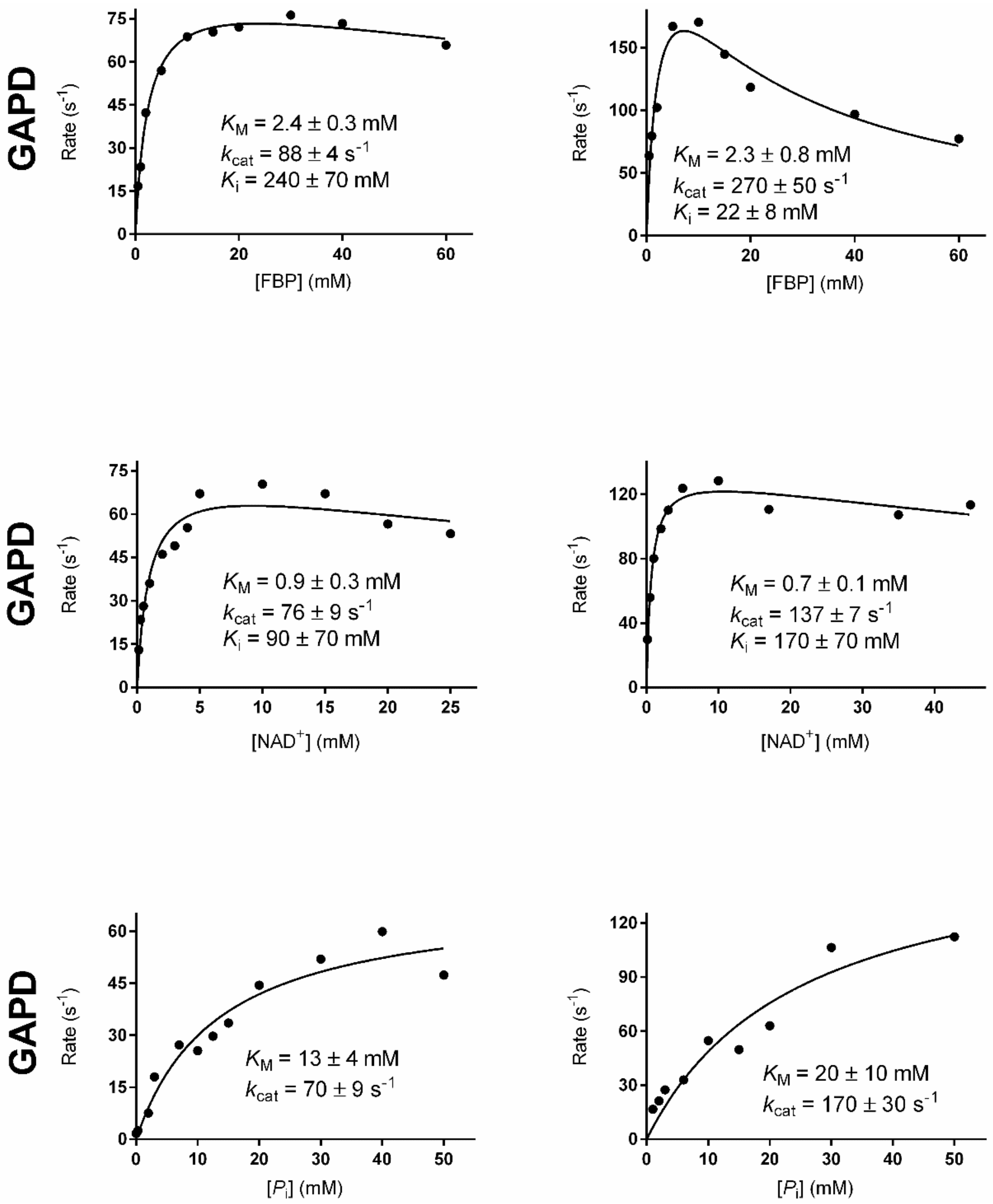
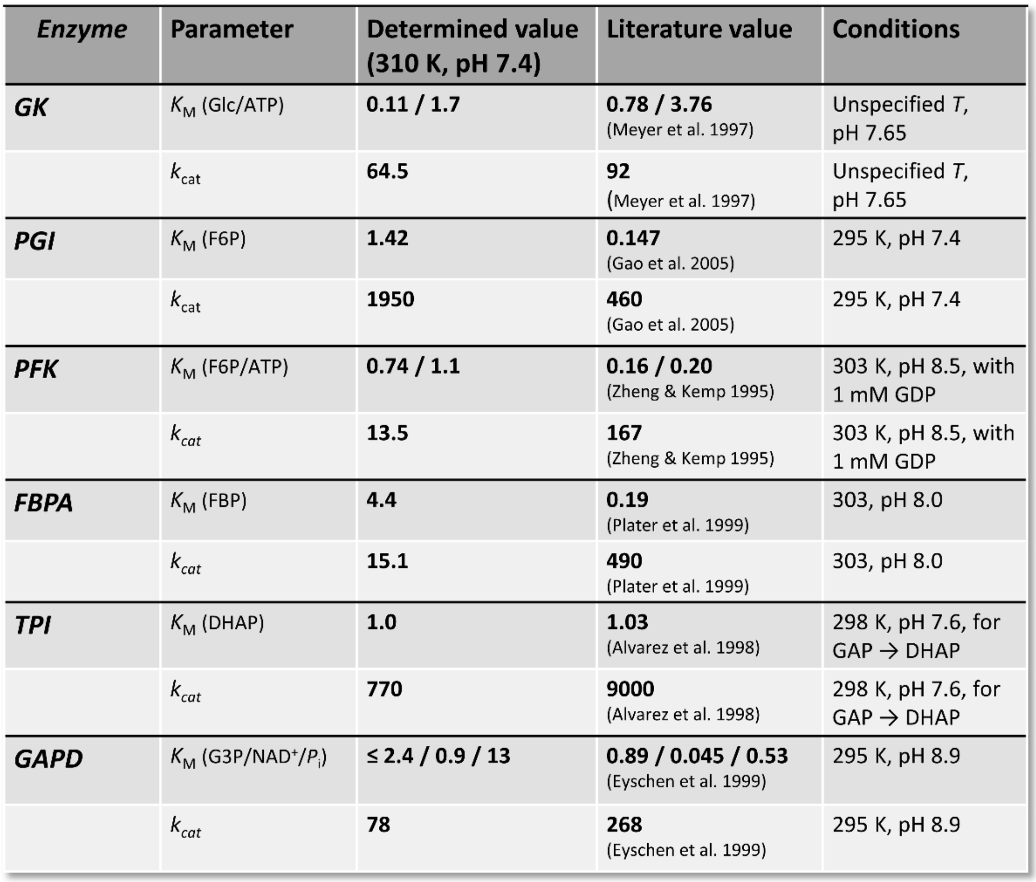


Figure 4.13: Michaelis Menten kinetics for the three substrates of GAPD. Plots at 37 °C are given on the left, and 44 °C on the right.

Table 4.1: Comparison of glycolytic enzyme kinetic parameters to literature examples from E. coli. *K_M values are given in mM, k_{cat} values in s^{-1}. Values are all in agreement within expected ranges given the differences in experimental conditions.*



Enzyme	**Parameter**	**Determined value (310 K, pH 7.4)**	**Literature value**	**Conditions**
GK	K_M (Glc/ATP)	**0.11 / 1.7**	**0.78 / 3.76** (Meyer et al. 1997)	Unspecified *T*, pH 7.65
	k_{cat}	**64.5**	**92** (Meyer et al. 1997)	Unspecified *T*, pH 7.65
PGI	K_M (F6P)	**1.42**	**0.147** (Gao et al. 2005)	295 K, pH 7.4
	k_{cat}	**1950**	**460** (Gao et al. 2005)	295 K, pH 7.4
PFK	K_M (F6P/ATP)	**0.74 / 1.1**	**0.16 / 0.20** (Zheng & Kemp 1995)	303 K, pH 8.5, with 1 mM GDP
	k_{cat}	**13.5**	**167** (Zheng & Kemp 1995)	303 K, pH 8.5, with 1 mM GDP
FBPA	K_M (FBP)	**4.4**	**0.19** (Plater et al. 1999)	303, pH 8.0
	k_{cat}	**15.1**	**490** (Plater et al. 1999)	303, pH 8.0
TPI	K_M (DHAP)	**1.0**	**1.03** (Alvarez et al. 1998)	298 K, pH 7.6, for GAP → DHAP
	k_{cat}	**770**	**9000** (Alvarez et al. 1998)	298 K, pH 7.6, for GAP → DHAP
GAPD	K_M (G3P/NAD$^+$/P_i)	**≤ 2.4 / 0.9 / 13**	**0.89 / 0.045 / 0.53** (Eyschen et al. 1999)	295 K, pH 8.9
	k_{cat}	**78**	**268** (Eyschen et al. 1999)	295 K, pH 8.9

Glucose-6-phosphate dehydrogenase

G6PDH (EC 1.1.1.49, Sigma-Aldrich, USA) was characterised as the indicator enzyme for the characterisation of both GK and PGI (Figure 4.10). G6PDH was used to oxidise G6P to 6-phospho-glucono-1,5-lactone, coupled to the reduction of NADP$^+$ to NADPH, the production of which was followed continuously at 340 nm. From these results, NADP$^+$ was added at a final concentration of one mM (nine times K_M of 0.11 mM at 37 and 44 °C) to subsequent assays. The k_{cat} values were also used to ensure G6PDH was added to reactions at a concentration high enough to ensure that the step of interest (GK or PGI) was rate limiting. Rough temperature profiles to ensure G6PDH was active at high temperatures were also performed (Appendix D.3).

Glucokinase

For GK, slight substrate inhibition was observed, occurring at substrate concentrations above two mM glucose. A minor deviation from Michaelis Menten kinetics was observed in the ATP characterisation, with data best fitting an allosteric sigmoidal model. The model suggests positive cooperativity of successive substrate binding events [a positive Hill coefficient (h)]. The deviations from Michaelis Menten kinetics are small, with the Hill coefficient only being significantly different from one at 44 °C.

Also relevant to experimental set up was the behaviour observed at concentrations above eight mM ATP for GK assays, where rapid decreases in rate occurred. For GK, the ATP-Mg complex is the active substrate, while ATP alone is inhibitory (Fornaini *et al.*, 1982; Magnani *et al.*, 1983; Cesar Mde *et al.*, 1997). Due to a concentration of ten mM $MgCl_2$ in the assay buffer, concentrations of ATP above eight mM result in enzymatic inhibition from free ATP (not shown). Due to this, in subsequent cases where ATP was included, Mg^{2+} was used at concentrations higher than that of ATP to force the equilibrium towards complex formation. Note a high ATP concentration (8 mM) and associated inhibition is the suspected reason that the k_{cat} for glucose at 44 °C is lower than expected compared to the equivalent value at 37 °C (6 mM ATP), or 44 °C ATP Michaelis Menten curve.

Phosphoglucose isomerase

To characterise PGI, the reaction was undertaken in reverse compared to the direction of the reaction during glycolysis, converting F6P to G6P. This was done so that the reaction could be followed in a continuous fashion via the production of G6P by coupling the reaction to G6PDH and NADPH (as described for GK assays; Section 2.3.3). Given the reaction is an isomerisation exchanging the alcohol and carbonyl at positions C_1 and C_2, there is a similar reaction difficulty in either direction, resulting in an expected similar curvature with temperature ($\Delta C^{\ddagger}_{p}$). At both temperatures, PGI followed Michaelis Menten kinetics, with a K_M between 1.2 and 1.4 mM F6P. It is acknowledged that the k_{cat} values are not commensurate with the temperature trend in the full temperature characterisation (Figure 4.14); a conclusive reason for this is not evident and requires further experimental work to clarify. K_M values between the two temperatures are in agreement, and the critical factor to the setup of temperature characterisations.

Phosphofructokinase

PFK was found to be inhibited by ATP at concentrations above five mM. Intercellular concentrations of ATP in non-starved cells sit between nine and 10 mM (Bennett *et al.*, 2009), resulting in slight substrate inhibition based on data presented here. Plots of F6P showed strong allosteric sigmoidal behaviour, especially at 44 °C with a Hill coefficient of 3.9 ± 0.8, demonstrating a strong positive cooperativity between substrate binding events.

Fructose bisphosphate aldose

In the characterisation of FBPA, highly different K_M values for the singular substrate (FBP) were observed between the two temperatures at which characterisations were undertaken. At 37 °C, a relatively low K_M concentration of 4.4 ± 0.4 mM was characterised, compared to 20 ± 10 at 44 °C. Full characterisations of K_M with temperature have shown rapid increases in K_M at certain temperatures after relative stability over lower temperature ranges (Thomas & Scopes, 1998); it appears this transition period occurs around the 37-44 °C range for FBPA. Substrate inhibition was present at both temperatures, however, there are large differences in the K_i values (43 ± 8 and 9 ± 6 mM at 37 and 44 °C respectively).

Triosephosphate isomerase

TPI showed different behaviour at 37 °C and 44 °C. At 37 °C classical Michaelis Menten behaviour was observed for DHAP, with a K_M of 1.0 ± 0.1 mM, and k_{cat} of 770 ± 20 s^{-1}. At 44 °C, rates at low concentration increased slowly with increased concentration, with a steep drop off in rates occurred above 40 mM DHAP. Although showing clear substrate inhibition, the highest substrate concentration did not fit a substrate inhibition model well. Thus in Figure 4.12, this point is shown (open circle); however, the substrate inhibition model is not fit to this data point. K_M and k_{cat} values from the illustrated fit were high due to the moderate gradient of increased rates at low concentration causing poor K_M extrapolation within the inhibition model. Therefore, rates and substrate concentrations are quoted for the point where maximum activity was observed.

Glyceraldehyde phosphate dehydrogenase

Of the three substrates of GAPD, only P_i followed Michaelis Menten kinetics. Possible substrate inhibition is evident at the highest P_i concentration at 37 °C, however this is not conclusive enough to determine. Substrate inhibition is evident for NAD^+. Fastest rates at both temperatures were gained at ten mM NAD^+. In both cases, this maximal activity is present at greater than ten times the K_M concentration (0.9 and 0.7 mM respectively at 37 and 44 °C).

The characterisation of FBP also indicated substrate inhibition. For this assay, FBP in combination with FBPA was used to make the substrate, GAP, prior to the assay (see Section 2.3.3.6 for methods). Assuming the reaction does not go to completion, both FPB and GAP are present in the reaction. Thus, the interpretation of the substrate inhibition in this case is ambiguous, as either high concentrations of FBP or GAP could be the inhibitory agent. Previously GAP inhibition has been reported in GAPD from rabbit muscle, (Orsi & Cleland, 1972), although K_i values in this case are smaller than those reported here (2.4 mM compared to 90-170 mM). Given it is unlikely that the substrate production reaction goes to completion (concentration FBP ≠ concentration formed GAP), the K_M for substrate (GAP) is likely to be lower than the quoted K_M for the concentration of FBP added to produce K_M concentrations of substrate. However, this K_M value was sufficient to determine the conditions for characterising the temperature dependence of GAPD.

4.3.1.2 $\Delta C^{\ddagger}_{p}$ determinations for single enzyme steps

The temperature behaviour of enzymes was characterised by the same methods as described in Section 4.3; substrate concentrations were used based on Michaelis Menten results (Section 4.3.1.1) to maintain saturating concentrations over the assays. Full assay composition details can be found in Appendix D.2.

In the case of PGI, PFK and GAPD, data fit MMRT with a temperature independent $\Delta C^{\ddagger}_{p}$ (equation **7**). GK, FBPA and TPI showed improved fit with a temperature dependent $\Delta C^{\ddagger}_{p}$ to fully explain the curvature in the data (Figure 4.14). Data fit for these enzymes are also shown fitted to the temperature independent $\Delta C^{\ddagger}_{\mathrm{p}}$ model, as this is carried through in later sections for simplicity of modelling based on the experimental data; data are fit over a narrower temperature range to approximate a temperature independent $\Delta C^{\ddagger}_{\mathrm{p}}$ where necessary in this data. These data, especially that of FBPA, show further evidence for the necessity of a temperature dependent $\Delta C^{\ddagger}_{\mathrm{p}}$ in describing data over a wide temperature range from that presented in Kraakman (2017).

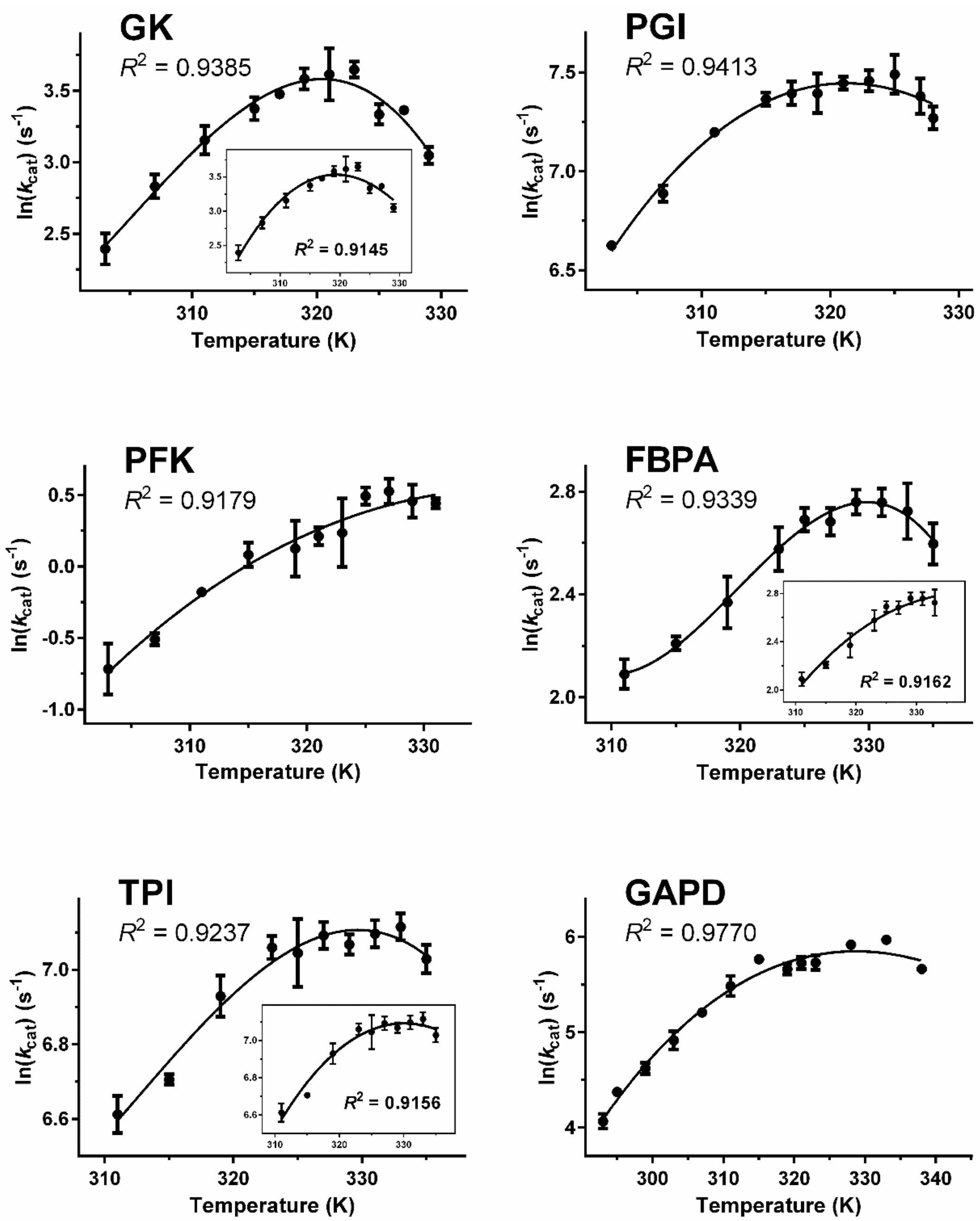
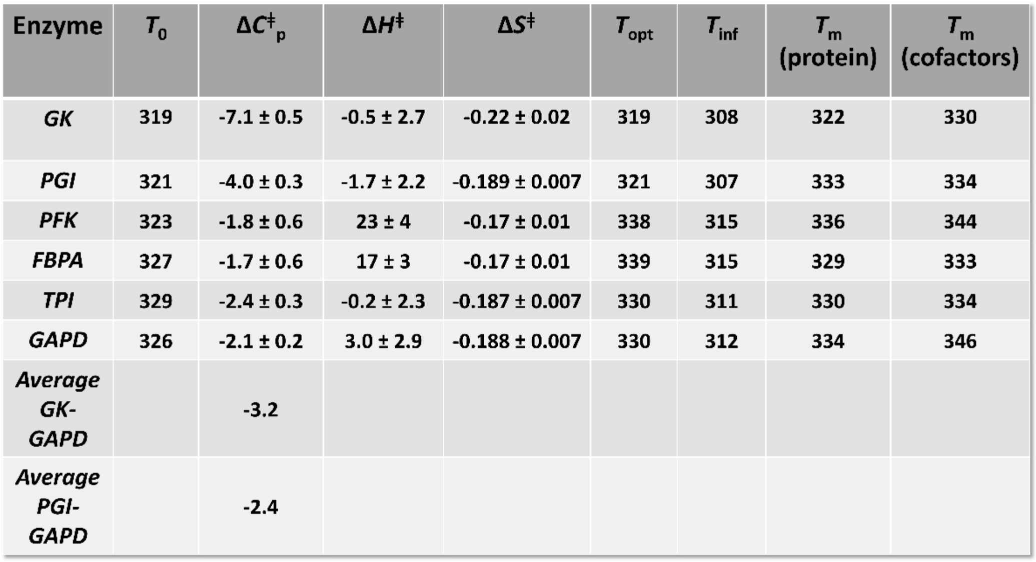


Figure 4.14: Temperature characterisation of the six glycolysis enzymes. Data are fit to the MMRT equation in either the standard form (equation 7; PGI, PFK and GAPD), or with the introduction of a temperature dependent $\Delta C^{\ddagger}_{p}$ (Section 1.4.4; GK, FBPA, TPI). Where data are better fit with the temperature dependent equation (non-zero slope outside of error values), the temperature independent fit used for simplicity in modelling is also shown. Error bars where visible are the SD of three replicates. The one exception to this is GAPD, where data points for the three highest points are taken as singletons. A full description of fit parameters including $\Delta C^{\ddagger}_{p}$ values is given in Table 4.2.

Table 4.2: MMRT parameters for individual enzymes as determined by fits in Figure 4.14 to the temperature independent MMRT equation ***(7)****. Values are given in kJ.mol*$^{-1}$*.K*$^{-1}$ *(*$\Delta C^{\ddagger}_{p}$*), and kJ.mol*$^{-1}$ *(*$\Delta H^{\ddagger}$ *and* $\Delta S^{\ddagger}$*) ± SE.* T_{opt} *and* T_{inf} *(K) values are calculated from equations* ***(14)*** *and* ***(15)*** *respectively (Section 2.3.4) based on the unrounded values.* T_m *(K) values are calculated from Figure 4.16 data from differential scanning fluorimetry.* T_0*, consistent with previous work (Hobbs et al., 2013; Schipper et al., 2014), is taken as 4 degrees below the temperature of the highest experimental rate and given as the reference temperature for* $\Delta H^{\ddagger}$ *and* $\Delta S^{\ddagger}$ *values.*



Enzyme	T_0	$\Delta C^{\ddagger}_{p}$	$\Delta H^{\ddagger}$	$\Delta S^{\ddagger}$	T_{opt}	T_{inf}	T_m (protein)	T_m (cofactors)
GK	**319**	**-7.1 ± 0.5**	**-0.5 ± 2.7**	**-0.22 ± 0.02**	**319**	**308**	**322**	**330**
PGI	**321**	**-4.0 ± 0.3**	**-1.7 ± 2.2**	**-0.189 ± 0.007**	**321**	**307**	**333**	**334**
PFK	**323**	**-1.8 ± 0.6**	**23 ± 4**	**-0.17 ± 0.01**	**338**	**315**	**336**	**344**
FBPA	**327**	**-1.7 ± 0.6**	**17 ± 3**	**-0.17 ± 0.01**	**339**	**315**	**329**	**333**
TPI	**329**	**-2.4 ± 0.3**	**-0.2 ± 2.3**	**-0.187 ± 0.007**	**330**	**311**	**330**	**334**
GAPD	**326**	**-2.1 ± 0.2**	**3.0 ± 2.9**	**-0.188 ± 0.007**	**330**	**312**	**334**	**346**
Average GK-GAPD		**-3.2**						
Average PGI-GAPD		**-2.4**						

4.3.1.2.1 Enzyme adaptation to environmental temperature

It is evident in the literature that T_{opt} values for enzyme catalysis are not well correlated to environmental growth temperature for the host (Peterson *et al.*, 2004), with T_{opt} values consistently higher than environmental temperatures. Lee *et al.* (2007) considered this issue under the equilibrium model (Section 1.3.2), and found stronger correlations between growth temperature and T_{eq} (the temperature mid-point for the transition to the folded-inactive enzyme state) when compared to any other parameter from the equilibrium model (including measures of stability and T_{opt}). Despite this, there is a systematic deviation from an ideal relationship, with psychrophilic enzymes displaying a greater difference between T_{eq} and growth temperature, as noted by the authors.

This disconnect between T_{opt} and growth temperature (310 K for *E. coli*), and between individual enzyme's T_{opt} from the same organism, is evident in the glycolysis data presented here (Table 4.3). T_{opt} values from MMRT fits range from 319 to 339 K (all at least ten degrees above the organism growth temperature). Using MMRT, there is a congruence between the lower inflection point and growth temperature; these values range from 307 to 315 K, encompassing the growth temperature of 310 K (Figure 4.15).

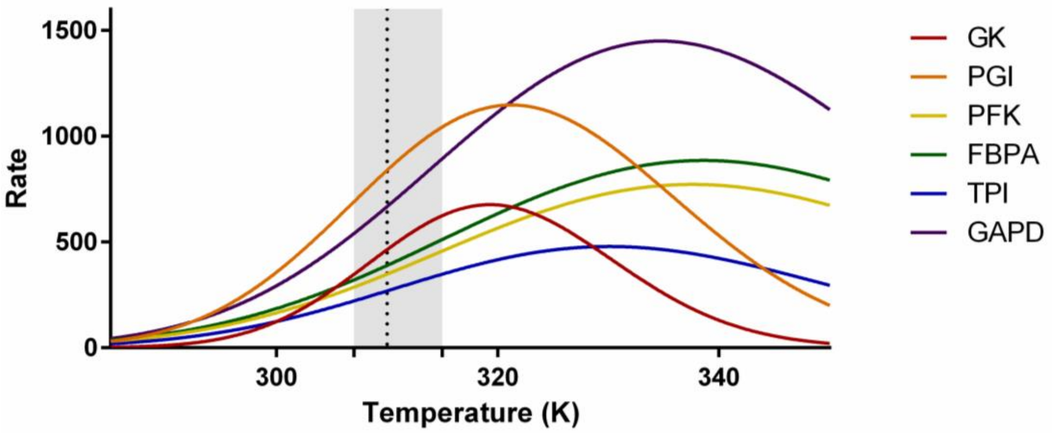


*Figure 4.15: Relative change in rates with temperature of glycolysis enzymes (*E. coli*) modelled by MMRT fit to experimental data (Section 4.3.1.2). Dotted line represents the ideal growth temperature (310 K); grey portion the range of T_{inf} values for the individual enzymes. From around 310 K and below, enzyme rates track relatively equally with changes in temperature; above this point, rates with temperature between different enzymes of the same pathway vary greatly.*

The inflection point (T_{inf}) is an important parameter for MMRT as it defines the point where rate increase with temperature starts to reduce ($d^2k/dT^2 = 0$). Intuitively, it makes more sense to align an inflection point to environmental temperature than the T_{opt}, as either side of T_{opt}, reaction rates decline. In terms of a cell, coordination over all enzymatically driven pathways is arguably a more dominant evolutionary parameter than enzyme function as isolated entities (Zhao *et al.*, 2007). This is especially true in coordinating cellular processes to all scale at the same rate with changes in temperature. Enzymatic T_{opt} is linked to $\Delta C^{\ddagger}_{\mathrm{p}}$ (Section 1.4.1), which in turn is hypothesised to be intimately linked to the catalytic proficiency of enzymes, as determined by the difficulty of a reaction chemistry (Section 3.2). If this is the case, it is not possible to coordinate T_{opt} values without compromising catalytic efficiencies [which are constrained to a narrow window (Wolfenden & Snider, 2001)] over the wide range of cellular chemistries. Regardless of T_{opt} differences however, rates about inflection points scale approximately equally with changes in temperature (Figure 4.15). Even the small selection of enzymes characterised from *E. coli* illustrate in Figure 4.15 illustrate the divergence in rates over these enzymes at temperatures above the inflection point. This is a critical point for enzymes functioning in a cascade in such a core metabolic pathway. Thus, fixing the inflection point at growth temperature offers a means to coordinate cellular processes over a range of growth temperatures to either side of the ideal.

Under the equilibrium model, a systematic deviation is noted as data moves from enzymes from thermophilic through to psychrophilic organisms (Lee *et al.*, 2007). The more psychrophilic the enzyme, the further the alignment between T_{eq} and growth temperature. Considering this, the idea that T_{eq}, as a measure of thermal adaptation, is evolutionarily lagging behind growth temperature in the adaptation of enzymes to temperature from a thermophilic common ancestor was proposed. Given the correlation between T_{inf} and growth temperature (Figure 4.15), with the use of an inflection point to describe temperature adaptation, no such argument is necessary as enzymes appear to be consistently adapted to growth temperature. This is unsurprising given the long evolutionary time scales, the importance of tuning enzymatic rates with temperature, and the highly adapted nature of many other cellular processes.

4.3.1.3 Ruling out the effect of denaturation on curvature

The temperature at which denaturation starts having a significant impact on rates was investigated by the differential scanning fluorimetry (Lo *et al.*, 2004). It is acknowledged that differential scanning fluorimetry is generally used only as a relative measure (ΔT_m), and the method is not the most accurate in terms of determining T_m information. However here, as data were collected in the assay buffer and thus conditions for the melts are as close as possible to the true assay conditions, melt data are presented as a close approximation of the true T_m values for each protein. T_m values were taken into account in the pathway assays (Figure 4.18), which required a relatively long assay time over which denaturation may have an effect. By comparison, single enzyme characterisations were assayed over 10 seconds, thus denaturation was less of a factor.

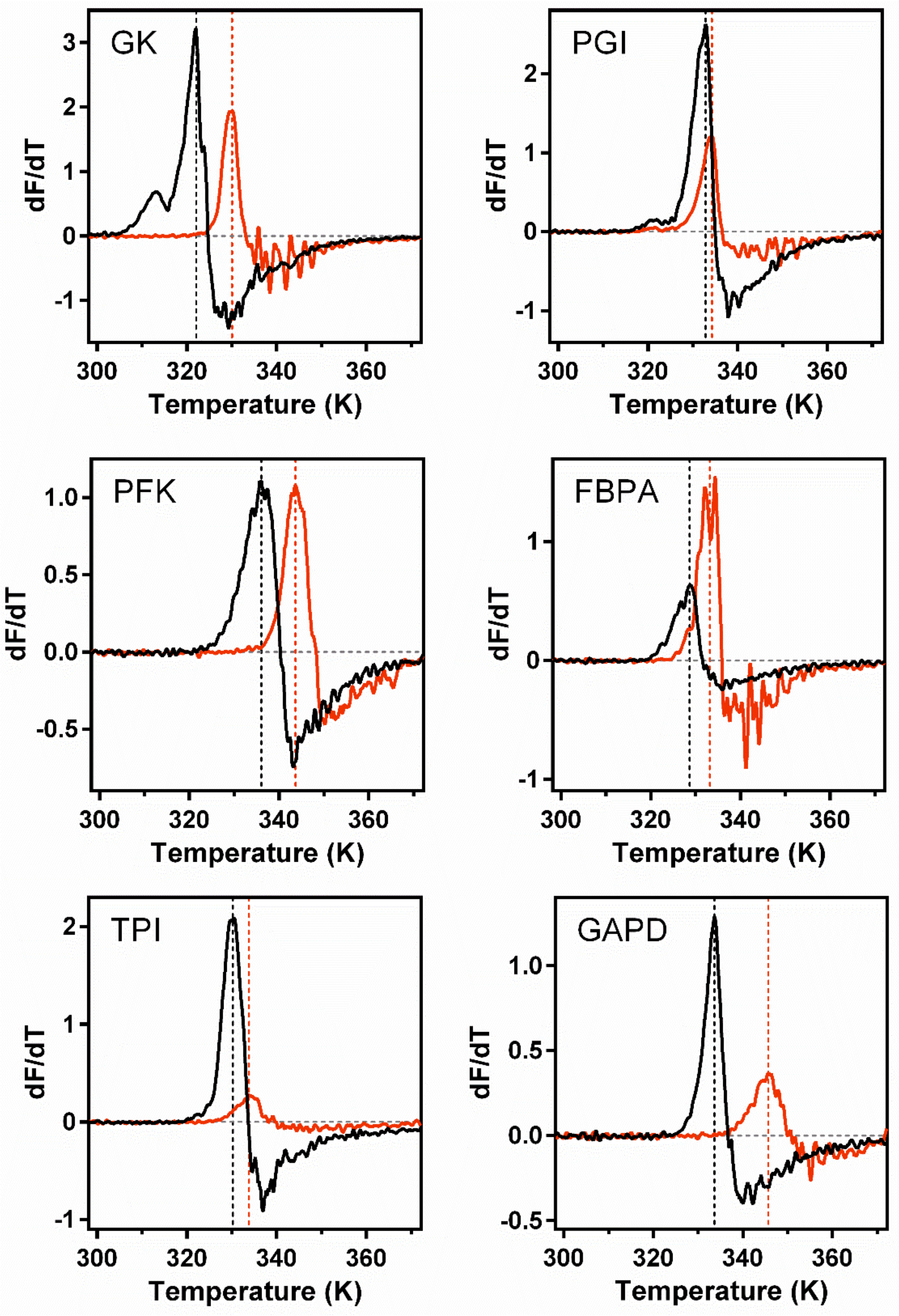


Figure 4.16: T_m determination for glycolytic enzymes in the absence (black) and presence (orange) of pathway ligands (P_i, NAD^+, ATP) as determined by differential scanning fluorimetry. Data are presented as the first derivative of the fluorescence output. Melt conditions (buffer and ligand concentrations) are the same as those used in the pathway assays (Section 2.3.5 and 2.3.6). T_m values are given in dashed lines (values in Table 4.2) as determined by the maximum value in the first derivative plot. The one exception to this is FBPA, where the average temperature between the two peaks is given for the T_m with ligands. This peak pattern suggests the presence of two populations of FBPA; this was not explored further.

4.3.2 Pathway characterisation

The values of $\Delta C^{\ddagger}_{p}$ for organism growth rates (Hobbs *et al.*, 2013) and ecosystem processes (Schipper *et al.*, 2014) were found to be in the same range as the values for individual enzymes. This suggests that the $\Delta C^{\ddagger}_{p}$ for a pathway of consecutive steps may be roughly approximated by the arithmetic mean of the $\Delta C^{\ddagger}_{p}$ values for the individual steps. The overall rate for a pathway is affected by the rate of the individual steps; the slower a reaction at an individual step, the longer the transition time through that step ($1/k_{cat}$; equation **16**). Similarly, this suggests that there may be a relationship between the average $\Delta C^{\ddagger}_{p}$ of all steps and the overall pathway $\Delta C^{\ddagger}_{p}$ as extremes in rates get balanced by other steps in the pathway. The natural extreme of this is to invoke a "master enzyme" scenario whereby a single enzyme determines the rate for cellular growth, as postulated by Corkrey *et al.* (2012). Here, a more complex scenario is envisaged whereby parameters from each step are considered to contribute to the rate of the process overall. From the $\Delta C^{\ddagger}_{p}$ values determined for each enzyme (Table 4.2), the average $\Delta C^{\ddagger}_{p}$ for the overall pathway is -3.2 kJ.mol^{-1}.K^{-1}. Given the large negative $\Delta C^{\ddagger}_{p}$ of GK (-7.1 kJ.mol^{-1}.K^{-1}) compared to the average for the pathway, removal of this first step increases the average $\Delta C^{\ddagger}_{p}$ to -2.4 kJ.mol^{-1}. K^{-1}. The effect of removing GK provides a strong test of the influence of individual enzymes on the overall pathway $\Delta C^{\ddagger}_{p}$.

Theoretical rates for an enzymatic pathway (k_{tot}) of *n* steps can be estimated from the sum of the inverse rate constants for the individual steps (equation **16**). This is based on a transit time approximation through each pathway step.

$$\frac{1}{k_{tot}} = \frac{1}{k_1} + \frac{1}{k_2} + \cdots \frac{1}{k_n} \quad \text{.................................}\textit{overall rate of } n \textit{ step pathway} \quad \textbf{(16)}$$

Given the experimentally determined individual MMRT curve fits (Figure 4.14), rates of the pathway at various temperatures can be simply modelled with this equation **(16)** to corroborate the expected pathway $\Delta C^{\ddagger}_{p}$ values. Modelled pathway rates show curvature which is modelled well by MMRT, consistent with the scalability of MMRT (Figure 4.17). Passage through the pathway from G6P via the last five enzymes of the pathway increases rates (Figure 4.17 A), and increases the $\Delta C^{\ddagger}_{p}$ to -1.7810 ± 1x10^{-4} kJ.mol^{-1}.K^{-1}, compared to -3.7 ± 0.1 kJ.mol^{-1}.K^{-1} when

GK is included in the pathway (Figure 4.17 B). The relative values of these estimates are the same as predicted by the arithmetic mean (Table 4.2). Removal of the GK step also incurs a shift in T_{opt} under this model, with T_{opt} shifting from 326 to 338 K. This is consistent with the removal of the enzyme with the lowest T_{opt}, and commensurate with the relationship between T_{opt} and $\Delta C^{\ddagger}_{\mathrm{p}}$ (Section 1.4.1)

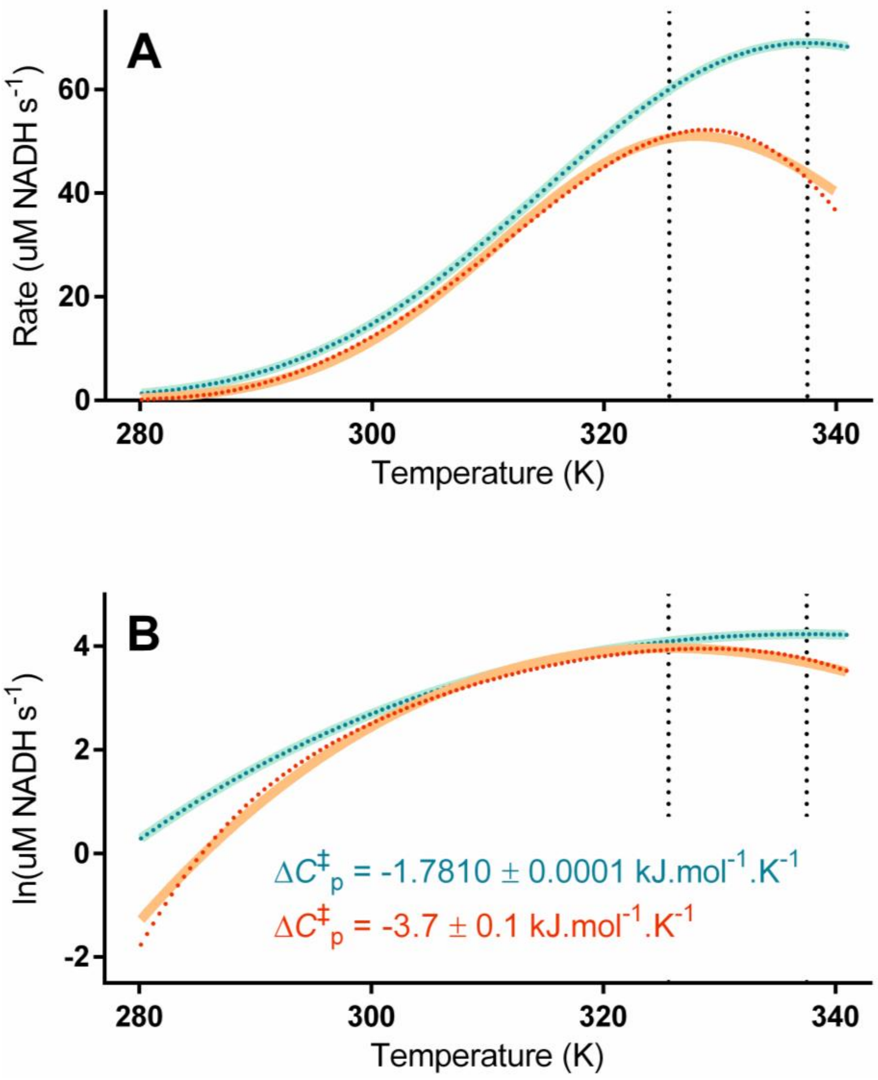


Figure 4.17: Modelled flux through the glycolytic pathway from glucose to BPG (orange) and G6P through BPG (blue) based on equation ***(16)****. Enzyme concentrations used in the analysis are the same as those used experimentally (Appendix D.2), while k_{cat} values are determined from MMRT fits to single enzyme T_{opt} data (Figure 4.14). Modelled data points are fit to a single MMRT curve as continuous lines for* ***(A)*** *raw rates and* ***(B)*** *ln(rate). The two pathways differ significantly in both $\Delta C^{\ddagger}_{p}$ and T_{opt} (vertical dashed lines). Errors are given as SE of the fit of the MMRT equation to the calculated rates.*

Predicted $\Delta C^{\ddagger}_{\mathrm{p}}$ values correlate with those observed experimentally (Figure 4.18). Experiments were set up with the same enzyme concentrations as those used for modelling, in the same assay buffer in which individual enzymes were characterised

(Figure 4.17; Appendix D.2). Over two replicates of the full pathway, $\Delta C^{\ddagger}_{\mathrm{p}}$ varied from -4.9 and -4.3 ± 0.7 kJ.mol^{-1}.K^{-1}. With the removal of the step catalysed by GK, $\Delta C^{\ddagger}_{\mathrm{p}}$ values increased to -2.3 ± 0.3 kJ.mol^{-1}.K^{-1}. The trend in these $\Delta C^{\ddagger}_{\mathrm{p}}$ values corresponds with that expected from both the arithmetic mean and inverse rate predictions (Table 4.3). Experimental T_{opt} values (equation **14**) shift with the removal of GK as expected, increasing from 326 to 331 K.

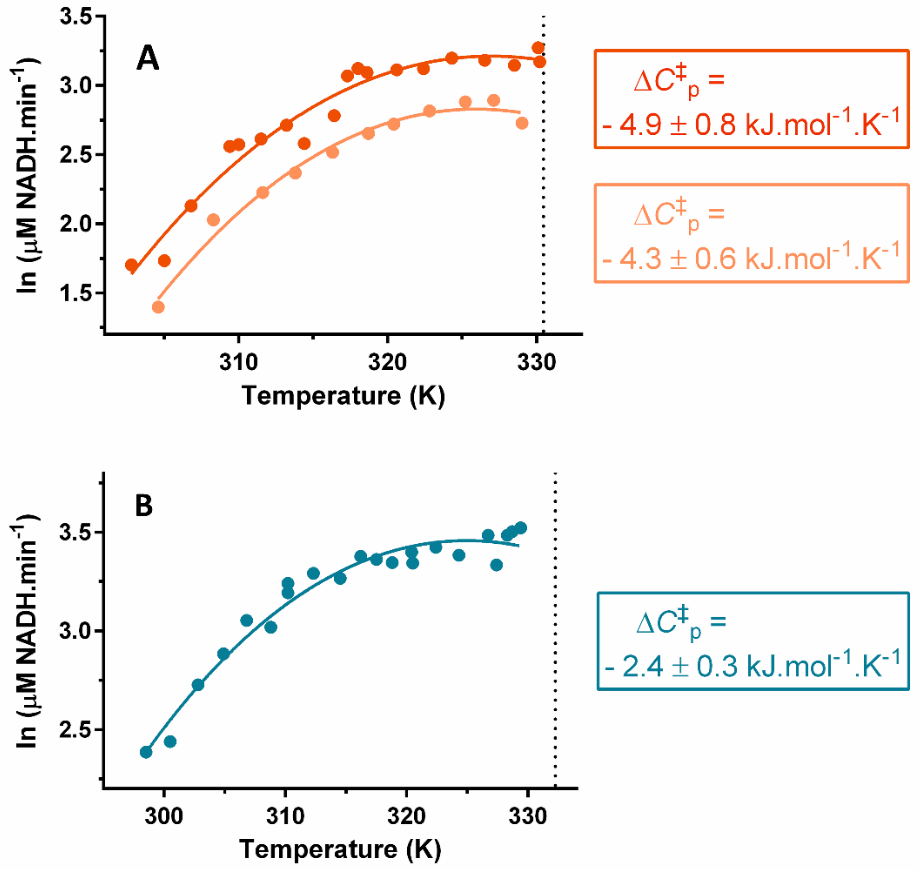


Figure 4.18: Experimental temperature profiles of the six-step pathway ***(A)*** *and five-step pathway* ***(B)****. Removal of the GK catalysed step in* ***(B)*** *increased the average* $\Delta C^{\ddagger}_{p}$ *calculated from the individual steps from -4.6 to -2.4 kJ.mol*$^{-1}$*. K*$^{-1}$*; this predicted change is observed outside of error range in the experimental data. Dotted vertical lines in both panes correspond to the lowest* T_m *value (GK and FBPA respectively) measured in the presence of ligands by differential scanning fluorimetry (Figure 4.16).*

The inverse rate model has an obvious discrepancy from the experimental set up in that it assumes saturating substrate concentrations. Data from T_{opt} curves were determined at saturating substrate concentrations, whereas all steps past the initial reaction in the experimental pathway are not known, but assumed to be well below saturating concentrations. To investigate this, the pathway was modelled using OpenCore CellML software (Cuellar *et al.*, 2003). Modelling to incorporate the effect of substrate concentration utilised K_M values from Michaelis Menten enzyme characterisations at 37 °C (Figure 4.11 through Figure 4.13). These were assumed to be invariable with temperature, as most examples are relatively stable over the increase to 44 °C. Variation in k_{cat} with temperature was taken from T_{opt} profiles (Figure 4.14) for 14 temperatures between 288 and 350 K for each enzyme. Details on the script developed for this, including simplifications and assumptions, are in Section 2.3.7.2; the full script is also given on the attached disk.

Modelling incorporating the effect of substrate concentrations based on measured K_M values highlighted the potential confounding effects of the ATP:glucose concentrations when the full pathway is assayed. In the modelling, if glucose is supplied in higher concentrations than ATP (as in the experimental set up), the quick flux through the first step results in rapid ATP depletion, limiting rates through the remainder of the pathway due to the dependence of PFK on ATP (Figure 4.19, column 2). Determination of rate of product formation (BPG, Figure 4.19, column 2) was also confounded by the sigmoidal output resultant from substrate depletion. This coupled to the low T_{opt} of GK (319 K), results in a region of depressed rates about the temperature where GK is working optimally and rapidly depleting ATP before PFK passes substrate onto the second part of the reaction (Figure 4.1). MMRT fits poorly to these data (22 % standard error, $R^2 = 0.94$), and importantly does not emulate the shape of the experimental data (Figure 4.20 A). By comparison, ATP never becomes limiting with the pathway from G6P, and the predicted temperature profile fits MMRT with low errors (-3.1 ± 0.1; $R^2 = 0.9971$; Figure 4.20 B). As this depression in the T_{opt} profile is not seen in the experimental data (Figure 4.18), it is assumed than ATP depletion is not occurring in the experimental set up; as such, ATP limitation was removed from the model (by altering starting concentrations so ATP > glucose). This model captures the profile of the experimental data (Figure 4.20), and predicts the relative $\Delta C^{\ddagger}_{p}$ values for the two pathways correctly.

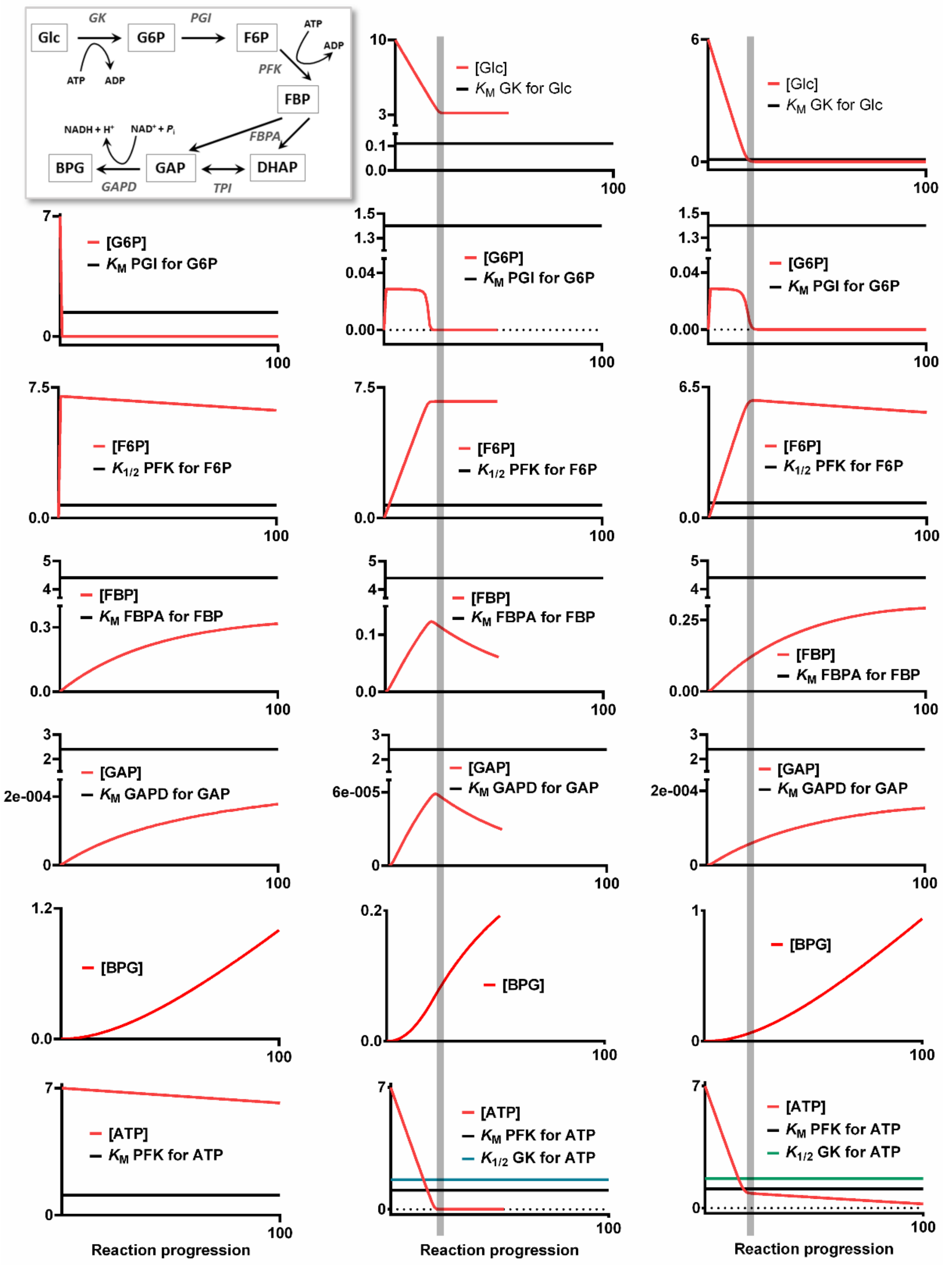


Figure 4.19: Changes in substrate concentrations (mM) over the course of the reaction as modelled by CellML for: ***(Left)*** *five enzyme steps from PGI;* ***(Middle)*** *six steps, with true assay ATP and glucose concentrations;* ***(Right)*** *six steps, with six mM glucose so that [ATP] > [Glc]. Relevant K_M values are given in reference to substrate concentrations in the assay. Vertical grey lines represent the time point at which glucose concentrations plateau (middle and right), and ATP becomes limiting (middle). The reaction pathway and enzymes are summarised top left.*

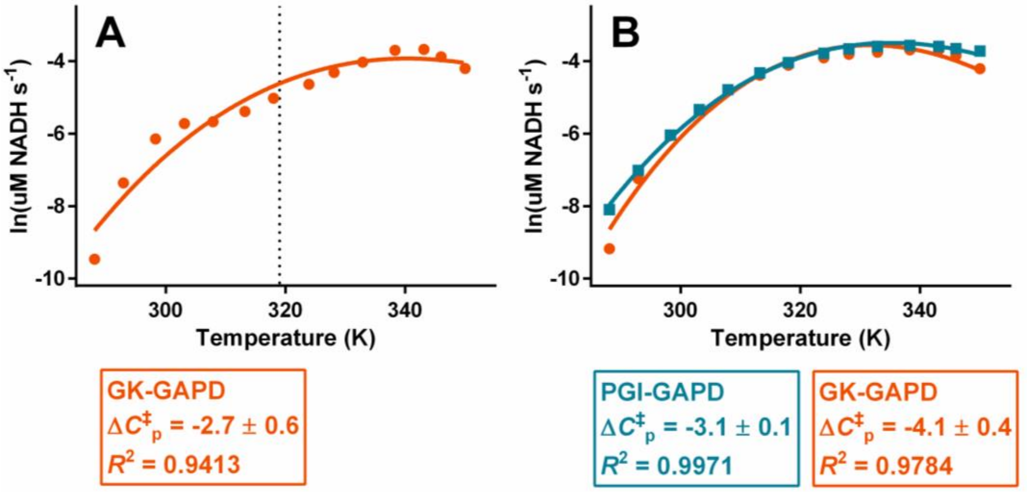


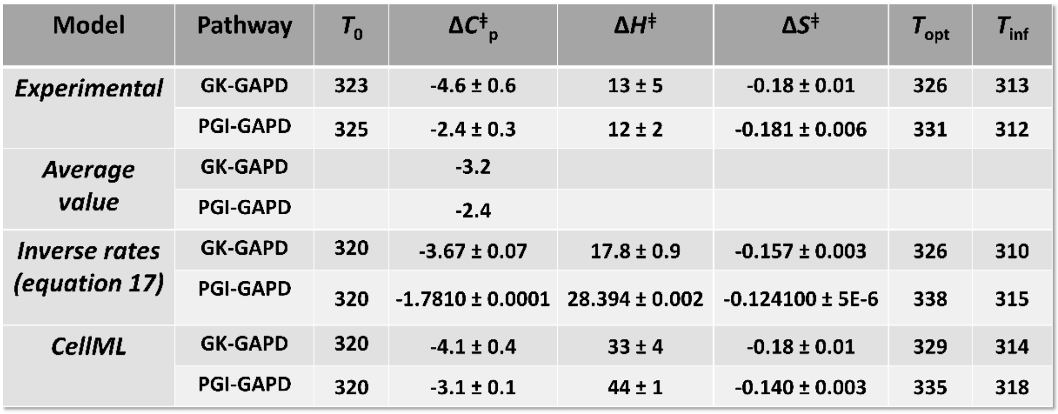
Figure 4.20: Modelled curvature from flux through the pathway in CellML. ***(A)*** *Modelling of GK-GAPD where glucose concentrations were greater than ATP concentrations (10 and 7 mM respectively) where ATP becomes limiting during the reaction. The T_{opt} of GK is indicated (dotted line) where fast GK turnover limits later steps in the pathway by depleting ATP. Poor agreement between the shape of these data and experimental data (see Figure 4.18) resulted from this approach. This suggested ATP limitation was not occurring experimentally, justifying the alteration of these relative concentrations in:* ***(B)*** *Modelling of the pathway from PGI (blue) and from GK-GAPD where ATP concentration does not become rate limiting (orange). Values for $\Delta C^{\ddagger}_p$ are in kJ.mol^{-1}.K^{-1} ± the SE.*

4.4 Conclusions

Overall, all three approaches were accurate predictors of the relative $\Delta C^{\ddagger}_{\mathrm{p}}$ between the two pathway fragments, and predicted $\Delta C^{\ddagger}_{\mathrm{p}}$ values in the correct range. Modelling using inverse rates from T_{opt} assays was a more reliable predictor of experimental curvature in this case, especially if considerations for complexity of set up are included. Of the averaging of individual $\Delta C^{\ddagger}_{\mathrm{p}}$ values and inverse rate modelling, both require the same experimental work (temperature profiles fit to MMRT), however, the inverse modelling approach estimated pathway $\Delta C^{\ddagger}_{\mathrm{p}}$ values more accurately. Despite the increased modelling complexity of the CellML model, removing the assumption of saturating substrate concentrations did not result in more accurate $\Delta C^{\ddagger}_{\mathrm{p}}$ predictions. This is likely due to the additional parameters going into this modelling ($K_{\mathrm{M}}/K_{½}$, K_{i}, *Hill coefficient*) which have been extrapolated from experimental data, all of which add error to the prediction. Inverse rate and CellML modelling had the added advantage of predicting T_{opt} and T_{inf} information;

in both cases, each method matched the experimentally gained values closely and neither method outperformed the other (Table 4.3).

*Table 4.3: Comparison of modelling approaches to the experimental data parameters. Temperature is given in K, $\Delta C^{\ddagger}_{p}$ in kJ.mol^{-1}.K^{-1}, and $\Delta S^{\ddagger}$/$\Delta H^{\ddagger}$ in kJ.mol^{-1}. T_0 is given as 4 degrees below the temperature at which highest rates were measured experimentally; T_{opt} is calculated from the curve fit (equation **14**). Errors are ± SE of the fit of equation **(7)**.*



Model	Pathway	T_0	$\Delta C^{\ddagger}_{p}$	$\Delta H^{\ddagger}$	$\Delta S^{\ddagger}$	T_{opt}	T_{inf}
Experimental	GK-GAPD	323	-4.6 ± 0.6	13 ± 5	-0.18 ± 0.01	326	313
	PGI-GAPD	325	-2.4 ± 0.3	12 ± 2	-0.181 ± 0.006	331	312
Average value	GK-GAPD		-3.2				
	PGI-GAPD		-2.4				
Inverse rates (equation 17)	GK-GAPD	320	-3.67 ± 0.07	17.8 ± 0.9	-0.157 ± 0.003	326	310
	PGI-GAPD	320	-1.7810 ± 0.0001	28.394 ± 0.002	-0.124100 ± 5E-6	338	315
CellML	GK-GAPD	320	-4.1 ± 0.4	33 ± 4	-0.18 ± 0.01	329	314
	PGI-GAPD	320	-3.1 ± 0.1	44 ± 1	-0.140 ± 0.003	335	318

Here, six enzymes working in succession (including a branch point) from the glycolytic pathway of *E. coli* have been cloned, expressed, purified and characterised as individual steps and as a pathway. Characterisation *in vitro* of individual enzymes allowed a full exploration of how individual enzymes contribute to kinetic characteristics of the complete pathway. This experimental characterisation of glycolysis as individual steps and as a pathway has demonstrated the scalability of MMRT. Previously, MMRT has been well established at the single enzyme level, and shown to describe organism growth rate data (Hobbs *et al.*, 2013). Here, the theoretical justification of the application of MMRT to metabolic pathways, and by extension organism growth rates is presented. The effects imposed by a non-zero $\Delta C^{\ddagger}_{p}$ for enzymatic catalysis on the temperature dependence of enzyme rates is carried through to the temperature profile of enzymatic pathways. Through the removal of a single step (GK) and the altered pathway $\Delta C^{\ddagger}_{p}$ imposed by this, it has been demonstrated that individual steps have a predictable impact of the temperature dependence of the pathway. Curvature observed in pathway rates with temperature is well modelled with MMRT, both in the experimental and modelled data for the pathway. Pathway modelling from various approaches has corroborated experimental data, and illustrated that the average $\Delta C^{\ddagger}_{p}$ for a set of individual steps approximates that of the pathway.

The structure of transition state analogue bound MalL WT and V200S

Previously, apo structures have been reported for MalL WT and the V200S, V200A, and G202P variants (Hobbs *et al.*, 2013). Valuable insights were gained into the potential mechanisms of $\Delta C^{\ddagger}_{p}$ change from the analysis of hydrogen bonding networks and through molecular dynamics (MD) simulations. From this, in conjunction with biochemical data, the mode of $\Delta C^{\ddagger}_{p}$ decrease in V200S and G202P mutants were proposed to be due to rigidification of the ground state, and increased flexibility of the transition state (TS) respectively (Section 1.4.2). Apo V200S displays reduced flexibility compared to WT, especially in loops about the active site in MD simulations. Protein-ligand C_p at the TS is inferred to be unaffected by the mutation, as k_{cat} values similar to WT indicate similar binding to the TS. By comparison, G202P has similar ground state flexibility to WT MalL in MD simulations, but C_p has been increased at the TS (as inferred by greatly compromised k_{cat}).

Here, this analysis is taken further through determination and analysis of the crystal structures of MalL WT and V200S bound to the TS analogue 1-deoxynijirimycin (DONM). From this, further insights into the interactions of MalL variants in complexes throughout the reaction cycle, and the potential effects of this on dynamics are investigated.

5.1 MalL transition state analogue selection

5.1.1 Mechanism of MalL catalysis

Glucosidases as a family are well studied enzymes, with a well understood general catalytic mechanism (Zechel & Withers, 2000). In addition to this, many TS analogues have been developed for a range of different glucosidases.

MalL catalyses the hydrolysis of isomaltose [D-Glc-α(1-6)-D-Glc] through a two-step double displacement retaining mechanism (Figure 5.1). In the first step, nucleophilic attack of the anomeric carbon of the -1 glucose by the catalytic base (Asp199) occurs in a concerted manner with the breaking of the glycosidic bond between the two glucose units, releasing the +1 glucose [glucose binding site nomenclature from (Davies *et al.*, 1997)]. Concomitantly, protonation of the glycosidic oxygen (of the leaving glucose unit) occurs as a hydrogen is donated from the general acid catalyst, Glu255. The -1 glucose is bound to Asp199 in a covalent glycosyl-enzyme intermediate (MacKenzie *et al.*, 1998). Nucleophilic attack by water releases the bound -1 glucose in the second step of the reaction, hydroxylating the anomeric carbon of the retained glucose. The α configuration of the -1 glucose is retained through the mechanism. Transition states for both steps are largely the same; the -1 glucose present in both TS's adopts a contorted conformation, flattening about the anomeric carbon, to allow for oxocarbenium ion character (Zechel & Withers, 2000). Oxocarbenium ion character about the O_5-anomeric carbon bond localises positive charge about the O_5 and C_1 atoms (Sinnott, 1990). In the first and second step transition states, bonds are partially breaking and forming respectively. In the first step, partial bonds link the anomeric carbon to Asp199. The glycolytic bond between the two glucose units is partially broken, and the bond to hydrogen from Glu255 is forming. The partial bonding arrangement in the second TS is similar, except water takes the place of the +1 glucose unit. The two transition states differ in that the bulky +1 glucose unit remains in the active site in the first TS. Compared to the substrate bound state, this +1 glucose is pushed further away from the -1 glucose due to elongation of the breaking glycosidic bond.

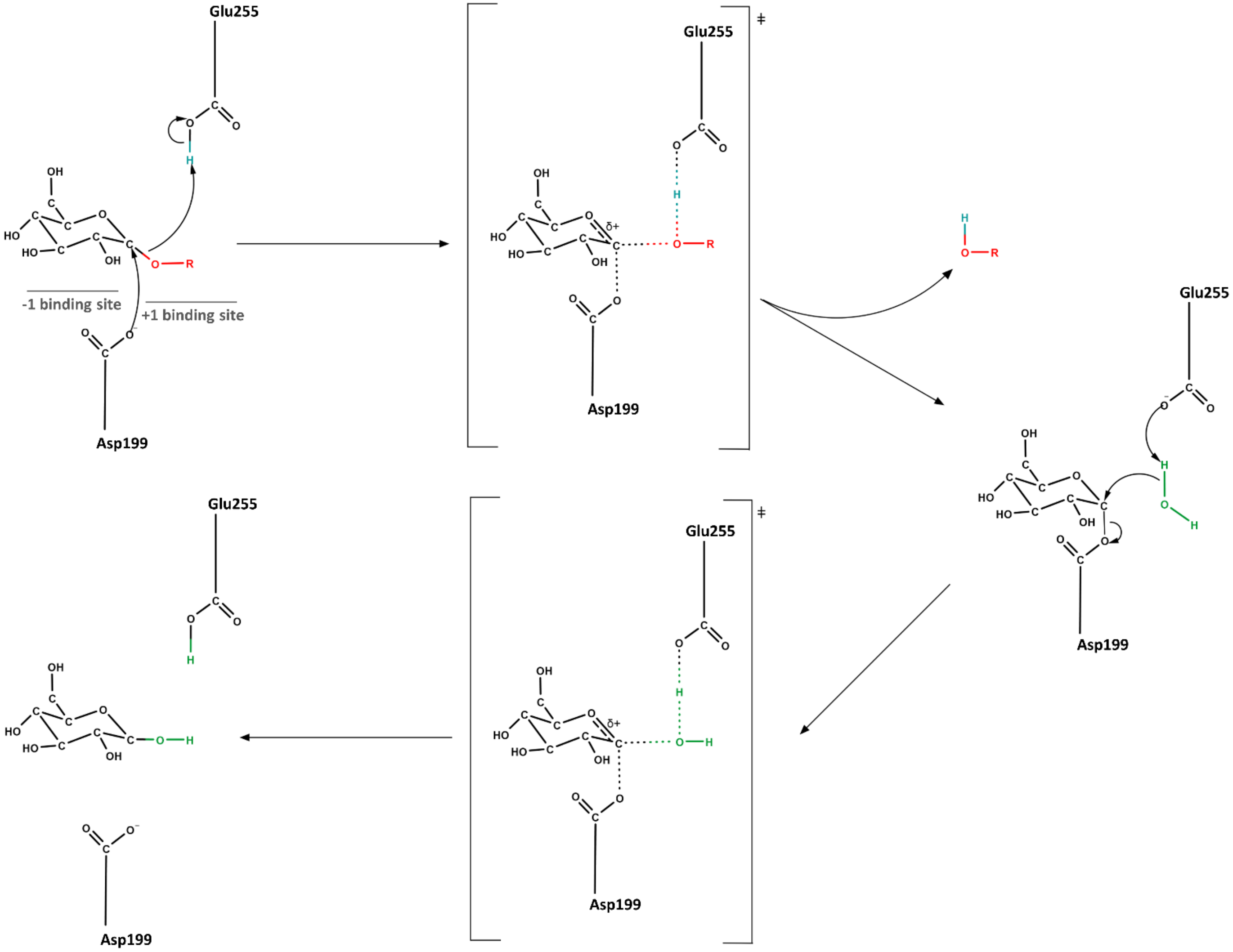


Figure 5.1: Mechanism of MalL catalysis [adapted from (Zechel & Withers, 2000)]. Where R = α-D-glucose. Note that the orientation at the anomeric carbon is in the down position (α) – angles have been altered slightly for clarity. MalL proceeds via a two-step mechanism to cleave isomaltose into two glucose units. In the first step, the +1 glucose unit is released, and a covalent glycosyl-enzyme intermediate is formed between Asp199 and the -1 glucose. The second step cleaves the covalent intermediate, releasing the second glucose via nucleophilic attack by water. The -1 glucose passes through a contorted half chair conformation as the anomeric carbon adopts a more planar conformation at the TS [not shown; (Rye & Withers, 2000; Vasella et al., 2002)].

5.1.2 Selection of transition state analogues of MalL

The literature reports a large array of glucosidase inhibitors from both natural and synthetic sources, such as nojirimycin and 1-deoxynijirimycin (DONM) iminosugars (natural) and 1-deoxy-L-idonojirimycin (synthetic) (Borges de Melo *et al.*, 2006; Compain *et al.*, 2009). The development of inhibitors has largely been driven by the commercial drug value of these compounds in treatments such as diabetes, HIV, and certain cancers. These inhibitors fall into several chemical groupings including disaccharides, iminosugar, thiosugars and amino sugars. Of these, the iminosugars are of interest here due to their charge distribution mimicking that of the TS (Bols, 1998). Iminosugars comprise a group of sugar analogues where a nitrogen atom has been substituted into the ring at either the C_1 position (e.g. noeuromycin) or ring oxygen (e.g. 1-deoxynojirimycin) – see Figure 5.2 for ring numbering of sugars. When protonated, this mimics the positive charge localised about the C_1-O_5 atoms of the sugar ring at the TS (Varrot *et al.*, 2003).

Investigations into iminosugars started from the fortuitous discovery of nojirimycin from various *Streptomyces* strains in the 1960's. Isolation and characterisation of the compound stemmed from the observed antimicrobial activity of *Streptomyces* against *Micrococcus luteus*, *Xanthomonas oryzae* and *Shigella flexneri* (Inouye *et al.*, 1968). However, as evident from the initial purification difficulties, nojirimycin is short lived due to the presence of an unstable N-C-O moiety. Removal of the hydroxyl group to alleviate this instability results in 1-deoxynojirimycin (Figure 5.2), which also shows broad spectrum inhibition against glucosidases hydrolysing sugars of α and β configuration (Hughes & Rudge, 1994; Dong *et al.*, 1996) . Since this time, the list of iminosugars has been expanded significantly through both natural isolates and synthetic derivatives (Borges de Melo *et al.*, 2006; Compain *et al.*, 2009). A range of synthetic glycosidase inhibitors has been developed that include extensions to the basic N-substituted glucose units such as the addition of various alkyl groups (including the indolizidine compounds where the alkyl chain loops back to form a second ring structure), addition and removal of various hydroxyl groups, methylation, pseudodisaccharides, and the incorporation of a different base sugar unit (Borges de Melo *et al.*, 2006; Compain *et al.*, 2009). As a general rule, these added complexities tend towards tighter, but more specific binders for a specific glucosidase (Borges de Melo *et al.*, 2006). For example, Dong *et al.* (1996) characterised broad inhibition by isofagomine (glucose with N-substituted anomeric centre) against five glucosidases with K_i values ranging from 800 to 0.1 μM. Pseudo-disaccharide derivatives of isofagomine with and without a zwitterion component showed increased inhibition against glucoamylase, while becoming a poorer inhibitor of the other tested glucosidases. Due to the wide range of inhibition of DONM across a range of glucosidases, along with commercial availability, DONM was chosen as a TS binder with which to crystallise MalL variants.

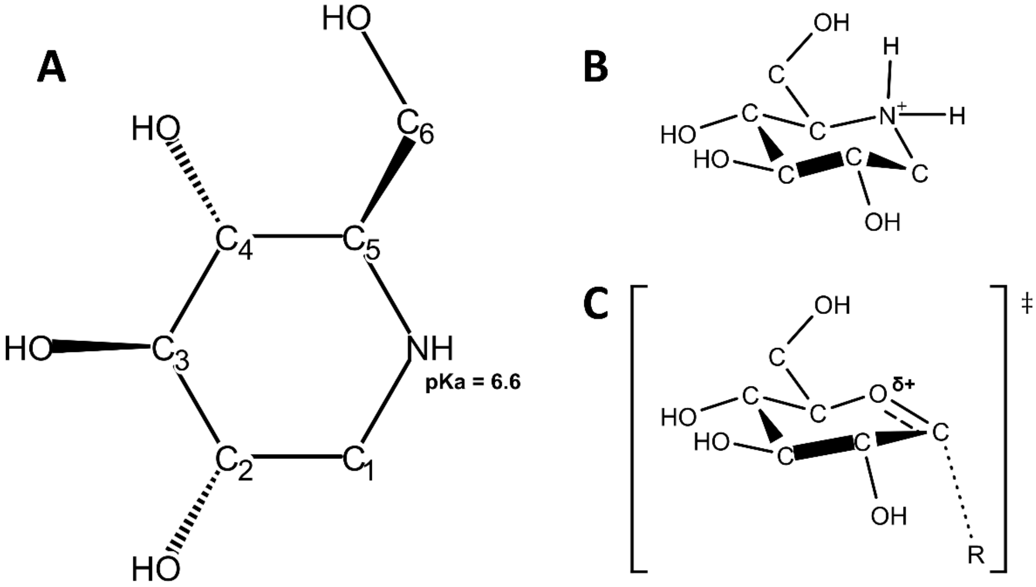


Figure 5.2: Chemical structure of DONM. ***(A)*** *Planar DONM view. Carbon atoms are numbered as they are referred to throughout the text, following the general protocols for numbering sugar atoms. Oxygens/hydroxyl groups are unlabelled, but throughout the text are referred to by the number of the carbon to which they are attached (e.g. O_2H). Nitrogen is in the position occupied by the ring oxygen in glucose. Nitrogen is protonated at pH's below 6.6 (Inouye et al., 1968).* ***(B)*** *Chair conformation of protonated DONM; charge distribution mimics* ***(C)*** *the reaction TS showing contortion away from the chair conformation, as the anomeric carbon becomes more planar. R is either glucose or water.*

5.2 MalL WT and V200S crystallisation and structure determination

5.2.1 Crystallisation

For both WT and V200S MalL variants, initial crystallisation conditions were determined from a total of 384 conditions using the Hampton Research (USA) screens via the sitting drop method (Section 2.2.3.1.2). Protein (between 7 and 11 mg.ml^{-1}) was mixed with DONM to give a final concentration of one mM DONM in the protein-mother liquor drop. Promising crystallisation conditions were carried through to hanging drop fine screens where pH and precipitant concentration were varied (Section 2.2.3.1.3). Final crystallisation conditions for WT MalL were 10 % (v/v) Tacsimate^{TM}, pH 6.0 (Hampton Research, USA), 100 mM MES, pH 6.4, and 25 % (w/v) PEG 4,000. Final crystallisation conditions for V200S were 100 mM BIS-TRIS, pH 5.3, with 23 (%) w/v PEG 3,350. Representative crystals are shown in Figure 5.3.

5.2.2 X-ray data collection

Crystals were sent to the Australian Synchrotron frozen in liquid nitrogen with 20 % (v/v) glycerol cryoprotectant for data collection. For both MalL variants, high and low resolution data sets were collected to define spots at both resolution extremes (Figure 5.3). High resolution data sets were collected first at a detector distance of 140 and 150 mm (for WT and V200S respectively) in one degree increments for one second for a full 360° rotation. Low resolution data sets were collected from the same crystal at a detector distance of 250 and 260 mm respectively, in one degree increments through 360° with an exposure time of one second.

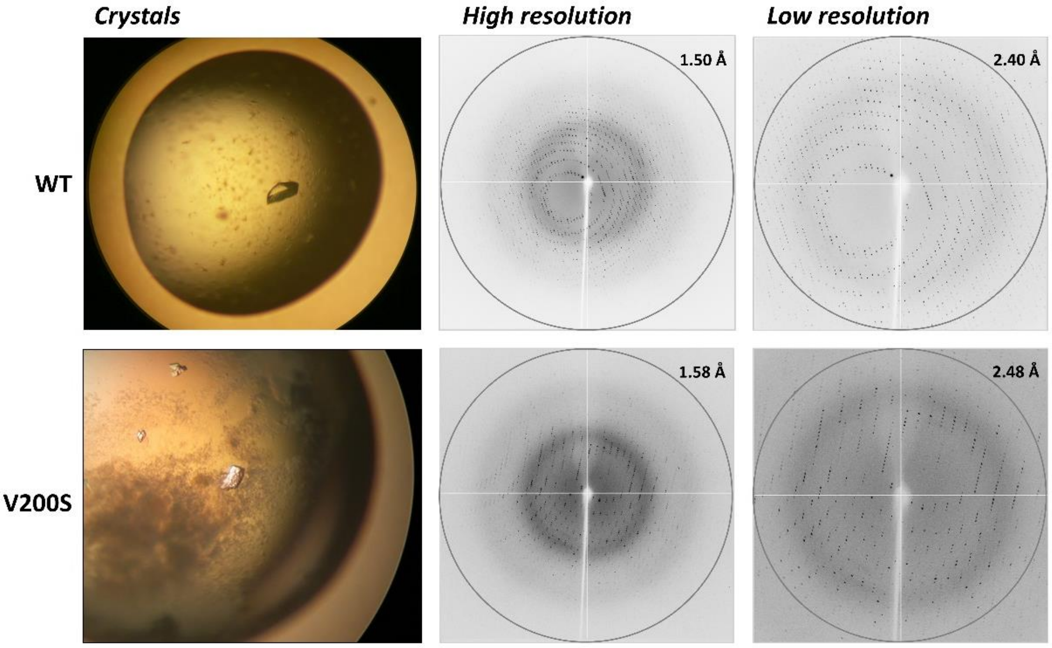


Figure 5.3: Crystallisation and diffraction of WT and V200S with bound DONM. High and low resolution data sets for each structure were collected for 360 degrees in one degree increments.

5.2.3 Data processing

Space group determination for both data sets was performed in MOSFLM (Leslie & Powell, 2007). High and low resolution data sets were integrated in MOSFLM separately, before being combined in Sortmtz (P. J. Daly, Daresbury) within the CCP4 data suite (Winn *et al.*, 2011). Integrated and combined reflections were scaled and merged in SCALA (Evans, 2006). Data were cut off at this stage (if necessary) at a resolution which gave a R_{merge} of less than 0.8 in the outer shell. Both WT and V200S structure were solved to 1.58 Å. Matthews coefficient analysis (Matthews, 1968) was used to determine the number of molecules in the asymmetric unit. Data from these steps is given in Table 5.1.

Table 5.1: Data collection statistics for MalL WT and V200S with DONM bound.

Data statistic	**MalL WT with DONM**		**MalL V200S with DONM**	
Space group	P 2_1		P 2_1	
Wavelength (Å)	0.9537		0.9537	
Cell dimensions				
a/b/c (Å)	61.6/98.8/101.4		48.5/100.4/61.8	
α/β/γ (°)	90/103.8/90		90/112.7/90	
Mosaicity	0.69		0.84	
Monomers in the asymmetric unit	2		1	
Resolution range (Å)	1.58-46.78	(1.58-1.67)	1.58-44.75	(1.58-1.67)
Number of observed reflections	1,437,833	(146,565)	685,746	(74,471)
Number of unique reflections	159,542	(22,758)	74,025	(10,563)
R_{merge}	0.078	(0.681)	0.094	(0.548)
Mean I/σI	16.4	(2.9)	15.3	(3.3)
Completeness	98.8	(97.2)	99.1	(96.8)
Multiplicity	9.0	(6.4)	9.3	(7.1)
Wilson B-factors	12.7		7.4	

*Values in brackets correspond to the outer resolution shell.

Molecular replacement was performed in PHENIX (Adams *et al.*, 2010) using the previously solved apo MalL WT structure (PDB code 4M56). Further model refinement was done automatically using AutoBuild within the Phenix software. The models were manually refined in COOT (Emsley & Cowtan, 2004) using $2|F_O|-|F_C|$ and $|F_O|-|F_C|$ maps contoured to 1 σ and 3 σ respectively. Refinement of the model was performed in Phenix (Adams *et al.*, 2010). A summary of the final statistics for MalL WT and V200S structures with DONM is given in Table 5.2. Electron density quality and model fit, including for the TS analogue, are illustrated in Figure 5.4.

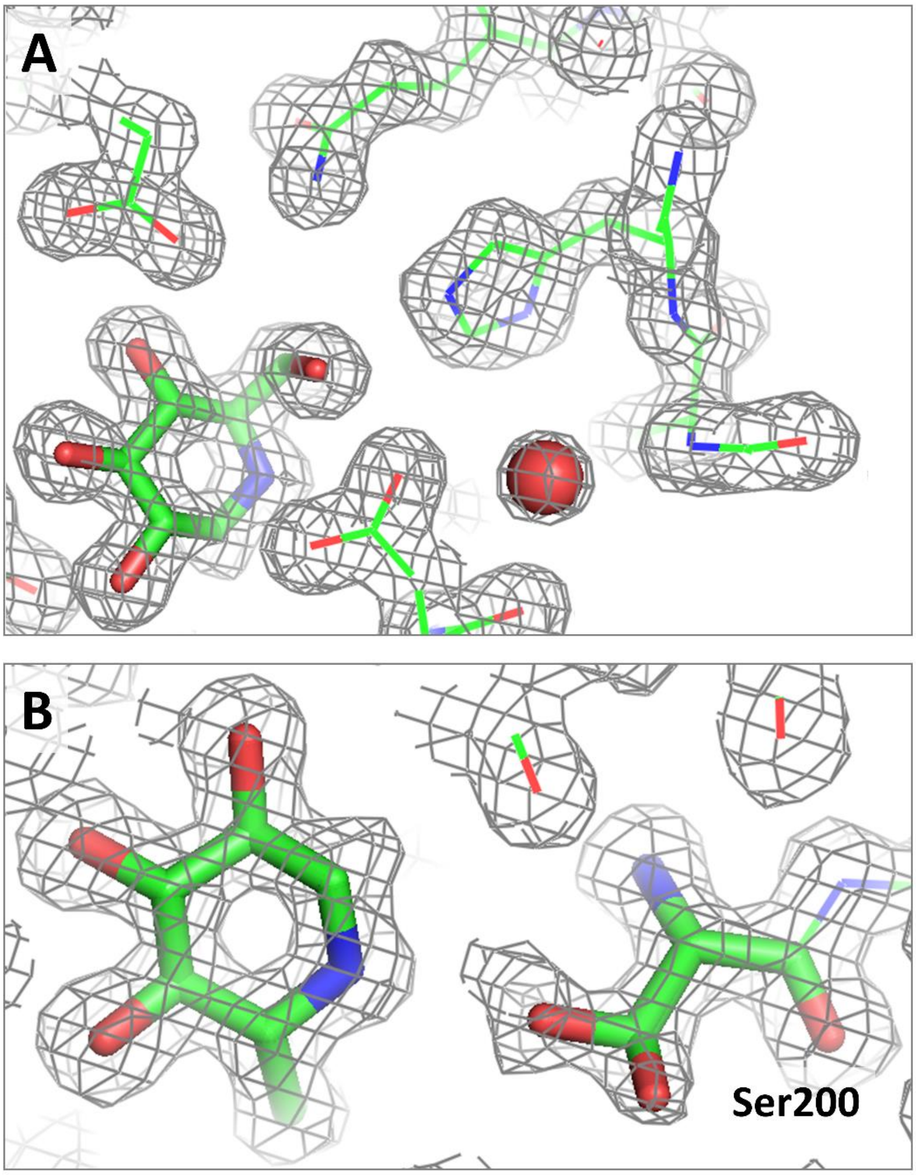


Figure 5.4: Electron density for MalL WT ***(A)*** *and V200S* ***(B)****. Residues and ligands are well defined throughout both structures. In the representation of V200S, Ser200 is shown in the two orientations represented in the electron density.*

In the WT structure, three loops were missing from the electron density and were thus unable to be modelled. These loops included residues Thr215-Arg219 in chain A, and loops Thr215-Ser220 and Asp133-Ser135 in chain B. From the N-terminal of the protein, 30 residues (taken from the cloned construct with hexa-His tag included) were not modelled in both chains; all residues were present at the C-terminal of the protein. In the V200S structure, all residues were defined in the electron density except for the 30 N-terminal residues. In addition to this, various side chains throughout both structures were missing electron density for the side chain of the residue. These were modelled in with zero occupancy.

In addition to DONM in the active site/s, electron density was also found to accommodate glycerol in the structures. The WT structure contained a total of three glycerol molecules, of which two are associated with chain A and one is associated with chain B. The V200S structure also contained three glycerol molecules, all associated with the one protein chain in the asymmetric unit. Glycerol was used in the crystallography procedure for both structures as the cryoprotectant during freezing.

Table 5.2: Refinement and model statistics for MalL WT and V200S with DONM bound

Refinement and model statistic	**MalL WT with DONM**	**MalL V200S with DONM**
R-factor	0.157	0.132
R_{free}	0.187	0.161
Total number of atoms	19,431	9,846
Total number of protein atoms	18,050	8,889
Other molecules	5	4
Number of waters	1,291	891
RMSD		
Bond lengths (Å)	0.1	0.009
Bond angles (°)	1.243	1.268
Average B-factors (Å^2)		
Protein monomer	15.6 (A) 20.1 (B)	11.3
Waters	27.5	22.8
Ramachandran analysis		
Percentage in favoured regions	98	98.4
Percentage in allowed regions	2	1.6
Percentage in disallowed regions	0	0

*Values in parentheses refer to the chain ID in the structure.

Final structure coordinates are available on the attached disk.

5.3 Analysis of MalL structures

5.3.1 Overall structure of MalL

The structure of MalL comprises three domains which are common to all members of the glycosidase hydrolase 13 family [(Yamamoto *et al.*, 2010); Figure 5.5 and Figure 5.6]. Domain A, the major globular portion of the protein, comprises a $(\beta/\alpha)_8$ TIM barrel. An extended sequence of 70 amino acids between the third β-strand

and α-helix of the TIM barrel forms domain B, comprising a three-strand antiparallel β-sheet, various loops and a short α-helical region positioned above domain A. The C-terminal end of the structure (domain C) forms two stacked antiparallel β-sheets in a double Greek key motif. Domain C lies to the side of domain A, parallel to α-helices four, five and six of the TIM barrel (Figure 5.6). The active site is positioned at the opening to the top of the TIM barrel (Figure 5.5). Catalytic residues (Asp199 and Glu255) are situated on C-terminal ends of β-strands four and five of domain A. A further conserved acidic residue (Glu332) among glycosidase hydrolase 13 family members (Janeček *et al.*, 1997) sits on the loop between β-strand and α-helix seven. The active site is enclosed by domain B, as well as an extended helix-loop-helix motif between β-strand eight and α-helix eight (herein called loop β_8-α_8). In general, loops to the top of domain A are longer than those at the base of the structure, creating the active site cleft.

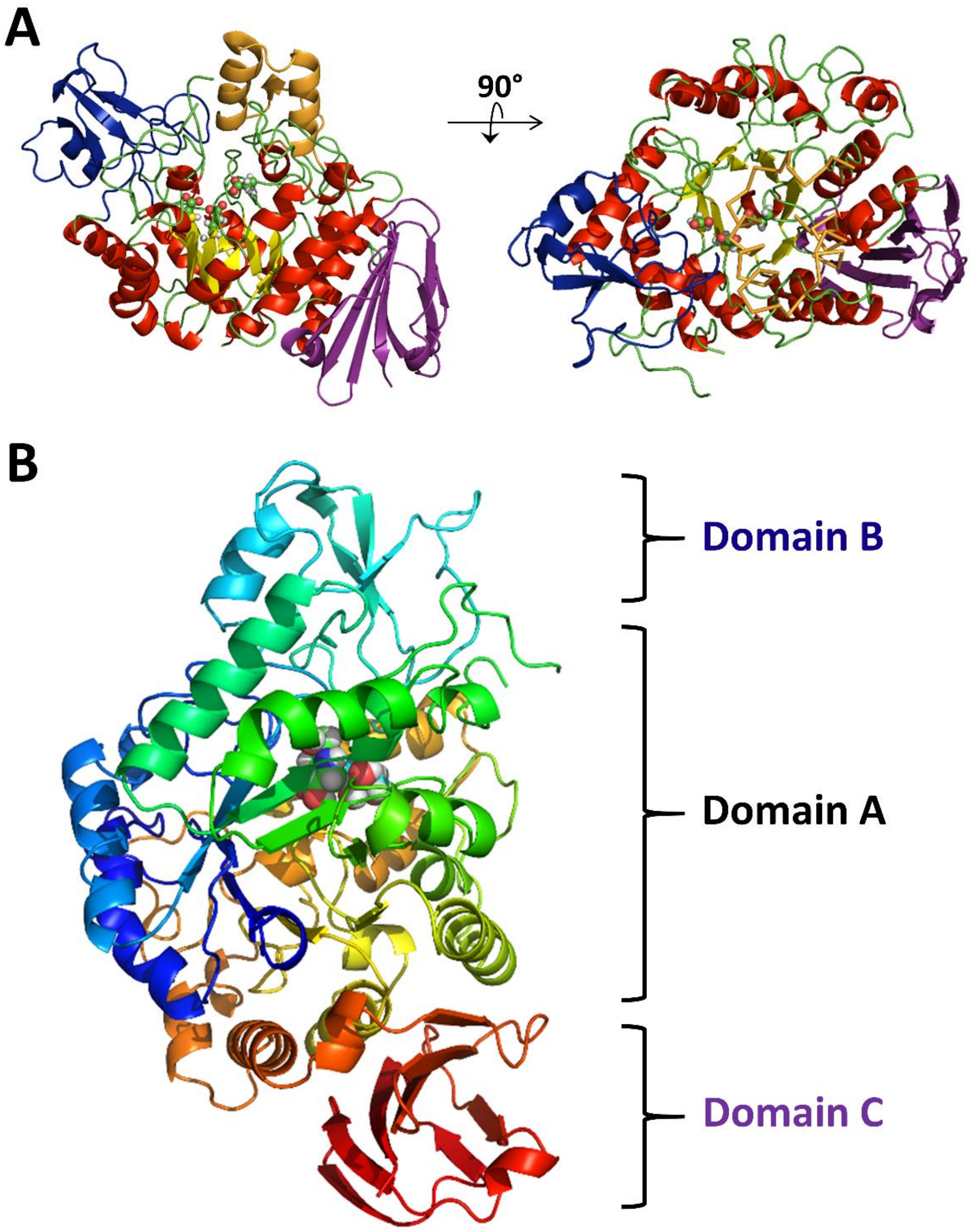


Figure 5.5: Cartoon structure of MalL. The position of the active site in all depictions is indicated by the three conserved catalytic residues, shown as spheres. ***(A)*** *Domains of MalL common to family 13 glycoside hydrolases. Domains A is coloured based on secondary structure; red α helices, yellow β sheets, and green loops. Domain B and C are blue and purple respectively. Loop β_8-α_8 of domain A is coloured orange – this constituent is shown as a backbone ribbon in the foreground of the rotated view.* ***(B)*** *MalL coloured rainbow (colouration red through blue from C to N-terminal) showing the active site (catalytic residues as spheres) between domain B and loop β_8-α_8 situated at the top of the TIM barrel comprising domain A.*

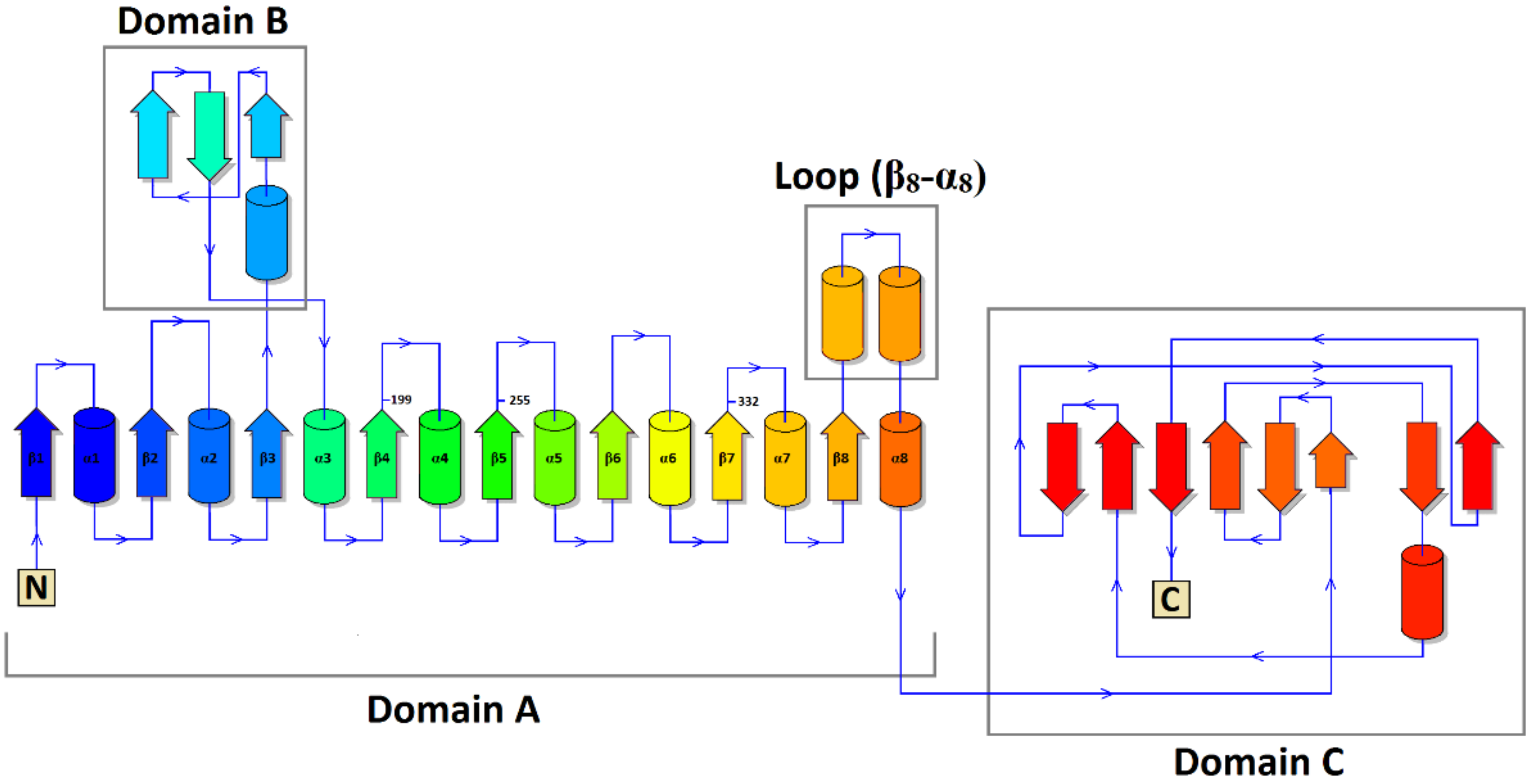


Figure 5.6: Topology structure of MalL. Sheet and helix colours correspond to colouration of the cartoon structure in Figure 5.5 B. A TIM barrel (domain A) comprises the bulk of the structure; elongated loops from the barrel form domain B and loop β_8-α_8, which provide the structures to enclose the active site. Domain C at the C-terminal of the protein chain sits to the side of the domain A, laying parallel to α-helices four, five and six of the TIM barrel. Figure created in PDBsum (Laskowski et al., 1997)

5.3.2 Structural comparison of apo MalL structures

MalL variants were first analysed by crystallography in Hobbs (2013) where apo structures of WT, V200S, G202P and V200A were presented. These structures are re-examined here, and presented along with the original findings to put the TS analogue bound structures in context.

Overall the apo variants of MalL were very similar, with RMSD values for Cα atoms from 0.18 to 0.21 Å between the WT and individual mutants (Hobbs *et al.*, 2013). The largest structural deviations occur between the WT and G202P structures. Mutation positions (positions 200 and 202) are located in the active site channel, around the interface between the -1 and +1 glucose binding sites (Figure 5.7 A).

Examination of the V200A mutant structure revealed obvious implications of the mutation. Val200 is positioned pointing into the active site cleft, not interacting with any other residues (Figure 5.7 B), thus it's substitution with a smaller, also non-polar side chain slightly alters the shape of the active site entrance without changing the chemistry. Given the limitations of analysing a static protein in a conformation constrained by crystallisation to gain information on dynamic processes, it is not unexpected that no obvious links between cause and effect are discernible.

Mutation of Val200 to serine introduces a polar group to the structure. As noted in Hobbs *et al.* (2013), Ser200 is rotated in the V200S structure, forming a hydrogen bond to the neighbouring Ser203. Ser203 is displaced by 0.8 Å (O-O distance) compared to the WT structure, pulled closer to residue 200 (Figure 5.7 B). In only the V200S structure, a water molecule is coordinating between Ser203 and the backbone carbonyl of His227. His227 is contained on the loop between α5 and β5 of domain A in which the main changes in flexibility were observed in molecular dynamics simulations [Section 1.4.2, Figure 1.6]. The additional hydrogen bond formed between the mutation and the base of this flexible loop explains the loss of flexibility (and reduction in C_p) in the V200S variant in the ground state (Section 1.4.2).

Due to the similar size and chemistry of threonine and serine, there is the potential for the V200T mutant to form the same hydrogen bond network and loop stabilisation as seen in V200S. This similarity is in line with the almost identical Michaelis Menten parameters of the two variants (Hobbs *et al.*, 2013). However due to the largely different $\Delta C^{\ddagger}_{p}$ behaviours of these two variants (-10.4 ± 0.6 and -5.9 ± 0.5 $kJ.mol^{-1}.K^{-1}$ for V200S and V200T respectively), there are likely differences in the significance of this loop stabilising interaction, and/or more complex factors affecting the two systems.

For the apo G202P mutant (Hobbs *et al.*, 2013), the introduction of the larger residue has forced Asn257 on a neighbouring loop to shift (Figure 5.7 B). Cα of Asn257 is displaced by 1.7 Å; the effect of this movement is carried over the two residues on to either side. The side chain itself is flipped at Cβ, orientated in the opposite direction to the other MalL variants. In the WT and V200 mutants, Asn257 hydrogen bonds with Ser203; in G202P this interaction is broken, and Asn257 is not within interaction distance with any other residues due to the orientation of the side chain into the active site cleft. In overlays with substrate bound MalL (PDB 3AXH), the displaced Asn257 clashes with a hydroxyl from the +1 glucose unit in the active site (Figure 5.11). This additional interaction may account for the low K_M of G202P compared to WT (92 and 205 mM respectively). The extra constraints introduced with the proline residue inclusion have also caused minor deviations of the backbone to the N-terminal side of the mutation position.

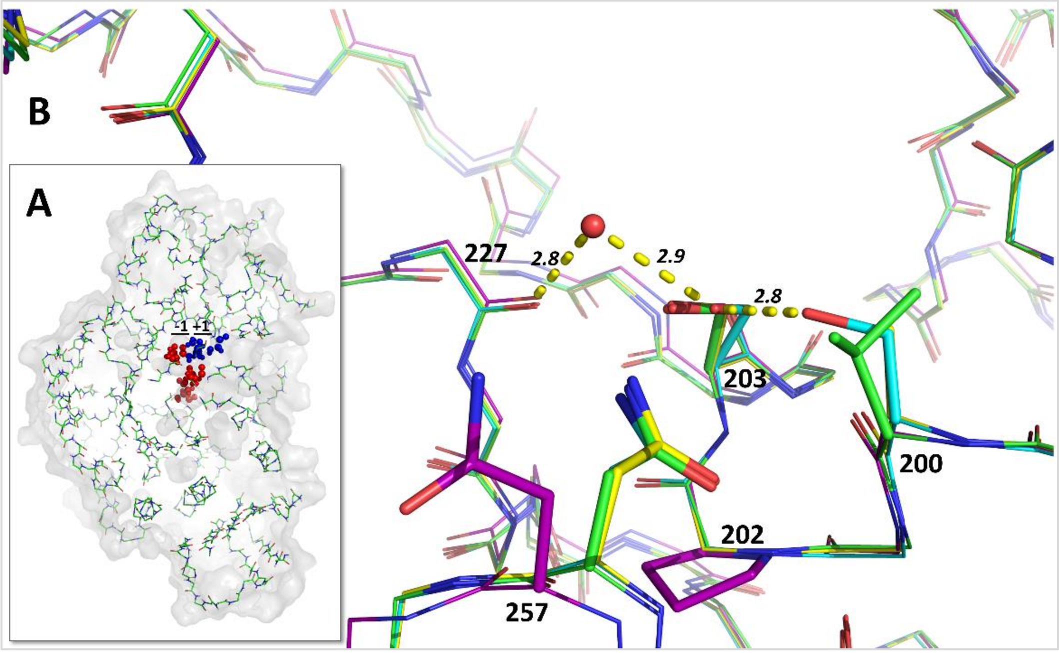


*Figure 5.7: **(A)** Position of mutations on the context of the full structure. The three conserved catalytic residues are shown in red, while the two mutation positions (200 and 202) are in blue. **(B)** Overlay of apo MalL WT, V200S, V200A and G202P. Carbons are coloured based on variant: WT green, V200S cyan, V200A yellow and G202P purple. The main chain is shown as the backbone only, while important residues side chains are illustrated. Adapted from (Hobbs et al., 2013).*

5.3.3 1-deoxynojirimycin (DONM) binding to MalL WT and V200S

5.3.3.1 General DONM binding site

In both the MalL WT and V200S structures, DONM binds in the -1 glucose binding pocket situated between domains A and B and loop β_8-α_8 (Figure 5.8). Standard binding site nomenclature (Davies *et al.*, 1997) is used here to describe the sugar units of isomaltose and binding sites in the protein. Ligand binding does not affect the overall structure of MalL; overlays between the apo and DONM bound structures of WT and V200S align with RMSD values of 0.16 and 0.099 Å respectively (where the WT RMSD value is an average of four alignments of the two molecules in the asymmetric unit for each variant). A lack of conformational change upon both substrate and competitive inhibitor (maltose) binding to isomaltase have previously been reported (Yamamoto *et al.*, 2011). This has also been noted as a general feature across the diverse members of the glycoside hydrolase 13 family (Kumar, 2010).

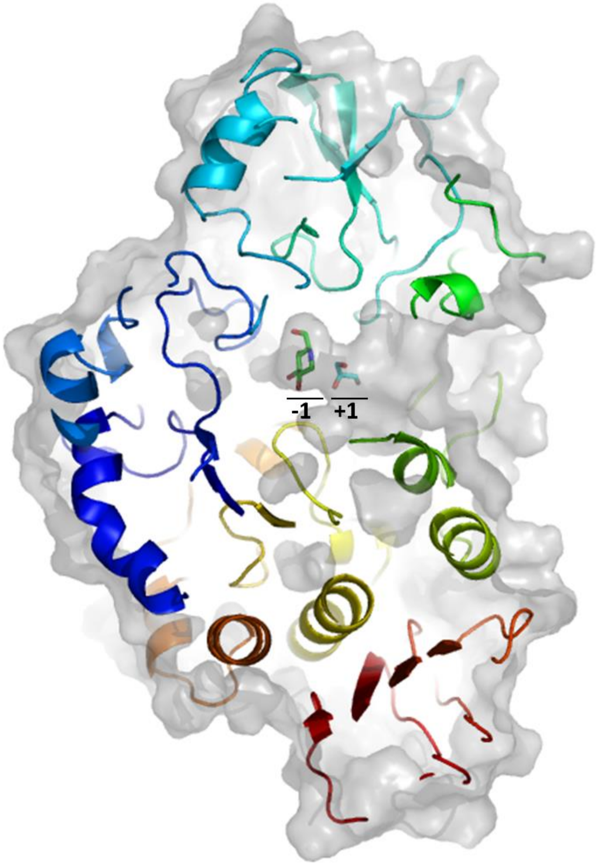


Figure 5.8: WT MalL with bound DONM in the -1 binding site. The +1 binding site is also indicated with a bound glycerol molecule. The binding pocket is visible as a channel extending into domain A. The structure is aligned in the same orientation as in Figure 5.5 B for standard nomenclature of the domains.

5.3.3.2 Specific binding interactions of DONM

DONM binds in a similar manner in MalL WT and V200S variants (Figure 5.9). The catalytic residue Asp199 forms a hydrogen bond to O_6H of DONM. This brings the residue 2.9 Å away from C_1 in both structures, in position for nucleophilic attack at the anomeric carbon of the substrate (Figure 5.9 B). O_2H is within interaction distance of Arg197, Asp332 and His331. His331 forms further interactions with the ligand via O_3H of DONM. A water is coordinated between O_3H and O_4H of DONM, and residues Asp59 and Arg422. The C_6-OH atoms of the ligand point towards the back of the binding site, hydrogen bonded to His102, in addition to the interaction with Asp199 (Figure 5.9). The back of the active site is composed of Tyr62; the rings of tyrosine and DONM stack parallel to one another approximately 4 Å apart, with the A face (Rose *et al.*, 1980) of the sugar ring facing the back of the active site. This ring stacking interaction between the A face of the sugar and a tyrosine residue in the -1 binding pocket is a highly preserved feature of glycoside hydrolase 13 family members, proposed as an important feature determining the binding and orientation of the substrate (Kumar, 2010).

The second catalytic residue, Glu255, sits on the opposite side of the DONM ring to Asp199, at a distance of 3.3 Å from the ligand (Figure 5.9). The two catalytic residues sit 5.2 and 5.4 Å apart in WT and V200S structures respectively, consistent with the mechanism of a retaining glucosidase (Wang *et al.*, 1994; Zechel & Withers, 2000). In both structures, Glu255 coordinates a catalytic water required for the second step in the reaction (Figure 5.1). This water is positioned between the anomeric carbon and Glu255 with respective distances of 3.5 and 2.7-2.6 Å in both structures (Figure 5.9 B).

A total of three acidic residues are conserved among glucoside hydrolase 13 family members; in addition to the catalytic nucleophile Asp199 and the general acid/base catalyst Glu255, Asp332 is also prevalent in sequence and structural alignments (Janeček *et al.*, 1997; Katsuya *et al.*, 1998; Yamamoto *et al.*, 2011). Asp332 assists the catalytic mechanism by increasing the pK_a of Glu255 via an interaction mediated though a water molecule (Yamamoto *et al.*, 2011). Thus, Asp332 is often included as a catalytic residue in the literature. Additionally, Asp332 is proposed to act in substrate recognition through the interactions with OH groups at C_2 and C_3 of the -1 sugar unit [Figure 5.9 B; (Hasegawa *et al.*, 1999)]. These interactions are also suggested to have a role in contorting the chair conformation slightly into a shape more commensurate for reaction, although the literature is still divided on this point (Hasegawa *et al.*, 1999). This disagreement stems from the fact that substrate/TS analogue distortion is not consistently seen in ligand bound structures of members of the glycoside hydrolase family 13. Regardless of the debate upon the universality and significance of this distortion, the DONM ring in both WT and V200S MalL structures is slightly distorted: in the plane formed by C_1-C_2-C_4-C_5, C_1 is lifted above, and C_2 dropped below the flat plane of a standard sugar chair conformation (not shown).

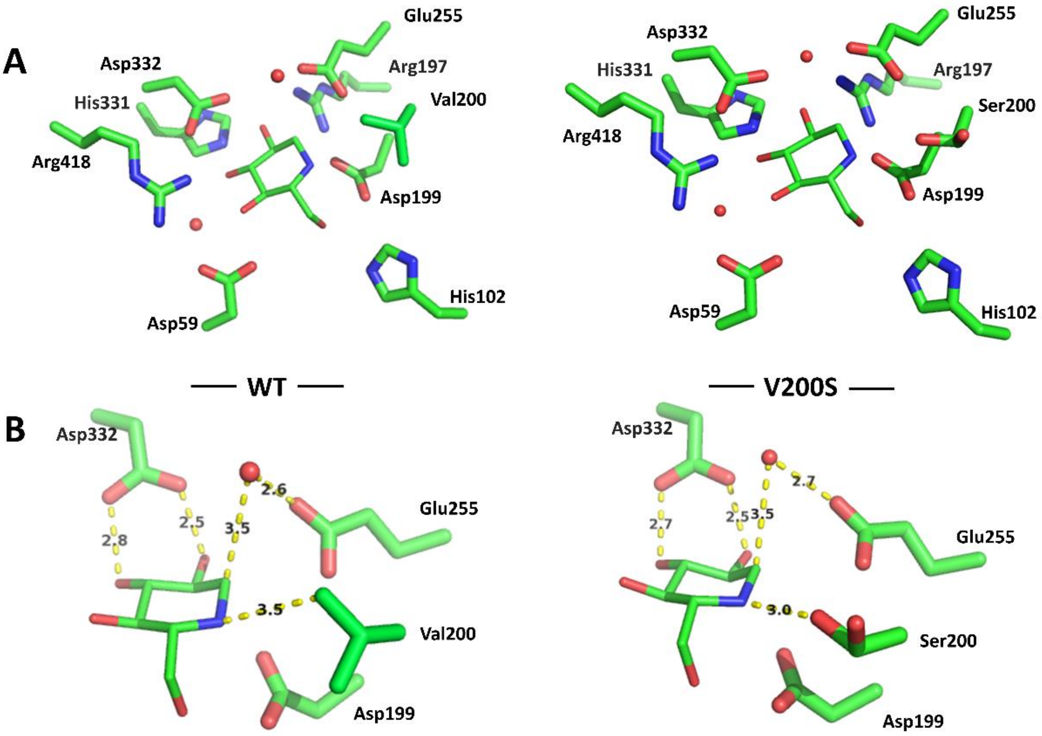


Figure 5.9: DONM binding in MalL WT and V200S structures. Ser200 in V200S is illustrated in the two alternate conformations represented in the electron density ***(A)*** *The main active site residues referred to in the text. The stacking residue Tyr62 forming the back of the active site is not visible as it sits below DONM in the depicted orientation.* ***(B)*** *Orientation of the catalytic nucleophile Asp199, general acid/base catalyst Glu255, conserved Asp332 and variant residue 200. The catalytic water coordinated by Glu255 is also illustrated. Distances between the catalytic residues (taken from the carbonyl C) Asp199 and Glu255 are 5.2 and 5.4 Å in WT and V200S respectively.*

In the WT structure, Val200 is situated 3.5 Å (CG1-N distance) from the ring nitrogen of DONM (Figure 5.9 B). In the V200S mutant, Ser200 comes into closer association with the DONM ring, forming a hydrogen bond of 3.0 Å between the nitrogen from DONM and the hydroxyl oxygen of serine in the closest of the two residue conformations in the structure. Despite the suggestion from these structures that the substitution of valine 200 for serine affords more TS stabilisation, k_{cat} for the two enzymes is similar (Section 1.4.1).. In a second orientation, the OH of Ser200 faces away from DONM in the binding pocket (Figure 5.9 B). In this second orientation, Ser200 makes the same hydrogen bonding network identified as the stabilising factor in the apo structures (Section 5.3.2). Ser200 interacts with Ser203 (3.2 Å O-O), Ser203 then coordinates a bridging water through to the backbone carbonyl of His227, as illustrated for the apo structure (Figure 5.7). This bridging water is not present in the WT structure. From the electron density map, occupation of these two orientations is roughly equal (Figure 5.4 B), with a slightly stronger electron density in the orientation interacting with DONM and the catalytic water (Figure 5.4). The bridging water between Ser203 and His227 also shows weaker electron density than coordinated waters in general in the structure. There are no steric effects introduced in V200S as Ser200 orientated toward the active site occupies a similar position to a methyl on Val200 in the WT structure.

In the original analysis of MalL mutants, G202P is concluded to have compromised binding to the TS, significantly affecting k_{cat} (Hobbs *et al.*, 2013). Due to the overall similarity between structures (Section 5.3.2), apo G202P was overlaid with WT to place DONM in the G202P active site. Despite the large differences in k_{cat} between the two variants (29 $\mathrm{s^{-1}}$ and 6 $\mathrm{min^{-1}}$), no significant implications of compromised TS binding are evident from the structures. Deviations in the backbone (RMSD of 0.23 Å for Cα atoms) in G202P do not translate to altered positioning of active site residues. Other residues in the active site pocket are in comparable positions between the two structures (Figure 5.10).

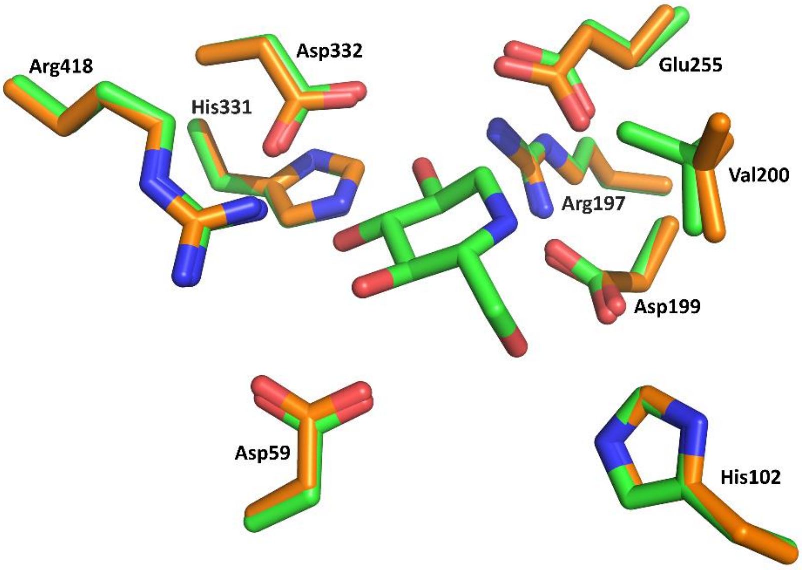


Figure 5.10: Active site overlay between MalL WT with DONM (green) and apo G202P (PDB 4MB1; orange). Little deviation in active site residues is evident to support arguments for the compromised catalytic ability of G202P, including in the catalytic residues (Asp199 and Glu255). For example, Glu255 is displaced by 0.4 Å at the carbonyl carbon. The only notable difference is the rotation of Val200. In the closest orientation to the ligand (WT), the shortest interaction spans 3.5 Å.

5.3.4 The +1 glucose binding site

Isomaltase (active site mutant E277A) from *Saccharomyces cerevisiae* has previously been solved with bound isomaltose (Yamamoto *et al.*, 2011). Binding at the +1 glucose site is weaker than at the -1 binding site, and is dominated by hydrophobic interactions. Comparison of this to the MalL structures shows the +1 glucose binding sites are largely similar (Figure 5.11 A). Of the five residues identified as important in the +1 glucose site in Yamamoto (2011), four are conserved in MalL [Phe144, Phe163, Phe281, and the hydrogen bond acceptor Glu389 (MalL sequence numbering)]. Only small deviations of these residues between the two structures are present. Most notably, Glu389 is further back from the binding pocket in the MalL structure, and Phe281 is rotated by roughly 70° (Figure 5.11 A). The sixth identified residue in the *S. cerevisiae* binding site, Tyr143, lies slightly displaced from parallel to the B face of the sugar ring. This interaction is reminiscent of the conserved stacking interaction described earlier for the -1 glucose binding site of MalL (Section 5.3.3.2). Previously, the presence of a stacking tyrosine in the +1 binding site has been noted to be randomly distributed over the glycoside hydrolase family 13 members (Kumar, 2010). In the MalL structure, this residue is Ile143, a significantly smaller residue. The space taken by tyrosine in the *S. cerevisiae* structure remains empty as part of the active site channel in the *Bacillus subtilis* MalL structure.

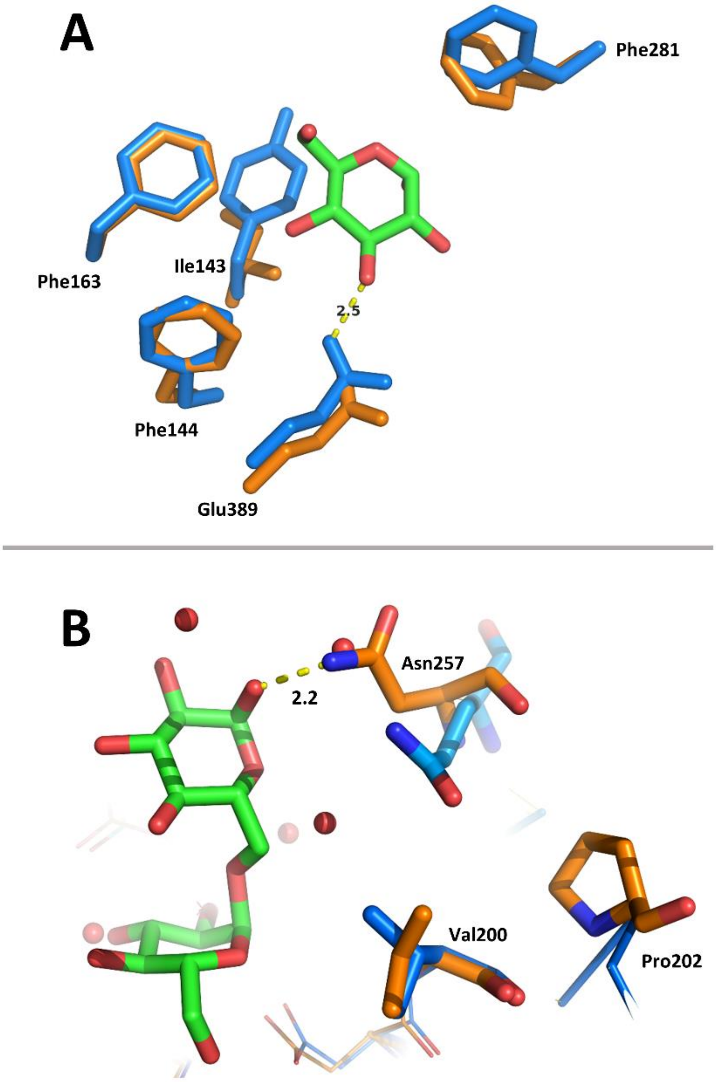


Figure 5.11: Alignment of isomaltose binding site of WT and G202P MalL based on an overlay of substrate bound E277A mutant from Saccharomyces cerevisiae (PDB code 3AXH). Residue numbering refers to the B. subtilis MalL sequence. ***(A)*** *Glucose in the +1 binding site of WT MalL. The WT structure is shown in orange, while 3AXH is in blue. Isomaltose from the 3AXH coordinates is coloured by atom type (green carbon, red oxygen); the -1 glucose, in the forefront of the +1 glucose is not shown.* ***(B)*** *The effect of G202P mutation on the interaction of Asn257 with substrate. Steric hindrance by Pro202 forces Asn257 within interference distance of the substrate binding site, as positioned by overlay with substrate bound MalL from S. cerevisiae.*

Yamamoto (2011) identified residues important to the specificity of isomaltose over maltose. These residues are a glutamine and valine which reside in the active site tunnel to the side of the +1 binding site. These two residues sit in a position where they clash with maltose if it is superimposed in the structure in a natural orientation with the -1 glucose correctly positioned in the active site. These residues are also present in the *B. subtilis* MalL structure in the same position (residue numbers

Asn257 and Val200 in MalL; not shown). Valine is present in both sequences, while asparagine is present in the MalL structure in place of glutamine, pulling the amide group back from the binding site by 1.9 Å.

In the MalL mutants studied (Hobbs *et al.*, 2013), the Val200 identified as important for substrate selectivity in (Yamamoto *et al.*, 2011) is altered to alanine, serine, and threonine. These changes have had little impact on the function of the enzymes, evident in the maintained low K_M and high k_{cat} values (Hobbs *et al.*, 2013). From overlays of these structures (not shown), these small residues occupy the same position in the structure, maintaining the active site in a shape conducive to the binding of isomaltose. In the G202P mutant, the second residue important for substrate identification (Asn257) is shifted in the apo structure (Section 5.3.2). This shift moves Asn257 further into the active site, into the substrate binding pocket (Figure 5.11 B). From kinetic analysis (Hobbs *et al.*, 2013), G202P has improved K_M and greatly compromised k_{cat} compared to WT MalL; this altered interaction with Asn257 may influence these parameters, and justifies further investigation.

5.3.4.1 Implications of binding in the +1 binding pocket

MalL WT bound with DONM was solved with two molecules in the asymmetric unit. The two chains are similar in conformation, with an RMSD of 0.15 Å. Both monomers in the unit cell contain DONM in the active site, however chain A also contains a glycerol molecule in the active site in the +1 binding pocket (Figure 5.8 and Figure 5.12). The single molecule in the asymmetric unit of V200S also contains a glycerol molecule in the +1 binding site. From overlays of MalL WT with the *S. cerevisiae* substrate bound structure (PDB code 3AXH), the three-carbon long glycerol carbon backbone sits slightly displaced (roughly 1 Å) from where one half of the sugar ring sits in the +1 binding site (not shown). WT Chain B, lacking a glycerol molecule, coordinates two water molecules linking between Arg418 and Glu389 in the positions where oxygens from the hydroxyls groups sit in chain A. The DONM ring sits in an identical position in the active sites of both chain A and B. The side chains surrounding the TS analogue are also in identical positions, and despite the additional ligand in the A chain, residues surrounding the +1 glucose site are in the same positions (Figure 5.12).

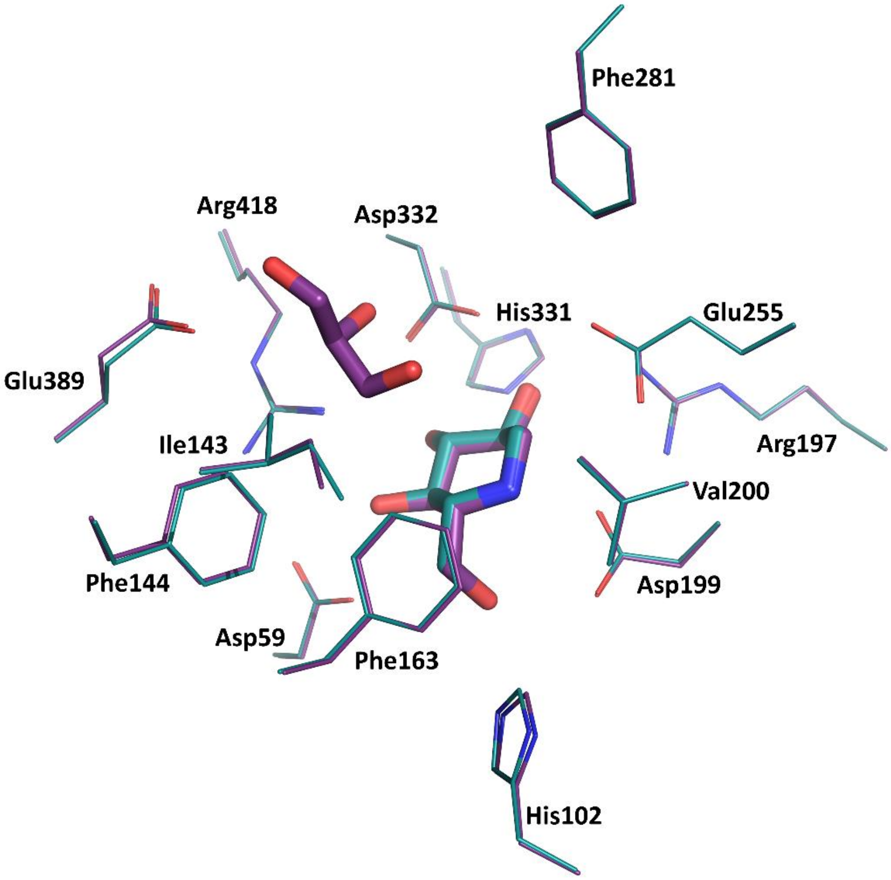


Figure 5.12: Active site overlay between WT MalL with DONM chains A (purple) and B (turquoise). Additional binding of glycerol in the -1 glucose binding pocket does not influence configuration in the residues surrounding the active site.

The additional binding of the glycerol as a proxy for the +1 glucose unit (as either the substrate or the TS of step one) appears to have rigidified the MalL WT structure (Figure 5.13). Average B-factors for the two chains are 15.6 for chain A and 20.1 for chain B. Different contacts between the two chains as a result of crystal packing can cause difference B-factors, however, these are similar over the two chains in this case, although not conclusively ruled out as a causative effect. Although lower B-factors are spread over the whole structure, significant effects are evident in exterior loops, especially those surrounding the active site. This is possibly indicative of the effects constituent binding in the +1 site has on the decreases in flexibility, and thus C_p, of the structure. Binding in the +1 glucose site is a closer approximation of the first reaction step TS (Figure 5.1) where the second glucose unit is still held in the active site. Given that the first step likely has a greater $\Delta C^{\ddagger}_p$ change (greater binding interactions), and also considering that this step is rate limiting (Zechel and Withers, 2000; Pudney, unpublished), this step and associated

TS are likely the predominant considerations when considering MalL in the context of MMRT.

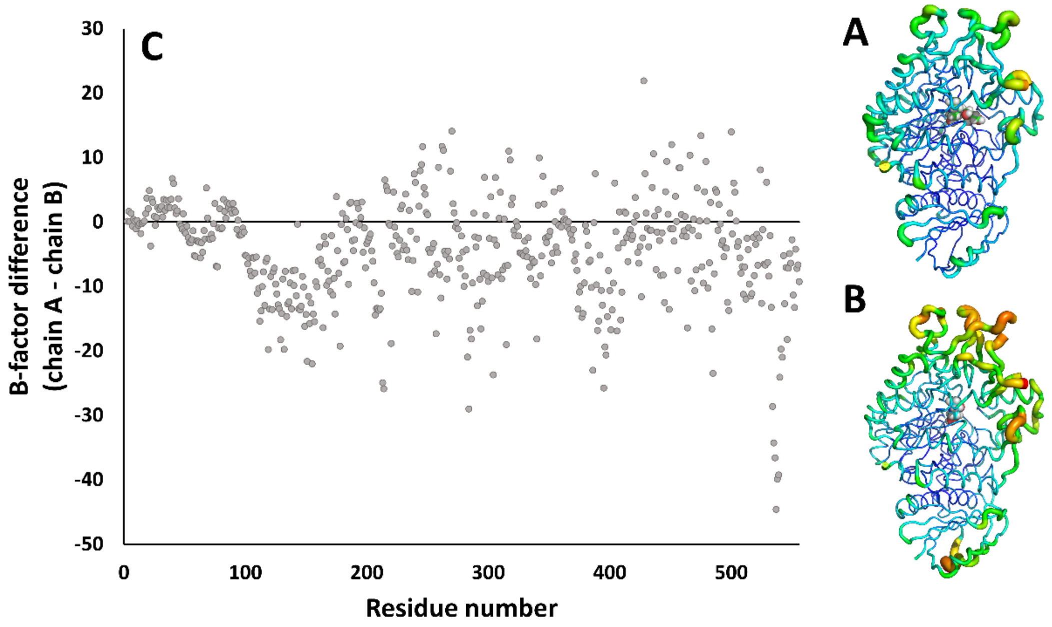


Figure 5.13: B-factor difference between chains A and B of MalL WT structure with DONM bound. These two structures differ in the presence of a glycerol molecule bound in the +1 binding site of chain A. ***(A)*** *and* ***(B)*** *B-factor putty structures of chain A and B respectively representing the differences between the two chains in the asymmetric unit. The position of the active site is indicated by the ligands visualised as spheres.* ***(C)*** *Plotted difference in B-factor between the two chains. Where the difference is negative, chain A is more rigid than chain B. This change is proposed to be due to the additional molecule binding in the +1 binding pocket as a proxy for the second glucose unit in either the Michaelis complex, or TS for the first step of the reaction.*

5.4 Conclusions from MalL structures

To date, a range of mutants identified with altered $\Delta C^{\ddagger}_{\mathrm{p}}$, and different states along the reaction coordinate have been analysed for MalL in the context of MMRT. Hobbs *et al.* (2013) first analysed the apo 3-dimensional structures of MalL variants WT, V200S, V200A and G202P. Here, this analysis has been extended to include TS analogue bound forms of WT and V200S MalL to 1.6 Å to further analyse the structural differences leading to the $\Delta C^{\ddagger}_{\mathrm{p}}$ difference between the two enzymes. This set of structures has also allowed analysis of the substrate bound MalL form through overlays with previously published structures.

Despite this wealth of crystallographic information, there is no definitive explanation evident to describe the differences in $\Delta C^{\ddagger}_{\mathrm{p}}$ between WT and V200S, or the other MalL variants. Given the close link between protein motion and C_p, the absence of definitive information from static structures is not surprising. The most significant finding so far is the extended hydrogen bonding network from Ser200 in V200S to a loop encapsulating the active site that is not present in WT MalL (Hobbs *et al.*, 2013). However, this idea only holds weight due to the agreement between the structures and associated molecular dynamics performed on the systems (Section 1.4.2). Further consideration of this is also required to account for differences between the apo enzyme and enzyme-substrate as the ground state. Given that Ser200 occupies two conformations in the TS-analogue bound structure, substrate binding may also induce this change, affecting the hydrogen bridging network. Further investigation is required to assess if, and the effect of, the possible Ser200 orientation differences in the substrate bound state.

The extended analysis undertaken here also gives insight into the potential importance of binding in the +1 glucose site to MalL stabilisation and thus C_p over the reaction mechanism. This finding is consistent with the first step of the reaction being rate limiting (Zechel & Withers, 2000), and also the findings that pseudo-disaccharide TS analogues bind better than their monosaccharide counterparts (Dong *et al.*, 1996).

In silico $\Delta C^{\ddagger}_{p}$ determination

Enzymatic C_p is inherently linked to protein dynamics. It is the combination of a multitude of low frequency vibrational modes from the unique size and structure of proteins which allows the potential to absorb energy, resulting in the unusually high heat capacity of proteins (Makhatadze, 1998; Arcus *et al.*, 2016). Molecular dynamics (MD) offers a means to access atomic level details of these vibrational motions. This has previously been applied in MMRT to elucidate the dynamic differences between variants of the model enzyme, MalL (Section 1.4.2).

There is a relationship between C_p and enthalpy [H; equation **17 a and b** (Prabhu & Sharp, 2005)] that suggests that $\Delta C^{\ddagger}_{p}$ might be calculated directly from *in silico* MD data. Enthalpy (H) is normally distributed (Figure 6.1), with a variance defined by equation **(17a)**. Definition of the variation of enthalpy values for a system in terms of heat capacity allows the rearrangement to calculate $\Delta C^{\ddagger}_{p}$ from the differences in $<\delta H^2>$ between two systems (equation **17b**). To define $\Delta C^{\ddagger}_{p}$ for enzymatic catalysis, the difference of the variance in enthalpy between the ground state (enzyme-substrate) and TS (enzyme-TS) need be considered.

$$< \delta H^2 > = \frac{-k_B}{\left(\frac{\delta^2 S}{\delta U^2}\right)_p} = k_B T^2 C_p \text{ .... } \textit{Variance in H distribution } \textbf{(17a)}$$

$$\Delta C_p = \frac{\Delta < \delta H^2 >}{k_B T^2} \text{........................ } \textit{Heat capacity as a function of enthalpy } \textbf{(17b)}$$

MMRT predicts a narrowing in the enthalpy distribution at the TS, reducing enthalpy variation and lowering C_p (Figure 6.1). This narrowing compared to the ground state results in a negative $\Delta C^{\ddagger}_{p}$, and curvature in $\Delta G^{\ddagger}$ and reaction rates with temperature (Section 1.1). Given the enthalpic information in MD data, in principles C_p information can be extracted from *in silico* data based on examination of the enthalpy distribution for each state (equation **17b**). To date in the literature, there is no published account of heat capacity information being determined for a protein system via this route.

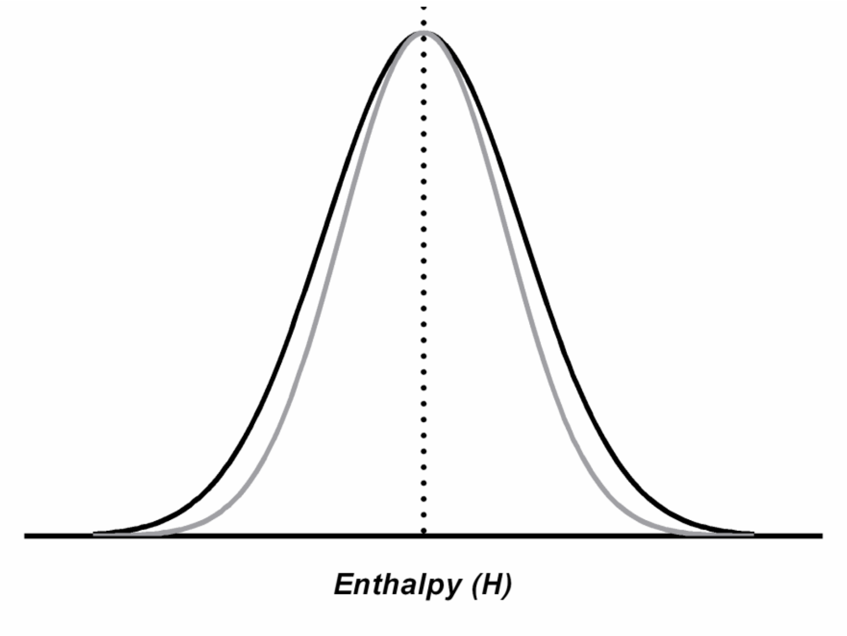


Figure 6.1: Narrowing of the enthalpy distribution at the TS (grey) as compared to the ground state (black). A reduced enthalpy distribution is associated with a reduced C_p (equation ***17b****). The mean enthalpy is indicated as a dashed line.*

Here, methods for determining an *in silico* $\Delta C^{\ddagger}_{\mathrm{p}}$ are explored in reference to experimental data. The major goal of this research was to determine if there is enough information in dynamics data to detect a meaningful enthalpy distribution and thus calculate C_{p} differences between substrate and TS bound enzyme systems. Further, would this calculation yield differences that are in agreement with experimental data. In collaboration with M. van der Kamp and A. Mulholland (Bristol University, UK), two systems were tested; KSI (Δ^5-3-keto steroid isomerase; EC 5.3.3.1) and the MalL variants WT and V200S. KSI, as a dimer of 14.9 kDa units, represents a small protein system, upon which extensive MD work has been performed previously (van der Kamp *et al.*, 2013). MalL, as a large protein of 69.5 kDa, presents a more challenging system to verify methods against a greater volume of experimental data. The V200S MalL mutant is also considered to further elucidate details of the changes in $\Delta C^{\ddagger}_{\mathrm{p}}$ in this system, and the sensitivity of methods for simulating $\Delta C^{\ddagger}_{\mathrm{p}}$ changes resultant from single amino acid differences. Data presented here largely comprise the MalL system, as analysed following methods developed for KSI in van der Kamp (unpublished). Finally, comparisons are drawn between the two systems to address the applicability of methods to different protein systems.

6.1 MalL structure preparation and set up

MD simulations were based on the crystallographic structures of MalL WT and V200S with TS analogue (DONM) bound (Section 5). From the final structures, loops missing in the crystallographic data were built using COOT (Emsley & Cowtan, 2004). If present, electron density too poor for absolute atomic placement was used to guide loop building; optimisation of ideal bond parameters was heavily relied on to place loops in a realistic orientation. This included loop Thr215-Arg219 in the WT structure (chain A), and two N-terminal residues in both structures. Where alternate residue conformations were present in the crystallographic data, the lesser occupied conformation was removed from the structure based on the intensity of electron density. The one exception to this was Ser200 in V200S. In the TS analogue bound structure, Ser200 occupies two orientations with equal electron density support (Section 5.3.3.2). These orientations align towards and away from the anomeric position of the ligand. In the apo structure, Ser200 solely orientates away from the active site. Due to the importance of this residue and potential for different orientation occupancies along the reaction coordinate, two structures were set up representing the two Ser200 orientations present in the crystallographic data. Additional molecules crystallised with the proteins (Table 5.2) were also removed from the structures.

To gain the ground state representation of MalL, substrate (isomaltose) was orientated in the active site through overlay with the E277A MalL structure from *Saccharomyces cerevisiae* (Yamamoto *et al.*, 2011). Alignment between atom positioning in DONM and the overlaid substrate was checked in the -1 glucose binding pocket to confirm correct substrate positioning. Crystallographic waters were not removed from the structures unless clashing with atoms of the ligand positioned in the active site. Building of the disaccharide TS analogue (Figure 6.2 C) was based on isomaltose; additional components were built in using PyMOL builder (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC). Further details of the TS analogue are given in Section 6.1.1.

PROPKA 3.1 (Olsson *et al.*, 2011; Søndergaard *et al.*, 2011) was used to determine the protonation state of protein residues with ligand in place; suggested protonation states were manually checked in the structure. Hydrogen bonding networks and side chain tautomer alterations were checked using the WHAT IF optimal hydrogen bonding network tool (Vriend, 1990); no side chain orientations were found to require alteration to optimise hydrogen bonding networks in the MalL structures. Charge distribution in the ligands were determined with the R.E.D Server (Vanquelef *et al.*, 2011).

MalL simulations used the General Amber Force Field [GAFF; (Wang *et al.*, 2004)] for simulation of the protein component of the system. For the ground state, substrate [isomaltose; D-Glc-α(1-6)-D-Glc] was modelled with the GLYCAM-06 force field specifically developed for carbohydrates (Kirschner *et al.*, 2008). For consistency, the TS analogue was also modelled with GLYCAM where possible. This was combined with GAFF where limitations in the GLYCAM pre-set sugar units did not allow these parameters to be applied to the whole TS molecule. Further details of the structure and modelling of the TS analogue are presented in Section 6.1.1. The system was solvated with TIP4P-Ew water molecules (Horn *et al.*, 2004) in a truncated octahedral periodic box, so that no atom of the protein was less than 11 Å from its edge., Water molecules were placed no less than 0.9 Å from the protein atoms. Charges were balanced with the addition of Na^+ ions.

To relax unfavourable contacts, systems were minimised by gradually reducing constraints on the system. Solvent and hydrogens were minimised followed by heating from 50 to 300 K. The system was minimised and heated (25-320 K) with constraints on Cα positions. Restrictions at Cα positions were reduced, followed by one ns of unrestrained simulation to equilibrate the system.

6.1.1 MalL TS development

To simulate the first step of the two-step hydrolysis reaction catalysed by MalL (Figure 5.1), a TS analogue was used to mimic the charge and shape distribution of the implied TS (Figure 6.2). A TS analogue was modelled based on the concepts implemented in the development of the 40 nM binder MT-DADMe-ImmA for methylthioadenosine phosphorylase (Guan *et al.*, 2013). MT-DADMe-ImmA imitates the charge distribution via a protonated nitrogen (as in DONM), and mimics the elongated breaking bond between the two substrate units by separating the two units with a methylene group. The ring nitrogen was moved (as compared to positioning in DONM; Figure 5.2) to the anomeric position which carries more of the delocalised positive charge at the TS, and is experimentally validated by the tighter binding of isofagomine to isomaltose. DONM and isofagomine are based on the same chemical skeleton except for the positioning of the ring nitrogen; positive charge localisation at the anomeric centre in isofagomine and noeuromycin decreases K_i from 11 μM (DONM) to 7.2 and 0.025 μM respectively in isomaltase from yeast (Liu *et al.*, 2001).

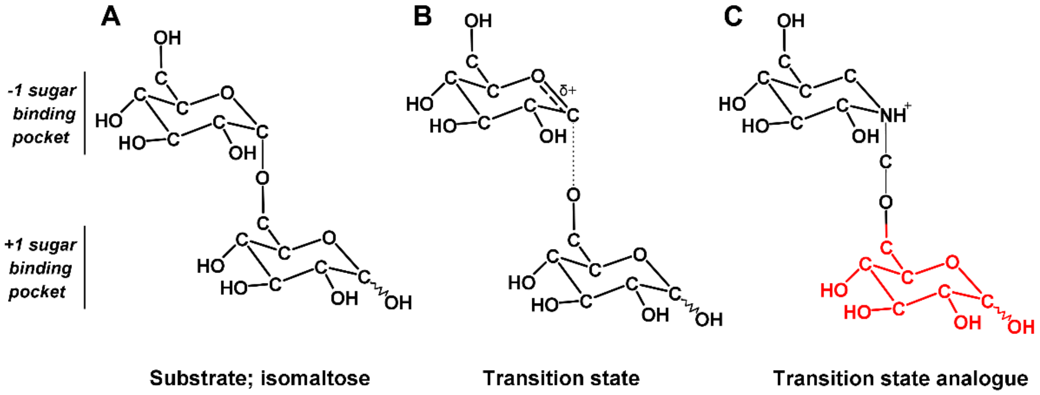


Figure 6.2: The TS analogue ***(C)*** *used in MD simulations for MalL, compared to the inferred TS* ***(B)*** *for step one of the hydrolysis reaction, and the substrate* ***(A)****. The TS analogue is coloured based on the force fields used in the MD (red GLYCAM, black GAFF). Nitrogen is placed in the TS analogue sugar ring to mimic the delocalised partial positive charge primarily located at the anomeric carbon in the TS, while a methylene group is added to mimic the bond elongation between the two sugar units.*

6.1.1.1 Force field details and validation for the TS analogue

The GLYCAM force field has been specifically developed for carbohydrates with parameters derived from quantum mechanical simulations of a range of small molecules (Kirschner *et al.*, 2008). The GLYCAM force field was used in substrate simulations, however, due to the addition of nitrogen to the -1 glucose ring and extended linker between the two units, GLYCAM was unsuitable for modelling of the TS analogue (Figure 6.2 C). GLYCAM was used for the +1 glucose of the TS analogue for consistency with the substrate. The rest of the molecule was parameterised with GAFF.

To validate the consistency in force fields between the two systems, simulations of glucose were run with GAFF and GLYCAM. These were run in a protein environment (i.e. modelled in the MalL -1 glucose binding site) to compare the two force fields in the final simulation environment. Two simulations for each force field were run (as described in Sections 6.1 and 6.2) and analysed over 60 ns. Overall, GAFF replicated bond lengths, angles, and dihedrals of glucose, as compared to GLYCAM, as well as variation about the mean value (Figure 6.3 A and B). In many cases, the difference between replicates of the same force field was greater than the difference between the two force fields (Figure 6.3 C). From this, it was concluded that GAFF was sufficient to model a portion of the TS analogue, with results directly comparable to the substrate modelled entirely by GLYCAM.

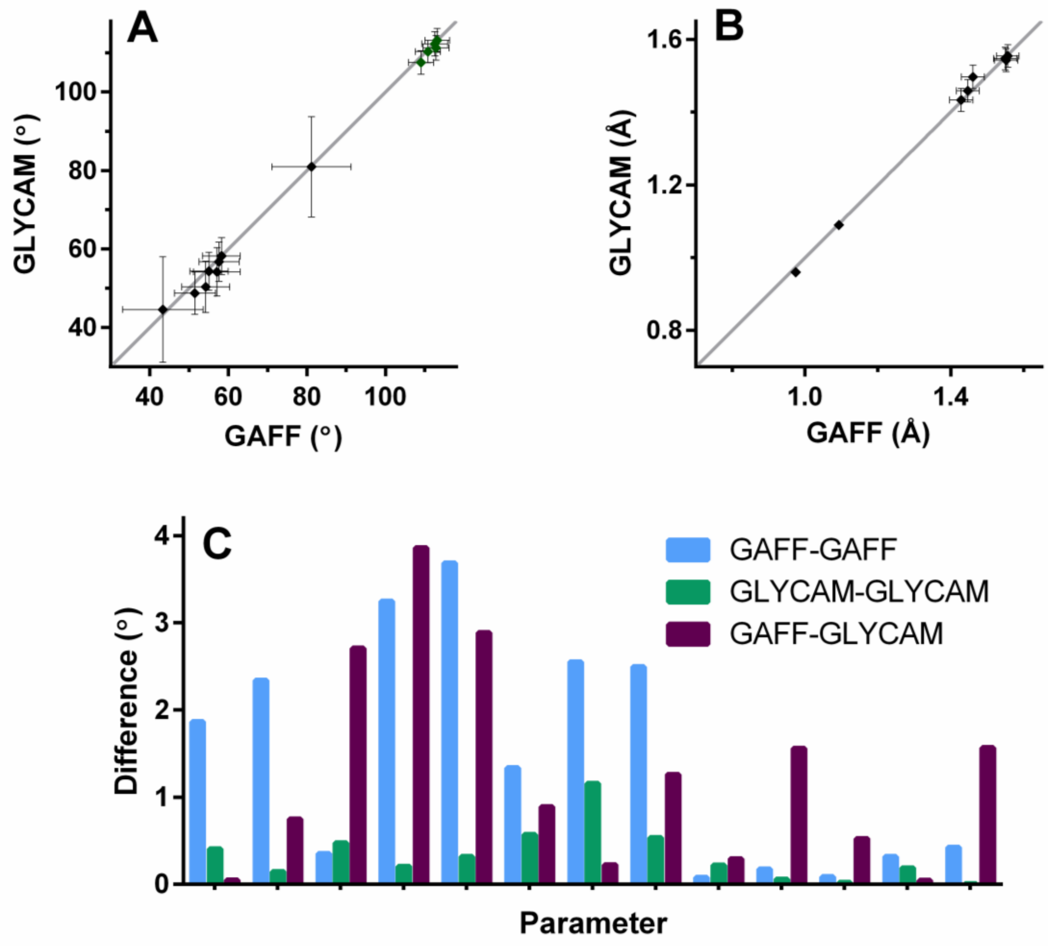


Figure 6.3: Comparison of GAFF and GLYCAM force fields for modelling glucose parameters over 60 ns simulation time. ***(A)*** *Comparison of mean values of dihedrals (black) and angles (green) as modelled by GAFF and GLYCAM. Error bars represent SD over the simulations.* ***(B)*** *Comparison of mean bond lengths. Error bars represent SD over the simulations. An ideal relationship between the two force fields is plotted in grey in* ***(A)*** *and* ***(B)****.* ***(C)*** *Differences in mean angles and dihedrals within two replicate GAFF and GLYCAM runs, and between the two force fields. Data, including the parameters sampled, are given in Appendix D.5.*

6.2 Dynamics simulation for MalL WT

Dynamics simulations were run for 500 ns at 320 K in steps of $2x10^{-6}$ ns. Initial simulation with substrate resulted in loss of ligand from the active site; final simulations included constraints between the substrate (centre of mass of the glucose unit in the -1 binding site) and active site residues in the binding pocket.

6.2.1 RMSD and conformational sampling of MalL WT

RMSD plots across both substrate and TS analogue bound runs show an initial increase in RMSD consistently over the first 50 ns of simulation time (Figure 6.4 and Figure 6.5 respectively). From this point, simulations trace a variety of RMSD trajectories; overall, simulations trend away from the starting structure. Examination of individual runs shows an opening about the MalL active site as loops surrounding the active site move apart. This is observed in simulations of both the substrate and TS analogue bound state (Figure 6.6 and Figure 6.7 respectively). Movement in loops 213-221, 387-417, and 287-302 away from the active site creates a cleft between domains A and B, creating an opened structure associated with larger RMSD.

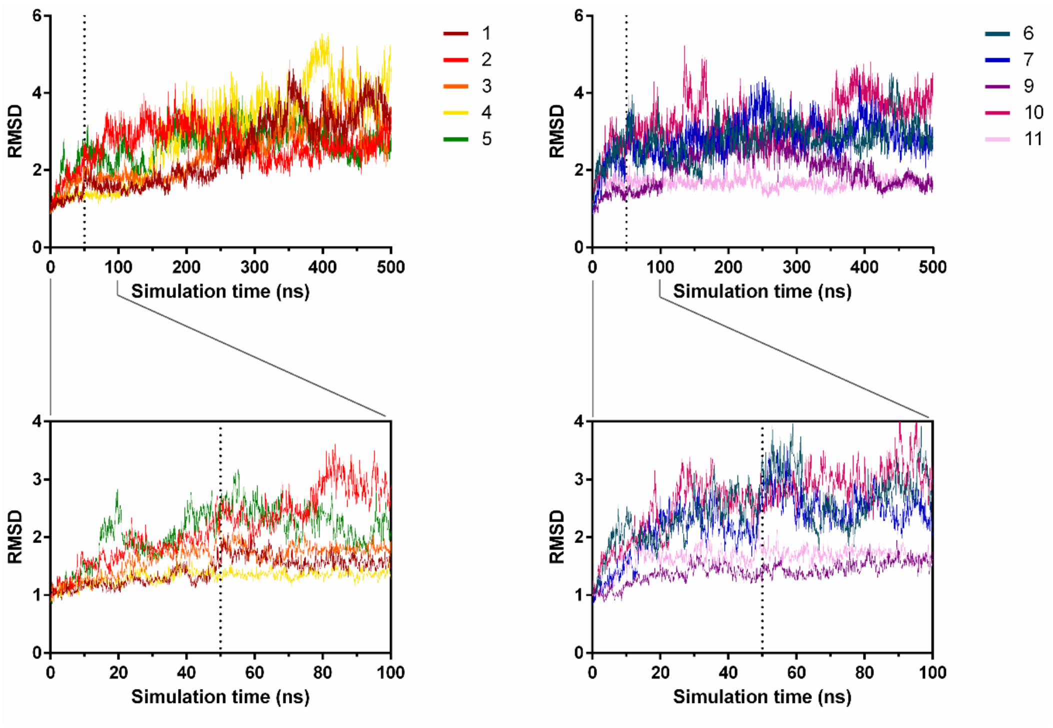


Figure 6.4: RMSD of ten replicate simulations for MalL WT with substrate. The initial 50 ns over which rapid changes in RMSD occur is indicated (vertical line). Run 8 (not shown) was replaced in the final analysis with run 11 (Section 6.2.3).

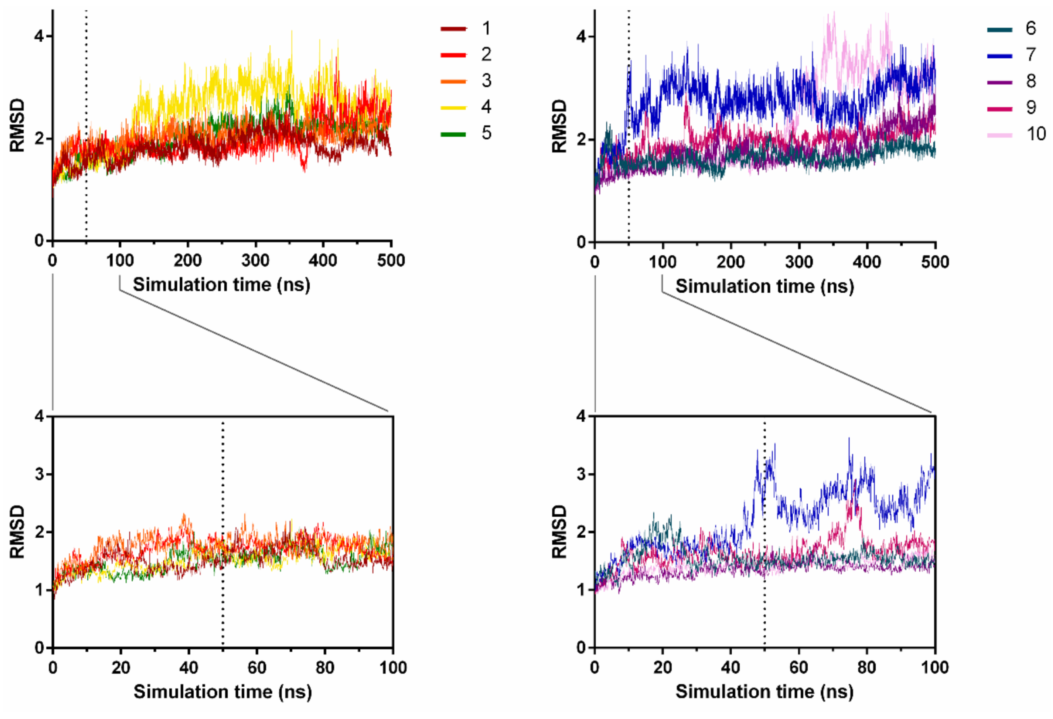


Figure 6.5: RMSD over ten replicate runs for MalL WT with TS analogue. The initial 50 ns which encompasses rapid changes in RMSD occur is indicated (vertical line).

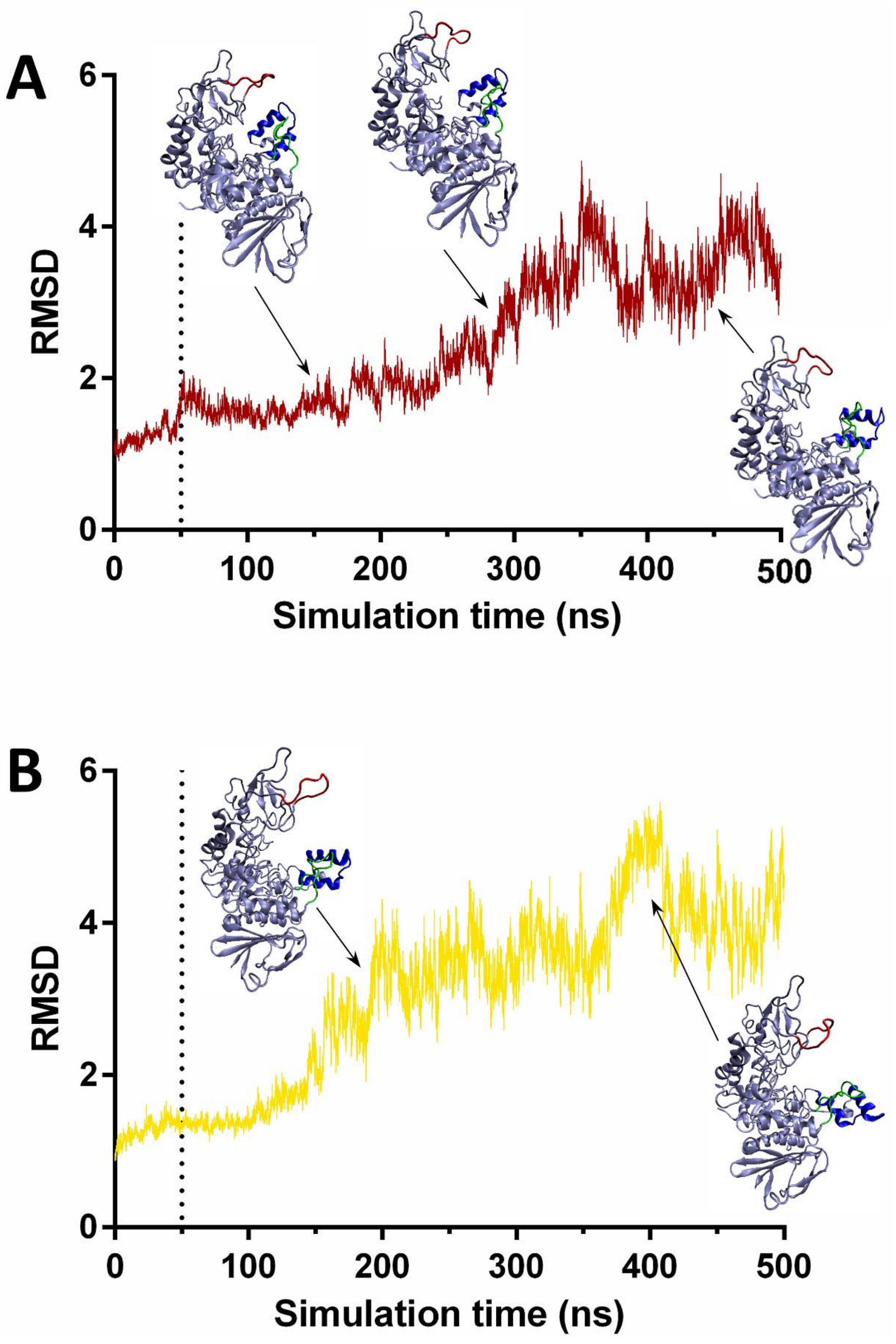


Figure 6.6: Representative WT with substrate bound simulations illustrating the active site opening for run one (maroon) and four (yellow). Loops contributing to the opening movement are indicated in red (residues 213-221), blue (387-417), and green (287-302). ***(A)*** *gradual increases in RMSD up to 350 ns are accompanied by active site opening; the largest changes and stabilisation occur with the large opening of the helix-loop-helix (blue).* ***(B)*** *Active site opening accompanies RMSD increases up to a plateau at 400 ns.*

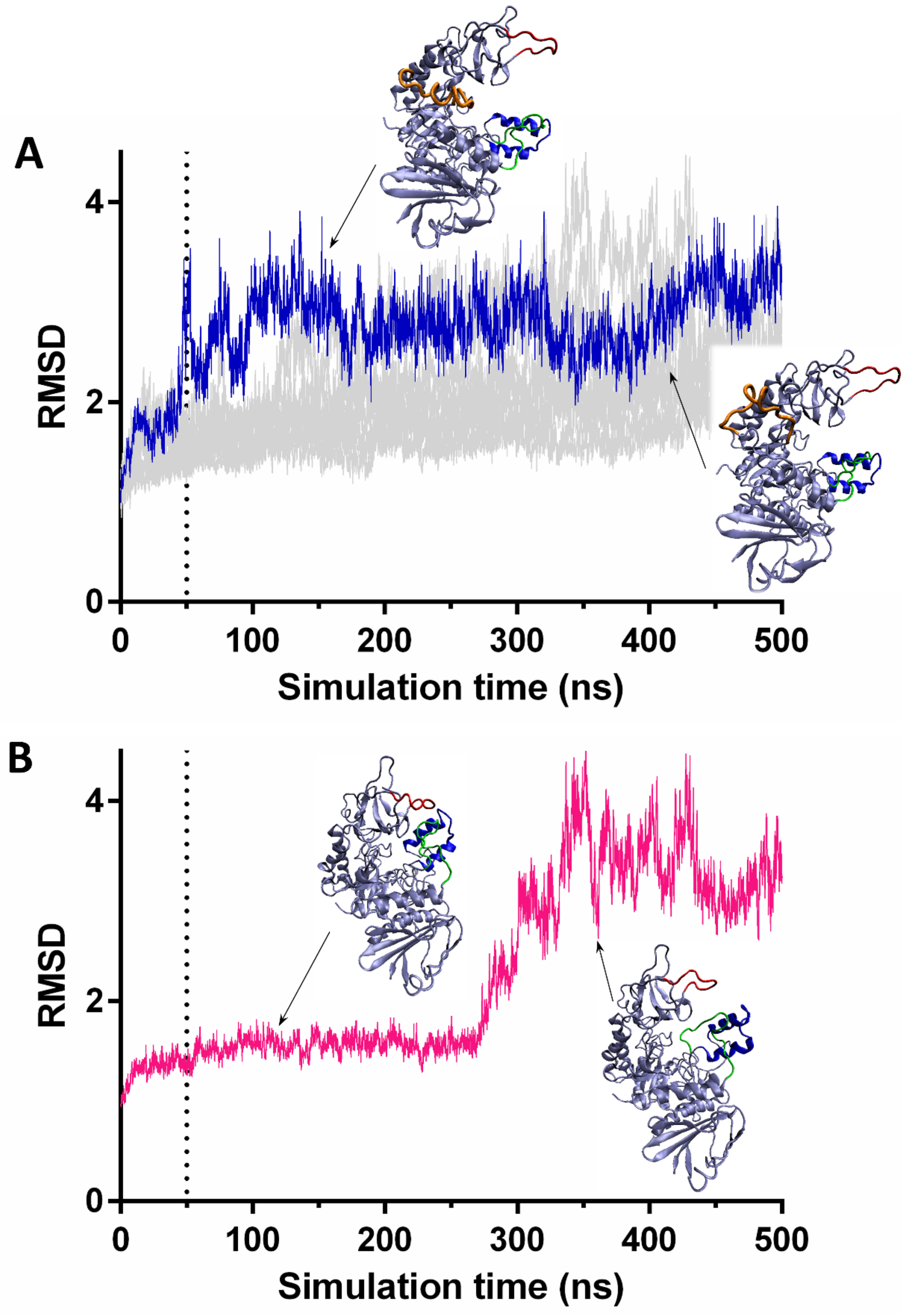


Figure 6.7: Representative WT with TS analogue bound simulations illustrating the active site opening for run seven (blue) and ten (pink). Major loops contributing to the opening movement are indicated in red (residues 213-221), blue (387-417), and green (287-302). ***(A)*** *RMSD plots of the remaining nine runs are traced as grey: run seven has high RMSD compared to these due to early opening of loops about the active site and high mobility in loop 256-272 (orange).* ***(B)*** *A RMSD jump (run ten) is accompanied by an opening motion in loops about the active site, especially the helix-loop-helix (blue).*

6.2.2 RMSF as an indication of protein flexibility changes

Enthalpy variation, and thus C_p (equation **17b**), depends on the range of conformations sampled. As such, RMSF data gives an indication of the effect of ligand binding, and protein regions where major differences are occurring between the substrate and TS bound forms of MalL. RMSF data indicates global rigidification at the TS compared to the ground state (Figure 6.8 A). Predominant differences occur in loops surrounding the active site (Figure 6.8 B). Global involvement in catalysis is consistent with the hypothesis presented in Section 3.2 where vibrational modes are considered as a reservoir of energy available to drive the catalytic process.

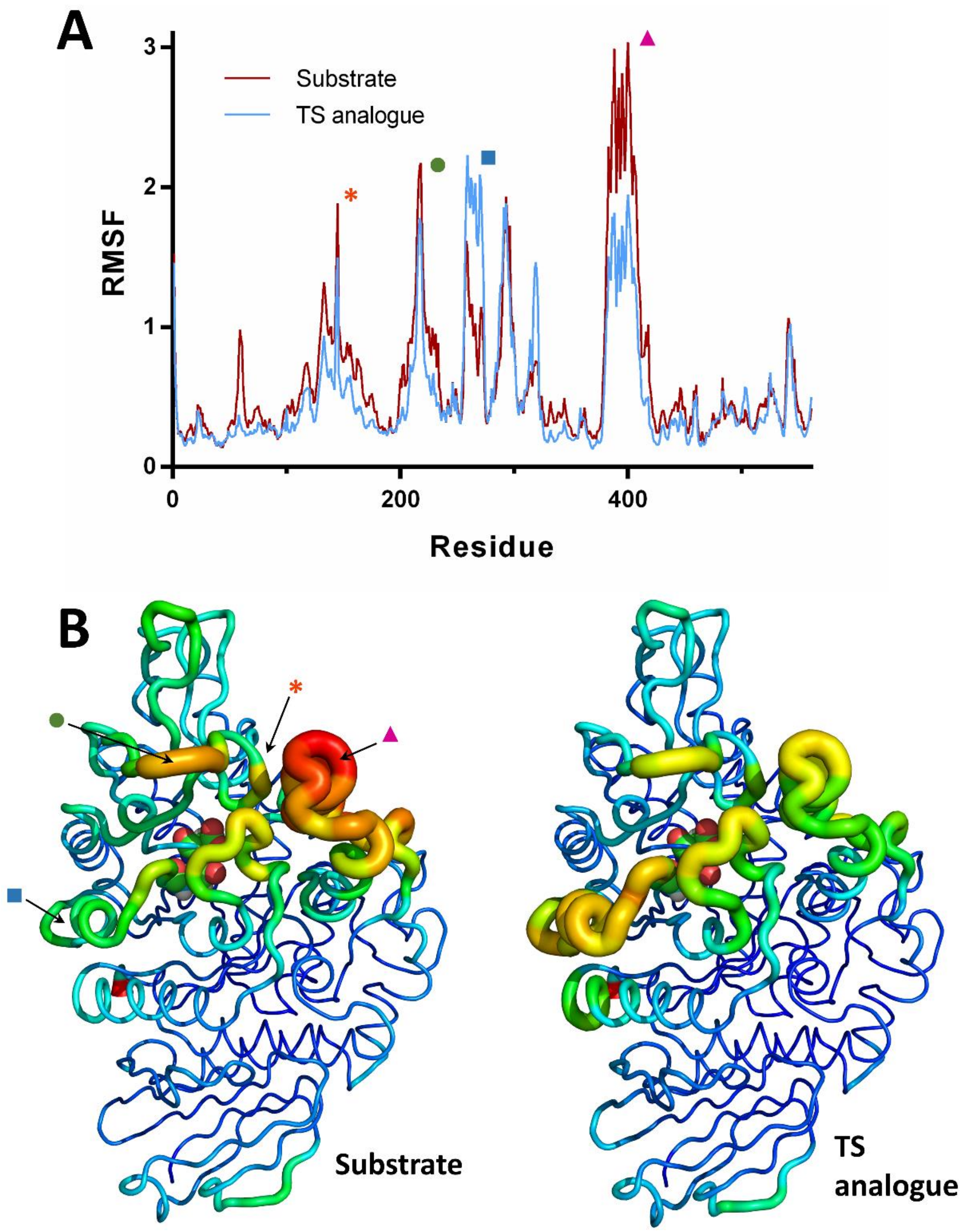


Figure 6.8: RMSF of Cα atoms for MalL WT as an average per residue over ten replicate runs for simulations with substrate and TS analogue. ***(A)*** *RMSF per residue between the substrate and TS analogue bound simulations.* ***(B)*** *RMSF values superimposed on the structure, coloured red through blue from the most flexible to rigid, with an accompanying decrease in cylinder width. Positioning of the active site is indicated with the two catalytic residues as spheres. Corresponding regions of interest are indicated with coloured symbols between the two representations.*

The most significant region of rigidification at the TS is the helix-loop-helix encompassing residues 380-420 (Figure 6.8). This structure forms one of the exterior loops encapsulating the active site. Two residues at either end of this loop (Glu389 and Arg418) are situated in the active site, interacting with the ligand

(Figure 6.9). Arg418 interacts in the cavity between the two glucose units of the ligand, while Glu389 interacts with hydroxyls of the +1 glucose. In TS analogue simulations, these interactions between ligand and these anchoring residues are much shorter and less variable, especially for Glu389 (Figure 6.9 B). Interaction distances show consistent interaction between Arg418 and Glu389 over simulations with TS analogue bound, compared to substrate simulations where interactions are lost from ~200 ns onwards (Figure 6.9 C). These improved interactions with the TS analogue, especially with Glu389, anchor loop 380-420, reducing flexibility, and thus C_p, compared to simulations with substrate.

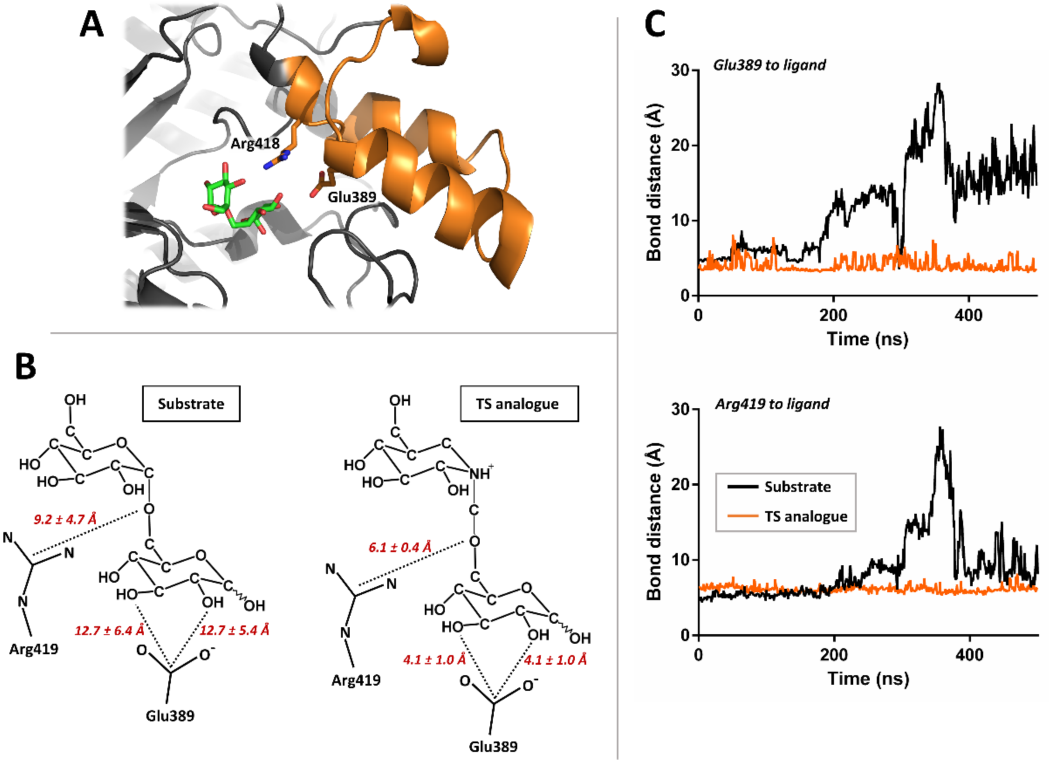


Figure 6.9: Rigidification of a key loop (380-420) at the TS in MalL WT simulations. ***(A)*** *Residues at either end of loop 380-420 (orange) interact with the ligand (isomaltose) in the active site.* ***(B)*** *Average bond lengths (± the SD) shorten with TS analogue in the active site, causing the anchoring of loop 380-420 at the TS.* ***(C)*** *Absolute bond lengths over a typical simulation between Arg418 and Glu389 between substrate and TS bound simulations. The release of loop 380-420 in substrate simulations causing the flexibility in the loop is evident from ~200 ns.*

The one exception to rigidification at the TS is in the loop from residue 253 to 275 which becomes more flexible at the TS (Figure 6.8). This section comprises of the C-terminal end of a β sheet (β5) of the TIM barrel below the active site, and extends along a short amphipathic helix at the solvent interface. Active site residue Glu255 is contained in this section, positioned at the C-terminal of β5. Further evidence is necessary to support conclusions around the significance of this behaviour, although it is of interest to consider the greater movement in one of the catalytic resides at a crucial stage in the reaction and how this may be relevant to catalysis.

6.2.3 Basis of outlier simulation replacement

Cluster analysis was performed on simulation sets to assess large scale conformational sampling across all runs, and highlight any simulations which solely occupied an outlier conformation cluster. Data are presented in Appendix D.6. The majority of replicates cluster together; outliers from this represent more open conformations in latter simulation stages, as illustrated in Figure 6.6 and Figure 6.7. As such are these are considered extremes on the closed-open continuum over which MalL samples.

The one exception to this is run eight with substrate bound, which is the sole occupant of a conformational cluster with 76 % occupancy. Removal of WT with substrate run eight was vindicated with further support from RMSD data and analysis of the loop conformations sampled during the simulation. The RMSD tracks high compared to replicate runs, especially in the latter half of the simulation (Figure 6.10 A). Increases in RMSD from 120 ns onwards correspond to anomalous orientation of loops 288-300 and 380-420 (Figure 6.10 B). Usually, the helix-loop-helix comprising of residues 380-420 sits into the active site; Glu389 in this region has been identified as significant for substrate recognition [Section 5.3.4; (Yamamoto *et al.*, 2011)]. However, for a majority of the run, this interaction is broken as loop 288-300 flips into the active site, excluding helix-loop-helix 380-420. Based on the outlier analysis, high RMSD and the trapping of the structure in an anomalous loop configuration, this simulation was replaced in the $\Delta C^{\ddagger}_{\mathrm{p}}$ analysis.

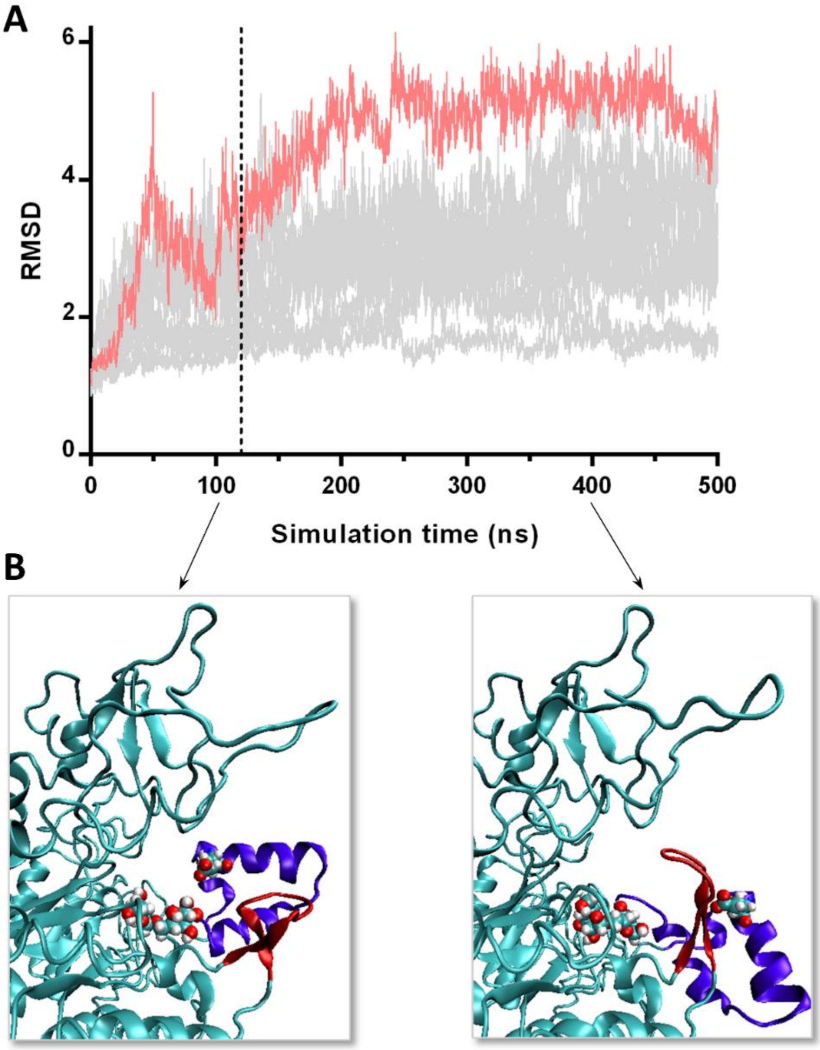


Figure 6.10: Data supporting the exclusion of MalL WT with substrate run 8. ***(A)*** *RMSD plot of run 8 (pink) against all other nine runs (grey). The point where loop 288-300 first flips into the active site is indicated (vertical dashed line).* ***(B)*** *Details of protein movements over the run. Loop 288-300 (red) flips into the active site channel around 120 ns, excluding helix-loop-helix 380-420 (blue) from the normal orientation (left). The substrate and substrate interacting residue Glu389 (Yamamoto et al., 2011) are shown as spheres.*

6.3 *In silico* $\Delta C^{\ddagger}_{p}$ for MalL WT

Processing methods to determine a consistent representative *H* variance from simulation data were determined for KSI (van der Kamp, unpublished). This method is summarised below and applied to MalL.

6.3.1 $\Delta C^{\ddagger}_{p}$ calculation process

Final methods for $\Delta C^{\ddagger}_{p}$ determination involve consideration of the variation in enthalpy in increasing sized windows to capture vibrational modes of decreasing frequency (van der Kamp, unpublished). As window sizes increase to capture the correct vibrational frequencies, final $\Delta C^{\ddagger}_{p}$ values plateau about a stable value.

Enthalpy calculations consider the protein-ligand system alone to avoid overwhelming of the signal difference between substrate and TS analogue bound systems with properties of the bulk water. Energies for the protein-ligand system were considered at 0.01 ns intervals. Of these, the first 50 ns was removed, consistent with the initial large changes relaxing the system from crystallographic constraints observed in RMSD values (Figure 6.4 and Figure 6.5).

Enthalpy variation was calculated with a sliding window of varying time length, shifting along the trajectory in 0.01 ns units. Variation values over the full 450 ns of considered simulation time were averaged to determine the average variance per simulation for each window size. For each window size, variance was averaged over ten simulation replicates for both the substrate and TS analogue bound systems, and the difference taken to get the difference in enthalpy variance ($<\Delta H^2>$; equation **17b**).

$\Delta C^{\ddagger}_{p}$ was calculated with equation **(17b)** for each window size. Final *in silico* $\Delta C^{\ddagger}_{p}$ values were taken from the point at which $\Delta C^{\ddagger}_{p}$ is observed to plateau over increases in window size.

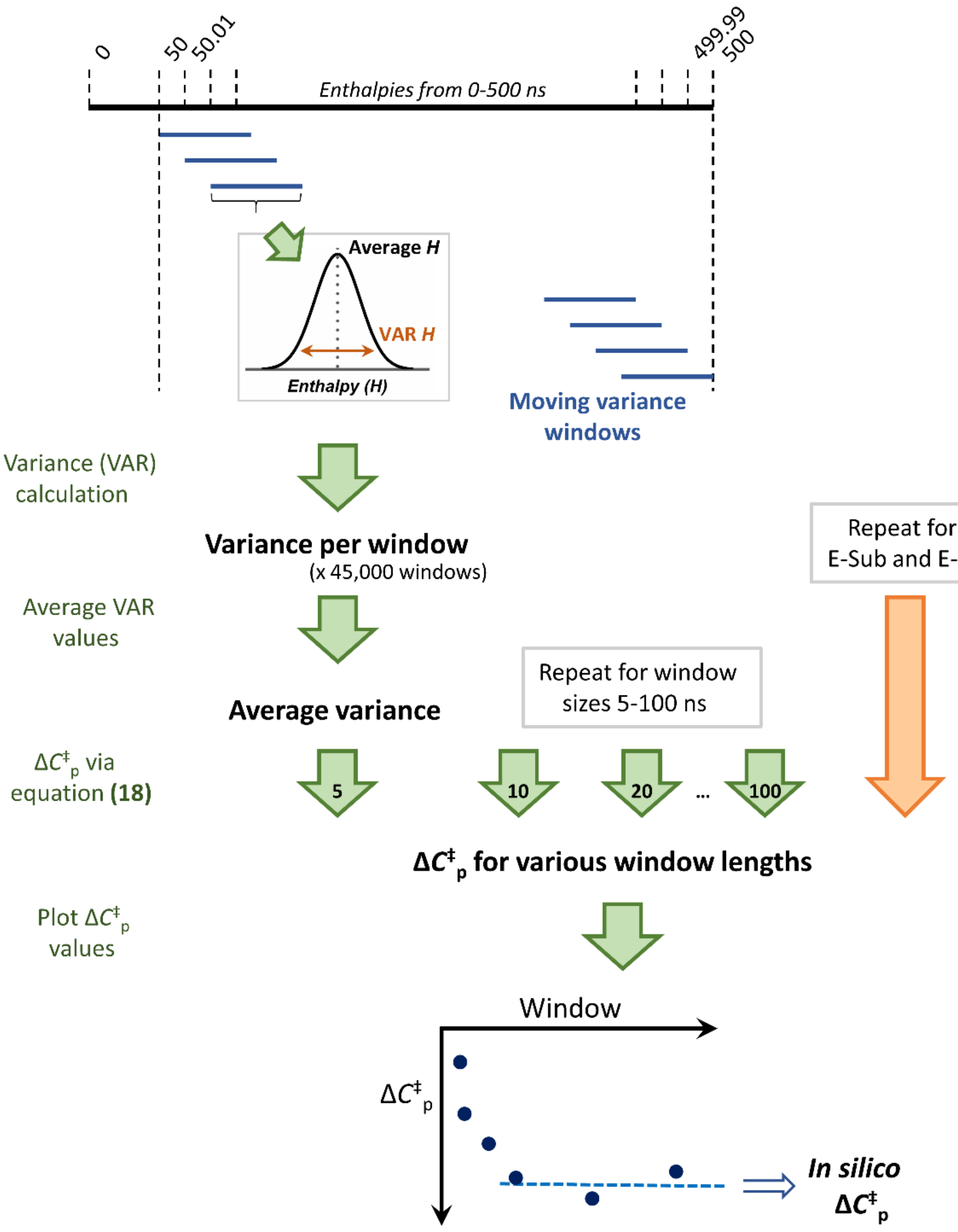


Figure 6.11: Summarised process to extract $\Delta C^{\ddagger}_{p}$ information from enthalpy information. Ten simulation replicates of 500 ns per state were considered in the final $\Delta C^{\ddagger}_{p}$ calculation.

6.3.2 $\Delta C^{\ddagger}_{p}$ results for MalL WT

The method described in Section 6.3.1 resulted in a stable $\Delta C^{\ddagger}_{p}$ value of -10.4 ± 1.8 kJ.mol^{-1}.K^{-1} over window sizes from 40 to 80 ns (Figure 6.12). $\Delta C^{\ddagger}_{p}$ values for smaller window sizes show increasingly negative $\Delta C^{\ddagger}_{p}$ as larger window sizes sample information from slower vibrational modes. Mean $\Delta C^{\ddagger}_{p}$ values are less negative compared to experimental values (-11.6 ± 0.4 kJ.mol^{-1}.K^{-1}), however the *in silico* values agrees with experimental data within error bounds.

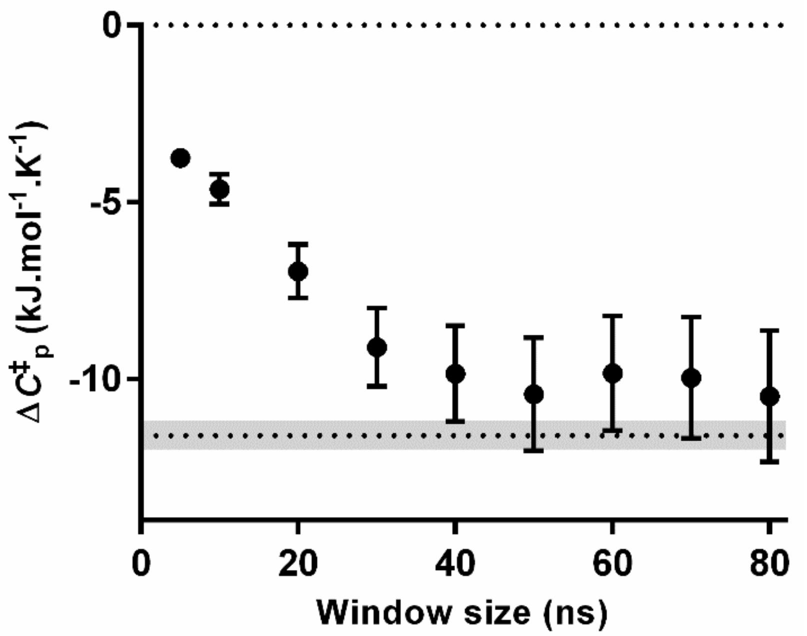


Figure 6.12: In silico $\Delta C^{\ddagger}_{p}$ calculated from simulations of substrate and TS bound MalL WT. Data points are given as the mean ± SD. The experimental $\Delta C^{\ddagger}_{p}$ [-11.6 ± 0.4 kJ.mol^{-1}.K^{-1}; (Hobbs et al., 2013)] is given as a dashed line with the error range in grey. In silico *$\Delta C^{\ddagger}_{p}$ stabilises at -10.4 ± 1.8 kJ.mol^{-1}.K^{-1} from a window sizes of 40 ns.*

Following the plateau in $\Delta C^{\ddagger}_{p}$ with window size, window sizes greater than 80 ns calculate increasingly negative $\Delta C^{\ddagger}_{p}$ values. This overestimation of $\Delta C^{\ddagger}_{p}$ is likely due to the inclusion of low frequency vibrational modes of structural rearrangements not related to the $\Delta C^{\ddagger}_{p}$ for catalysis. This behaviour is not seen in the smaller KSI system where there is less potential for these large structural variations in the conformational landscape (van der Kamp, unpublished).

6.4 $\Delta C^{\ddagger}_{p}$ determination for MalL V200S

MalL presents a valuable system for investigation due to the set of $\Delta\Delta C^{\ddagger}_{p}$ mutants which have been characterised (Hobbs *et al.*, 2013). These mutants comprise identical systems to MalL WT, with a single amino acid change which has altered $\Delta C^{\ddagger}_{p}$. The V200S variant is of interest here due to the significant $\Delta C^{\ddagger}_{p}$ change (-11.6 ± 0.4 to -5.6 ± 0.5 kJ.mol^{-1}.K^{-1}). MD of the apo MalL variants has previously implicated a loop near the active site in the $\Delta C^{\ddagger}_{p}$ change of V200S [Figure 1.6; (Hobbs *et al.*, 2013)]. Here, the methods applied to MalL WT (Section 6.1) are applied to V200S to replicate *in silico* the $\Delta C^{\ddagger}_{p}$ change in V200S, and examine the molecular level details for differences between the systems.

RMSD and RMSF analysis of the V200S simulation are overall similar to that of MalL WT. Large increases in RMSD consistently occur in the initial 50 ns of simulation time. RMSD values cover the same range, and sample varying degrees of open site conformations (Figure 6.13 and Figure 6.14). RMSF overall is lower in V200S with TS analogue than substrate (Figure 6.15). Rigidification at the TS occurs in the same regions between the two structures. The increases in RMSF observed at the TS in WT in loop 253 to 275 are not seen in V200S (Figure 6.15, blue squares).

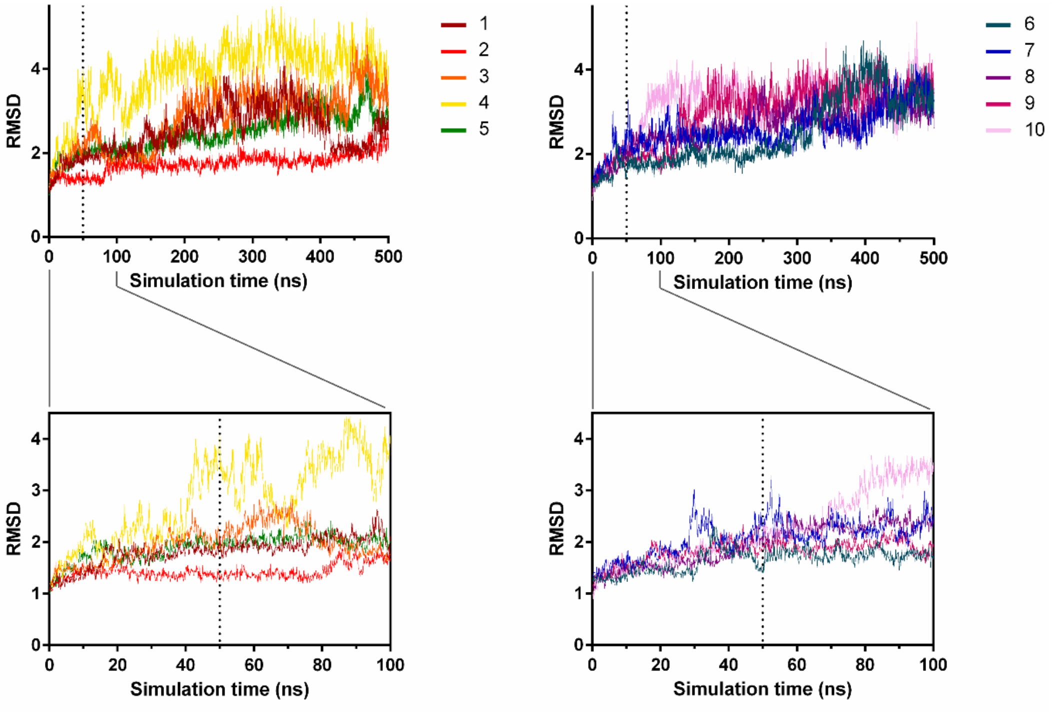


Figure 6.13: RMSD over ten replicate runs for V200S with substrate. The initial 50 ns which encompasses rapid changes in RMSD occur is indicated (vertical line).

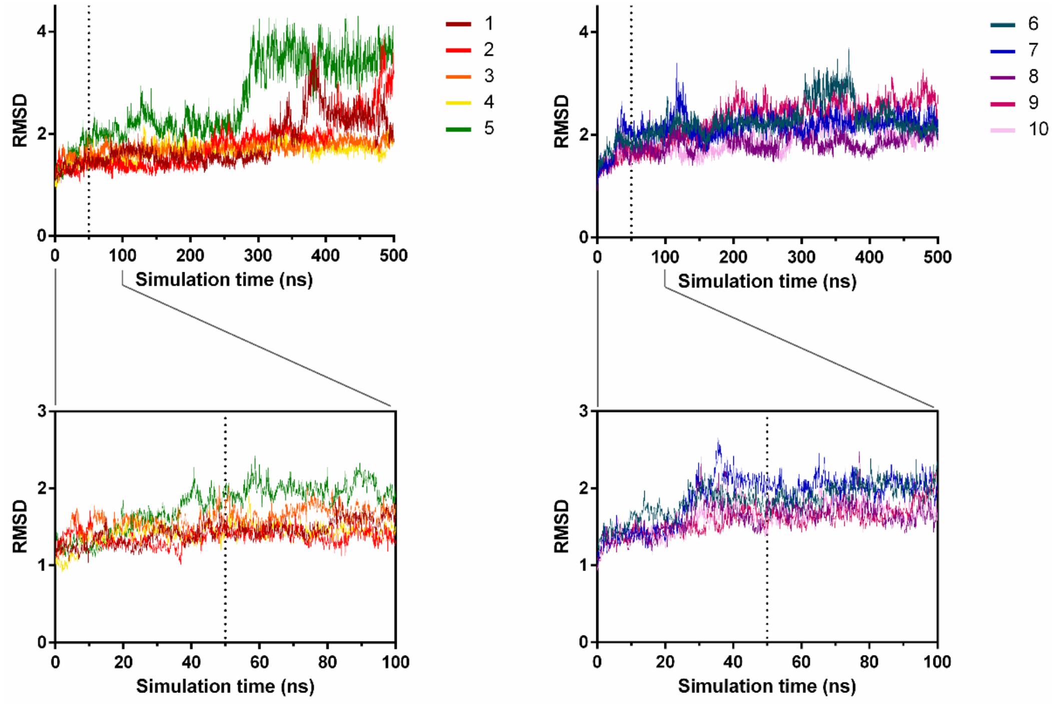


Figure 6.14: RMSD over ten replicate runs for V200S with TS analogue. The initial 50 ns which encompasses rapid changes in RMSD occur is indicated (vertical line).

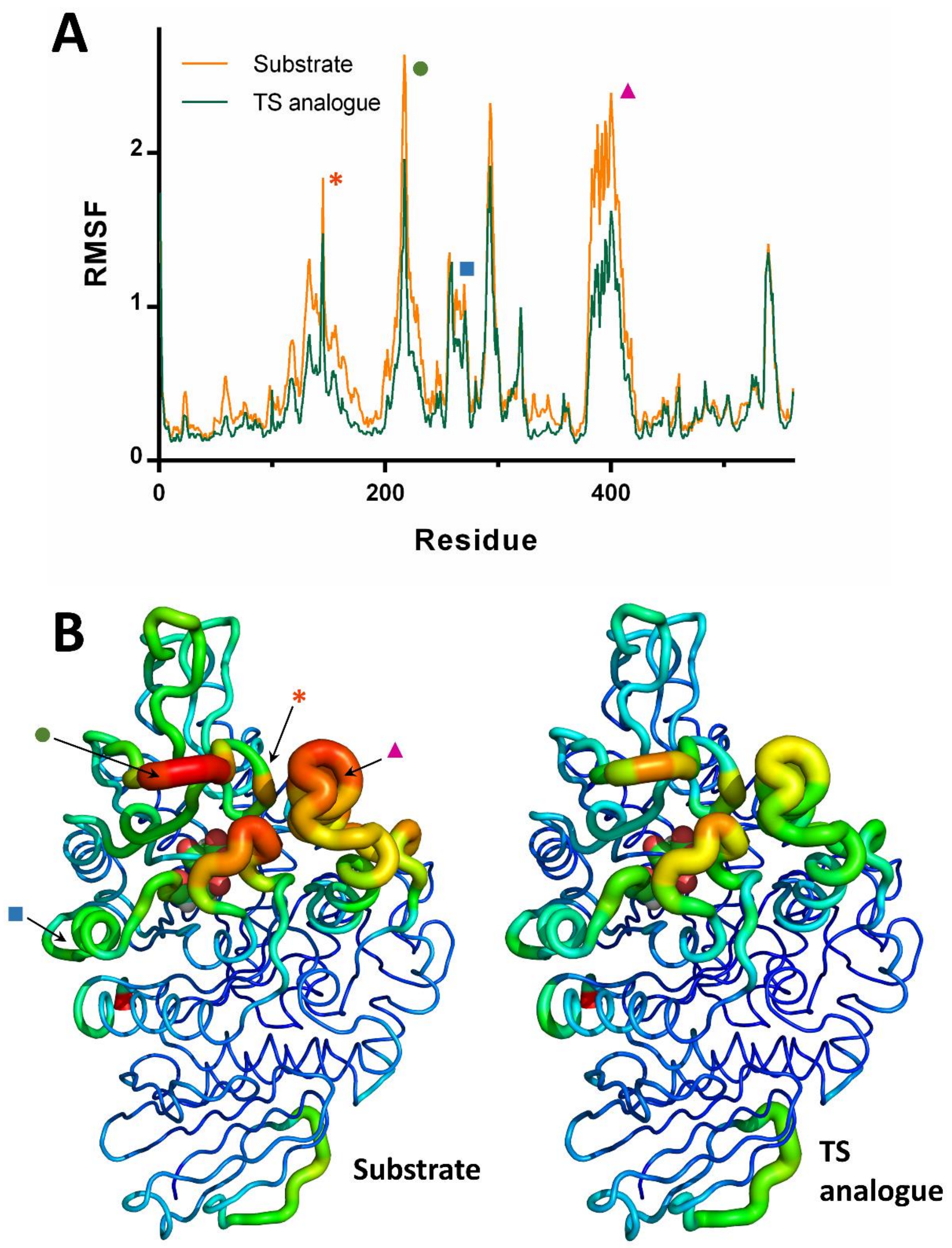


Figure 6.15: RMSF of Cα atoms for V200S as an average per residue over ten replicate runs for simulations with substrate and TS analogue. ***(A)*** *RMSF per residue between the substrate and TS analogue bound simulations.* ***(B)*** *RMSF superimposed on the structure, coloured red through blue from the most flexible to rigid, with an accompanying decrease in cylinder width. Positioning of the active site is indicated with the two catalytic residues as spheres. Corresponding regions of interest are indicated with coloured symbols between the two representations.*

Despite similarity to the WT system, *in silico* $\Delta C^{\ddagger}_{\mathrm{p}}$ approximations do not replicate the experimental value (-5.9 ± 0.5 kJ.mol^{-1}.K^{-1}). Calculated $\Delta C^{\ddagger}_{\mathrm{p}}$ values do not converge with increasing window size, and range from -15 to -40 kJ.mol^{-1}.K^{-1}, significantly more negative than the experimental value and opposite of the expected trend when compared to WT MalL (Figure 6.16).

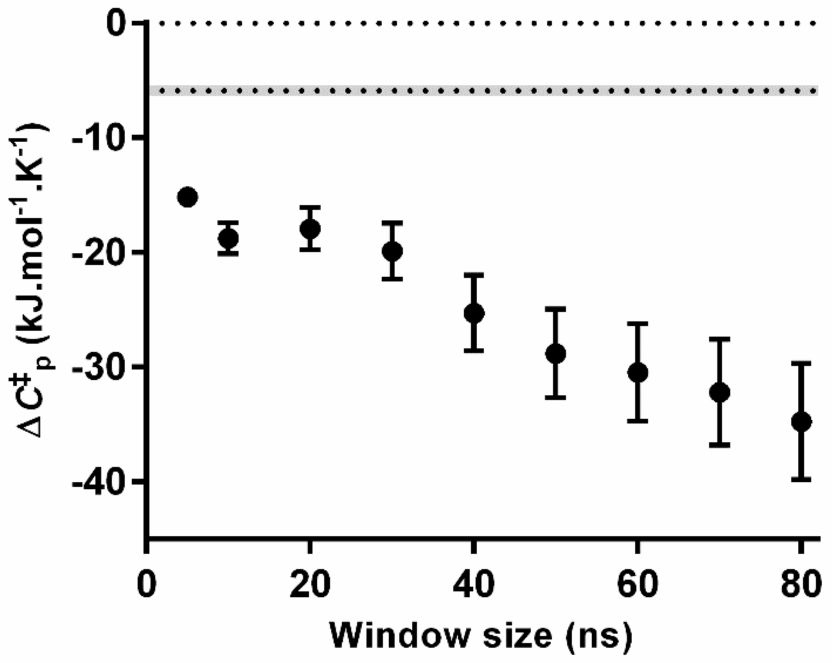


Figure 6.16: In silico $\Delta C^{\ddagger}_p$ *calculated from simulations of substrate and TS bound V200S. Data points are given as ± SD. The experimental* $\Delta C^{\ddagger}_p$ *[-5.9 ± 0.6 kJ.mol*$^{-1}$*.K*$^{-1}$*; (Hobbs et al., 2013)] is given as a dashed line with the error range in grey. In this case,* $\Delta C^{\ddagger}_p$ *failed to stabilise, and modelled a* $\Delta C^{\ddagger}_p$ *significantly larger than experimental data at all window sizes.*

6.4.1 Rationalisation of the unrealistic V200S results

Due to inconsistencies between experimental and modelled data for V200S, the trajectories were further investigated to elucidate the implications of the serine mutation. WT and V200S have overall similar flexibility, except in loop 380-420 where V200S is less flexible (Figure 6.17). This is discussed further in Section 6.4.2.

As evident in the RMSF values, the major discrepancy between the simulation of WT and V200S is in the loop encompassing residues 253 to 275. For the WT, this loop becomes more flexible at the TS (Section 6.2.2), however, this behaviour is not replicated for V200S (Figure 6.17; see Figure 6.15 A also).

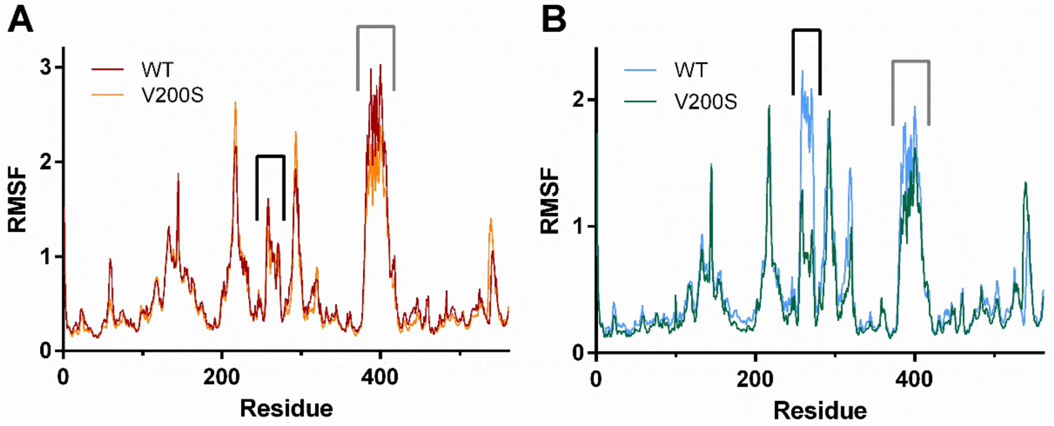


Figure 6.17: RMSF overlays between WT and V200S MalL, where RMSF is reported for Cα atoms as an average per residue over ten replicate runs per state. ***(A)*** *substrate bound simulations.* ***(B)*** *TS analogue bound simulations. Indicated are the protein sections which are rigidified in V200S compared to WT (grey), and hypothesised to be incorrectly simulated in the V200S variant at the TS (black).*

From the simulations, the source of this difference is evident as an anchoring bond network established from the ligand nitrogen, via Ser200, to residues of the loop 253-275 (Figure 6.18). The introduction of serine, hydrogen bonded to NH^+ in the TS analogue displaces Glu255. This movement breaks interactions between side chain CH_2 groups of Glu255 and Ile201, allowing greater flexibility in isoleucine (backbone dihedrals of 148 ± 8 and 136 ± 67 ° respectively in WT and V200S). Greater mobility of Ile201 allows the maintenance of interactions with Gly254, Ala256, the ring of Tyr267 and Leu275 in a hydrophobic pocket; residue-residue distances in this pocket are consistently shorter in V200S (Figure 6.18). Interactions via residues 254 and 256, along with 267 and 275 at either end of the loop-helix-loop provide anchor points rigidifying the encompassed region in V200S.

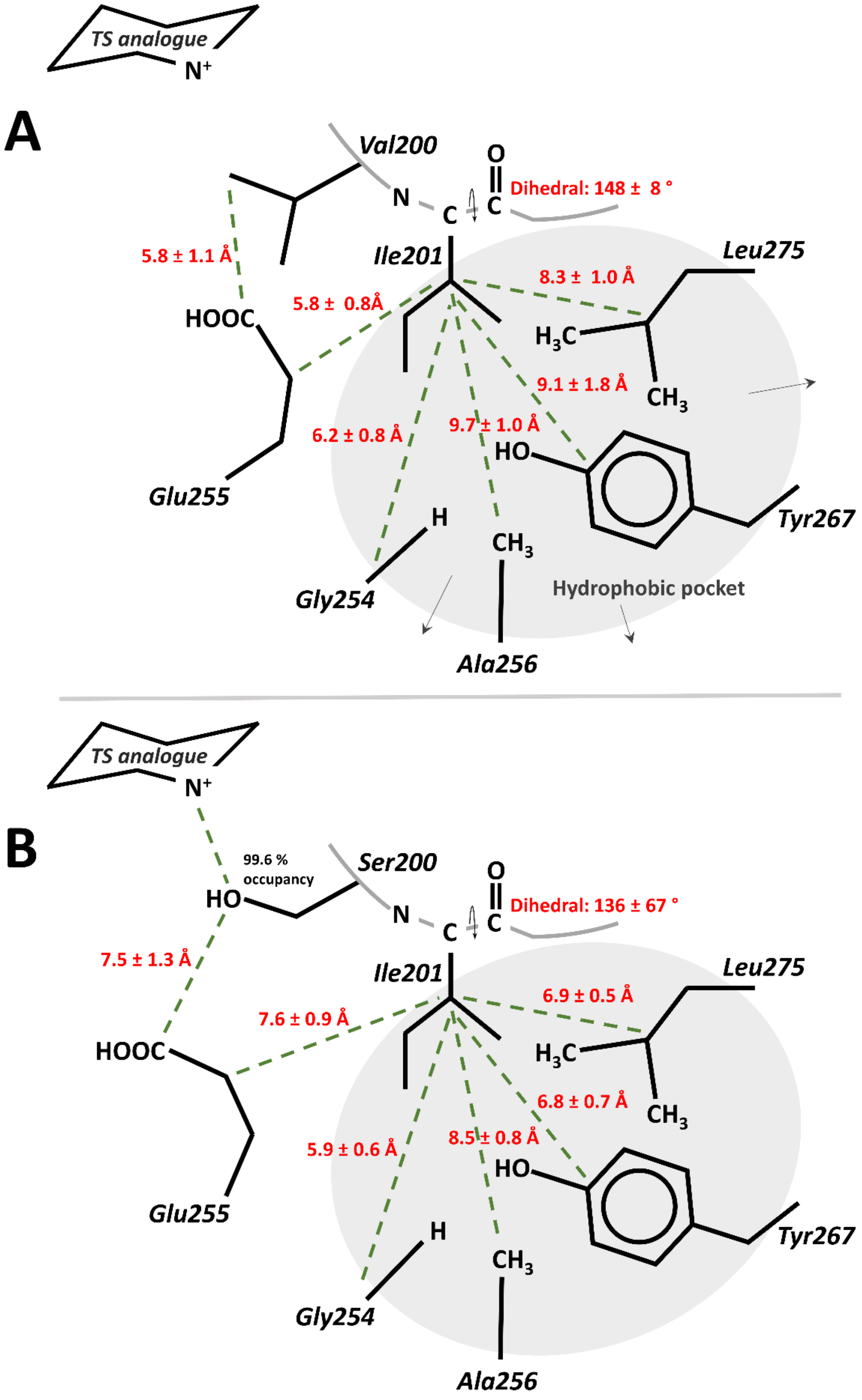


Figure 6.18: Network of interactions reducing flexibility in loop 253-275 in V200S ***(B)*** *compared to WT* ***(A)*** *in simulations with TS analogue. The network of parameters altered in V200S (as described in the text) are given in red as the mean distances/angles over a representative simulation (± the SD). The OH of Tyr267 orientates away from residues in the hydrophobic pocket (not shown).*

It is the introduction of a strong positive charge at the anomeric centre of the TS analogue that sets up this stable Ser200 orientation and rigidification network: over the ten simulations, Ser200 orientates towards the TS analogue 99.6 % of the time (Figure 6.19 C and D). The same effect is not seen in substrate simulations due to the lowered charge about the anomeric carbon, resulting in lower residence of Ser200 orientated towards the ligand (92 %; Figure 6.19 A and B). Given the partial charge in the implied TS (Figure 6.2), this system is likely to behave in a manner similar to the observed V200S substrate simulations, flipping Ser200 away from the ligand a percentage of time sufficient to establish the contacts through Glu255 and Ile201 necessary for release of the hydrophobic pocket and loop 253-275. This effect in the V200S TS analogue simulations is proposed as a factor contributing to the overestimation of the flexibility differences between V200S substrate and TS states, causing the inflated *in silico* $\Delta C^{\ddagger}_{\mathrm{p}}$ (Figure 6.16).

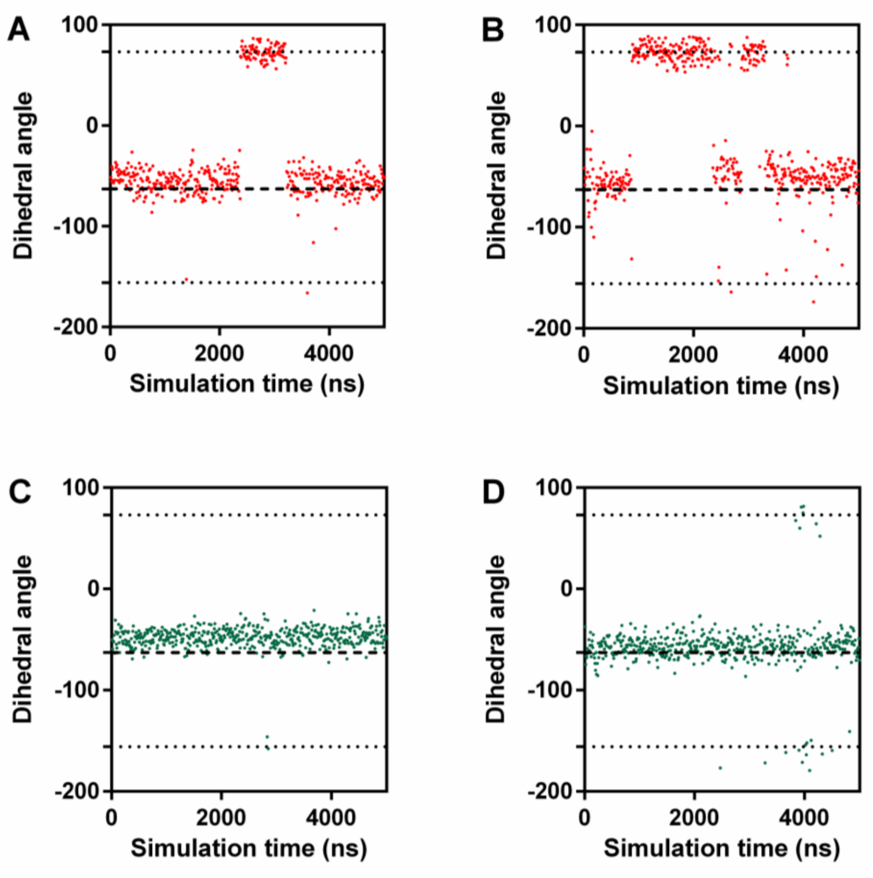


*Figure 6.19: Orientation of Ser200 in V200S simulations (as measured by the dihedral through N-Cα-Cβ-O of Ser200) for the substrate bound state (**A** and **B**) and TS bound state (**C** and **D**) for two representative simulations per state. Dashed lines give the dihedral angle when orientated towards the ligand, dotted lines away from the ligand. Over ten runs, Ser200 orientates towards the ligand 92 % of the time with substrate bound, and 99.6 % with TS analogue bound.*

Issues encountered in the $\Delta C^{\ddagger}_{\mathrm{p}}$ calculation for V200S illustrate the attention necessary in system setup, especially with respect to the choice of TS analogue. Here, localised positive charge in the TS analogue compared to a partial delocalised charge in the TS induces significant dynamics differences which are reflected in an inflated $\Delta C^{\ddagger}_{\mathrm{p}}$. Moving of the charge placement in the sugar ring in the TS analogue presents an option for testing this hypothesis and, if correct, rectifying this issue in V200S; this will be explored in the future.

Implication presented here for the consequences of flexibility about Ile201 at the TS potentially has bearing on the alternate MalL variant studied previously (Hobbs *et al.*, 2013). The variant G202P has altered $\Delta C^{\ddagger}_{\mathrm{p}}$ compared to WT MalL (-5.7 ± 0.6 and -11.6 ± 0.4 kJ.mol^{-1}.K^{-1} respectively). From compromised kinetics and previous MD results, $\Delta C^{\ddagger}_{\mathrm{p}}$ changes are proposed to be effected by increased C_{p}

at the TS (Section 1.4.2). If substitution of glycine with proline affects flexibility in neighbouring Ile201, the reduced capacity for interaction within the hydrophobic pocket can induce flexibility in loop 253-275, increasing C_p. This link offers further insight into the possible alterations of G202P at the TS, although requires further investigation with the G202P variant.

6.4.2 Sources of $\Delta C^{\ddagger}_p$ change in V200S

Comparison of simulations for WT and V200S with substrate indicate global rigidification in V200S, with major differences in loop 380-420 (Figure 6.17). The major loop (210-220) implicated in Hobbs *et al.* (2013) shows little difference in RMSF in simulations with substrate or TS analogue; loop 380-420 does show rigidification in V200S in the original apo simulations. As transition from the substrate to the first TS is considered the rate limiting step for MalL (Zechel and Withers, 2000; Pudney, unpublished), simulations with substrate are a more realistic state to consider as the ground state. Differences in loop 210-220 flexibility between the sets of MD simulations likely represent the rigidification evens upon substrate binding.

Due to the issues identified with the V200S simulations, no definite conclusion can be drawn about the source of $\Delta\Delta C^{\ddagger}_p$ in V200S. However, it is of interest that loop 380-420 has larger drops in RMSF between substrate and TS simulations for V200S, and that global rigidification at the ground state is indicated in simulations. These possibilities warrant further investigation as possible sources of $\Delta C^{\ddagger}_p$ change in V200S.

6.5 Conclusions

Here, initial efforts to determine $\Delta C^{\ddagger}_p$ from *in silico* MD data through analysis of enthalpy distributions indicate there is merit in the methodology. Simulations with substrate and TS analogue for WT MalL capture the correct decreases in flexibility (RMSF) at the TS, and result in $\Delta C^{\ddagger}_p$ values commensurate with experimental data. Simulations presented here predicted a $\Delta C^{\ddagger}_p$ of -10.4 ± 1.8 kJ.mol^{-1}.K^{-1}, within error bounds of the experimental value of -11.6 ± 0.4 kJ.mol^{-1}.K^{-1} for the MalL WT system. Global rigidification contributed to lowered C_p at the TS, with major rigidification contributions from one loop with anchoring interactions extending into the active site. Equivalent simulations with KSI with substrate and intermediate as a TS proxy predict an *in silico* $\Delta C^{\ddagger}_p$ of -1.16 ± 0.19, commensurate with the experimental value of -0.86 ± 0.13 kJ.mol^{-1}.K^{-1} (van der Kamp, unpublished).

KSI and MalL presented contrasting systems in which to trial *in silico* methods for determining $\Delta C^{\ddagger}_p$. MalL is a large system (69.5 kDa), whereas KSI is a dimer of 29.8 kDa total. Final *in silico* $\Delta C^{\ddagger}_p$ values for KSI have lower error than for MalL. Given the equal simulation lengths and replicates despite the considerably larger mass of MalL, this is not surprising. A range of conformations identified for MalL, predominantly varying along the open-closed continuum (Section 6.2.1) likely requires more, and longer, simulations to represent the full canonical ensemble of states. Consideration of minimum simulation time and replicates as a function of system size and conformation range is a logical extension to methodologies to generate reliable data for a range of systems. For large systems, achieving these simulation lengths and replicates is limited by computational power and storage. Given the continuing increases in computing power, application of methods to large systems with significant replicates to converge on a reliable final value is of increasing viability.

Application of the methods to the V200S MalL mutant highlighted pitfalls of the method, especially related to TS analogue selection. V200S has a less negative $\Delta C^{\ddagger}_p$ compared to WT MalL (-5.9 ± 0.5 and -11.6 ± 0.4 kJ.mol^{-1}.K^{-1} respectively); however, *in silico* $\Delta C^{\ddagger}_p$ values greatly overestimated this. Further investigation implicated a channel of interactions from the artificially high positive charge in the TS analogue to a loop which remains constrained at the TS simulation for V200S. Simulations with V200S serve as an example of the care required in selecting an appropriate TS analogue, and the importance of comparison to reliable experimental data.

Overall, MD proves a crucial tool for investigating atomic level detail of the changes occurring in C_p over a reaction, and changes in $\Delta C^{\ddagger}_p$ between different enzyme variants. Agreement between *in silico* and experimental $\Delta C^{\ddagger}_p$ values in the case of MalL WT and KSI lends weight to conclusions drawn from these systems.

Despite the potential confounding factors highlighted in simulations with V200S, MD provides critical information concerning atomic level fluctuations not accessible by other means.

Discussion

Curvature of rates with temperature is a defining characteristic of biological systems from enzymes, to organism growth rates, and the fluxes of whole ecosystems. MMRT as a thermodynamic description of this curvature in enzyme rates was first presented in Hobbs *et al.* (2013). MMRT expands on the established thermodynamics of small molecule reactions to account for the unusually high heat capacity (C_p) of enzymes. The change in heat capacity associated with enzyme catalysis ($\Delta C^{\ddagger}_p$) due to the high affinity for the reaction transition state imparts curvature to reaction rates in the absence of denaturation. Since 2013, MMRT has found applicability for elucidating a range of other biological phenomena including temperature adaptation, vibrational coupling reactions, organism growth rates and the temperature responses of ecosystems (Hobbs *et al.*, 2013; Schipper *et al.*, 2014; Arcus & Pudney, 2015; Arcus *et al.*, 2016).

Here, understanding of the atomic level details of MMRT has been extended by examining crystallographic states along the reaction coordinate, and through the development of methods to calculate an *in silico* $\Delta C^{\ddagger}_p$ from MD data using the model enzyme, MalL. How the temperature dependence of individual enzymes affect enzymatic pathways has been examined through the study of the glycolytic pathway in *E. coli*, and provides support for the applicability of MMRT to organism growth rate data. Finally, the reasons enzymes evolved to be such large macromolecules and thus exhibit unusually high heat capacities has been examined.

Crystallography and MD have both been utilised previously to elucidate atomic level details of MalL WT, and the mutants V200S and G202P that show increased $\Delta C^{\ddagger}_p$. Here, these techniques were extended to consider states along the reaction coordinate. In structures, little variation in atomic positions occurs in either MalL WT or V200S with a TS analogue bound compared to apo structures. Whether this is due to crystallographic constraints such as crystal packing or an indication of no significant rearrangements occurring upon ligand binding is unclear from the structural information. Previous analysis of apo structures indicated a bonding network potentially responsible for rigidification in the V200S mutant (Hobbs *et al.*, 2013). Addition of the TS analogue bound structures highlighted inconsistencies with this argument. With the TS analogue bound, Ser200 occupies two conformations, one maintaining the contacts present in apo V200S, the other interacting with the positive charge of the ligand. Overall, structural analysis of MalL WT and V200S with TS analogue bound revealed little conclusive evidence about the nature of $\Delta C^{\ddagger}_{\mathrm{p}}$ changes along the reaction coordinate, or the mechanism of $\Delta\Delta C^{\ddagger}_{\mathrm{p}}$ in V200S. Given the intrinsic link between vibrational modes and C_{p}, the absence of definitive information pertaining to MMRT from static structures is unsurprising.

MD provides considerably more evidence in terms of vibrational modes and regions of flexibility over the whole protein. Further, the relationship between enthalpy distributions and C_{p} suggests that $\Delta C^{\ddagger}_{\mathrm{p}}$ may be calculated directly from MD data. Development of methods to calculate an *in silico* $\Delta C^{\ddagger}_{\mathrm{p}}$ for KSI and MalL WT (in collaboration with M. van der Kamp and A. Mulholland, Bristol University, UK) suggested that there is sufficient information in MD data to calculate $\Delta C^{\ddagger}_{\mathrm{p}}$. The *in silico* value of -10.4 ± 1.8 $\mathrm{kJ.mol^{-1}.K^{-1}}$ for MalL is within error range of the experimental value (-11.6 ± 0.4 $\mathrm{kJ.mol^{-1}.K^{-1}}$). Flexibility changes (RMSF) indicate that global rigidification contributes to decreases in C_{p} at the TS, although some loops have a more significant impact in the reduction of C_{p} at the TS. The most significant loop rigidification at the TS occurs due to two residues through which interaction with the ligand is optimised at the TS, thus acting as an anchor point. Application of the methods successful for KSI and WT MalL to V200S showed that the methods are not infallible. Calculated *in silico* $\Delta C^{\ddagger}_{\mathrm{p}}$ values for V200S were significantly larger in magnitude than experimental values. A potential reason for this overestimation was tracked back to the anchoring of a loop at the TS due to the artificially positive charge introduced in the TS analogue. This serves to illustrate the attention necessary in system setup, especially with respect to the choice of an appropriate TS analogue.

Despite the issues in simulations, MD did provide further information about the behaviour of the V200S mutant. In simulations, the orientation of Ser200 towards the ligand [i.e. away from the bonding network identified in Hobbs *et al.* (2013)] was greatly preferred. In substrate simulations, this orientation was occupied 92 % of the time. This was increased with the TS analogue (99.6 %), although due to the artificially high positive charge with which Ser200 was interacting, this number is likely inflated. Conflicting evidence is also present between the two sets of MD data with and without ligand. MD of the apo states identified a different region of major rigidification in V200S (compared to WT) than simulations here with substrate and TS analogue. This likely represents the stepwise rigidification occurring from apo to substrate bound to TS bound enzyme states. The inconsistencies with dynamics results for V200S indicates further investigation is necessary to confirm the origin of $\Delta\Delta C^{\ddagger}_{\mathrm{p}}$ in V200S. The first means of approaching this is to repeat the V200S TS simulations with the positive charge moved in the ring to the O_5 position to reduce the positive charge interacting with Ser200. In the inferred TS, positive charge is delocalised over O_5 and the anomeric carbon in the -1 sugar ring, so this shift is chemically realistic.

The contribution of global vibrational modes to catalysis is a contentious issue in the literature (Schwartz & Schramm, 2009; Kamerlin & Warshel, 2010). The importance of the chemical environment in the active site has obvious importance for catalysis, but the involvement of vibrational modes over a range of time scales remains controversial. Critically, the extent of dynamic contribution at and leading to the TS, and the contribution of these motions to lowering the energetic barrier to reaction, as opposed to the effects of active site preorganisation alone, is not well understood (Kandathil *et al.*, 2014). Several aspects of MMRT argue for a global contribution to catalysis. Arcus and Pudney (2015) showed that $\Delta C^{\ddagger}_{\mathrm{p}}$ is a reliable predictor for vibrational coupling in systems where there is strong evidence for the phenomenon. The correlation discovered here between enzyme mass and efficiency also indicates a global contribution to catalysis. Although further work is necessary to prove the correlation, the trend suggests the mass of an enzyme acts as an energy reservoir in the multitude of low frequency vibrational modes, some of which is available for catalysis (Arcus *et al.*, 2016). It is this drop in stored energy that is measured as the $\Delta C^{\ddagger}_{\mathrm{p}}$ for catalysis. Protein wide flexibility changes between enzyme-substrate and enzyme-TS complexes in MD also suggest there is a global contribution to rigidification at the TS, although some protein regions are more significant than others.

Previous publications have noted a fit of the MMRT function to temperature response data from both organism growth rates (Hobbs *et al.*, 2013) and ecosystem fluxes (Schipper *et al.,* 2014). Accurate descriptions of temperature responses at the ecosystem level are especially pertinent given the current changes occurring at a global scale (IPCC, 2013). Schipper *et al.* (2014) showed an improved descriptive nature of MMRT for a range of ecosystem level data compared to currently used models, especially at high and low temperature ranges. Work here put into context the application of MMRT at the ecosystem scale by providing evidence that the temperature response of individual enzymes, independent of denaturation, regulates organism growth rates. Characterisation of six individual enzyme responses to temperature in conjunction with the enzymes working in a pathway showed that the overall $\Delta C^{\ddagger}_{\mathrm{p}}$ is an average of the constituent parts. This indicates that organism temperature responses are affected by all enzymes is the system, rather than by a single master enzyme (Corkrey *et al.*, 2012) and demonstrates that the effect of enzyme $\Delta C^{\ddagger}_{\mathrm{p}}$ is relevant at the cellular level. This justifies the application of MMRT at larger scales where large number of enzymes contribute to overall rates.

Organisms inhabit a wide range of temperature environments, to which enzymes must be finely adapted. In thermal adaptation, not only is the average environmental temperature important, but also the temperature sensitivity and coordination of enzyme responses over metabolic pathways. Hobbs *et al.* (2013) discussed the relationship between $\Delta C^{\ddagger}_{\mathrm{p}}$ and T_{opt}; in order to increase or decrease T_{opt}, $\Delta C^{\ddagger}_{\mathrm{p}}$ must become less or more negative respectively. Achieving these shifts has implications for biochemical parameters such as K_{M} and k_{cat} (Arcus *et al.*, 2016). This concept is important to the adaptation of psychrophiles; in order to adapt to lower temperatures, $\Delta C^{\ddagger}_{\mathrm{p}}$ must become more negative, imparting greater curvature and thus an inherent sensitivity to small changes in environmental temperatures (Arcus *et al.*, 2016). Here, work with glycolytic enzymes also highlighted the relative importance of the inflection point of curvature with temperature (T_{inf}) over T_{opt}. Enzymatic T_{opt} is consistently higher than environmental temperatures (Peterson *et al.*, 2004), and for the glycolytic enzymes, found to be highly variable over enzymes from the same pathway of the same organism. However, T_{inf} was found to be correlated over the enzymes studied to the optimal growth temperature. This is proposed as a mechanism to synchronise changes in enzyme rates with temperature over the cell, to coordinate metabolic fluxes through changes in temperature.

7.1 Further work

As alluded to previously, there is a wealth of further information accessible through MD simulations. Critical to this is the correction of the V200S simulations to model experimental $\Delta C^{\ddagger}_{\mathrm{p}}$ values correctly. The apparent fix for this is to move the charge imparted by introduction of a nitrogen to the ring oxygen position rather than at the anomeric carbon. This is chemically relevant as, in the TS, positive charge is delocalised over these two atoms, and the alterations will lessen the positive influence fixing Ser200 in a single orientation. Validation of this altered TS analogue in simulations with WT MalL will also help validate methods and allow direct comparison between simulations of the two mutants. From this, more conclusive evidence will be accessible to establish the mode of $\Delta C^{\ddagger}_{\mathrm{p}}$ change in V200S compared to WT. Enzymes exhibiting significant activity at elevated temperatures are important to biotechnology (Coker, 2016). However, manipulating T_{opt} without compromising activity has proved a difficult hurdle. MMRT suggests that not only must thermostability be increased to withstand elevated temperatures, but $\Delta C^{\ddagger}_{\mathrm{p}}$ must be manipulated to increase activity at high temperatures. The altered thermal activity of the MalL mutant V200S established this is possible, without negatively effecting the enzymes catalytic proficiency. However, predicting the nature of molecular changes necessary to selectively alter $\Delta C^{\ddagger}_{\mathrm{p}}$ to produce a desired effect still presents a challenge. An improved general understanding of the molecular details underpinning changes in $\Delta C^{\ddagger}_{\mathrm{p}}$ and thus T_{opt}, starting with fully understanding the MalL system and mutants through further MD work, is the first step to more guided approaches to altering enzyme activity for commercial applications.

Dynamics simulations also provide a platform to assess the changes in vibrational modes associated with reduced C_{p} at the TS through normal mode analysis. Turton *et al.* (2014) identified shifts in the range of 20 to 100 cm^{-1} upon inhibitor binding to lysozyme which has guided theoretical models used to describe MMRT. MD provides an alternate route to assess these changes to determine which modes are important for catalysis in MalL and KSI, and if the same frequency modes are important to catalysis over different enzyme systems.

Enzymes interact with their aqueous environment, forming solvation shells of ordered water molecules (Ebbinghaus *et al.*, 2007). Enzyme catalysis is associated with the rearrangement of these waters, especially desolvation of the active site. For example, the binding of transcription factors to DNA, associated with a negative ΔC_p, has a contribution to ΔC_p from desolvation (Bergqvist *et al.*, 2004). Consideration of specific waters into *in silico* $\Delta C^{\ddagger}_p$ calculations provides a starting point to assessing the extent and effect of water shells on $\Delta C^{\ddagger}_p$ and catalysis.

MMRT is presented as a general description of the temperature behaviour of a range of enzymes. To date, a variety of enzymes have been analysed which conform to MMRT kinetics including MalL (seven variants), barnase (two variants), KSI, IPMDH (10 variants), PRTNR, MR, SLO-1, DHFR, AADG, HLADH and the six glycolytic enzymes examined here (Hobbs *et al.*, 2013; Arcus & Pudney, 2015; Arcus *et al.*, 2016). Various other enzymes exhibit curvature with temperature including the 21 enzymes examined in Daniel and Danson (2010) with ‘zero time’ rates independent of denaturation. There are also examples where no curvature is observed with temperature; if this is a true phenomenon or due to data limitations (e.g. a narrow temperature range) requires further investigation. The extensive characterisation of more enzymes with comprehensive temperature plots, denaturation data, dynamics simulations, and a detailed understanding of reactions mechanisms from a range of organisms and reaction types in the future will continue to further our understanding of enzyme catalysed reactions in the context of MMRT.

References

Adams P.D., Afonine P.V., Bunkoczi G., Chen V.B., Davis I.W., Echols N., Headd J.J., Hung L.-W., Kapral G.J., Grosse-Kunstleve R.W. and others (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D*, **66**(2), 213-221.

Afonine P.V., Grosse-Kunstleve R.W., Echols N., Headd J.J., Moriarty N.W., Mustyakimov M., Terwilliger T.C., Urzhumtsev A., Zwart P.H., Adams P.D. (2012). Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallographica Section D*, **68**(4), 352-367.

Alvarez M., Zeelen J.P., Mainfroid V., Rentier-Delrue F., Martial J.A., Wyns L., Wierenga R.K., Maes D. (1998). Triose-phosphate isomerase (TIM) of the psychrophilic bacterium *Vibrio marinus*. Kinetic and structural properties. *Journal of Biological Chemistry*, **273**(4), 2199-2206.

Andrews P.R., Smith G.D., Young I.G. (1973). Transition-State Stabilization and Enzymic Catalysis. Kinetic and Molecular Orbital Studies of the Rearrangement of Chorismate to Prephenate. *Biochemistry*, **12**(18), 3492-3498.

Arcus V.L., Pudney C.R. (2015). Change in heat capacity accurately predicts vibrational coupling in enzyme catalyzed reactions. *FEBS Letters*, **589**(17), 2200-2206.

Arcus V.L., Prentice E.J., Hobbs J.K., Mulholland A.J., Van der Kamp M.W., Pudney C.R., Parker E.J., Schipper L.A. (2016). On the Temperature Dependence of Enzyme-Catalyzed Rates. *Biochemistry*, **55**(12), 1681-1688.

Baicu S.C., Taylor M.J. (2002). Acid–base buffering in organ preservation solutions as a function of temperature: new parameters for comparing buffer capacity and efficiency. *Cryobiology*, **45**(1), 33-48.

Balasubramanian A., Durairajpandian V., Elumalai S., Mathivanan N., Munirajan A.K., Ponnuraj K. (2013). Structural and functional studies on urease from pigeon pea (*Cajanus cajan*). *International Journal of Biological Macromolecules*, **58**, 301-309.

Balls A.K., Walden M.K., Thompson R.R. (1948). A CRYSTALLINE β-AMYLASE FROM SWEET POTATOES. *Journal of Biological Chemistry*, **173**(1), 9-19.

Bar-Even A., Milo R., Noor E., Tawfik D.S. (2015). The Moderately Efficient Enzyme: Futile Encounters and Enzyme Floppiness. *Biochemistry*, **54**(32), 4969-4977.

Basran J., Sutcliffe M.J., Scrutton N.S. (1999). Enzymatic H-Transfer Requires Vibration-Driven Extreme Tunneling. *Biochemistry*, **38**(10), 3218-3222.

Bearne S.L., Wolfenden R. (1995). Enzymatic Hydration of an Olefin: the Burden Borne by Fumarase. *Journal of the American Chemical Society*, **117**(37), 9588-9589.

Bearne S.L., Wolfenden R. (1997). Mandelate Racemase in Pieces: Effective Concentrations of Enzyme Functional Groups in the Transition State. *Biochemistry*, **36**(7), 1646-1656.

Bell J.B., Jones M.E. (1991). Purification and Characterization of Yeast Orotidine 5'-Monophosphate Decarboxylase Overexpressed from Plasmid PGU2. *Journal of Biological Chemistry*, **266**(19), 12662-12667.

ben-Avraham D. (1993). Vibrational normal-mode spectrum of globular proteins. *Physical Review B*, **47**(21), 14559-14560.

Bennett B.D., Kimball E.H., Gao M., Osterhout R., Van Dien S.J., Rabinowitz J.D. (2009). Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nature Chemical Biology*, **5**(8), 593-599.

Bergqvist S., Williams M.A., O'Brien R., Ladbury J.E. (2004). Heat Capacity Effects of Water Molecules and Ions at a Protein–DNA Interface. *Journal of Molecular Biology*, **336**(4), 829-842.

Blethen S.L., Boeker E.A., Snell E.E. (1968). Arginine Decarboxylase from *Escherichia coli* : I. PURIFICATION AND SPECIFICITY FOR SUBSTRATES AND COENZYME. *Journal of Biological Chemistry*, **243**(8), 1671-1677.

Bols M. (1998). 1-Aza Sugars, Apparent Transition State Analogues of Equatorial Glycoside Formation/Cleavage. *Accounts of Chemical Research*, **31**(1), 1-8.

Borges de Melo E., da Silveira Gomes A., Carvalho I. (2006). α- and β-Glucosidase inhibitors: chemical structure and biological activity. *Tetrahedron*, **62**(44), 10277-10302.

Brant D.A., Barnett L.B., Alberty R.A. (1963). The Temperature Dependence of the Steady State Kinetic Parameters of the Fumarase Reaction. *Journal of the American Chemical Society*, **85**(15), 2204-2209.

Britt B.M. (1997). For enzymes, bigger is better. *Biophysical Chemistry*, **69**(1), 63-70.

Brückner R., Titgemeyer F. (2006). Carbon catabolite repression in bacteria: choice of the carbon source and autoregulatory limitation of sugar utilization. *FEMS Microbiology Letters*, **209**(2), 141-148.

Bruice T.C., Benkovic S.J. (2000). Chemical Basis for Enzyme Catalysis. *Biochemistry*, **39**(21), 6267-6274.

Caratzoulas S., Mincer J.S., Schwartz S.D. (2002). Identification of a Protein-Promoting Vibration in the Reaction Catalyzed by Horse Liver Alcohol

Dehydrogenase. *Journal of the American Chemical Society*, **124**(13), 3270-3276.

Case D.A., Babin V., Berryman J.T., Betz R.M., Cai Q., Cerutti D.S., Cheatham III T.E., Darden T.A., Duke R.E., Gohlke H. and others (2014). AMBER 14. *University of California, San Francisco*.

Cesar Mde C., Rosa C.D., Rosa R. (1997). Purification and properties of buffalo (*Bubalus bubalis*) erythrocyte hexokinase. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, **118**(2), 395-401.

Chen H., Tian H.-Q. (2005). Does a General Temperature-Dependent Q_{10} Model of Soil Respiration Exist at Biome and Global Scale? *Journal of Integrative Plant Biology*, **47**(11), 1288-1302.

Chipman D.M. (1971). A Kinetic Analysis of the Reaction of Lysozyme with Oligosaccharides from Bacterial Cell Walls. *Biochemistry*, **10**(9), 1714-1722.

Coker J.A. (2016). Extremophiles and biotechnology: current uses and prospects. *F1000 Research*, **5**, 396-403.

Compain P., Chagnault V., Martin O.R. (2009). Tactics and strategies for the synthesis of iminosugar C-glycosides: a review. *Tetrahedron: Asymmetry*, **20**(6), 672-711.

Corazza A., Rosano C., Pagano K., Alverdi V., Esposito G., Capanni C., Bemporad F., Plakoutsi G., Stefani M., Chiti F. and others (2006). Structure, Conformational Stability, and Enzymatic Properties of Acylphosphatase from the Hyperthermophile *Sulfolobus solfataricus*. *Proteins: Structure, Function, and Bioinformatics*, **62**(1), 64-79.

Corkrey R., Olley J., Ratkowsky D., McMeekin T., Ross T. (2012). Universality of Thermodynamic Constants Governing Biological Growth Rates. *PLoS One*, **7**(2), e32003.

Cox R.A., Cook G.M. (2007). Growth Regulation in the Mycobacterial Cell. *Current Molecular Medicine*, **7**(3), 231-245.

Cuellar A.A., Lloyd C.M., Nielsen P.F., Bullivant D.P., Nickerson D.P., Hunter P.J. (2003). An Overview of CellML 1.1, a Biological Model Description Language. *SIMULATION*, **79**(12), 740-747.

D'Amico S., Sohier J.S., Feller G. (2006). Kinetics and Energetics of Ligand Binding Determined by Microcalorimetry: Insights into Active Site Mobility in a Psychrophilic α-Amylase. *Journal of Molecular Biology*, **358**(5), 1296-1304.

D'Amico S., Marx J.-C., Gerday C., Feller G. (2003). Activity-Stability Relationships in Extremophilic Enzymes. *Journal of Biological Chemistry*, **278**(10), 7891-7896.

Dametto M., Antoniou D., Schwartz S.D. (2012). Barrier Crossing in Dihydrofolate Reductase does not involve a rate-promoting vibration. *Molecular Physics*, **110**(9-10), 531-536.

Daniel R.M., Danson M.J. (2010). A new understanding of how temperature affects the catalytic activity of enzymes. *Trends in Biochemical Sciences*, **35**(10), 584-591.

Daniel R.M., Danson M.J., Eisenthal R. (2001). The temperature optima of enzymes: a new perspective on an old phenomenon. *Trends in Biochemical Sciences*, **26**(4), 223-225.

Daniel R.M., Peterson M.E., Danson M.J., Price N.C., Kelly S.M., Monk C.R., Weinberg C.S., Oudshoorn M.L., Lee C.K. (2009). The molecular basis of the effect of temperature on enzyme activity. *Biochemical Journal*, **425**(2), 353-360.

Davies G.J., Wilson K.S., Henrissat B. (1997). Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochemical Journal*, **321**(Letter), 557-559.

DeWolf W.E., Jr., Fullin F.A., Schramm V.L. (1979). The catalytic site of AMP nucleosidase. Substrate specificity and pH effects with AMP and formycin 5'-PO_4. *Journal of Biological Chemistry*, **254**(21), 10868-10875.

Dong W., Jespersen T., Bols M., Skrydstrup T., Sierks M.R. (1996). Evaluation of Isofagomine and Its Derivatives As Potent Glycosidase Inhibitors. *Biochemistry*, **35**(8), 2788-2795.

Dumas D.P., Caldwell S.R., Wild J.R., Raushel F.M. (1989). Purification and Properties of the Phosphotriesterase from *Pseudomonas diminuta*. *Journal of Biological Chemistry*, **264**(33), 19659-19665.

Ebbinghaus S., Kim S.J., Heyden M., Yu X., Heugen U., Gruebele M., Leitner D.M., Havenith M. (2007). An extended dynamical hydration shell around proteins. *PNAS*, **104**(52), 20749-20752.

Edwards D.R., Lohman D.C., Wolfenden R. (2011). Catalytic Proficiency: The Extreme Case of S–O Cleaving Sulfatases. *Journal of the American Chemical Society*, **134**(1), 525-531.

Elias M., Wieczorek G., Rosenne S., Tawfik D.S. (2014). The universality of enzymatic rate–temperature dependency. *Trends in Biochemical Sciences*, **39**(1), 1-7.

Emsley P., Cowtan K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D*, **60**, 2126-2132.

Eswari Pandaranayaka P.J., Kanagavalli J., Krishnadas S.R., Sundaresan P., Krishnaswamy S. (2008). Over expression and purification of recombinant human myocilin. *World Journal of Microbiology and Biotechnology*, **24**(6), 903-907.

Evans P. (2006). Scaling and assessment of data quality. *Acta Crystallographica Section D*, **62**(1), 72-82.

Eyschen J., Vitoux B., Marraud M., Cung M.T., Branlant G. (1999a). Engineered Glycolytic Glyceraldehyde-3-Phosphate Dehydrogenase Binds the *Anti*Conformation of NAD^+ Nicotinamide but Does Not Experience A-Specific Hydride Transfer. *Archives of Biochemistry and Biophysics*, **364**(2), 219-227.

Eyschen J., Vitoux B., Marraud M., Cung M.T., Branlant G. (1999b). Engineered Glycolytic Glyceraldehyde-3-Phosphate Dehydrogenase Binds the *Anti* Conformation of NAD^+ Nicotinamide but Does Not Experience A-Specific Hydride Transfer. *Archives of Biochemistry and Biophysics*, **364**(2), 219-227.

Fedor M.J., Uhlenbeck O.C. (1992). Kinetics of Intermolecular Cleavage by Hammerhead Ribozymes. *Biochemistry*, **31**(48), 12042-12054.

Feller G., Gerday C. (2003). Psychrophilic enzymes: hot topics in cold adaptation. *Nature Reviews Microbiology*, **1**(3), 200-208.

Feller G., Narinx E., Arpigny J.L., Aittaleb M., Baise E., Genicot S., Gerday C. (1996). Enzymes from psychrophilic organisms. *FEMS Microbiology Reviews*, **18**(2-3), 189-202.

Fields P.A., Somero G.N. (1998). Hot spots in cold adaptation: Localized increases in conformational flexibility in lactate dehydrogenase A_4 orthologs of Antarctic notothenioid fishes. *PNAS*, **95**(19), 11476-11481.

Fornaini G., Dachá M., Magnani M., Stocchi V. (1982). Hexokinase from rabbit red blood cells. *Methods in Enzymology*, **90**(E), 3-10.

Forterre P. (2005). The two ages of the RNA world, and the transition to the DNA world: a story of viruses and cells. *Biochimie*, **87**(9–10), 793-803.

Frick L., Mac Neela J.P., Wolfenden R. (1987). Transition State Stabilization by Deaminases: Rates of Nonenzymatic Hydrolysis of Adenosine and Cytidine. *Bioorganic Chemistry*, **15**(2), 100-108.

Friedlingstein P., Cox P., Betts R., Bopp L., W. v.B., Brovkin V., Cadule P., Doney S., Eby M., Fung I. and others (2006). Climate-carbon cycle feedback analysis: Results from the (CMIP)-M-4 model intercomparison. *Journal of Climate*, **19**, 3337-3353.

Galzigna L., Manani G., Giron G.P., Burlina A. (1977). Enzymatic assay of fructose-1,6-diphosphate for the measurement of its utilization by tissues. *International Journal for Vitamin and Nutrition Research*, **47**(1), 88-91.

Gao H., Chen Y., Leary J.A. (2005). Kinetic measurements of phosphoglucose isomerase and phosphomannose isomerase by direct analysis of phosphorylated aldose–ketose isomers using tandem mass spectrometry. *International Journal of Mass Spectrometry*, **240**(3), 291-299.

Garrett E.R., Mehta P.J. (1972). Solvolysis of Adenine Nucleosides. I. Effects of Sugars and Adenine Substituents on Acid Solvolyses. *Journal of the American Chemical Society*, **94**(24), 8532-8541.

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. 2005. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker JM ed. The Proteomics Protocols Handbook, Humana Press Pp. 571-607.

Guan R., Tyler P.C., Evans G.B., Schramm V.L. (2013). Thermodynamic Analysis of Transition State Features in Picomolar Inhibitors of Human 5′ Methylthioadenosine Phosphorylase. *Biochemistry*, **52**(46), 8313-8322.

Guan R., Ho M.C., Brenowitz M., Tyler P.C., Evans G.B., Almo S.C., Schramm V.L. (2011). Entropy-Driven Binding of Picomolar Transition State Analogue Inhibitors to Human 5'-Methylthioadenosine Phosphorylase. *Biochemistry*, **50**(47), 10408-10417.

Hall A., Knowles J.R. (1975). The Uncatalyzed Rates of Enolization of Dihydroxyacetone Phosphate and of Glyceraldehyde 3-Phosphate in Neutral Aqueous Solution. Quantitative Assessment of the Effectiveness of an Enzyme Catalyst. *Biochemistry*, **14**(19), 4348-4352.

Hamdi S., Moyano F., Sall S., Bernoux M., Chevallier T. (2013). Synthesis analysis of the temperature sensitivity of soil respiration from laboratory studies in relation to incubation methods and soil conditions. *Soil Biology and Biochemistry*, **58**, 115-126.

Hansen T., Wendorff D., Schönheit P. (2004). Bifunctional Phosphoglucose/Phosphomannose Isomerases from the Archaea *Aeropyrum pernix* and *Thermoplasma acidophilum* Constitute a Novel Enzyme Family within the Phosphoglucose Isomerase Superfamily. *Journal of Biological Chemistry*, **279**(3), 2262-2272.

Harel M., Quinn D.M., Nair H.K., Silman I., Sussman J.L. (1996). The X-ray Structure of a Transition State Analog Complex Reveals the Molecular Origins of the Catalytic Power and Substrate Specificity of Acetylcholinesterase. *Journal of the American Chemical Society*, **118**(10), 2340-2346.

Harris P., Navarro Poulsen J.-C., Jensen K.F., Larsen S. (2000). Structural Basis for the Catalytic Mechanism of a Proficient Enzyme: Orotidine 5‘-Monophosphate Decarboxylase. *Biochemistry*, **39**(15), 4217-4224.

Hasegawa K., Kubota M., Matsuura Y. (1999). Roles of catalytic residues in α-amylases as evidenced by the structures of the product-complexed mutants of a maltotetraose-forming amylase. *Protein Engineering* **12**(10), 819-824.

Hawkinson D.C., Eames T.C., Pollack R.M. (1991). Energetics of 3-Oxo-Δ^5-steroid Isomerase: Source of the Catalytic Power of the Enzyme. *Biochemistry*, **30**(45), 10849-10858.

Hay S., Sutcliffe M.J., Scrutton N.S. (2007). Promoting motions in enzyme catalysis probed by pressure studies of kinetic isotope effects. *PNAS*, **104**(2), 507-512.

Hay S., Johannissen L.O., Hothi P., Sutcliffe M.J., Scrutton N.S. (2012). Pressure Effects on Enzyme-Catalyzed Quantum Tunneling Events Arise from Protein-Specific Structural and Dynamic Changes. *Journal of the American Chemical Society*, **134**(23), 9749-9754.

Henzler-Wildman K.A., Lei M., Thai V., Kerns S.J., Karplus M., Kern D. (2007). A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature*, **450**(7171), 913-916.

Hobbs J.K., Jiao W., Easter A.D., Parker E.J., Schipper L.A., Arcus V.L. (2013). Change in Heat Capacity for Enzyme Catalysis Determines Temperature Dependence of Enzyme Catalyzed Rates. *ACS Chemical Biology*, **8**(11), 2388–2393.

Hobbs J.K., Shepherd C., Saul D.J., Demetras N.J., Haaning S., Monk C.R., Daniel R.M., Arcus V.L. (2012). On the Origin and Evolution of Thermophily: Reconstruction of Functional Precambrian Enzymes from Ancestors of *Bacillus*. *Molecular Biology and Evolution*, **29**(2), 825-835.

Horn H.W., Swope W.C., Pitera J.W., Madura J.D., Dick T.J., Hura G.L., Head-Gordon T. (2004). Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *The Journal of Chemical Physics*, **120**(20), 9665-9678.

Hughes A.B., Rudge A.J. (1994). Deoxynojirimycin: Synthesis and Biological Activity. *Natural Product Reports*, **11**, 135-162.

Humphrey W., Dalke A., Schulten K. (1996). VMD - Visual Molecular Dynamics. *Journal of Molecular Graphics*, **14**, 33-38.

Inouye S., Tsuruoka T., Ito T., Niida T. (1968). Structure and synthesis of nojirimycin. *Tetrahedron*, **23**(5), 2125-2144.

IPCC 2013. Climate Change 2013: The Physical Science Basis. Cambridge University Press.

Jagadish M.N., Carter B.L. (1978). Effects of temperature and nutritional conditions on the mitotic cell cycle of *Saccharomyces cerevisiae*. *Journal of Cell Science*, **31**, 71-78.

Janeček S., Svensson B., Henrissat B. (1997). Domain Evolution in the α-Amylase Family. *Journal of Molecular Evolution*, **45**(3), 322-331.

Jen-Jacobson L., Engler L.E., Jacobson L.A. (2000). Structural and Thermodynamic Strategies for Site-Specific DNA Binding Proteins. *Structure*, **8**(10), 1015-1023.

Johannissen L.O., Hay S., Scrutton N.S., Sutcliffe M.J. (2007). Proton Tunneling in Aromatic Amine Dehydrogenase is Driven by a Short-Range Sub-Picosecond Promoting Vibration: Consistency of Simulation and Theory with Experiment. *The Journal of Physical Chemistry A*, **111**(10), 2631-2638.

Kamemoto E.S., Mansour T.E. (1986). Phosphofructokinase in the Liver Fluke *Fasciola hepatica*. PURIFICATION AND KINETIC CHANGES BY PHOSPHORYLATION. *Journal of Biological Chemistry*, **261**(9), 4346-4351.

Kamerlin S.C., Warshel A. (2010). At the Dawn of the 21st Century: Is Dynamics the Missing Link for Understanding Enzyme Catalysis? *Proteins*, **78**(6), 1339-1375.

Kandathil S.M., Driscoll M.D., Dunn R.V., Scrutton N.S., Hay S. (2014). Proton tunnelling and promoting vibrations during the oxidation of ascorbate by ferricyanide? *Physical Chemistry Chemical Physics*, **16**(6), 2256-2259.

Katsuya Y., Mezaki Y., Kubota M., Matsuura Y. (1998). Three-dimensional Structure of *Pseudomonas Isoamylase* at 2.2 Å Resolution. *Journal of Molecular Biology*, **281**(5), 885-897.

Kell D.B. (1982). Enzymes as energy 'funnels'? *Trends in Biochemical Sciences*, **7**(10), 349.

Kikovska E., Mikkelsen N.E., Kirsebom L.A. (2005). The naturally *trans*-acting ribozyme RNase P RNA has leadzyme properties. *Nucleic Acids Research*, **33**(21), 6920-6930.

Kirschner K.N., Yongye A.B., Tschampel S.M., González-Outeiriño J., Daniels C.R., Foley B.L., Woods R.J. (2008). GLYCAM06: A Generalizable Biomolecular Force Field. Carbohydrates. *Journal of Computational Chemistry*, **29**(4), 622-655.

Kleeb A.C., Kast P., Hilvert D. (2006). A Monofunctional and Thermostable Prephenate Dehydratase from the Archaeon *Methanocaldococcus jannaschii*. *Biochemistry*, **45**(47), 14101-14110.

Knapp M.J., Rickert K., Klinman J.P. (2002). Temperature-Dependent Isotope Effects in Soybean Lipoxygenase-1: Correlating Hydrogen Tunneling with Protein Dynamics. *Journal of the American Chemical Society*, **124**(15), 3865-3874.

Kraakman K. (2017). *Testing Macromolecular Rate Theory*. Masters thesis, University of Waikato, Hamilton, New Zealand.

Kumar V. (2010). Analysis of the key active subsites of glycoside hydrolase 13 family members. *Carbohydrate Research*, **345**(7), 893-898.

Laidler K.J., King M.C. (1983). The Development of Transition-State Theory. *The Journal of Physical Chemistry*, **87**(15), 2657-2664.

Laskowski R.A., Hutchinson E.G., Michie A.D., Wallace A.C., Jones M.L., Thornton J.M. (1997). PDBsum: a Web-based database of summaries and analyses of all PDB structures. *Trends Biochem Sci*, **22**(12), 488-90.

Lee C.K., Daniel R.M., Shepherd C., Saul D., Cary S.C., Danson M.J., Eisenthal R., Peterson M.E. (2007). Eurythermalism and the temperature dependence of enzyme activity. *Faseb Journal* **21**(8), 1934-1941.

Leslie A.W., Powell H. 2007. Processing diffraction data with mosflm. Evolving Methods for Macromolecular Crystallography, Springer Netherlands. Pp. 41-51.

Liu H., Liang X., Søhoel H., Bülow A., Bols M. (2001). Noeuromycin, a Glycosyl Cation Mimic that Strongly Inhibits Glycosidases. *Journal of the American Chemical Society*, **123**(21), 5116-5117.

Lloyd J., Taylor J.A. (1994). On the temperature dependence of soil respiration. *Functional Ecology*, **8**(3), 315-323.

Lo M.-C., Aulabaugh A., Jin G., Cowling R., Bard J., Malamas M., Ellestad G. (2004). Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Analytical Biochemistry*, **332**(1), 153-159.

Lovell S.C., Davis I.W., Arendall W.B., 3rd, de Bakker P.I., Word J.M., Prisant M.G., Richardson J.S., Richardson D.C. (2003). Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. *Proteins: Structure, Function, and Bioinformatics*, **50**(3), 437-450.

Loveridge E.J., Tey L.-H., Behiry E.M., Dawson W.M., Evans R.M., Whittaker S.B.M., Günther U.L., Williams C., Crump M.P., Allemann R.K. (2011). The Role of Large-Scale Motions in Catalysis by Dihydrofolate Reductase. *Journal of the American Chemical Society*, **133**(50), 20561-20570.

Luk L.Y.P., Javier Ruiz-Pernía J., Dawson W.M., Roca M., Loveridge E.J., Glowacki D.R., Harvey J.N., Mulholland A.J., Tuñón I., Moliner V. and others (2013). Unraveling the role of protein dynamics in dihydrofolate reductase catalysis. *PNAS*, **110**(41), 16344-16349.

MacKenzie L.F., Sulzenbacher G., Divne C., Jones T.A., Wöldike H.F., Schülein M., Withers S.G., Davies G.J. (1998). Crystal structure of the family 7 endoglucanase I (Cel7B) from *Humicola insolens* at 2.2 Å resolution and identification of the catalytic nucleophile by trapping of the covalent glycosyl-enzyme intermediate. *Biochemical Journal*, **335 (Pt 2)**, 409-416.

Magnani M., Stocchi V., Serafini N., Piatti E., Dachá M., Fornaini G. (1983). Pig red blood cell hexokinase: regulatory characteristics and possible physiological role. *Archives of Biochemistry Biophysics*, **226**(1), 377-387.

Makarov A.A., Yakovlev G.I., Mitkevich V.A., Higgin J.J., Raines R.T. (2004). Zinc(II)-mediated inhibition of ribonuclease Sa by an *N*-hydroxyurea nucleotide and its basis. *Biochemical and Biophysical Research Communications*, **319**(1), 152-156.

Makhatadze G.I. (1998). Heat capacities of amino acids, peptides and proteins. *Biophysical Chemistry*, **71**(2–3), 133-156.

Makhatadze G.I., Privalov P.L. (1990). Heat capacity of proteins. I. Partial molar heat capacity of individual amino acid residues in aqueous solution: Hydration effect. *Journal of Molecular Biology*, **213**(2), 375-384.

Masgrau L., Roujeinikova A., Johannissen L.O., Hothi P., Basran J., Ranaghan K.E., Mulholland A.J., Sutcliffe M.J., Scrutton N.S., Leys D. (2006). Atomic Description of an Enzyme Reaction Dominated by Proton Tunneling. *Science*, **312**(5771), 237-241.

Matthews B.W. (1968). Solvent content of protein crystals. *Journal of Molecular Biology*, **33**(2), 491-497.

Mayweather D., Danyal K., Dean D.R., Seefeldt L.C., Hoffman B.M. (2012). Temperature Invariance of the Nitrogenase Electron Transfer Mechanism. *Biochemistry,* **51**(42), 8391-8398.

McCoy A.J., Grosse-Kunstleve R.W., Adams P.D., Winn M.D., Storoni L.C., Read R.J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography*, **40**(4), 658-674.

Meyer D., Schneider-Fresenius C., Horlacher R., Peist R., Boos W. (1997). Molecular Characterization of Glucokinase from *Escherichia coli* K-12. *Journal of Bacteriology,* **179**(4), 1298-1306.

Meyer M.P., Tomchick D.R., Klinman J.P. (2008). Enzyme structure and dynamics affect hydrogen tunneling: The impact of a remote side chain (I553) in soybean lipoxygenase-1. *PNAS*, **105**(4), 1146-1151.

Miller B.G., Raines R.T. (2004). Identifying Latent Enzyme Activities: Substrate Ambiguity within Modern Bacterial Sugar Kinases. *Biochemistry,* **43**(21), 6387-6392.

Mincer J.S., Schwartz S.D. (2004). Rate-promoting vibrations and coupled hydrogen–electron transfer reactions in the condensed phase: A model for enzymatic catalysis. *The Journal of Chemical Physics* **120**(6), 7755-7760.

Mitra B., Kallarakal A.T., Kozarich J.W., Gerlt J.A., Clifton J.R., Petsko G.A., Kenyon G.L. (1995). Mechanism of the Reaction Catalyzed by Mandelate Racemase: Importance of Electrophilic Catalysis by Glutamic Acid 317. *Biochemistry*, **34**(9), 2777-2787.

Mohr P.W., Krawiec S. (1980). Temperature Characteristics and Arrhenius Plots for Nominal Psychrophiles, Mesophiles and Thermophiles. *Journal of General Microbiology*, **121**(2), 311-317.

Na H., Song G., ben-Avraham D. (2016). Universality of vibrational spectra of globular proteins. *Physical biology* **13**(1), 016008.

Nicholas P.C. (1988). Determination of fructose-1,6-diphosphate aldolase activity with glyceraldehyde-3-phosphate dehydrogenase and diformazan formation. *Biochemical Society Transactions*, **16**(5), 753-754.

Ohmichi T., Okumoto Y., Sugimoto N. (1998). Effect of substrate RNA sequence on the cleavage reaction by a short ribozyme. *Nucleic Acids Research*, **26**(24), 5655-5661.

Olguin L.F., Askew S.E., O'Donoghue A.C., Hollfelder F. (2008). Efficient Catalytic Promiscuity in an Enzyme Superfamily: An Arylsulfatase Shows a Rate Acceleration of 10^{13} for Phosphate Monoester Hydrolysis. *Journal of the American Chemical Society*, **130**(49), 16547-16555.

Olsen K., Svensson B., Christensen U. (1992). Stopped-flow fluorescence and steady-state kinetic studies of ligand-binding reactions of glucoamylase from *Aspergillus niger*. *European Journal of Biochemistry*, **209**(2), 777-784.

Olsson M.H.M., Søndergaard C.R., Rostkowski M., Jensen J.H. (2011). PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p*Ka* Predictions. *Journal of Chemical Theory and Computation*, **7**(2), 525-537.

Orsi B.A., Cleland W.W. (1972). Inhibition and Kinetic Mechanism of Rabbit Muscle Glyceraldehyde-3-phosphate Dehydrogenase. *Biochemistry*, **11**(1), 102-109.

Page M.I., Jencks W.P. (1971). Entropic Contributions to Rate Accelerations in Enzymic and Intramolecular Reactions and the Chelate Effect. *PNAS*, **68**(8), 1678-1683.

Pang J., Hay S., Scrutton N.S., Sutcliffe M.J. (2008). Deep Tunneling Dominates the Biologically Important Hydride Transfer Reaction from NADH to FMN in Morphinone Reductase. *Journal of the American Chemical Society*, **130**(22), 7092-7097.

Paoli P., Cirri P., Camici L., Manao G., Cappugi G., Moneti G., Pieraccini G., Camici G., Ramponi G. (1997). Common-type acylphosphatase: steady-state kinetics and leaving-group dependence. *Biochemical Journal*, **327**(Pt 1), 177-184.

Pauling L. (1946). Molecular Architecture and Biological Reactions. *Chemical and Engineering News* **24**(10), 1375-1377.

Payens T.A.J. (1983). Why are enzymes so large? *Trends in Biochemical Sciences*, **8**(2), 46.

Pearson M.A., Park I.-S., Schaller R.A., Michel L.O., Karplus P.A., Hausinger R.P. (2000). Kinetic and Structural Characterization of Urease Active Site Variants. *Biochemistry*, **39**(29), 8575-8584.

Peterson M.E., Eisenthal R., Danson M.J., Spence A., Daniel R.M. (2004). A new intrinsic thermal parameter for enzymes reveals true temperature optima. *Journal of Biological Chemistry*, **279**(20), 20717-20722.

Plater A.R., Zgiby S.M., Thomson G.J., Qamar S., Wharton C.W., Berry A. (1999). Conserved Residues in the Mechanism of the *E. coli* class II FBP-aldolase. *Journal of Molecular Biology*, **285**(2), 843-855.

Pocker Y., Meany J.E. (1967). A Comparative Study of Enzymatic and Metal Ion Catalyzed Hydration of Pyridine Aldehydes. *Journal of the American Chemical Society*, **89**(3), 631-636.

Pollack R.M., Zeng B., Mack J.P.G., Eldin S. (1989). Determination of the Microscopic Rate Constants for the Base Catalyzed Conjugation of 5-Androstene-3,17-dione. *Journal of the American Chemical Society*, **111**(16), 6419-6423.

Prabhu N.V., Sharp K.A. (2005). Heat Capacity in Proteins. *Annual Review of Physical Chemistry*, **56**(1), 521-548.

Privalov P.L., Makhatadze G.I. (1990). Heat capacity of proteins. II. Partial molar heat capacity of the unfolded polypeptide chain of proteins: Protein unfolding effects. *Journal of Molecular Biology*, **213**(2), 385-391.

Pudney C.R., Johannissen L.O., Sutcliffe M.J., Hay S., Scrutton N.S. (2010). Direct Analysis of Donor−Acceptor Distance and Relationship to Isotope Effects and the Force Constant for Barrier Compression in Enzymatic H-Tunneling Reactions. *Journal of the American Chemical Society*, **132**(32), 11329-11335.

Pudney C.R., Hay S., Levy C., Pang J., Sutcliffe M.J., Leys D., Scrutton N.S. (2009). Evidence To Support the Hypothesis That Promoting Vibrations Enhance the Rate of an Enzyme Catalyzed H-Tunneling Reaction. *Journal of the American Chemical Society*, **131**(47), 17072-17073.

Pudney C.R., Guerriero A., Baxter N.J., Johannissen L.O., Waltho J.P., Hay S., Scrutton N.S. (2013). Fast Protein Motions Are Coupled to Enzyme H-Transfer Reactions. *Journal of the American Chemical Society*, **135**(7), 2512-2517.

Radzicka A., Wolfenden R. (1995). A Proficient Enzyme. *Science*, **267**(5194), 90-93.

Radzicka A., Wolfenden R. (1996). Rates of Uncatalyzed Peptide Bond Hydrolysis in Neutral Solution and the Transition State Affinities of Proteases. *Journal of the American Chemical Society*, **118**(26), 6105-6109.

Reger A.S., Wu R., Dunaway-Mariano D., Gulick A.M. (2008). Structural Characterization of a 140° Domain Movement in the Two-Step Reaction Catalyzed by 4-Chlorobenzoate:CoA Ligase. *Biochemistry*, **47**(31), 8016-8025.

Reyes A.C., Koudelka A.P., Amyes T.L., Richard J.P. (2015). Enzyme Architecture: Optimization of Transition State Stabilization from a Cation–Phosphodianion Pair. *Journal of the American Chemical Society*, **137**(16), 5312-5315.

Robinson J.L., Pyzyna B., Atrasz R.G., Henderson C.A., Morrill K.L., Burd A.M., DeSoucy E., Fogleman R.E., 3rd, Naylor J.B., Steele S.M. and others (2005). Growth Kinetics of Extremely Halophilic *Archaea* (Family *Halobacteriaceae*) as Revealed by Arrhenius Plots. *Journal of Bacteriology*, **187**(3), 923-929.

Roe D.R., Cheatham T.E. (2013). PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation*, **9**(7), 3084-3095.

Rose I.A., Hanson K.R., Wilkinson K.D., Wimmer M.J. (1980). A suggestion for naming faces of ring compounds. *PNAS*, **77**(5), 2439-2441.

Rubach J.K., Ramaswamy S., Plapp B.V. (2001). Contributions of Valine-292 in the Nicotinamide Binding Site of Liver Alcohol Dehydrogenase and Dynamics to Catalysis. *Biochemistry*, **40**(42), 12686-12694.

Rye C.S., Withers S.G. (2000). Glycosidase mechanisms. *Current Opinion in Chemical Biology*, **4**(5), 573-580.

Schimel J.P., Mikan C. (2005). Changing microbial substrate use in Arctic tundra soils through a freeze-thaw cycle. *Soil Biology and Biochemistry*, **37**(8), 1411-1418.

Schipper L.A., Hobbs J.K., Rutledge S., Arcus V.L. (2014). Thermodynamic Theory explains the Temperature Optima of Soil Microbial Processes and High Q_{10} Values at Low Temperatures. *Global Change Biology*, **20**(11), 3578–3586.

Schönert S., Buder T., Dahl M.K. (1998). Identification and Enzymatic Characterization of the Maltose-Inducible α-glucosidase MalL (Sucrase-Isomaltase-Maltase) of *Bacillus subtilis*. *Journal of Bacteriology*, **180**(9), 2574-2578.

Schulte P.M. (2015). The effects of temperature on aerobic metabolism: towards a mechanistic understanding of the responses of ectotherms to a changing environment. *The Journal of Experimental Biology*, **218**(12), 1856-1866.

Schwartz S.D., Schramm V.L. (2009). Enzymatic transition states and dynamic motion in barrier crossing. *Nature Chemical Biology* **5**(8), 551-558.

Seewald M.J., Pichumani K., Stowell C., Tibbals B.V., Regan L., Stone M.J. (2000). The role of backbone conformational heat capacity in protein stability: temperature dependent dynamics of the B1 domain of *Streptococcal* protein G. *Protein Science*, **9**(6), 1177-1193.

Sierra C.A. (2012). Temperature sensitivity of organic matter decomposition in the Arrhenius equation: some theoretical considerations. *Biogeochemistry*, **108**(1), 1-15.

Sinnott M.L. (1990). Catalytic Mechanism of Enzymic Glycosyl Transfer. *Chemical Reviews*, **90**(7), 1171-1202.

Snider M.J., Wolfenden R. (2000). The Rate of Spontaneous Decarboxylation of Amino Acids. *Journal of the American Chemical Society*, **122**(46), 11507-11508.

Søndergaard C.R., Olsson M.H.M., Rostkowski M., Jensen J.H. (2011). Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of p*Ka* Values. *Journal of Chemical Theory and Computation*, **7**(7), 2284-2295.

Srere P.A. (1984). Why are enzymes so big? *Trends in Biochemical Sciences*, **9**(9), 387-390.

Stage-Zimmermann T.K., Uhlenbeck O.C. (1998). Hammerhead ribozyme kinetics. *RNA*, **4**(8), 875-889.

Stevens R.L., Micalizzi E.R., Fessler D.C., Pals D.T. (1972). Angiotensin I Converting Enzyme of Calf Lung. Method of Assay and Partial Purification. *Biochemistry*, **11**(16), 2999-3007.

Sullivan B.J., Durani V., Magliery T.J. (2011). Triosephosphate Isomerase by Consensus Design: Dramatic Differences in Physical Properties and Activity of Related Variants. *Journal of Molecular Biology*, **413**(1), 195-208.

Swanwick R.S., Maglia G., Tey L.H., Allemann R.K. (2006). Coupling of protein motions and hydrogen transfer during catalysis by *Escherichia coli* dihydrofolate reductase. *Biochemical Journal*, **394**(Pt 1), 259-265.

Tansey M.R. (1972). Effect of Temperature on Growth Rate and Development of the Thermophilic Fungus *Chaetomium thermophile*. *Mycologia*, **64**(6), 1290-1299.

Tejero I., Garcia-Viloca M., González-Lafont A., Lluch J.M., York D.M. (2006). Enzyme dynamics and tunneling enhanced by compression in the hydrogen abstraction catalyzed by soybean lipoxygenase-1. *The Journal of Physical Chemistry B*, **110**(48), 24708-24719.

Terwilliger T.C., Grosse-Kunstleve R.W., Afonine P.V., Moriarty N.W., Zwart P.H., Hung L.-W., Read R.J., Adams P.D. (2008). Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallographica Section D*, **64**(1), 61-69.

Teusink B., Passarge J., Reijenga C.A., Esgalhado E., van der Weijden C.C., Schepper M., Walsh M.C., Bakker B.M., van Dam K., Westerhoff H.V. and others (2000). Can yeast glycolysis be understood in terms of *in vitro* kinetics of the constituent enzymes? Testing biochemistry. *European Journal of Biochemistry*, **267**(17), 5313-5329.

Thomas T.M., Scopes R.K. (1998). The effects of temperature on the kinetics and stability of mesophilic and thermophilic 3-phosphoglycerate kinases. *Biochemical Journal* **330**(3), 1087-1095.

Thompson M.J., Eisenberg D. (1999). Transproteomic Evidence of a Loop-Deletion Mechanism for Enhancing Protein Thermostability. *Journal of Molecular Biology*, **290**(2), 595-604.

Totir M., Echols N., Nanao M., Gee C.L., Moskaleva A., Gradia S., Iavarone A.T., Berger J.M., May A.P., Zubieta C. and others (2012). Macro-to-Micro Structural Proteomics: Native Source Proteins for High-Throughput Crystallization. *PLoS One*, **7**(2), e32498-e32498.

Tuomi M., Vanhala P., Karhu K., Fritze H., Liski J. (2008). Heterotrophic soil respiration—Comparison of different models describing its temperature dependence. *Ecological Modelling*, **211**(1–2), 182-190.

Turton D.A., Senn H.M., Harwood T., Lapthorn A.J., Ellis E.M., Wynne K. (2014). Terahertz underdamped vibrational motion governs protein-ligand binding in solution. *Nature Communications*, **5**, 3999.

van der Kamp M.W., Chaudret R., Mulholland A.J. (2013). QM/MM modelling of ketosteroid isomerase reactivity indicates that active site closure is integral to catalysis. *Febs Journal*, **280**(13), 3120-3131.

Van Wazer J., Callis C. (1958). Metal Complexing by Phosphates. *Chemical Reviews*, **58**(6), 1011-1046.

Vanquelef E., Simon S., Marquant G., Garcia E., Klimerak G., Delepine J.C., Cieplak P., Dupradeau F.Y. (2011). R.E.D. Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Research*, **39**(Web Server issue), W511-W517.

Varrot A., Tarling C.A., Macdonald J.M., Stick R.V., Zechel D.L., Withers S.G., Davies G.J. (2003). Direct Observation of the Protonation State of an Imino Sugar Glycosidase Inhibitor upon Binding. *Journal of the American Chemical Society*, **125**(25), 7496-7497.

Vasella A., Davies G.J., Böhm M. (2002). Glycosidase mechanisms. *Current Opinion in Chemical Biology*, **6**(5), 619-629.

Vieille C., Zeikus G.J. (2001). Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiology and Molecular Biology Reviews*, **65**(1), 1-43.

Villà J., Štrajbl M., Glennon T.M., Sham Y.Y., Chu Z.T., Warshel A. (2000). How important are entropic contributions to enzyme catalysis? *PNAS*, **97**(22), 11899-11904.

Vriend G. (1990). WHAT IF: a molecular modeling and drug design program. *Journal of Molecular Graphics*, **8**(1), 52-56.

Wang J., Wolf R.M., Caldwell J.W., Kollman P.A., Case D.A. (2004). Development and Testing of a General Amber Force Field. *Journal of Computational Chemistry*, **25**(9), 1157-1174.

Wang Q., Graham R.W., Trimbur D., Warren R.A.J., Withers S.G. (1994). Changing Enzymic Reaction Mechanisms by Mutagenesis: Conversion of a Retaining Glucosidase to an Inverting Enzyme. *Journal of the American Chemical Society*, **116**(25), 11594-11595.

Warshel A., Sharma P.K., Kato M., Xiang Y., Liu H., Olsson M.H. (2006). Electrostatic basis for enzyme catalysis. *Chemical Reviews*, **106**(8), 3210-3235.

Watney J.B., Agarwal P.K., Hammes-Schiffer S. (2003). Effect of Mutation on Enzyme Motion in Dihydrofolate Reductase. *Journal of the American Chemical Society*, **125**(13), 3745-3750.

Wilks J.C., Slonczewski J.L. (2007). pH of the Cytoplasm and Periplasm of *Escherichia coli*: Rapid Measurement by Green Fluorescent Protein Fluorimetry. *Journal of Bacteriology*, **189**(15), 5601-5607.

Winn M.D., Ballard C.C., Cowtan K.D., Dodson E.J., Emsley P., Evans P.R., Keegan R.M., Krissinel E.B., Leslie A.G.W., McCoy A. and others (2011). Overview of the CCP4 suite and current developments. *Acta Crystallographica*, **67**(4), 235-242.

Wolfenden R. (1969). Transition State Analogues for Enzyme Catalysis. *Nature*, **223**(5207), 704-705.

Wolfenden R. (2014). Massive Thermal Acceleration of the Emergence of Primordial Chemistry, the Incidence of Spontaneous Mutation, and the Evolution of Enzymes. *Journal of Biological Chemistry*, **289**(44), 30198-30204.

Wolfenden R., Snider M.J. (2001). The Depth of Chemical Time and the Power of Enzymes as Catalysts. *Accounts of Chemical Research*, **34**(12), 938-945.

Wolfenden R., Lu X., Young G. (1998). Spontaneous Hydrolysis of Glycosides. *Journal of the American Chemical Society*, **120**(27), 6814-6815.

Yamamoto K., Miyake H., Kusunoki M., Osaki S. (2010). Crystal structures of isomaltase from *Saccharomyces cerevisiae* and in complex with its competitive inhibitor maltose. *Febs Journal*, **277**(20), 4205-4214.

Yamamoto K., Miyake H., Kusunoki M., Osaki S. (2011). Steric hindrance by 2 amino acid residues determines the substrate specificity of isomaltase from *Saccharomyces cerevisiae*. *Journal of Bioscience and Bioengineering*, **112**(6), 545-550.

Yu T., Lloyd C.M., Nickerson D.P., Cooling M.T., Miller A.K., Garny A., Terkildsen J.R., Lawson J., Britten R.D., Hunter P.J. and others (2011). The Physiome Model Repository 2. *Bioinformatics*, **27**(5), 743-744.

Zechel D.L., Withers S.G. (2000). Glycosidase Mechanisms: Anatomy of a Finely Tuned Catalyst. *Accounts of Chemical Research*, **33**(1), 11-18.

Zechel D.L., Boraston A.B., Gloster T., Boraston C.M., Macdonald J.M., Tilbrook D.M.G., Stick R.V., Davies G.J. (2003). Iminosugar Glycosidase Inhibitors: Structural and Thermodynamic Dissection of the Binding of Isofagomine and 1-Deoxynojirimycin to β-Glucosidases. *Journal of the American Chemical Society*, **125**(47), 14313-14323.

Zhao J., Ding G.-H., Tao L., Yu H., Yu Z.-H., Luo J.-H., Cao Z.-W., Li Y.-X. (2007). Modular co-evolution of metabolic networks. *BMC Bioinformatics*, **8**(1), 311-323.

Zheng R.L., Kemp R.G. (1995). Phosphofructo-1-kinase: Role of Charge Neutralization in the Active Site. *Biochemical and Biophysical Research Communications*, **214**(3), 765-770.

Zhu A., Romero R., Petty H.R. (2011). An enzymatic colorimetric assay for glucose-6-phosphate. *Analytical Biochemistry*, **419**(2), 266-270.

Appendices

Appendix A: Media compositions

All media were heat sterilised at 121 °C for 20 minutes. In instances where heating was potentially degradative to a certain component, individual chemicals were added after heating and sterilised by filtration [0.2 μM Minisart syringe filters (Sartorius AG, Germany)].

LB media

10 $g.L^{-1}$ NaCl, 10 $g.L^{-1}$ tryptone, 5 $g.L^{-1}$ yeast extract

LB glucose media

10 $g.L^{-1}$ NaCl, 10 $g.L^{-1}$ tryptone, 5 $g.L^{-1}$ yeast extract, 2 $g.L^{-1}$ glucose

Agar recipe

LB/LB + glucose media was set with the addition of 15 $g.L^{-1}$ agar.

Appendix B: GMO details

B.1: Cell strains

Table B.9.1: Cell strains used in study.

DH5α	fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17

B.2: Plasmids

Table B.9.2: Plasmids used in study.

pPROEX Htb	*E. coli* expression vector; N-terminal hexa-His tag; Amp$_r$

B.3: GMOs used and generated in study

Table B.9.3: GMO's used and generated in this study.

Name	**Plasmid**	**Cell strain**	**Source**
MalL WT	pPROEX	DH5α	(Hobbs *et al.*, 2013)
MalL V200S	pPROEX	DH5α	(Hobbs *et al.*, 2013)
MalL G202P	pPROEX	DH5α	(Hobbs *et al.*, 2013)
GK	pPROEX	DH5α	Generated
PGI	pPROEX	DH5α	Generated
PFK	pPROEX	DH5α	Generated
FBPA	pPROEX	DH5α	Generated
TPI	pPROEX	DH5α	Generated
GAPD	pPROEX	DH5α	Generated

B.4: Gene and protein sequences of GMOs used in this study

MalL WT and variants | pPROEX | N-terminal hexa-his tag

Nucleotide sequence – WT | 1771 nucleotides

ATGTCGTACTACCATCACCATCACCATCACGATTACGATATCCCAACG

ACCGAAAACCTGTATTTTCAGGGCGCGGCGCGCCCATGGGTGAATGG

TGGAAAGAAGCCGTTGTTTATCAGATTTATCCGCGTAGCTTTTATGAT

GCCAATGGTGATGGTTTTGGTGATCTGCAAGGTGTTATTCAGAAACTG

GATTACATCAAAAATCTGGGTGCCGATGTGATTTGGCTGAGTCCGGTT

TTTGATAGTCCGCAGGATGATAATGGTTATGATATTAGCGATTACAAA

AACATGTATGAAAAATTTGGCACCAATGAAGATATGTTTCAGCTGATT

GATGAAGTGCATAAACGCGGTATGAAAATTGTTATGGATCTGGTGGTT

AATCATACCAGTGATGAACATGCATGGTTTGCAGAAAGCCGTAAAAGC

AAAGATAATCCGTATCGCGATTATTATCTGTGGAAAGATCCGAAACCG

GATGGTAGCGAACCGAATAATTGGGGTAGCATTTTTAGCGGTAGCGCA

TGGACCTATGATGAAGGCACCGGTCAGTATTATCTGCATTACTTTAGC

AAAAAACAGCCGGATCTGAATTGGGAAAATGAAGCAGTTCGTCGTGAA

GTGTATGATGTTATGCGTTTTTGGATGGATCGTGGTGTTGATGGTTGG

CGTATGGAT**GTT**ATT**GGT**AGCATTAGCAAATATACCGATTTTCCGGAT

TATGAAACCGATCATAGCCGTAGCTATATTGTGGGTCGTTATCATAGC

AATGGTCCGCGTCTGCATGAATTTATTCAGGAAATGAATCGCGAAGTT

CTGAGCCATTATGATTGTATGACCGTTGGTGAAGCAAATGGCAGCGAT

ATTGAAGAAGCCAAAAAATACACAGATGCCAGTCGCCAAGAACTGAAT

ATGATTTTTACCTTTGAACATATGGATATTGATAAAGAACAGAATAGC

CCGAATGGCAAATGGCAGATTAAACCGTTTGATCTGATTGCCCTGAAA

AAAACCATGACCCGTTGGCAGACAGGTCTGATGAATGTTGGTTGGAAT

ACCCTGTATTTTGAAAATCATGATCAGCCTCGTGTTATTAGCCGTTGG

GGTAATGATCGTAAACTGCGTAAAGAATGTGCAAAAGCATTTGCAACC

GTTCTGCATGGTATGAAAGGCACCCCGTTTATCTATCAGGGTGAAGAA

ATTGGTATGGTGAATAGCGATATGCCGCTGGAAATGTATGATGATCTG

GAAATCAAAAATGCCTATCGCGAACTGGTGGTGGAAAACAAAACCATG

AGCGAAAAAGAATTTGTGAAAGCCGTGATGATTAAAGGTCGTGATCAT

GCACGTACCCCGATGCAGTGGGATGCAGGTAAACATGCAGGTTTTACC

GCAGGCGATCCGTGGATTCCGGTTAATAGCCGTTATCAGGATATTAAT
GTGAAAGAAAGCCTGGAAGATCAGGATAGCATTTTCTTTTACTATCAG
AAACTGATTCAGCTGCGCAAACAGTATAAAATCATGATTTATGGCGAT
TATCAGCTGCTGCAAGAAAATGATCCGCAGGTTTTTAGCTATCTGCGT
GAATATCGTGGTGAAAAACTGCTGGTTGTTGTTAATCTGAGCGAAGAA
AAAGCACTGTTTGAAGCACCTCCGGAACTGATTCATGAACGTTGGAAA
GTTCTGATTAGCAATTATCCGCAGGAACGTGCAGATCTGAAAAGCATT
AGCCTGAAACCGTATGAAGCAGTTATGGGCATTAGCATCTAA

Mutation positions for variants V200S and G202P are bolded and underlined. Mutations in these codons are GTT → TCC for V200S and GGT → CCU/A for G202P.

Protein sequence – WT | 69.47 kDa | 589 amino acids | Ɛ = 2.2 M^{-1}.L^{-1}

MSYYHHHHHHDYDIPTTENLYFQGAARPMGEWWKEAVVYQIYPRSFYD
ANGDGFGDLQGVIQKLDYIKNLGADVIWLSPVFDSPQDDNGYDISDYK
NMYEKFGTNEDMFQLIDEVHKRGMKIVMDLVVNHTSDEHAWFAESRKS
KDNPYRDYYLWKDPKPDGSEPNNWGSIFSGSAWTYDEGTGQYYLHYFS
KKQPDLNWENEAVRREVYDVMRFWMDRGVDGWRMD**V**I**G**SISKYTDFPD
YETDHSRSYIVGRYHSNGPRLHEFIQEMNREVLSHYDCMTVGEANGSD
IEEAKKYTDASRQELNMIFTFEHMDIDKEQNSPNGKWQIKPFDLIALK
KTMTRWQTGLMNVGWNTLYFENHDQPRVISRWGNDRKLRKECAKAFAT
VLHGMKGTPFIYQGEEIGMVNSDMPLEMYDDLEIKNAYRELVVENKTM
SEKEFVKAVMIKGRDHARTPMQWDAGKHAGFTAGDPWIPVNSRYQDIN
VKESLEDQDSIFFYYQKLIQLRKQYKIMIYGDYQLLQENDPQVFSYLR
EYRGEKLLVVVNLSEEKALFEAPPELIHERWKVLISNYPQERADLKSI
SLKPYEAVMGISI*

Positions 200 and 202 are bolded and underlined to illustrate the positions of V200S and G202P mutations.

GK | pPROEX | N-terminal hexa-his tag

Nucleotide sequence | 1050 nucleotides

ATGTCGTACTACCATCACCATCACCATCACGATTACGATATCCCAACG

ACCGAAAACCTGTATTTTCAGGGCGCCATGGGATCCATGACAAAGTAT

GCATTAGTCGGTGATGTGGGCGGCACCAACGCACGTCTTGCTCTGTGT

GATATTGCCAGTGGTGAAATCTCGCAGGCTAAGACCTATTCAGGGCTT

GATTACCCCAGCCTCGAAGCGGTCATTCGCGTTTATCTTGAAGAACAT

AAGGTCGAGGTGAAAGACGGCTGTATTGCCATCGCTTGCCCAATTACC

GGTGACTGGGTGGCGATGACCAACCATACCTGGGCGTTCTCAATTGCC

GAAATGAAAAAGAATCTCGGTTTTAGCCATCTGGAAATTATTAACGAT

TTTACCGCTGTATCGATGGCGATCCCGATGCTGAAAAAAGAGCATCTG

ATTCAGTTTGGTGGCGCAGAACCGGTCGAAGGTAAGCCTATTGCGGTT

TACGGTGCCGGAACGGGGCTTGGGGTTGCGCATCTGGTCCATGTCGAT

AAGCGTTGGGTAAGCTTGCCAGGCGAAGGCGGTCACGTTGATTTTGCG

CCGAATAGTGAAGAAGAGGCCATTATCCTCGAAATATTGCGTGCGGAA

ATTGGTCATGTTTCGGCGGAGCGCGTGCTTTCTGGCCCTGGGCTGGTG

AATTTGTATCGCGCAATTGTGAAAGCTGACAACCGCCTGCCAGAAAAT

CTCAAGCCAAAAGATATTACCGAACGCGCGCTGGCTGACAGCTGCACC

GATTGCCGCCGCGCATTGTCGCTGTTTTGCGTCATTATGGGCCGTTTT

GGCGGCAATCTGGCGCTCAATCTCGGGACATTTGGCGGCGTGTTTATT

GCGGGCGGTATCGTGCCGCGCTTCCTTGAGTTCTTCAAAGCCTCCGGT

TTCCGTGCCGCATTTGAAGATAAAGGGCGCTTTAAAGAATATGTCCAT

GATATTCCGGTGTATCTCATCGTCCATGACAATCCGGGCCTTCTCGGT

TCCGGTGCACATTTACGCCAGACCTTAGGTCACATTCTGTAA

Protein sequence | 38.09 kDa | 349 amino acids | Ɛ = 0.9 $M^{-1}.L^{-1}$

MSYYHHHHHHDYDIPTTENLYFQGAMGSMTKYALVGDVGGTNARLALC

DIASGEISQAKTYSGLDYPSLEAVIRVYLEEHKVEVKDGCIAIACPIT

GDWVAMTNHTWAFSIAEMKKNLGFSHLEIINDFTAVSMAIPMLKKEHL

IQFGGAEPVEGKPIAVYGAGTGLGVAHLVHVDKRWVSLPGEGGHVDFA

PNSEEEAIILEILRAEIGHVSAERVLSGPGLVNLYRAIVKADNRLPEN

LKPKDITERALADSCTDCRRALSLFCVIMGRFGGNLALNLGTFGGVFI

AGGIVPRFLEFFKASGFRAAFEDKGRFKEYVHDIPVYLIVHDNPGLLG

SGAHLRQTLGHIL*

PGI F433L | pPROEX | N-terminal hexa-his tag

Nucleotide sequence | 1734 nucleotides

ATGTCGTACTACCATCACCATCACCATCACGATTACGATATCCCAACG

ACCGAAAACCTGTATTTTCAGGGCGCCATGGGATCCATGAAAAACATC

AATCCAACGCAGACCGCTGCCTGGCAGGCACTACAGAAACACTTCGAT

GAAATGAAAGACGTTACGATCGCCGATCTTTTTGCTAAAGACGGCGAT

CGTTTTTCTAAGTTCTCCGCAACCTTCGACGATCAGATGCTGGTGGAT

TACTCCAAAAACCGCATCACTGAAGAGACGCTGGCGAAATTACAGGAT

CTGGCGAAAGAGTGCGATCTGGCGGGCGCGATTAAGTCGATGTTCTCT

GGCGAGAAGATCAACCGCACTGAAAACCGCGCCGTGCTGCACGTAGCG

CTGCGTAACCGTAGCAATACCCCGATTTTGGTTGATGGCAAAGACGTA

ATGCCGGAAGTCAACGCGGTGCTGGAGAAGATGAAAACCTTCTCAGAA

GCGATTATTTCCGGTGAGTGGAAAGGTTATACCGGCAAAGCAATCACT

GACGTAGTGAACATCGGGATCGGCGGTTCTGACCTCGGCCCATACATG

GTGACCGAAGCTCTGCGTCCGTACAAAAACCACCTGAACATGCACTTT

GTTTCTAACGTCGATGGGACTCACATCGCGGAAGTGCTGAAAAAAGTA

AACCCGGAAACCACGCTGTTCCTGGTAGCATCTAAAACCTTCACCACT

CAGGAAACTATGACCAACGCCCATAGCGCGCGTGACTGGTTCCTGAAA

GCGGCAGGTGATGAAAAACACGTTGCAAAACACTTTGCGGCGCTTTCC

ACCAATGCCAAAGCCGTTGGCGAGTTTGGTATTGATACTGCCAACATG

TTCGAGTTCTGGGACTGGGTTGGCGGCCGTTACTCTTTGTGGTCAGCG

ATTGGCCTGTCGATTGTTCTCTCCATCGGCTTTGATAACTTCGTTGAA

CTGCTTTCCGGCGCACACGCGATGGACAAGCATTTCTCCACCACGCCT

GCCGAGAAAAACCTGCCTGTACTGCTGGCGCTGATTGGCATCTGGTAC

AACAATTTCTTTGGTGCGGAAACTGAAGCGATTCTGCCGTATGACCAG

TATATGCACCGTTTCGCGGCGTACTTCCAGCAGGGCAATATGGAGTCC

AACGGTAAGTATGTTGACCGTAACGGTAACGTTGTGGATTACCAGACT

GGCCCGATTATCTGGGGTGAACCAGGCACTAACGGTCAGCACGCGTTC

TACCAGCTGATCCACCAGGGAACCAAAATGGTACCGTGCGATTTCATC

GCTCCGGCTATCACCCATAACCCGCTCTCTGATCATCACCAGAAACTG

CTGTCTAACTTCTTCGCCCAGACCGAAGCGCTGGCG**CTT**GGTAAATCC

CGCGAAGTGGTTGAGCAGGAATATCGTGATCAGGGTAAAGATCCGGCA
ACGCTTGACTACGTGGTGCCGTTCAAAGTATTCGAAGGTAACCGCCCG
ACCAACTCCATCCTGCTGCGTGAAATCACTCCGTTCAGCCTGGGTGCG
TTGATTGCGCTGTATGAGCACAAAATCTTTACTCAGGGCGTGATCCTG
AACATCTTCACCTTCGACCAGTGGGGCGTGGAACTGGGTAAACAGCTG
GCGAACCGTATTCTGCCAGAGCTGAAAGATGATAAAGAAATCAGCAGC
CACGATAGCTCGACCAATGGTCTGATTAACCGCTATAAAGCGTGGCGC
GGTTAA

The mutated codon is bolded and highlighted. The original codon for this in the WT PGI sequence is TTT.

Protein sequence | 64.90 kDa | 577 amino acids | Ɛ = 1.3 $M^{-1}.L^{-1}$

MSYYHHHHHHDYDIPTTENLYFQGAMGSMKNINPTQTAAWQALQKHFD
EMKDVTIADLFAKDGDRFSKFSATFDDQMLVDYSKNRITEETLAKLQD
LAKECDLAGAIKSMFSGEKINRTENRAVLHVALRNRSNTPILVDGKDV
MPEVNAVLEKMKTFSEAIISGEWKGYTGKAITDVVNIGIGGSDLGPYM
VTEALRPYKNHLNMHFVSNVDGTHIAEVLKKVNPETTLFLVASKTFTT
QETMTNAHSARDWFLKAAGDEKHVAKHFAALSTNAKAVGEFGIDTANM
FEFWDWVGGRYSLWSAIGLSIVLSIGFDNFVELLSGAHAMDKHFSTTP
AEKNLPVLLALIGIWYNNFFGAETEAILPYDQYMHRFAAYFQQGNMES
NGKYVDRNGNVVDYQTGPIIWGEPGTNGQHAFYQLIHQGTKMVPCDFI
APAITHNPLSDHHQKLLSNFFAQTEALA**L**GKSREVVEQEYRDQGKDPA
TLDYVVPFKVFEGNRPTNSILLREITPFSLGALIALYEHKIFTQGVIL
NIFTFDQWGVELGKQLANRILPELKDDKEISSHDSSTNGLINRYKAWR
G*

The mutated amino acid (F433L) is bolded and underlined.

PFK | pPROEX | N-terminal hexa-his tag

Nucleotide sequence | 1047 nucleotides

ATGTCGTACTACCATCACCATCACCATCACGATTACGATATCCCAACG
ACCGAAAACCTGTATTTTCAGGGCGCCATGGGATCCATGATTAAGAAA
ATCGGTGTGTTGACAAGCGGCGGTGATGCGCCAGGCATGAACGCCGCA
ATTCGCGGGGTTGTTCGTTCTGCGCTGACAGAAGGTCTGGAAGTAATG
GGTATTTATGACGGCTATCTGGGTCTGTATGAAGACCGTATGGTACAG
CTAGACCGTTACAGCGTTTCTGACATGATCAACCGTGGCGGTACGTTC
CTCGGTTCTGCGCGTTTCCCGGAATTCCGCGACGAGAACATCCGCGCC
GTGGCTATCGAAAACCTGAAAAAACGTGGTATCGACGCGCTGGTGGTT
ATCGGCGGTGACGGTTCCTACATGGGTGCAATGCGTCTGACCGAAATG
GGCTTCCCGTGCATCGGCCTGCCGGGCACTATCGACAACGACATCAAA
GGCACTGACTACACTATCGGTTTCTTCACTGCGCTGAGCACCGTTGTA
GAAGCGATCGACCGTCTGCGTGACACCTCTTCTTCTCACCAGCGTATT
TCCGTGGTGGAAGTGATGGGCCGTTATTGTGGCGATCTGACGTTGGCT
GCGGCCATTGCCGGTGGCTGTGAATTCGTTGTGGTTCCGGAAGTTGAA
TTCAGCCGTGAAGACCTGGTAAACGAAATCAAAGCGGGTATCGCGAAA
GGTAAAAAACACGCGATCGTGGCGATTACCGAACATATGTGTGATGTT
GACGAACTGGCGCATTTCATCGAGAAAGAAACCGGTCGTGAAACCCGC
GCAACTGTGCTGGGCCACATCCAGCGCGGTGGTTCTCCGGTGCCTTAC
GACCGTATTCTGGCTTCCCGTATGGGCGCTTACGCTATCGATCTGCTG
CTGGCAGGTTACGGCGGTCGTTGCGTAGGTATCCAGAACGAACAGCTG
GTTCACCACGACATCATCGACGCTATCGAAAACATGAAGCGTCCGTTC
AAAGGCGACTGGCTAGACTGCGCGAAAAAACTGTATTAA

Protein sequence | 38.21 kDa | 348 amino acids | Ɛ = 0.7 $M^{-1}.L^{-1}$

MSYYHHHHHHDYDIPTTENLYFQGAMGSMIKKIGVLTSGGDAPGMNAA
IRGVVRSALTEGLEVMGIYDGYLGLYEDRMVQLDRYSVSDMINRGGTF
LGSARFPEFRDENIRAVAIENLKKRGIDALVVIGGDGSYMGAMRLTEM
GFPCIGLPGTIDNDIKGTDYTIGFFTALSTVVEAIDRLRDTSSSHQRI
SVVEVMGRYCGDLTLAAAIAGGCEFVVVPEVEFSREDLVNEIKAGIAK
GKKHAIVAITEHMCDVDELAHFIEKETGRETRATVLGHIQRGGSPVPY

DRILASRMGAYAIDLLLAGYGGRCVGIQNEQLVHHDIIDAIENMKRPF
KGDWLDCAKKLY*

FBPA | pPROEX | N-terminal hexa-his tag

Nucleotide sequence | 1164 nucleotides

ATGTCGTACTACCATCACCATCACCATCACGATTACGATATCCCAACG
ACCGAAAACCTGTATTTTCAGGGCGCCATGGGATCCATGTCTAAGATT
TTTGATTTCGTAAAACCTGGCGTAATCACTGGTGATGACGTACAGAAA
GTTTTCCAGGTAGCAAAAGAAAACAACTTCGCACTGCCAGCAGTAAAC
TGCGTCGGTACTGACTCCATCAACGCCGTACTGGAAACCGCTGCTAAA
GTTAAAGCGCCGGTTATCGTTCAGTTCTCCAACGGTGGTGCTTCCTTT
ATCGCTGGTAAAGGCGTGAAATCTGACGTTCCGCAGGGTGCTGCTATC
CTGGGCGCGATCTCTGGTGCGCATCACGTTCACCAGATGGCTGAACAT
TATGGTGTTCCGGTTATCCTGCACACTGACCACTGCGCGAAGAAACTG
CTGCCGTGGATCGACGGTCTGTTGGACGCGGGTGAAAAACACTTCGCA
GCTACCGGTAAGCCGCTGTTCTCTTCTCACATGATCGACCTGTCTGAA
GAATCTCTGCAAGAGAACATCGAAATCTGCTCTAAATACCTGGAGCGC
ATGTCCAAAATCGGCATGACTCTGGAAATCGAACTGGGTTGCACCGGT
GGTGAAGAAGACGGCGTGGACAACAGCCACATGGACGCTTCTGCACTG
TACACCCAGCCGGAAGACGTTGATTACGCATACACCGAACTGAGCAAA
ATCAGCCCGCGTTTCACCATCGCAGCGTCCTTCGGTAACGTACACGGT
GTTTACAAGCCGGGTAACGTGGTTCTGACTCCGACCATCCTGCGTGAT
TCTCAGGAATATGTTTCCAAGAAACACAACCTGCCGCACAACAGCCTG
AACTTCGTATTCCACGGTGGTTCCGGTTCTACTGCTCAGGAAATCAAA
GACTCCGTAAGCTACGGCGTAGTAAAAATGAACATCGATACCGATACC
CAATGGGCAACCTGGGAAGGCGTTCTGAACTACTACAAAGCGAACGAA
GCTTATCTGCAGGGTCAGCTGGGTAACCCGAAAGGCGAAGATCAGCCG
AACAAGAAATACTACGATCCGCGCGTATGGCTGCGTGCCGGTCAGACT
TCGATGATCGCTCGTCTGGAGAAAGCATTCCAGGAACTGAACGCGATC
GACGTTCTGTAA

Protein sequence | 42.52 kDa | 387 amino acids | Ɛ = 1.1 $M^{-1}.L^{-1}$

```
MSYYHHHHHHDYDIPTTENLYFQGAMGSMSKIFDFVKPGVITGDDVQK
VFQVAKENNFALPAVNCVGTDSINAVLETAAKVKAPVIVQFSNGGASF
IAGKGVKSDVPQGAAILGAISGAHHVHQMAEHYGVPVILHTDHCAKKL
LPWIDGLLDAGEKHFAATGKPLFSSHMIDLSEESLQENIEICSKYLER
MSKIGMTLEIELGCTGGEEDGVDNSHMDASALYTQPEDVDYAYTELSK
ISPRFTIAASFGNVHGVYKPGNVVLTPTILRDSQEYVSKKHNLPHNSL
NFVFHGGSGSTAQEIKDSVSYGVVKMNIDTDTQWATWEGVLNYYKANE
AYLQGQLGNPKGEDQPNKKYYDPRVWLRAGQTSMIARLEKAFQELNAI
DVL*
```

TPI | pPROEX | N-terminal hexa-his tag

Nucleotide sequence | 852 nucleotides

```
ATGTCGTACTACCATCACCATCACCATCACGATTACGATATCCCAACG
ACCGAAAACCTGTATTTTCAGGGCGCCATGGGATCCATGCGACATCCT
TTAGTGATGGGTAACTGGAAACTGAACGGCAGCCGCCACATGGTTCAC
GAGCTGGTTTCTAACCTGCGTAAAGAGCTGGCAGGTGTTGCTGGCTGT
GCGGTTGCAATCGCACCACCGGAAATGTACATCGATATGGCGAAGCGC
GAAGCTGAAGGCAGCCACATCATGCTGGGTGCGCAAAACGTGGACCTG
AACCTGTCCGGCGCATTCACCGGTGAAACCTCTGCTGCTATGCTGAAA
GACATCGGCGCACAGTACATCATCATCGGCCACTCTGAACGTCGTACT
TACCACAAAGAGTCTGACGAACTGATCGCGAAAAAATTCGCGGTGCTG
AAAGAGCAGGGCCTGACTCCGGTTCTGTGCATCGGTGAAACCGAAGCT
GAAAACGAAGCGGGCAAAACTGAAGAAGTTTGCGCACGTCAGATCGAC
GCGGTACTGAAAACTCAGGGTGCTGCGGCATTCGAAGGTGCGGTTATC
GCTTACGAACCCGTATGGGCAATCGGTACTGGCAAATCTGCAACTCCG
GCTCAGGCACAGGCTGTTCACAAATTCATTCGTGACCACATCGCTAAA
GTTGACGCTAACATCGCTGAACAAGTGATCATTCAGTACGGCGGCTCT
GTAAACGCGTCTAACGCTGCAGAACTGTTTGCTCAGCCAGACATCGAC
GGCGCGCTGGTTGGCGGTGCTTCTCTGAAAGCTGACGCTTTCGCAGTA
ATCGTTAAAGCTGCAGAAGCGGCTAAACAGGCTTAA
```

Protein sequence | 30.34 kDa | 283 amino acids | Ɛ = 0.8 $M^{-1}.L^{-1}$

MSYYHHHHHHDYDIPTTENLYFQGAMGSMRHPLVMGNWKLNGSRHMVH
ELVSNLRKELAGVAGCAVAIAPPEMYIDMAKREAEGSHIMLGAQNVDL
NLSGAFTGETSAAMLKDIGAQYIIIGHSERRTYHKESDELIAKKFAVL
KEQGLTPVLCIGETEAENEAGKTEEVCARQIDAVLKTQGAAAFEGAVI
AYEPVWAIGTGKSATPAQAQAVHKFIRDHIAKVDANIAEQVIIQYGGS
VNASNAAELFAQPDIDGALVGGASLKADAFAVIVKAAEAAKQA*

GAPD | pPROEX | N-terminal hexa-his tag

Nucleotide sequence | 1080 nucleotides

ATGTCGTACTACCATCACCATCACCATCACGATTACGATATCCCAACG
ACCGAAAACCTGTATTTTCAGGGCGCCATGGGATCCATGACTATCAAA
GTAGGTATCAACGGTTTTGGCCGTATCGGTCGCATTGTTTTCCGTGCT
GCTCAGAAACGTTCTGACATCGAGATCGTTGCAATCAACGACCTGTTA
GACGCTGATTACATGGCATACATGCTGAAATATGACTCCACTCACGGC
CGTTTCGACGGTACCGTTGAAGTGAAAGACGGTCATCTGATCGTTAAC
GGTAAAAAAATCCGTGTTACCGCTGAACGTGATCCGGCTAACCTGAAA
TGGGACGAAGTTGGTGTTGACGTTGTCGCTGAAGCAACTGGTCTGTTC
CTGACTGACGAAACTGCTCGTAAACACATCACCGCTGGTGCGAAGAAA
GTGGTTATGACTGGTCCGTCTAAAGACAACACTCCGATGTTCGTTAAA
GGCGCTAACTTCGACAAATATGCTGGCCAGGACATCGTTTCCAACGCT
TCCTGCACCACCAACTGCCTGGCTCCGCTGGCTAAAGTTATCAACGAT
AACTTCGGCATCATCGAAGGTCTGATGACCACCGTTCACGCTACTACC
GCTACTCAGAAAACCGTTGATGGCCCGTCTCACAAAGACTGGCGCGGC
GGCCGCGGCGCTTCCCAGAACATCATCCCGTCCTCTACCGGTGCTGCT
AAAGCTGTAGGTAAAGTACTGCCAGAACTGAATGGCAAACTGACTGGT
ATGGCGTTCCGCGTTCCGACCCCGAACGTATCTGTAGTTGACCTGACC
GTTCGTCTGGAAAAAGCTGCAACTTACGAGCAGATCAAAGCTGCCGTT
AAAGCTGCTGCTGAAGGCGAAATGAAAGGCGTTCTGGGCTACACCGAA
GATGACGTAGTATCTACCGATTTCAACGGCGAAGTTTGCACTTCCGTG
TTCGATGCTAAAGCTGGTATCGCTCTGAACGACAACTTCGTGAAACTG
GTATCCTGGTACGACAACGAAACCGGTTACTCCAACAAAGTTCTGGAC
CTGATCGCTCACATCTCCAAATAA

Protein sequence | 38.90 kDa | 359 amino acids | Ɛ = 0.9 $M^{-1}.L^{-1}$

MSYYHHHHHHDYDIPTTENLYFQGAMGSMTIKVGINGFGRIGRIVFRA
AQKRSDIEIVAINDLLDADYMAYMLKYDSTHGRFDGTVEVKDGHLIVN
GKKIRVTAERDPANLKWDEVGVDVVAEATGLFLTDETARKHITAGAKK
VVMTGPSKDNTPMFVKGANFDKYAGQDIVSNASCTTNCLAPLAKVIND
NFGIIEGLMTTVHATTATQKTVDGPSHKDWRGGRGASQNIIPSSTGAA
KAVGKVLPELNGKLTGMAFRVPTPNVSVVDLTVRLEKAATYEQIKAAV
KAAAEGEMKGVLGYTEDDVVSTDFNGEVCTSVFDAKAGIALNDNFVKL
VSWYDNETGYSNKVLDLIAHISK*

B.5: Primer sequences used in study

Glycolysis cloning primers

Table B.5I: Cloning primers used to extract glycolysis genes from BL21 genomic DNA. Restriction cut sites are indicated in bold.

Gene	**Forward primer (5’-3’)**	**Reverse primer (5’-3’)**	**Restriction sites**
GK	CGC**GGATCC**ATGACAAAGTATGCATTAGT	AAAA**CTGCAG**TTACAGAATGTGACCTAAGG	BamHI/pstI
PGI	CGC**GGATCC**ATGAAAAACATCAATCCAAC	AAAA**CTGCAG**TTAACCGCGCCACGCTTTAT	BamHI/pstI
PFK_I	CGC**GGATCC**ATGATTAAGAAAATCGGTGT	AAAA**CTGCAG**TTAATACAGTTTTTTCGCGC	BamHI/pstI
FBPA_II	CGC**GGATCC**ATGTCTAAGATTTTTGATTT	CCG**CTCGAG**TTACAGAACGTCGATCGCGT	BamHI/XhoI
TPI	CGC**GGATCC**ATGCGACATCCTTTAGTGAT	CCC**AAGCTT**TTAAGCCTGTTTAGCCGCTT	BamHI/HindIII
GAPD	CGC**GGATCC**ATGACTATCAAAGTAGGTAT	AAAA**CTGCAG**TTATTTGGAGATGTGAGCGA	BamHI/pstI

Vector sequencing primers

Table B.5II: Vector specific sequencing primers used in this study.

Primer	**Forward primer (5’-3’)**	**Reverse primer (5’-3’)**
pPROEX	GCGGATAACAATTTCACACA	ATCTTCTCTCATCCGCCAAA

Appendix C: Summarised protein purification and characterisation buffers

Table C.1: Summarised purification and assay buffers for enzymes characterised in this study.

Enzyme	**MalL variants**	**Glycolysis enzymes**
IMAC lysis buffer	10 mM KCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 25 mM imidazole, pH 7.0	50 mM HEPES, 25 mM imidazole, 150 mM NaCl, pH 7.4
IMAC elution buffer	10 mM KCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 1 M imidazole, pH 7.0	50 mM HEPES, 1 M imidazole, 150 mM NaCl, pH 7.4
Size exclusion buffer	20 mM HEPES, pH 7.0	50 mM HEPES, 150 mM NaCl, 10 mM $MgCl_2$, pH 7.4
Assay buffer	200 mM sodium phosphate buffer, 150 mM NaCl, pH 6.5	50 mM HEPES, 150 mM NaCl, 15 mM $MgCl_2$, 2.5 mM $(NH_4)_2SO_4$, pH 7.4

Appendix D: Supplementary data

D.1: Enzyme size and efficiency graph details

Table D.1: Catalyst size versus efficiency details and references pertaining to Figure 3.3.

Enzyme	Organism	MW of the catalytic domain (kDa)	$ln(k_{cat}/k_{non})$	MW source	References
Acetylcholinesterase	*Torpedo californica*	65.9	30.7	1AMN	(Harel *et al.*, 1996; Britt, 1997)
Glucoamylase	*Aspergillus niger*	54.1	26.5	3EQA	(Olsen *et al.*, 1992; Britt, 1997)
Lysozyme	*Gallus gallus*	16.2	20.6	4I8S	(Chipman, 1971; Britt, 1997)
Cytidine deaminase	*Escherichia coli*	31.5	26.6	1ALN	(Frick *et al.*, 1987; Britt, 1997)
Chorismate mutase	*Enterobacter aerogenes*	17.5	14.4	2GBB	(Andrews *et al.*, 1973; Britt, 1997)
Urease	*Klebsiella aerogenes*	41.0	32.2	(Balasubramanian *et al.*, 2013)	(Britt, 1997; Pearson *et al.*, 2000)
Adenosine deaminase	*Rattus*	40.0	16.7	3KM8	(Frick *et al.*, 1987; Britt, 1997)
Triosephosphate isomerase	*Gallus gallus*	26.6	21.0	8TIM	(Hall & Knowles, 1975; Britt, 1997)
Carbonic anhydrase	*Bovine*	28.8	20.7	1V9E	(Pocker & Meany, 1967; Britt, 1997)
Fumarase	*Sus*	50.0	35.7	3EO4	(Brant *et al.*, 1963; Bearne & Wolfenden, 1995; Britt, 1997)
Arginine decarboxylase	*Escherichia coli*	69.1	45.7	2VYC	(Blethen *et al.*, 1968; Snider & Wolfenden, 2000; Wolfenden & Snider, 2001)
OMP decarboxylase	*Saccharomyces cerevisiae*	58.4	39.5	3GDL	(Bell & Jones, 1991; Radzicka & Wolfenden, 1995; Wolfenden & Snider, 2001)
Beta-amylase	*Ipomoea batatas*	56.1	41.1	1FA2	(Balls *et al.*, 1948; Wolfenden *et al.*, 1998; Wolfenden & Snider, 2001)
Mandelate racemase	*Pseudomonas aeruginosa*	44.2	35.0	1MDL	(Mitra *et al.*, 1995; Bearne & Wolfenden, 1997; Wolfenden & Snider, 2001)
Peptidyl-dipeptidase A	*Bovine*	70.1	30.2	2X8Z	(Stevens *et al.*, 1972; Radzicka & Wolfenden, 1996; Wolfenden & Snider, 2001)

Table D.1 continued.

Enzyme	**Organism**	**MW of the catalytic domain (kDa)**	***ln*(k_{cat} /k_{non})**	**MW source**	**References**
Ketosteroid isomerase	*Pseudomonas testosteroni*	14.5	26.7	3VGN	(Pollack *et al.*, 1989; Hawkinson *et al.*, 1991; Wolfenden & Snider, 2001)
Arylsulfatase	*Pseudomonas aeruginosa*	59.9	24.1	1HDH	(Olguin *et al.*, 2008; Edwards *et al.*, 2011)
AMP nucleosidase	*Azotobacter vinelandii*	38.6	29.4	1T8S	(Garrett & Mehta, 1972; DeWolf *et al.*, 1979; Radzicka & Wolfenden, 1995)
Phosphotriesterase	*Pseudomonas diminuta*	39.0	26.4	4LE6	(Dumas *et al.*, 1989; Radzicka & Wolfenden, 1995)
Barnase	*Streptomyces aureofaciens*	17.1	24.6	3DGY	(Wolfenden & Snider, 2001; Makarov *et al.*, 2004)
Acylphosphatase	*Sulfolobus solfataricus*	11.6	11.9	2BJD	(Paoli *et al.*, 1997; Corazza *et al.*, 2006; Elias *et al.*, 2014)
Prephenate dehydrogenase	*Methanocaldococcus jannaschii*	31.4	12.2	4LUB	(Kleeb *et al.*, 2006; Elias *et al.*, 2014)

Ribozyme	**Organism**	**MW of the catalytic domain (Da)**	***ln*(k_{cat} /k_{non})**	**MW source**	**References**
Leadzyme	In vitro evolution	4.1	12.3	(Ohmichi *et al.*, 1998)	(Wolfenden & Snider, 2001)
Hammerhead ribozyme	Artificial construct	15.9	15.3	(Fedor & Uhlenbeck, 1992)	(Stage-Zimmermann & Uhlenbeck, 1998; Wolfenden & Snider, 2001)

D.2: Glycolysis assay details and summarised results

Table D.2: Specific set up details for glycolysis characterization assays of individual enzymes and pathways pertaining to Section 4.3. MM = Michaelis Menten.

Enzyme	Parameter	T (K)	Assay details	Value	
G6PDH	MM, NADP$^+$	310	G6PDH 0.003125 mg.ml^{-1}, G6P 5 mM	K_M	0.11 ± 0.04 mM
				k_{cat}	900 ± 60 s^{-1}
	MM, G6P	310	G6PDH 0.003125 mg.ml^{-1}, NADP$^+$ 1 mM	K_M	0.16 ± 0.05 mM
				k_{cat}	830 ± 50 s^{-1}
	MM, NADP$^+$	317	G6PDH 0.0015625 mg.ml^{-1}, G6P 2 mM	K_M	0.11 ± 0.02 mM
				k_{cat}	1440 ± 50 s^{-1}
	MM, G6P	317	G6PDH 0.0015625 mg.ml^{-1}, NADP$^+$ 1 mM	K_M	0.35 ± 0.02 mM
				k_{cat}	1550 ± 30 s^{-1}
	T_{opt}, rough	-	G6PDH 0.0015625 mg.ml^{-1}, G6P 1.6 mM, NADP$^+$ 1.1 mM	T_{opt}	~ 323 K
GK	MM, ATP	310	GK 0.001875 mg.ml^{-1}, Glc 1.5 mM, G6PDH 0.00625 mg.ml^{-1}, NADP$^+$ 1.1 mM	$K_{½}$	1.7 ± 0.4 mM
				h	1.2 ± 0.2
				k_{cat}	66 ± 7 s^{-1}
	MM, Glc	310	GK 0.001875 mg.ml^{-1}, ATP 6 mM, G6PDH 0.00625 mg.ml^{-1}, NADP$^+$ 1.1 mM	K_M	0.11 ± 0.02 mM
				K_i	50 ± 30 mM
				k_{cat}	65 ± 3 s^{-1}
	MM, ATP	317	GK 0.0009375 mg.ml^{-1}, Glc 4 mM, G6PDH 0.003125 mg.ml^{-1}, NADP$^+$ 1 mM	$K_{½}$	1.3 ± 0.2 mM
				h	1.5 ± 0.2
				k_{cat}	86 ± 6 s^{-1}
	MM, Glc	317	GK 0.0009375 mg.ml^{-1}, ATP 8 mM, G6PDH 0.003125 mg.ml^{-1}, NADP$^+$ 1 mM	K_M	0.12 ± 0.02 mM
				K_i	30 ± 10 mM
				k_{cat}	64 ± 3 s^{-1}
	T_{opt}	-	GK 0.0009375 mg.ml^{-1}, Glc 1 mM, ATP 6 mM, G6PDH 0.003125 mg.ml^{-1}, NADP$^+$ 1.1 mM	T_{opt}	319 K
				$\Delta C_p^{\ddagger}$	-7.1 ± 0.5 kJ.mol^{-1}.K^{-1}
				$\Delta S^{\ddagger}$	-0.22 ± 0.02 kJ.mol^{-1}
				$\Delta H^{\ddagger}$	-0.5 ± 2.7 kJ.mol^{-1}
PGI	MM, F6P	310	PGI 0.003125 mg.ml^{-1}, G6PDH 0.0075 mg.ml^{-1}, NADP$^+$ 1.1 mM	K_M	1.42 ± 0.09 mM
				k_{cat}	1950 ± 60 s^{-1}
	MM, F6P	318	PGI 0.005625 mg.ml^{-1}, G6PDH 0.01125 mg.ml^{-1}, NADP$^+$ 1 mM	K_M	1.2 ± 0.1 mM
				k_{cat}	1540 ± 50 s^{-1}
	T_{opt}	-	PGI 0.0025 mg.ml^{-1}, F6P 7 mM, G6PDH 0.00625 mg.ml^{-1}, NADP$^+$ 1.1 mM	T_{opt}	321 K
				$\Delta C_p^{\ddagger}$	-4.0 ± 0.3 kJ.mol^{-1}.K^{-1}
				$\Delta S^{\ddagger}$	-0.189 ± 0.007 kJ.mol^{-1}
				$\Delta H^{\ddagger}$	-1.7 ± 2.2 kJ.mol^{-1}

Table D.2 continued.

Enzyme	Parameter	T (K)	Assay details	Value	
PFK	MM, ATP	310	PFK 0.025 mg.ml^{-1}, F6P 5 mM, FBPA 0.14875 mg.ml^{-1}, GAPD 0.0125 mg.ml^{-1}, NAD^{+} 10 mM, P_i 86 mM	K_M	1.1 ± 0.4 mM
				K_i	30 ± 20 mM
				k_cat	15 ± 2 s^{-1}
	MM, F6P	310	PFK 0.025 mg.ml^{-1}, ATP 5 mM, FBPA 0.14875 mg.ml^{-1}, GAPD 0.0125 mg.ml^{-1}, NAD^{+} 10 mM, P_i 86 mM	$K_{½}$	0.74 ± 0.09 mM
				h	1.7 ± 0.3
				k_cat	12.0 ± 0.8 s^{-1}
	MM, ATP	317	PFK 0.025 mg.ml^{-1}, F6P 5 mM, FBPA 0.14875 mg.ml^{-1}, GAPD 0.0125 mg.ml^{-1}, NAD^{+} 10 mM, P_i 86 mM	K_M	1.5 ± 0.4 mM
				K_i	30 ± 20 mM
				k_cat	22 ± 3 s^{-1}
	MM, F6P	317	PFK 0.025 mg.ml^{-1}, ATP 5 mM, FBPA 0.14875 mg.ml^{-1}, GAPD 0.0125 mg.ml^{-1}, NAD^{+} 10 mM, P_i 86 mM	$K_{½}$	0.60 ± 0.04 mM
				h	3.9 ± 0.8
				k_cat	15.5 ± 0.6 s^{-1}
	T_opt	-	PFK 0.025 mg.ml^{-1}, ATP 8 mM, F6P 12 mM, FBPA 0.14875 mg.ml^{-1}, GAPD 0.0125 mg.ml^{-1}, NAD^{+} 10 mM, P_i 83 mM	T_opt	338 K
				$\Delta C^{\ddagger}_\mathrm{p}$	-1.8 ± 0.6 kJ.mol^{-1}.K^{-1}
				$\Delta S^{\ddagger}$	-0.17 ± 0.001 kJ.mol^{-1}
				$\Delta H^{\ddagger}$	23 ± 4 kJ.mol^{-1}
FBPA	MM, FBP	310	FBPA 0.0265 mg.ml^{-1}, GAPD 0.077 mg.ml^{-1}, NAD^{+} 10 mM, P_i 83 mM	K_M	4.4 ± 0.4 mM
				K_i	43 ± 8 mM
				k_cat	15.1 ± 0.8 s^{-1}
	MM, FBP	317	FBPA 0.025 mg.ml^{-1}, GAPD 0.0625 mg.ml^{-1}, NAD^{+} 10 mM, P_i 86 mM	K_M	20 ± 10 mM
				K_i	9 ± 6 mM
				k_cat	50 ± 20 s^{-1}
	T_opt	-	FBPA 0.025 mg.ml^{-1}, FBP 25 mM, GAPD 0.05 mg.ml^{-1}, NAD^{+} 10 mM, P_i 46 mM	T_opt	339 K
				$\Delta C^{\ddagger}_\mathrm{p}$	-1.7 ± 0.6 kJ.mol^{-1}.K^{-1}
				$\Delta S^{\ddagger}$	-0.17 ± 0.01 kJ.mol^{-1}
				$\Delta H^{\ddagger}$	17 ± 3 kJ.mol^{-1}
TPI	MM, DHAP	310	TPI 0.00065 mg.ml^{-1}, GAPD 0.01594 mg.ml^{-1}, NAD^{+} 10 mM, P_i 46 mM	K_M	1.0 ± 0.1 mM
				k_cat	770 ± 20 s^{-1}
	MM, DHAP	317	TPI 0.0003125 mg.ml^{-1}, GAPD 0.0125 mg.ml^{-1}, NAD^{+} 10 mM, P_i 83 mM	Max activity at:	30-40 mM
				Max rate	~ 650 s^{-1}
	T_opt	-	TPI 0.0003086 mg.ml^{-1}, DHAP 10 mM, GAPD 0.03086 mg.ml^{-1}, NAD^{+} 10 mM, P_i 46 mM	T_opt	330 K
				$\Delta C^{\ddagger}_\mathrm{p}$	-2.4 ± 0.3 kJ.mol^{-1}.K^{-1}
				$\Delta S^{\ddagger}$	-0.187 ± 0.007 kJ.mol^{-1}
				$\Delta H^{\ddagger}$	-0.2 ± 2.3 kJ.mol^{-1}

Table D.2 continued.

Enzyme	Parameter	T (K)	Assay details	Value	
GAPD	MM, P_i	310	GAPD 0.003 mg.ml^{-1}, NAD$^+$ 15 mM, FBP 40 mM	K_M	13 ± 4 mM
				k_{cat}	70 ± 9 s^{-1}
	MM, NAD$^+$	310	GAPD 0.003 mg.ml^{-1}, P_i 83 mM, FBP 40 mM	K_M	0.9 ± 0.3 mM
				K_i	90 ± 70 mM
				k_{cat}	76 ± 9 s^{-1}
	MM, 'FBP'	310	GAPD 0.003 mg.ml^{-1}, P_i 83 mM, NAD$^+$ 10 mM	K_M	2.4 ± 0.3 mM
				K_i	240 ± 70 mM
				k_{cat}	88 ± 4 s^{-1}
	MM, P_i	317	GAPD 0.003 mg.ml^{-1}, NAD$^+$ 10 mM, FBP 10 mM	K_M	20 ± 10 mM
				k_{cat}	170 ± 30 s^{-1}
	MM, NAD$^+$	317	GAPD 0.003 mg.ml^{-1}, P_i 86 mM, FBP 10 mM	K_M	0.7 ± 0.1 mM
				K_i	170 ± 70 mM
				k_{cat}	137 ± 7 s^{-1}
	MM, 'FBP'	317	GAPD 0.003 mg.ml^{-1}, P_i 83 mM, NAD$^+$ 10 mM	K_M	2.3 ± 0.8 mM
				K_i	22 ± 8 mM
				k_{cat}	270 ± 50 s^{-1}
	T_{opt}	-	GAPD 0.003 mg.ml^{-1}, P_i 83 mM, NAD$^+$ 10 mM, FBP 10 mM	T_{opt}	330 K
				$\Delta C^{\ddagger}_{p}$	-2.1 ± 0.2 kJ.mol^{-1}.K^{-1}
				$\Delta S^{\ddagger}$	-0.188 ± 0.007 kJ.mol^{-1}
				$\Delta H^{\ddagger}$	3.0 ± 2.9 kJ.mol^{-1}
Pathway GK-GAPD	T_{opt}	-	GK 0.005308 mg.ml^{-1}, PGI 0.06175 mg.ml^{-1}, PFK 0.1083 mg.ml^{-1}, FBPA 0.1073 mg.ml^{-1}, TPI 0.03434 mg.ml^{-1}, GAPD 0.02708 mg.ml^{-1}, glucose 2 mM, ATP 7 mM, NAD$^+$ 10 mM, P_i 100 mM	T_{opt}	326 K
				$\Delta C^{\ddagger}_{p}$	-4.6 ± 0.6 kJ.mol^{-1}.K^{-1}
				$\Delta S^{\ddagger}$	-0.18 ± 0.01 kJ.mol^{-1}
				$\Delta H^{\ddagger}$	13 ± 5 kJ.mol^{-1}
Pathway PGI-GAPD	T_{opt}	-	GK 0.005308 mg.ml^{-1}, PGI 0.06175 mg.ml^{-1}, PFK 0.1083 mg.ml^{-1}, FBPA 0.1073 mg.ml^{-1}, TPI 0.03434 mg.ml^{-1}, GAPD 0.02708 mg.ml^{-1}, G6P 7 mM, ATP 7 mM, NAD$^+$ 10 mM, P_i 100 mM	T_{opt}	331 K
				$\Delta C^{\ddagger}_{p}$	-2.4 ± 0.3 kJ.mol^{-1}.K^{-1}
				$\Delta S^{\ddagger}$	-2.4 ± 0.3 kJ.mol^{-1}
				$\Delta H^{\ddagger}$	2 kJ.mol^{-1}

D.3: G6PDH temperature characterisation

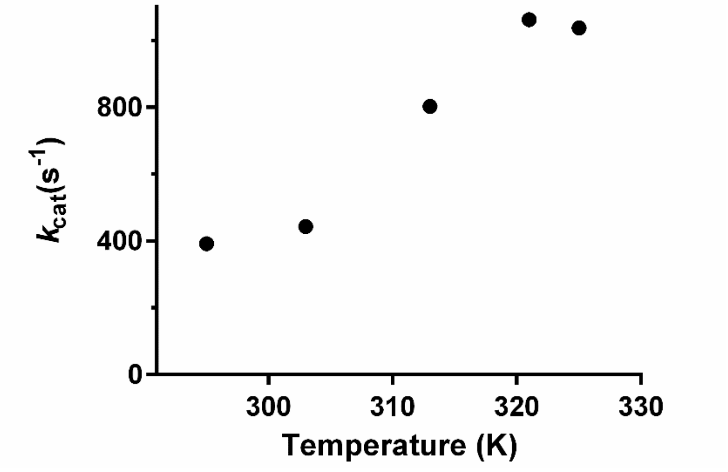


Figure D.3: Rough temperature profile for G6PDH used in the characterisation of GK and PGI (Section 4.3.1.1). G6PDH showed activity over the temperature range relevant for characterising GK and PGI (Figure 4.14).

D.4: Gel filtration S200 10/300 calibration curve

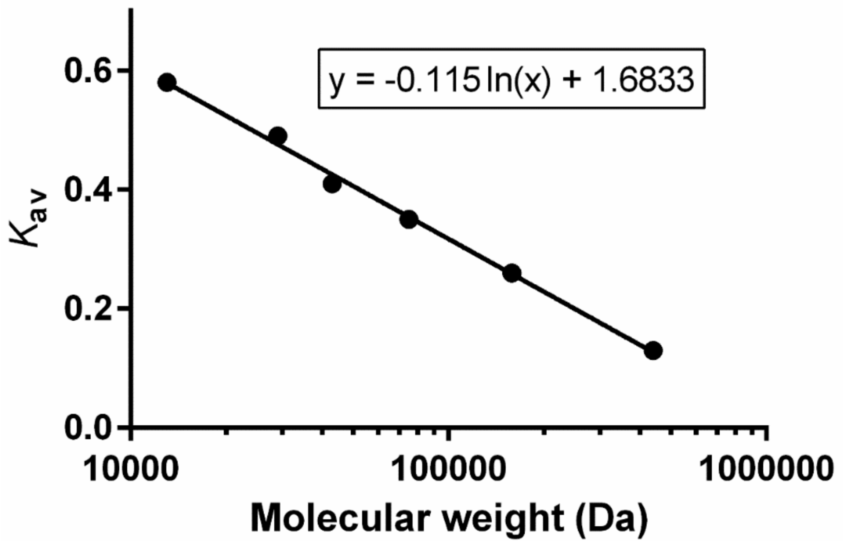


Figure D.4: Previously determined calibration curve for the gel filtration S200 10/300 column used for molecular weight determination of glycolytic enzymes (Section 4.2.2). Protein molecular weight and oligomeric state was estimated from this calibration.

D.5: GAFF-GLYCAM force field comparison

Table D.5I: Average values and variation (SD) about the mean for variation parameters about glucose over 60 ns of simulation time as modelled by GAFF, and averaged values across the two replicate runs (Section 6.1.1.1). Overall values for the two runs are given as absolute values.

Parameter	**GAFF run 1**		**GAFF run 2**		**Overall values**	
	Average	*SD*	*Average*	*SD*	*Average*	*SD*
Dihedrals (°)						
C1-C2-C3-C4	-57.3	4.9	-59.2	4.6	58.3	4.8
C2-C3-C4-C5	53.9	4.9	56.2	4.8	55.1	4.8
C3-C4-C5-O5	-51.6	5.1	-51.3	5.4	51.5	5.2
C4-C5-O5-C1	55.9	6.0	52.6	6.2	54.2	6.1
C5-O5-C1-C2	-58.9	5.7	-55.2	6.2	57.1	6.0
O5-C1-C2-C3	58.3	4.9	57.0	5.2	57.6	5.1
O5-C5-C6-O6	-79.9	12.4	-82.4	7.6	81.1	10.0
C4-C5-C6-O6	44.5	13.0	42.0	7.6	43.3	10.3
Angles (°)						
C2-C1-O1	112.5	3.4	112.6	3.2	112.5	3.3
C5-C4-O4	109.1	3.2	108.9	3.3	109.0	3.2
C4-C3-O3	110.7	3.1	110.8	3.2	110.8	3.2
C3-C2-O2	113.3	3.1	113.0	3.1	113.2	3.1
C5-C6-O6	112.6	3.3	113.0	3.2	112.8	3.3
Distances (Å)						
C1-C2	1.5	0.0	1.6	0.0	1.6	0.0
C3-C4	1.6	0.0	1.6	0.0	1.6	0.0
C5-O5	1.5	0.0	1.5	0.0	1.5	0.0
C2-H	1.1	0.0	1.1	0.0	1.1	0.0
C4-H	1.1	0.0	1.1	0.0	1.1	0.0
C5-H	1.1	0.0	1.1	0.0	1.1	0.0
C6-H	1.1	0.0	1.1	0.0	1.1	0.0
C1-O1	1.4	0.0	1.4	0.0	1.4	0.0
C3-O3	1.4	0.0	1.4	0.0	1.4	0.0
C5-C6	1.6	0.0	1.6	0.0	1.6	0.0
O2-H	1.0	0.0	1.0	0.0	1.0	0.0
O4-H	1.0	0.0	1.0	0.0	1.0	0.0
O6-H	1.0	0.0	1.0	0.0	1.0	0.0

Table D.5II: Average values and variation (SD) about the mean for variation parameters about glucose over 60 ns of simulation time as modelled by GLYCAM, and averaged values across the two replicate runs (Section 6.1.1.1). Overall values for the two runs are given as absolute values.

Parameter	**GLYCAM run 1**		**GLYCAM run 2**		**Overall values**	
	Average	*SD*	*Average*	*SD*	*Average*	*SD*
Dihedrals (°)						
C1-C2-C3-C4	-58.0	4.7	-58.4	4.7	58.2	4.7
C2-C3-C4-C5	54.4	4.8	54.3	4.9	54.3	4.9
C3-C4-C5-O5	-49.0	5.4	-48.5	5.6	48.8	5.5
C4-C5-O5-C1	50.5	6.5	50.3	6.7	50.4	6.6
C5-O5-C1-C2	-54.0	6.0	-54.4	6.3	54.2	6.2
O5-C1-C2-C3	56.5	5.0	57.0	5.1	56.8	5.0
O5-C5-C6-O6	-81.5	9.3	-80.3	16.2	80.9	12.7
C4-C5-C6-O6	44.3	9.4	44.8	17.5	44.6	13.4
Angles (°)						
C2-C1-O1	112.3	3.1	112.1	3.1	112.2	3.1
C5-C4-O4	107.5	3.1	107.4	3.0	107.5	3.1
C4-C3-O3	110.2	2.9	110.2	2.9	110.2	2.9
C3-C2-O2	113.2	3.1	113.0	3.0	113.1	3.0
C5-C6-O6	111.2	3.2	111.2	3.2	111.2	3.2
Distances (Å)						
C1-C2	1.5	0.0	1.5	0.0	1.5	0.0
C3-C4	1.6	0.0	1.6	0.0	1.6	0.0
C5-O5	1.5	0.0	1.5	0.0	1.5	0.0
C2-H	1.1	0.0	1.1	0.0	1.1	0.0
C4-H	1.1	0.0	1.1	0.0	1.1	0.0
C5-H	1.1	0.0	1.1	0.0	1.1	0.0
C6-H	1.1	0.0	1.1	0.0	1.1	0.0
C1-O1	1.4	0.0	1.4	0.0	1.4	0.0
C3-O3	1.5	0.0	1.5	0.0	1.5	0.0
C5-C6	1.5	0.0	1.5	0.0	1.5	0.0
O2-H	1.0	0.0	1.0	0.0	1.0	0.0
O4-H	1.0	0.0	1.0	0.0	1.0	0.0
O6-H	1.0	0.0	1.0	0.0	1.0	0.0

D.6: MD cluster analysis for outlier identification

Table D.6I: Cluster analysis for identification of outlier runs for MalL WT simulations with substrate and TS analogue. Fractions of time occupied in six difference conformational clusters (sieves) are given. Two analyses with substrate are given: with outlier run 8 included (Sub I) and after run 8 replacement (Sub II). Removal of run 8 was validated with further RMSD and conformation data (Section 6.2.3).

WT Sub I	**Run**									
Cluster	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	0.604	0.967	0.535	0.312	0.413	0.34	0.931	0.044	0.973	0.081
2	0.361	0.033	0.455	0.276	0.587	0.446	0.069	0.14	0.024	0.535
3	0.001	0	7E-04	0	0	0	0	0.756	0	0.001
4	0.031	0	0.007	7E-04	0	0.034	0	0.053	0	0.345
5	0.004	0	0.002	0.411	0	0	0	0.007	0	0.033
6	0	0	4E-04	0	0	0.18	0	0	0.003	0.005

WT Sub II	**Run**									
Cluster	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**9**	**10**	**11**
1	0.334	0.865	0.028	0.25	0.868	0.605	0.992	0.461	0.486	0
2	0.449	0.135	0.4	0.224	0.075	0.073	0.008	0.539	0.002	1
3	0.194	0	0.036	0.07	0.034	0.079	0	0	0.48	0
4	0.023	0	0.535	0.023	0.023	0.014	0	2E-04	0.003	0
5	0	0	0	0.432	0	7E-04	0	0	0.01	0
6	2E-04	0	4E-04	7E-04	2E-04	0.229	0	0	0.019	0

WT TS	**Run**									
Cluster	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	1	0.923	0.966	0.17	0.98	1	0.011	0.985	0.972	0.554
2	4E-04	0.046	0.034	0.005	7E-04	0	0.773	0.015	0.028	0
3	0	0.02	0	0.826	0.019	2E-04	0	0	0	0
4	0	0.01	0	0	0	0	0	0	0	0.446
5	0	2E-04	0	0	0	0	0.118	0	0	0
6	0	4E-04	0	0	4E-04	0	0.098	0	0	0

Table D.6II: Cluster analysis for identification of outlier runs for V200S simulations with substrate and TS analogue pertaining to Section 6.2.3. Fractions of time occupied in six difference conformational clusters are given.

V200S Sub	**Run**									
Cluster	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	0.509	0.998	0.282	0.002	0.887	0.583	0.562	0.421	0.187	0.069
2	0.469	0.002	0.619	0.138	2E-04	0.023	0.426	0.578	0.802	0.924
3	0.021	0	0.056	0.851	0	0.003	0.004	0.001	0.011	0.008
4	9E-04	0	0.043	0.005	0	0.391	0.008	0	0	0
5	0	0	0	0	0.113	0	0	0	0	0
6	0	0	0	0.003	0	0	0	0	0	0

V200S TS	**Run**									
Cluster	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	0.607	0.904	1	1	0.407	0.641	0.876	0.984	0.141	0.984
2	0	0	0	0	0.069	0.108	9E-04	0.013	0.859	0
3	0.007	0	0	0	0.493	0	0.001	0	0	0
4	0.386	0.011	0	0	0.004	0	0.053	0	0	0.009
5	0	0.016	0	0	0.027	0.25	0.067	0.003	0	0.006
6	2E-04	0.07	0	0	0	0	0.002	0	0	0

Appendix E: Publications

BIOCHEMISTRY
including biophysical chemistry & molecular biology
New Concept
pubs.acs.org/biochemistry

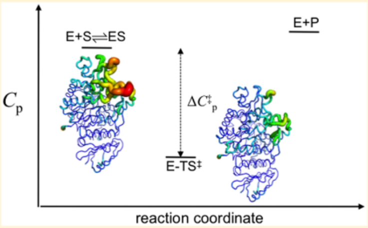
On the Temperature Dependence of Enzyme-Catalyzed Rates

Vickery L. Arcus,[*,†] Erica J. Prentice,[†] Joanne K. Hobbs,[†,#] Adrian J. Mulholland,[§]
Marc W. Van der Kamp,[§,∥] Christopher R. Pudney,[⊥] Emily J. Parker,[‡] and Louis A. Schipper[†]

[†]School of Science, University of Waikato, Hamilton 3240, New Zealand
[‡]Biomolecular Interaction Centre and Department of Chemistry, University of Canterbury, Christchurch 8041, New Zealand
[§]School of Chemistry and [∥]School of Biochemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, United Kingdom
[⊥]Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, United Kingdom

Supporting Information

ABSTRACT: One of the critical variables that determine the rate of any reaction is temperature. For biological systems, the effects of temperature are convoluted with myriad (and often opposing) contributions from enzyme catalysis, protein stability, and temperature-dependent regulation, for example. We have coined the phrase "macromolecular rate theory (MMRT)" to describe the temperature dependence of enzyme-catalyzed rates independent of stability or regulatory processes. Central to MMRT is the observation that enzyme-catalyzed reactions occur with significant values of $\Delta C_p^\ddagger$ that are in general negative. That is, the heat capacity (C_p) for the enzyme−substrate complex is generally larger than the C_p for the enzyme-transition state complex. Consistent with a classical description of enzyme catalysis, a negative value for $\Delta C_p^\ddagger$ is the result of the enzyme binding relatively weakly to the substrate and very tightly to the transition state. This observation of negative $\Delta C_p^\ddagger$ has important implications for the temperature dependence of enzyme-catalyzed rates. Here, we lay out the fundamentals of MMRT. We present a number of hypotheses that arise directly from MMRT including a theoretical justification for the large size of enzymes and the basis for their optimum temperatures. We rationalize the behavior of psychrophilic enzymes and describe a "psychrophilic trap" which places limits on the evolution of enzymes in low temperature environments. One of the defining characteristics of biology is catalysis of chemical reactions by enzymes, and enzymes drive much of metabolism. Therefore, we also expect to see characteristics of MMRT at the level of cells, whole organisms, and even ecosystems.



The rate of any chemical reaction is a function of the temperature (T) and the energy difference between the reactants and the transition state, the so-called activation energy (E_a). Arrhenius was the first to formalize this relationship in the 19th century (based on empirical observations) with his famous eponymous equation $k = A\ \exp(-E_a/RT)$, where k is the rate constant and R is the universal gas constant. Early in the 20th century, the development of transition state theory (TST) by Eyring, Polanyi, and others led to the Eyring equation for rate constants (eq 1 for a first order rate constant, where $\Delta G^\ddagger$ is the change in Gibbs free energy between reactants and the transition state, k_B and h are Boltzmann's and Planck's constants respectively, and κ is the transmission coefficient, hereafter assumed to be 1 for simplicity. For a more general definition in terms of partition functions see, e.g., ref 1).[1] This led to an understanding of, and statistical mechanical justification of, the terms in the Arrhenius expression. The Arrhenius and Eyring equations are found in most modern (bio)chemistry textbooks and provide an excellent description of the temperature dependence of a wide array of chemical processes.[1] The Eyring equation in its simplest form is sufficient for our purposes here (eq 1). An assumption often made with respect to eq 1 is that $\Delta H^\ddagger$ and $\Delta S^\ddagger$ are independent of temperature (and hence that $\Delta G^\ddagger$ varies with temperature according to the Gibbs equation: $\Delta G^\ddagger = \Delta H^\ddagger - T\Delta S^\ddagger$). Indeed, this assumption holds well for many reactions involving small molecules in standard solvents. However, a number of investigators have noted deviations from eq 1 when plotting temperature versus enzyme-catalyzed rates, suggesting that the above assumption is not valid and that there is a more complex temperature dependence for these systems.[1,2]

Enzymes are flexible macromolecules of high molecular weight and with correspondingly high heat capacities (C_p). For example, the heat capacity for folded proteins is estimated to be ~45 J mol^{-1}·K^{-1} per amino acid,[3] and thus a typical enzyme of molecular mass 65 kDa will have a heat capacity in water of ~25 300 J mol^{-1}·K^{-1} (for comparison, liquid water has C_p = 76 J mol^{-1}·K^{-1} at 25 °C).

$$k = \frac{\kappa k_B T}{h} e^{(-\Delta G^\ddagger/RT)} = \frac{\kappa k_B T}{h} e^{(\Delta S^\ddagger/R)} e^{(-\Delta H^\ddagger/RT)} \tag{1}$$

Received: October 6, 2015
Revised: February 12, 2016
Published: February 16, 2016

$$\Delta G^{\ddagger} = [\Delta H^{\ddagger}_{T_0} + \Delta C^{\ddagger}_{p}(T - T_0)] - T[\Delta S^{\ddagger}_{T_0} + \Delta C^{\ddagger}_{p}(\ln T - \ln T_0)] \tag{2}$$

$$k = \frac{k_B T}{h} e^{\left[\frac{-\Delta H^{\ddagger}_{T_0} - \Delta C^{\ddagger}_{p}(T - T_0)}{RT} + \frac{\Delta S^{\ddagger}_{T_0} + \Delta C^{\ddagger}_{p}(\ln T - \ln T_0)}{R}\right]} \tag{3}$$

$$\ln k = \ln \frac{k_B T}{h} - \left[\frac{\Delta H^{\ddagger}_{T_0} + \Delta C^{\ddagger}_{p}(T - T_0)}{RT}\right] + \left[\frac{\Delta S^{\ddagger}_{T_0} + \Delta C^{\ddagger}_{p}(\ln T - \ln T_0)}{R}\right] \tag{4}$$

The C_p of a system is a fundamental thermodynamic parameter that quantifies the temperature dependence of the enthalpy (H) and entropy (S) according to eq 2, and incorporation of a $\Delta C_p^{\ddagger}$ term into the Eyring eq (eq 1) gives eqs 3 and 4. If $\Delta C_p^{\ddagger} = 0$, eq 3 collapses into eq 1. However, for reactions catalyzed by enzymes with high heat capacities, the $\Delta C_p^{\ddagger}$ term may be nonzero, and eqs 3 and 4 should be implemented. Is there a difference in heat capacity between the enzyme−substrate and enzyme−transition state species for enzyme-catalyzed reactions (i.e., is $\Delta C_p^{\ddagger}$ nonzero for enzyme-catalyzed reactions)? If so, what are the consequences for the temperature dependence of enzyme catalyzed rates?

We have previously demonstrated that enzymatically catalyzed rates proceed with negative values of $\Delta C_p^{\ddagger}$ ranging from −1 to −12 kJ mol^{-1}·K^{-1} (independent of denaturation),[4] and we have coined the phrase macromolecular rate theory (MMRT) to reflect this unusual thermodynamic property of biochemical reactions.[5,6] MMRT has significant implications for the temperature dependence of enzyme-catalyzed rates and for the rates of biologically driven processes in general, such as microbial growth, respiration, and photosynthesis. MMRT unifies a number of disparate observations with respect to thermophilic, mesophilic, and psychrophilic enzymes and also presents some surprising hypotheses. Here, we lay out the basis for MMRT and present new hypotheses based on MMRT that warrant further experimental verification. We also use MMRT to provide a theoretical argument for relative size of enzymes according to the chemistry that they catalyze.

■ MACROMOLECULAR HEAT CAPACITY

The internal energy of a system is partitioned between translational, rotational, vibrational, and electronic modes. The heat capacity C is formally the change in internal energy with a change in temperature and is a measure of the capacity for the translational, rotational, vibrational, and electronic modes to absorb energy. For systems in water at biologically relevant temperatures ($-20 < T < 100$ °C), electronic modes above the ground state are generally inaccessible, and thus electronic modes do not contribute to heat capacity in this context. It has been shown experimentally that the greatest contribution to the heat capacity of a folded protein in water is the number of accessible vibrational modes.[3] Molecular dynamics simulations over long time-scales paint a vivid picture of the low-frequency modes for proteins for example.[7] These numerous modes contribute an estimated 82% of the C_p term for globular proteins with a further 15% contribution from the interaction between the protein and water.[3,8] Indeed, this is the origin of the increase in C_p per amino acid described above—the addition of each amino acid to a protein chain adds an additional number of vibrational modes to the molecule (increasing C_p by ~45 J mol^{-1}·K^{-1} per amino acid). The change in heat capacity, ΔC_p, for a macromolecular system in equilibrium between two states is then most significantly due to the change in the frequencies of the vibrational modes of the molecule between the two states. The simplest example of macromolecular ΔC_p can be seen in protein−ligand binding:

$$\mathrm{P_{apo}} + \mathrm{L} \rightleftharpoons \mathrm{PL}$$

The apo form of the protein ($\mathrm{P_{apo}}$) is often more flexible (i.e., more low frequency vibrational modes) than the ligand-bound form (which has increased rigidity and fewer low frequency modes). There is also a desolvation term that contributes to ΔC_p and is the result of displacement of bound water molecules by the ligand.[9] If we consider the contributions from the protein alone, ΔC_p for such an interaction is generally negative. Protein crystallographers have been exploiting this phenomenon for decades: proteins are more likely to crystallize in the presence of a bound ligand because the protein becomes more "ordered" or more rigid in the ligand-bound state. In general, any process that limits the number of low frequency vibrational modes for macromolecules will be accompanied by a corresponding negative ΔC_p for that process. ΔC_p has been experimentally determined for many protein−ligand interactions. For example, the interaction between the enzyme methylthioadenosine phosphorylase (MTAP) and a tight binding inhibitor has $\Delta C_p = -2.5$ kJ mol^{-1}·K^{-1}.[10] The binding of transcription factors to DNA are characterized by ΔC_p values ranging from −1.4 to −3.4 kJ mol^{-1}·K^{-1}.[11] In this latter case, the role of changing solvation upon binding is thought to make a significant contribution to ΔC_p.[11] In these and other examples where ΔC_p has been measured directly, it has been found to be largely independent of temperature (i.e., linear ΔH versus T plots), and we assume this to be the case for MMRT.

The classical description of enzyme catalysis invokes relatively weak binding of the substrate to the enzyme (K_M) and tight binding of the transition state to the enzyme. The tight binding of the transition state significantly lowers $\Delta G^{\ddagger}$ for the reaction, leading to the extraordinary rate enhancements we see for enzyme catalyzed reactions.[12] By analogy with the description for ligand binding at equilibrium above, we would expect a negative value for $\Delta C_p^{\ddagger}$ for enzyme catalysis—very tight binding of the transition state will reduce the number of low frequency vibrational modes for this state when compared to the relatively weakly bound enzyme−substrate complex.[4]

It must be acknowledged that the classical description for enzyme catalysis, based on tight binding of the transition state, is hotly debated, and the precise origin of enzyme catalysis is currently the subject of controversy in the literature, in particular, the role of protein dynamics in catalysis.[13−16] Various arguments for and against the role of protein dynamics in enzyme catalysis have been presented based on statistical thermodynamics,[1] molecular dynamics,[17] transition state barrier crossing,[14] Marcus theory and preorganization,[18] promoting vibrations,[19] and quantum tunnelling,[20] among others. Notwithstanding this controversy, the sign and magnitude of $\Delta C_p^{\ddagger}$ for enzyme catalysis are a statistical thermodynamic property for the reaction that describes the difference in heat capacity between the ensemble in the ground state and that at the transition state. This property does not report on enzyme dynamics with respect to catalysis. $\Delta C_p^{\ddagger}$ can be mathematically defined in terms of statistical thermodynamics according to eq

5.[9] For any particular state on the reaction coordinate, the heat capacity is proportional to the mean squared fluctuation of the enthalpy divided by kT^2. Thus, $\Delta C_p^\ddagger$ is the difference in the mean squared fluctuation in the enthalpy at the transition state compared to the ground state.

$$\Delta C_p^\ddagger = \frac{\Delta \langle \partial H^2 \rangle^\ddagger}{kT^2} \tag{5}$$

IMPLICATIONS OF NEGATIVE $\Delta C_p^\ddagger$ VALUES FOR ENZYME CATALYSIS

We have previously determined $\Delta C_p^\ddagger$ values for several enzymes.[4] The $\Delta C_p^\ddagger$ values range from −1 to −12 kJ $\text{mol}^{-1}\cdot\text{K}^{-1}$. When the MMRT function is fitted to data collected by others, the data are also well described and $\Delta C_p^\ddagger$ values also lie in this range (see Figure S2). This has significant implications for the temperature dependence of enzyme-catalyzed rates. First, $\Delta H^\ddagger$ and $\Delta S^\ddagger$ are steeply dependent on temperature according to eq 2 (Figure 1A). $\Delta G^\ddagger$ is curved as a result and the curvature is determined by the magnitude of $\Delta C_p^\ddagger$ (to illustrate, we have set $\Delta C_p^\ddagger = -3.0$ kJ $\text{mol}^{-1}\cdot\text{K}^{-1}$ in Figure 1A−C). In this context, it makes little sense to discuss in general terms the relative contributions of enthalpy and/or entropy to enzyme catalysis without precisely defining the temperature. For example, the scheme presented in Figure 1A shows that the entropic contribution ($-T\Delta S^\ddagger$) to $\Delta G^\ddagger$ is precisely zero at 289 K. In contrast, at 312 K, the enthalpic term is zero and $T\Delta S^\ddagger = -70.5$ kJ mol^{-1}. The steep temperature dependence of these terms may account for the disparate observations in the literature regarding the relative contributions of enthalpy and entropy to ligand binding events (though for binding, solvation changes will make a large contribution),[10] and the relative importance of entropy and enthalpy to enzyme catalysis.[21] Suffice to say that at low temperatures the $\Delta G^\ddagger$ barrier for the reaction is predominantly enthalpic, and at high temperatures, the $\Delta G^\ddagger$ barrier is predominantly entropic. $\Delta G^\ddagger$ is a minimum when $\Delta S^\ddagger = 0$. Empirical examples of the temperature dependence of $\Delta H^\ddagger$ and $\Delta S^\ddagger$ have been observed and discussed previously.[16]

It follows that for $\Delta C_p^\ddagger < 0$, the rate of an enzyme catalyzed reaction initially rises with temperature and then reaches an optimum temperature (T_{opt}) after which the rate falls again, in contrast to simple Arrhenius and Eyring kinetics (Figure 1B,C). The increase in rate at temperatures up to T_{opt} is driven by the enthalpic term in eq 4 ($-\Delta H^\ddagger/RT$). However, this term is slowly overcome by the entropic term ($\Delta S^\ddagger/R$) at temperatures above T_{opt}, leading to a reduction in the reaction rate (Figure 1C).

The steep temperature dependence of $-\Delta H^\ddagger/RT$ and $\Delta S^\ddagger/R$ also makes it difficult to describe a general mechanism regarding how enzymes reduce the activation barrier for the uncatalyzed reaction. What can be said is that at temperatures close to the optimum temperature for the enzyme catalyzed rate (T_{opt}) $\Delta H^\ddagger$ for the enzyme catalyzed reaction is near zero (Figure 1C).

Curvature in a rate-versus-temperature plot is thus a generic property of enzyme-catalyzed rates where the heat capacity of the enzyme−substrate complex is greater than the heat capacity of the enzyme-transition-state species (i.e., $\Delta C_p^\ddagger < 0$). This curvature has previously been attributed to Arrhenius-like behavior below the optimum temperature (T_{opt}) and denaturation above T_{opt}. Of course, the process of denaturation impacts on enzyme-catalyzed rates at high temperatures. However, decreases in rate above T_{opt} occur even in the absence of denaturation/unfolding, as a direct consequence of MMRT (when $\Delta C_p^\ddagger < 0$). For example, psychrophilic enzymes have long been considered enigmatic as they display decreases in rate above T_{opt} in the absence of denaturation.[22]

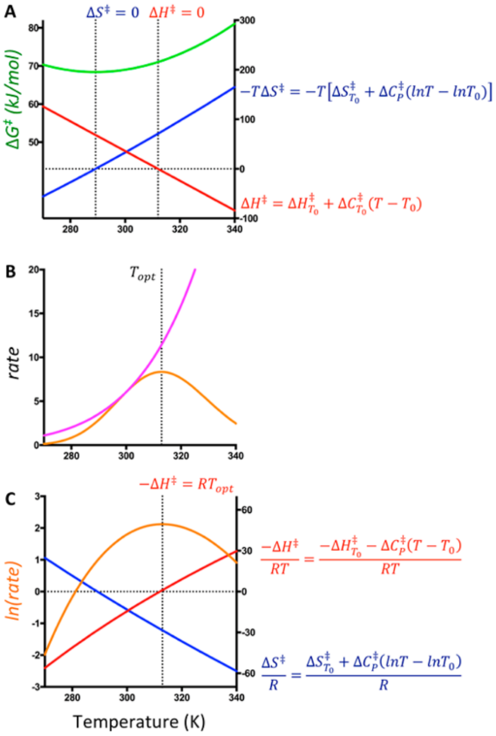


Figure 1. The temperature dependence of rates where $\Delta C_p^\ddagger$ is nonzero. (A) The temperature dependence of $\Delta G^\ddagger$ (left-hand *y*-axis) is shown in green. Individual contributions to $\Delta G^\ddagger$ from enthalpy and entropy are shown in red and blue, respectively (right-hand *y*-axis). Vertical dotted lines denote temperatures where $\Delta S^\ddagger = 0$ and $\Delta H^\ddagger = 0$. $\Delta G^\ddagger$ is a minimum when $\Delta S^\ddagger = 0$. $\Delta C_p^\ddagger$ is set to −3.0 kJ $\text{mol}^{-1}\cdot\text{K}^{-1}$. (B) The temperature dependence of the reaction rate according to simple Eyring (purple, eq 1) and MMRT (orange, eq 3, $\Delta C_p^\ddagger = -3.0$ kJ $\text{mol}^{-1}\cdot\text{K}^{-1}$) predictions. The optimum temperature for MMRT is shown by a vertical dotted line. (C) The temperature dependence of ln(rate) showing MMRT (orange, left-hand *y*-axis) and the individual contributions to ln(rate) from the enthalpy term ($-\Delta H^\ddagger/RT$) and the entropy term ($\Delta S^\ddagger/R$), red and blue lines, respectively (right-hand *y*-axis). The vertical dotted line shows T_{opt} and $-\Delta H^\ddagger = RT_{\text{opt}}$.

OPTIMUM TEMPERATURE FOR ENZYME CATALYSIS

When $\Delta C_p^\ddagger < 0$, T_{opt} is mathematically defined when the first derivative of eq 3, with respect to temperature, is equal to 0 ($dk/dT = 0$, see Figure 1B) Setting $dk/dT = 0$ for eq 3 gives eq 6. Equations 7 and 8 follow directly (Figure 1C).

$$\Delta H^\ddagger = -RT_{\text{opt}} \tag{6}$$

DOI: 10.1021/acs.biochem.5b01094
Biochemistry 2016, 55, 1681−1688

$$\Delta H^{\ddagger}_{T_0} + \Delta C^{\ddagger}_{p}(T_{\mathrm{opt}} - T_0) = -RT_{\mathrm{opt}} \qquad (7)$$

$$T_{\mathrm{opt}} = \frac{\Delta H^{\ddagger}_{T_0} - \Delta C^{\ddagger}_{p} T_0}{-\Delta C^{\ddagger}_{p} - R} \approx T_0 - \frac{\Delta H^{\ddagger}_{T_0}}{\Delta C^{\ddagger}_{p}} \quad (\text{for } |\Delta C_p^{\ddagger}| \gg R) \qquad (8)$$

Equation 6 places exacting constraints on the thermodynamics of enzyme-catalyzed reactions at the optimum temperature and limits $\Delta H^{\ddagger}$ to a very narrow range of values at T_{opt}. Enzyme T_{opt} values typically range from ~15 °C up to ~100 °C and this limits $\Delta H^{\ddagger}$ to values between −2.4 and −3.1 kJ mol^{-1} at T_{opt}. In contrast, the entropic term is almost always the dominant term at T_{opt}. It is noteworthy that at temperatures below T_{opt} (where many enzyme kinetic studies are performed), $\Delta S^{\ddagger}/R$ approaches zero and the enthalpic term dominates at these temperatures (Figure 1C). Feller and coworkers have described these trends in the enthalpic and entropic contributions to enzyme catalyzed rates when homologous psychrophilic, mesophilic, and thermophilic enzymes are compared,[23,24] and their observations are a direct consequence of eqs 3 and 6. Åqvist and colleagues have shown lower enthalpies of activation and more negative entropies of activation for psychrophilic enzymes which also follows from eqs 3 and 6 (Figure 1).[25,26]

Equation 8 is significant for the evolutionary adaptation of enzymes to thermophilic or psychrophilic environments. In order to evolve to higher or lower T_{opt} values, $\Delta C_p^{\ddagger}$ for the enzyme-catalyzed reaction must increase or decrease, respectively. We have previously demonstrated the correlation between T_{opt} and $\Delta C_p^{\ddagger}$ for a range of enzymes and their mutants.[4] Let us take as an example the evolution of a mesophilic enzyme to a psychrophilic enzyme requiring a shift of T_{opt} downward. Equation 8 implies that $\Delta C_p^{\ddagger}$ must become more negative to shift T_{opt} downward ($\Delta H^{\ddagger}$ is fixed at T_{opt}). In molecular terms, decreasing $\Delta C_p^{\ddagger}$ can be achieved by increasing the C_p of the enzyme−substrate complex (more low frequency vibrational modes), and/or decreasing the C_p of the enzyme-transition-state species (fewer low frequency modes, i.e. making it more rigid). The effect of both of these strategies is evident in the increased K_{M} and k_{cat} values for psychrophilic enzymes compared to their mesophilic counterparts (at low temperatures),[27] indicative of decreased binding affinity for the substrate and increased binding affinity for the transition state. Åqvist has described this as "protein surface softness" for psychrophilic enzymes.[26] An important consequence of an increasingly negative $\Delta C_p^{\ddagger}$ (with decreasing T_{opt}) is the increased curvature in the ln(rate)-versus-temperature plot (Figure 2). Figure 2A shows ln(k) versus temperature for homologous isopropylmalate dehydrogenase (IPMDH) enzymes from thermophilic, mesophilic, and psychrophilic species of *Bacillus*.[28] Over evolutionary time, the migration of T_{opt} values has tracked with $\Delta C_p^{\ddagger}$ in a manner consistent with eq 8. The same trend is seen for mutants of the *Bacillus subtilis* enzyme MalL (Figure 2B) where single point mutations simultaneously change T_{opt} of the enzyme and $\Delta C_p^{\ddagger}$ of the enzyme-catalyzed reaction.[4] The enzyme rate for each of these MalL variants is almost exactly the same at 37 °C, and it is the temperature dependence of the rates that has been altered (i.e., an altered $\Delta C_p^{\ddagger}$). Combining the temperature-rate data for 10 IPMDH enzymes and seven MalL enzymes from *Bacillus* clearly demonstrates the relationship between T_{opt} and $\Delta C_p^{\ddagger}$ (Figure 2C).

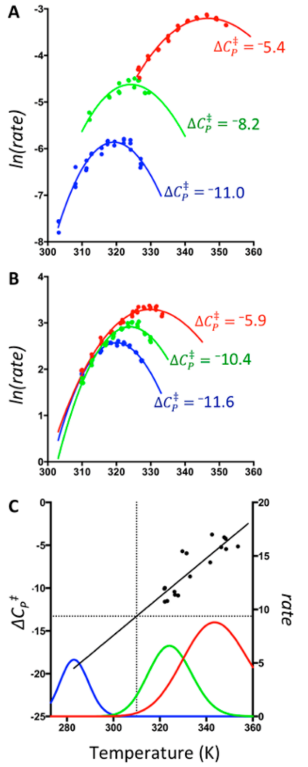


Figure 2. $\Delta C_p^{\ddagger}$, T_{opt}, and the psychrophilic trap. (A) The temperature dependence of IPMDH enzymes from thermophilic (red), mesophilic (green), and psychrophilic *Bacillus* species (blue) illustrating the increasing curvature in the ln(rate) versus temperature plot as T_{opt} decreases. (B) The temperature dependence of MalL enzymes from *Bacillus subtilis*. The wild type enzyme is shown in blue, the mutant V200T is shown in green, and the mutant V200S is shown in red. (C) The correlation between $\Delta C_p^{\ddagger}$ (left-hand *y*-axis) and T_{opt} (*x*-axis) for 10 IPMDH and 7 MalL Bacillus enzymes (black dots). A linear correlation is also shown. Colored curves are plots of rate (right-hand *y*-axis) versus temperature with $\Delta C_p^{\ddagger}$ values that are indicative of those falling on the line. This illustrates the extreme curvature for an enzyme whose T_{opt} lies at 10 °C for which the line predicts $\Delta C_p^{\ddagger} \approx -18.0$ kJ mol^{-1}·K^{-1}.

A PSYCHROPHILIC TRAP

The increasing curvature for the rate-versus-temperature plot as T_{opt} falls results in a "psychrophilic trap" whereby extreme curvature will render enzymes unfit for environments where the temperature varies even by a few degrees (Figure 2C). This explains the rarity of genuine psychrophiles whose T_{opt} matches the mean temperature for the environment for low temperatures ($-10 < T < 10$ °C). Indeed, IPMDH from *Bacillus psychrosaccarolyticus* (with a T_{opt} of 47 °C) points toward this organism being psychrotolerant as opposed to a genuine psychrophile (Figure 2A).

The large, negative values of $\Delta C_p^{\ddagger}$ required to set T_{opt} at, say, 10 °C (blue line in Figure 2C), creates large curvature in a rate-versus-temperature plot so as to make the enzyme rates highly—and probably prohibitively—sensitive to small changes in temperature. For example, a change in temperature of 10 °C away from T_{opt} for the hypothetical "psychrophile" in Figure 2C (blue line) causes a > 70% reduction in rate, whereas for the hypothetical mesophile and thermophile (green and red lines in Figure 2C) rates fall by 39% and 17% over the same temperature range. To avoid the psychrophilic trap, there appears to be a natural minimum T_{opt} of ~20 °C.[23] The only context where this restriction can be overcome and true psychrophiles may evolve is in environments where the ambient temperature is low and almost invariant over both short and long (evolutionary) time scales. Studying enzymes from deep-sea organisms where temperatures are very stable (and relatively low) may be the best way to test this hypothesis.

In the context of hot environments, $\Delta C_p^\ddagger$ will approach zero for increasingly thermophilic enzymes (as T_{opt} approaches 100 °C). Indeed, the linear correlation shown in Figure 2C crosses the *x*-axis at 369 (±13) K, very close to 100 °C. As such, the temperature dependence of thermophilic enzymes will approach Arrhenius behavior and the enthalpic and entropic contributions to the rate are much less temperature sensitive. Only under these circumstances ($\Delta C_p^\ddagger \approx 0$, extreme thermophiles) may it be possible to tease out the relative contributions of enthalpy and entropy to catalysis (as $\Delta H^\ddagger$ and $\Delta S^\ddagger$ will be constant over a relatively wide temperature range).

■ ENZYME MASS AND REACTION CHEMISTRY

Protein synthesis is one of the most energetically costly processes in the cell. It has been estimated that approximately 80% of cellular ATP is given over to protein synthesis.[29] Given the acute selection pressures that drive energy efficiency in cellular metabolism, this implies that enzymes are a minimum size to carry out their cellular function(s). Further, it implies that there is an important functional purpose for the large mass of an enzyme.

A number of concepts have previously been suggested in an effort to rationalize enzyme size, and these have included the possibility of the enzyme surface acting to concentrate substrate molecules near the active site.[31] Other researchers have suggested that the enzyme acts to channel thermal energy to the active site,[32] or that the enzyme/substrate mass ratio is linked to the level of stabilization of the transition state,[30] or that there is a requirement for additional transition state-specific interactions in order to stabilize this state relative to the unbound transition state.[42] The general need for a preorganized active site (which may place catalytic groups in unfavorable environments) has been suggested to impose a size requirement to give the requisite folding energy.[43,44]

MMRT and its consideration of the role of $\Delta C_p^\ddagger$ for enzymes can provide some insight into the relationship between enzyme mass and enzyme catalysis as $\Delta C_p^\ddagger$ reports directly on the change in the frequencies of the vibrational modes between the enzyme−substrate complex and the enzyme-transition state complex.

The internal energy of an enzyme in water is partitioned across kinetic, vibrational, and rotational modes (ignoring the electronic modes for the purposes of this discussion). The potential energy residing in the vibrational modes represents a significant energy reservoir. The observation of $\Delta C_p^\ddagger = -11.6$ kJ mol^{-1}·K^{-1} for MalL,[4] for example, suggests that this energy reservoir changes on progression from the enzyme−substrate complex to the enzyme-transition state complex (at a given temperature). Let us make the conjecture that this change in energy has contributed toward the stabilization of the transition state. It follows that should there be a requirement for a greater stabilization of the transition state to either increase the reaction rate or to catalyze more "difficult" chemistry, we would expect to see an increase in the mass of the enzyme as this is a mechanism to increase the molecular heat capacity (the potential energy reservoir).

To illustrate this idea, let us adopt Wolfenden's classic formalism[12] and use the hydrolysis of oligosaccarides, DNA, and RNA as examples. In water at 25 °C oligosaccarides are very stable. For example, the half-life ($t_{1/2}$) for the hydrolysis of maltose at 25 °C is ~11 million years. DNA is less stable to hydrolysis with $t_{1/2} \approx 140\,000$ years and RNA hydrolysis has $t_{1/2} \approx 4$ years.[12] Thus, the uncatalyzed rates of hydrolysis of these compounds vary by 7 orders of magnitude. In contrast, the rate constants for all three enzyme-catalyzed reactions vary by just 2 orders of magnitude and are dramatically larger: $k_{cat} \approx 0.01{-}1.0$ s^{-1}. If a contribution to catalysis were to originate from the molecular heat capacity reservoir, we would expect to see enzymes increase in size from RNases to DNases to glycosidases. This is what we observe (Figure 3). We have surveyed a range of enzymes that catalyze pseudo-first order reactions (hydrolases, esterases, decarboxylases, and isomerases), and we find a significant correlation ($p = 0.0004$) between enzyme molecular weight and the log(ratio) of enzyme-catalyzed to noncatalyzed rate constants, $\ln(k_{cat}/k_{non})$. This indicates that larger enzymes are required to catalyze increasingly difficult chemistries. For this data set, we deliberately chose pseudo-first order reactions, and we computed the molecular weight for isolated monomeric catalytic domains. Thus, the data are not confounded by oligomeric states or by entropic considerations for bimolecular and termolecular reactions. We have also added two ribozymes to the data set (in orange diamonds in Figure 3), and these are also consistent with the correlation.

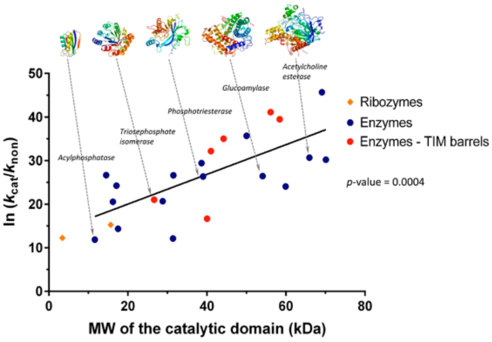


Figure 3. Molecular weight of the catalytic domain of an enzyme versus the observed rate enhancement, $\ln(k_{cat}/k_{uncat})$. Selected structures for the catalytic domains are illustrated across the top of the graph and labeled (arrows point to the corresponding point on the graph). Ribozymes are shown with orange diamonds and enzymes are shown with filled circles. The TIM-domain enzymes are shown in red. A linear regression is also shown ($y = 0.34x + 13.3$, $R^2 = 0.48$). Values are taken from references detailed in Table S1.

An intriguing observation from these data is that enzymes with the same fold increase in mass to catalyze more difficult chemistries. For example, there are six TIM barrel enzymes in the data set (in red circles in Figure 3), and their increasing masses are consistent with the correlation.

The relationship between enzyme mass and rate enhancement in Figure 3 suggests an approximate lower bound on enzyme catalysis where the linear correlation crosses the *y*-axis at $\ln(k_{cat}/k_{non}) = 13.3$, $(k_{cat}/k_{non}) \approx 6.0 \times 10^5$. Rate enhancements that lie below this figure are certainly accessible via intramolecular small molecule catalysis at ambient temperatures, for example, and thus the *y*-axis intercept suggests an approximate transition between small molecule catalysis and macromolecular catalysis.

The correlation shown in Figure 3 does not prove causality, and further experiments are required to verify the relationship between the macromolecular heat capacity potential energy reservoir and enzyme catalysis. If this behavior is borne out and found to be generally important then it may have implications for enzyme design.

The position of the ribozymes at the left-hand end of the scatter of points may say something about the suitability of ribozymes for catalysis of increasingly difficult chemistry (i.e., very slow chemical reactions). Ribozymes are typically poor enzymes, and even given large mass increases ribozymes do not realize the significant rate enhancements of enzymes. RNase P for example, with a mass of 120 kDa, only achieves a comparatively paltry rate enhancement of $\ln(k_{cat}/k_{non}) = 15$, which is comparable to the two small ribozymes included in Figure 3. This suggests that increasing ribozyme mass has very little effect on catalysis and will not bring slow chemical reactions into the correct time frame for life. This may point to a possible explanation for the supplanting of ribozymes by proteins as biological catalysts over the course of evolutionary history.

■ A NOTE ON ENZYME DENATURATION

The significance of ΔC_p for the temperature dependence of *protein stability* at equilibrium has been understood for 40 years.[33,34] The role of $\Delta C_p^{\ddagger}$ in the temperature dependence of protein folding kinetics has also been well described.[35] For protein folding, $\Delta C_p^{\ddagger}$ is generally negative as the unfolded state is very dynamic and hydrophobic side chains have a characteristic solvation shell in this state, whereas the transition state for protein folding is generally relatively compact and more closely resembles the native state.[36] For folding, desolvation of hydrophobic residues on the folding pathway is thought to make a significant contribution to $\Delta C_p^{\ddagger}$.[37] In contrast, the similarity between the folded state and the transition state for folding leads to small positive values of $\Delta C_p^{\ddagger}$ for protein unfolding and almost no curvature in the temperature dependence of unfolding rates.[35]

MMRT describes the relationship between temperature and the rate of enzyme catalysis in the absence of denaturation. For a complete description of the temperature dependence of enzyme catalysis, denaturation must be included. As unfolding occurs with small positive values of $\Delta C_p^{\ddagger}$ and shows little curvature over a wide temperature range,[38] it is relatively straightforward in many cases to apply a simple correction term for unfolding to the observed enzyme catalyzed rate. In the case of MalL, the unfolding rate (k_u) is 3 orders of magnitude slower than (k_{cat}), and this allows accurate deconvolution of the temperature dependence of k_{cat} and k_u.[4] Indeed, we would advocate that enzymologists routinely measure unfolding rates at the temperature(s) of interest.

■ DISCUSSION

The importance of $\Delta C_p^{\ddagger}$ for the temperature dependence of enzyme kinetics was first described by Hobbs et al. 2013.[4] The $\Delta C_p^{\ddagger}$ term for enzyme kinetics has been hidden in plain sight as enzyme denaturation has always been thought to be the cause of decreasing rates for enzymes above their optimum temperature of activity. In this context, psychrophilic enzymes were seen as enigmatic insofar as they display a curved temperature dependence of their catalytic rates despite relatively high unfolding temperatures.[27] MMRT now resolves this enigma by making explicit the difference in heat capacity between the enzyme substrate complex and the enzyme transition state species: this difference dictates curvature in the temperature-ln(rate) plot for enzymes independent of denaturation.

The extent to which MMRT is generalizable to all enzymes remains to be seen and in many cases relies on the deconvolution of the temperature dependence of enzymatic rates from denaturation rates (which is not common practice). Nevertheless, here we review and analyze data for 6 enzymes and 22 mutants that conform to MMRT kinetics. Further, Daniel and Danson[2] have shown curved temperature-rate plots at "zero time" (i.e., independent of denaturation) for 21 enzymes (see Figure S2 for three examples), and the psychrophilic enzymes (as reviewed by Feller and Gerday[22]) also show this same curvature. In a separate paper we have recently analyzed the temperature dependent behavior of three model enzymes in the context of MMRT.[6] On the other side of the coin, there are many historical examples of linear Arrhenius plots for enzymes, and this suggests that in many cases, $\Delta C^{\ddagger}_{p}$ is close to zero (collapsing eq 3 into eq 1). There are pragmatic reasons why, in some cases, curvature may be obscured. For example, where data are collected over a narrow temperature range or where there are too few data points to assess the presence or absence of curvature. The generalizability of MMRT will be tested by the careful collection of data over a sufficient temperature range (to calculate accurate k_{cat} values) and deconvolution of these rates from denaturation rates.

The origin of a negative value of $\Delta C_p^{\ddagger}$ for enzyme kinetics is a shift in the frequencies of the vibrational modes that leads to reduced fluctuation in energies at the transition state (eq 5). In a classical description of enzyme kinetics, negative $\Delta C_p^{\ddagger}$ values are the result of relatively weak binding of the enzyme to its substrate and tight binding to the transition state. We have described this phenomenon as "macromolecular rate theory" as $\Delta C_p^{\ddagger}$ is enabled by the high intrinsic heat capacity of macromolecules, in contrast to catalysis by small molecules. The high heat capacity for proteins in particular (in contrast to ribozymes for example, see Figure 3) appears to be significant in the context of enzyme catalysis.

MMRT is based on a quasi-two-state model where the kinetics of the rate determining chemical step are determined by the energy difference between the enzyme−substrate complex and the enzyme-transition state complex. Other investigators have postulated quasi-three-state models to account for the curvature in the temperature dependence of some enzyme reaction rates.[1,2] These models postulate an equilibrium between conformations with different activation energies. In an analysis of enzymatic Q_{10} values Elias and Tawfik also find evidence for a three-state model.[39] On the basis of temperature-rate data alone, it is not possible to discriminate between a two-state model incorporating $\Delta C^{\ddagger}_{p}$

DOI: 10.1021/acs.biochem.5b01094
Biochemistry 2016, 55, 1681−1688

and three-state models. However, the three-state models assume $\Delta C_p^\ddagger = 0$, and this parameter may be measured directly for enzyme kinetics using isothermal titration calorimetry (ITC). These experiments are the subject of ongoing research in our laboratories. Should these experiments reveal a nonzero $\Delta C^\ddagger_p$ for enzyme catalyzed rates then MMRT will be sufficient to describe the curvature in temperature-ln(rate) data without recourse to three-state models (Figure 2 and Figure S2).

MMRT uses a greatly simplified scheme (like all models), and there are a number of potentially confounding factors that must be acknowledged. First, we assume that there is no change in the chemical mechanism associated with catalysis, as temperature changes. Possible sources of mechanistic change could arise as a result of changes in pK_a with temperature for example. Second, we assume that the heat capacities of the ground state and transition state ensembles are essentially constant over the biological temperature range. Previous studies on protein folding provide some evidence that justifies this assumption where a very small temperature dependence of $\Delta C_p^\ddagger$ for protein unfolding is observed.[35] Linear plots of ΔH versus T for protein−ligand binding also point toward a very small temperature dependence of ΔC_p.[10]

MMRT prescribes a number of traits for the temperature dependence of enzyme catalysis. First, $\Delta H^\ddagger$ and $\Delta S^\ddagger$ are steeply temperature dependent and range from large positive values to large negative values across temperatures relevant to biology (0−100 °C). Second, the optimum temperature for activity, T_{opt}, is defined by $\Delta H^\ddagger$ (at T_{opt}) and $\Delta C_p^\ddagger$ where $\Delta H^\ddagger(T_{\mathrm{opt}}) = -RT_{\mathrm{opt}}$ (eqs 6−8). This dictates that for lower values of T_{opt}, $\Delta C_p^\ddagger$ becomes increasingly negative and curvature of the temperature-rate profile becomes extreme. This places a practical limit on T_{opt} for psychrophilic enzymes, and we have called this the psychrophilic trap. In contrast, $\Delta C_p^\ddagger$ approaches zero as T_{opt} approaches 100 °C.

The observation of negative $\Delta C_p^\ddagger$ values for enzyme catalysis implies that there is a change in the distribution of vibrational modes between the ground state ensemble and the transition state ensemble. Equivalently (according to eq 5) there is a narrowing of the distribution of enthalpies on going from the ground state ensemble to the transition state ensemble. The large number of enzyme vibrational modes and the associated heat capacity may be viewed as an energy reservoir, and the correlation between enzyme mass and k_{uncat} suggests that a larger energy reservoir (in enzymes with greater mass) may be required to help stabilize the transition state. Indeed, this is consistent with the hypothesis of Warshel and colleagues who have suggested the folding energies of enzymes may need to be reduced so that a small reorganization energy and thus a low energy barrier for reaction can be realized. In order to obtain low energy barriers for reaction for difficult chemistries, enzymes of increasing size may be required so that folding energies can be reduced to “pay the price” for low reorganization energies.[40]

Any contribution of vibrational modes to the catalytic effect of enzymes is extremely controversial at present,[16] and no inference concerning the microscopic picture of enzyme catalysis can be drawn from MMRT. Interestingly, however, it was recently observed that the magnitude of $\Delta C_p^\ddagger$ appears to be correlated with the presence or absence of vibrational modes that may be coupled to the reaction coordinate for very well-studied, small model enzyme systems[6] (although, when this coupling has been investigated directly, it has been shown to be small).[41] Suffice to say that MMRT incorporates $\Delta C_p^\ddagger$ to describe enzyme kinetics, and as such $\Delta C_p^\ddagger$ may influence kinetic behavior, but as a statistical thermodynamic property $\Delta C_p^\ddagger$ does not inform on any vibrational coupling to the reaction coordinate.

The signature of MMRT (curvature for temperature versus rate in Arrhenius or Eyring plots) has also been described for biological systems at increasing scales—from organism growth rates to ecosystems.[4,5] Extrapolation from the behavior of a few enzymes up to networks of thousands of enzymes or even thousands of organisms is risky—fools rush in where angels fear to tread. However, many schemes that currently describe the temperature dependence of organism growth rates or of ecosystem processes, such as respiration and photosynthesis, use functions based on the Arrhenius relationship. Curvature is observed for these processes, and it remains to be determined whether or not this curvature is a signature of MMRT in these higher order biological systems.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.5b01094.

Table S1: Numbers and references to accompany Figure 3 (PDF)

Figure S2: Fit of the MMRT function to data taken from papers by Daniel and Danson.[2] (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: varcus@waikato.ac.nz.

Present Address

[#]Department of Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia, Canada.

Funding

E.J.P. is supported by University of Waikato Ph.D. and Graduate Women Education Trust scholarships. Much of the work described here was part of a project funded by the Marsden Fund of New Zealand (08-UOW-57). A.J.M. and M.W.v.d.K. acknowledge funding from the BBSRC (Grant Numbers BB/L01386X/1 and BB/L018756/1) and EPSRC (EP/M022609/1). M.W.v.d.K. is funded by a BBSRC David Phillips Fellowship. L.A.S. acknowledges funding from Core Funding to New Zealand Crown Research Institutes as a subcontract from Landcare Research.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

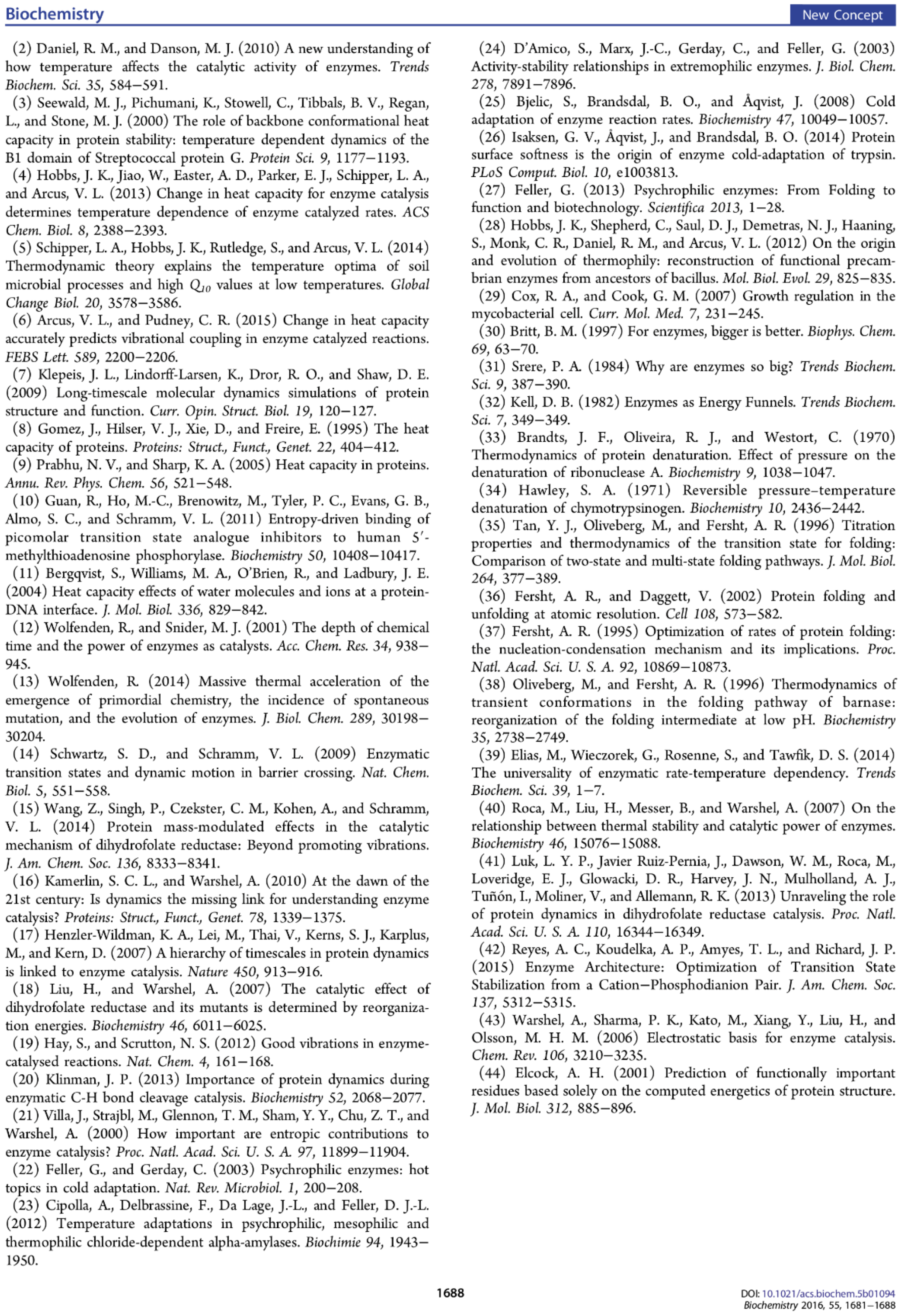
We would like to thank Jeremy Harvey for useful discussions and critical reading of the manuscript.

■ ABBREVIATIONS

MMRT, macromolecular rate theory; TIM, triose phosphate isomerase; MTAP, methylthioadenosine phosphorylase; IPMDH, isopropylmalate dehydrogenase

■ REFERENCES

(1) Glowacki, D. R., Harvey, J. N., and Mulholland, A. J. (2012) Taking Ockham's razor to enzyme dynamics and catalysis. *Nat. Chem.* *4*, 169−176.



(2) Daniel, R. M., and Danson, M. J. (2010) A new understanding of how temperature affects the catalytic activity of enzymes. *Trends Biochem. Sci. 35*, 584−591.

(3) Seewald, M. J., Pichumani, K., Stowell, C., Tibbals, B. V., Regan, L., and Stone, M. J. (2000) The role of backbone conformational heat capacity in protein stability: temperature dependent dynamics of the B1 domain of Streptococcal protein G. *Protein Sci. 9*, 1177−1193.

(4) Hobbs, J. K., Jiao, W., Easter, A. D., Parker, E. J., Schipper, L. A., and Arcus, V. L. (2013) Change in heat capacity for enzyme catalysis determines temperature dependence of enzyme catalyzed rates. *ACS Chem. Biol. 8*, 2388−2393.

(5) Schipper, L. A., Hobbs, J. K., Rutledge, S., and Arcus, V. L. (2014) Thermodynamic theory explains the temperature optima of soil microbial processes and high Q_{10} values at low temperatures. *Global Change Biol. 20*, 3578−3586.

(6) Arcus, V. L., and Pudney, C. R. (2015) Change in heat capacity accurately predicts vibrational coupling in enzyme catalyzed reactions. *FEBS Lett. 589*, 2200−2206.

(7) Klepeis, J. L., Lindorff-Larsen, K., Dror, R. O., and Shaw, D. E. (2009) Long-timescale molecular dynamics simulations of protein structure and function. *Curr. Opin. Struct. Biol. 19*, 120−127.

(8) Gomez, J., Hilser, V. J., Xie, D., and Freire, E. (1995) The heat capacity of proteins. *Proteins: Struct., Funct., Genet. 22*, 404−412.

(9) Prabhu, N. V., and Sharp, K. A. (2005) Heat capacity in proteins. *Annu. Rev. Phys. Chem. 56*, 521−548.

(10) Guan, R., Ho, M.-C., Brenowitz, M., Tyler, P. C., Evans, G. B., Almo, S. C., and Schramm, V. L. (2011) Entropy-driven binding of picomolar transition state analogue inhibitors to human 5′-methylthioadenosine phosphorylase. *Biochemistry 50*, 10408−10417.

(11) Bergqvist, S., Williams, M. A., O'Brien, R., and Ladbury, J. E. (2004) Heat capacity effects of water molecules and ions at a protein-DNA interface. *J. Mol. Biol. 336*, 829−842.

(12) Wolfenden, R., and Snider, M. J. (2001) The depth of chemical time and the power of enzymes as catalysts. *Acc. Chem. Res. 34*, 938−945.

(13) Wolfenden, R. (2014) Massive thermal acceleration of the emergence of primordial chemistry, the incidence of spontaneous mutation, and the evolution of enzymes. *J. Biol. Chem. 289*, 30198−30204.

(14) Schwartz, S. D., and Schramm, V. L. (2009) Enzymatic transition states and dynamic motion in barrier crossing. *Nat. Chem. Biol. 5*, 551−558.

(15) Wang, Z., Singh, P., Czekster, C. M., Kohen, A., and Schramm, V. L. (2014) Protein mass-modulated effects in the catalytic mechanism of dihydrofolate reductase: Beyond promoting vibrations. *J. Am. Chem. Soc. 136*, 8333−8341.

(16) Kamerlin, S. C. L., and Warshel, A. (2010) At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis? *Proteins: Struct., Funct., Genet. 78*, 1339−1375.

(17) Henzler-Wildman, K. A., Lei, M., Thai, V., Kerns, S. J., Karplus, M., and Kern, D. (2007) A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature 450*, 913−916.

(18) Liu, H., and Warshel, A. (2007) The catalytic effect of dihydrofolate reductase and its mutants is determined by reorganization energies. *Biochemistry 46*, 6011−6025.

(19) Hay, S., and Scrutton, N. S. (2012) Good vibrations in enzyme-catalysed reactions. *Nat. Chem. 4*, 161−168.

(20) Klinman, J. P. (2013) Importance of protein dynamics during enzymatic C-H bond cleavage catalysis. *Biochemistry 52*, 2068−2077.

(21) Villa, J., Strajbl, M., Glennon, T. M., Sham, Y. Y., Chu, Z. T., and Warshel, A. (2000) How important are entropic contributions to enzyme catalysis? *Proc. Natl. Acad. Sci. U. S. A. 97*, 11899−11904.

(22) Feller, G., and Gerday, C. (2003) Psychrophilic enzymes: hot topics in cold adaptation. *Nat. Rev. Microbiol. 1*, 200−208.

(23) Cipolla, A., Delbrassine, F., Da Lage, J.-L., and Feller, D. J.-L. (2012) Temperature adaptations in psychrophilic, mesophilic and thermophilic chloride-dependent alpha-amylases. *Biochimie 94*, 1943−1950.

(24) D'Amico, S., Marx, J.-C., Gerday, C., and Feller, G. (2003) Activity-stability relationships in extremophilic enzymes. *J. Biol. Chem. 278*, 7891−7896.

(25) Bjelic, S., Brandsdal, B. O., and Åqvist, J. (2008) Cold adaptation of enzyme reaction rates. *Biochemistry 47*, 10049−10057.

(26) Isaksen, G. V., Åqvist, J., and Brandsdal, B. O. (2014) Protein surface softness is the origin of enzyme cold-adaptation of trypsin. *PLoS Comput. Biol. 10*, e1003813.

(27) Feller, G. (2013) Psychrophilic enzymes: From Folding to function and biotechnology. *Scientifica 2013*, 1−28.

(28) Hobbs, J. K., Shepherd, C., Saul, D. J., Demetras, N. J., Haaning, S., Monk, C. R., Daniel, R. M., and Arcus, V. L. (2012) On the origin and evolution of thermophily: reconstruction of functional precambrian enzymes from ancestors of bacillus. *Mol. Biol. Evol. 29*, 825−835.

(29) Cox, R. A., and Cook, G. M. (2007) Growth regulation in the mycobacterial cell. *Curr. Mol. Med. 7*, 231−245.

(30) Britt, B. M. (1997) For enzymes, bigger is better. *Biophys. Chem. 69*, 63−70.

(31) Srere, P. A. (1984) Why are enzymes so big? *Trends Biochem. Sci. 9*, 387−390.

(32) Kell, D. B. (1982) Enzymes as Energy Funnels. *Trends Biochem. Sci. 7*, 349−349.

(33) Brandts, J. F., Oliveira, R. J., and Westort, C. (1970) Thermodynamics of protein denaturation. Effect of pressure on the denaturation of ribonuclease A. *Biochemistry 9*, 1038−1047.

(34) Hawley, S. A. (1971) Reversible pressure−temperature denaturation of chymotrypsinogen. *Biochemistry 10*, 2436−2442.

(35) Tan, Y. J., Oliveberg, M., and Fersht, A. R. (1996) Titration properties and thermodynamics of the transition state for folding: Comparison of two-state and multi-state folding pathways. *J. Mol. Biol. 264*, 377−389.

(36) Fersht, A. R., and Daggett, V. (2002) Protein folding and unfolding at atomic resolution. *Cell 108*, 573−582.

(37) Fersht, A. R. (1995) Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. *Proc. Natl. Acad. Sci. U. S. A. 92*, 10869−10873.

(38) Oliveberg, M., and Fersht, A. R. (1996) Thermodynamics of transient conformations in the folding pathway of barnase: reorganization of the folding intermediate at low pH. *Biochemistry 35*, 2738−2749.

(39) Elias, M., Wieczorek, G., Rosenne, S., and Tawfik, D. S. (2014) The universality of enzymatic rate-temperature dependency. *Trends Biochem. Sci. 39*, 1−7.

(40) Roca, M., Liu, H., Messer, B., and Warshel, A. (2007) On the relationship between thermal stability and catalytic power of enzymes. *Biochemistry 46*, 15076−15088.

(41) Luk, L. Y. P., Javier Ruiz-Pernia, J., Dawson, W. M., Roca, M., Loveridge, E. J., Glowacki, D. R., Harvey, J. N., Mulholland, A. J., Tuñón, I., Moliner, V., and Allemann, R. K. (2013) Unraveling the role of protein dynamics in dihydrofolate reductase catalysis. *Proc. Natl. Acad. Sci. U. S. A. 110*, 16344−16349.

(42) Reyes, A. C., Koudelka, A. P., Amyes, T. L., and Richard, J. P. (2015) Enzyme Architecture: Optimization of Transition State Stabilization from a Cation−Phosphodianion Pair. *J. Am. Chem. Soc. 137*, 5312−5315.

(43) Warshel, A., Sharma, P. K., Kato, M., Xiang, Y., Liu, H., and Olsson, M. H. M. (2006) Electrostatic basis for enzyme catalysis. *Chem. Rev. 106*, 3210−3235.

(44) Elcock, A. H. (2001) Prediction of functionally important residues based solely on the computed energetics of protein structure. *J. Mol. Biol. 312*, 885−896.

DOI: 10.1021/acs.biochem.5b01094
Biochemistry 2016, 55, 1681−1688