Understanding Allosteric Arginine Mutations Using Macromolecular Rate Theory

A thesis

submitted in fulfilment

of the requirements for the degree

of

Doctor of Philosophy in Molecular and Cellular Biology

at

The University of Waikato

by

Carlin James Hamill

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2024

Abstract

Enzyme catalysed reaction rates have been traditionally modelled with the Arrhenius and Eyring-Polanyi equations. These models assume that the reaction rate is exponential with temperature, and thus the natural log of the reaction rate versus 1/T is linear. Significant deviations from these models at high temperatures has traditionally been attributed to thermal denaturation. An increasing body of evidence has shown that denaturation alone is insufficient to account for these deviations from Arrhenius behaviour. Macromolecular rate theory (MMRT) accounts for these deviations with the introduction of the activation heat capacity ($\Delta C_P^{\ddagger}$) to the rate equation. The activation heat capacity is a consequence of the restriction in conformational freedom along the reaction coordinate, as an enzyme moves from the enzyme-substrate complex to the transition state complex – thus for an enzymatic reaction the activation heat capacity is negative. A non-zero activation heat capacity imparts temperature dependence to the activation entropy and activation enthalpy, introducing curvature to the rate equation independent of thermal denaturation. The activation heat capacity may itself be temperature dependent. MMRT equations of increasing complexity have been developed to reflect this and are suitable for different applications.

This thesis explores the effects of allosteric arginine mutations on the temperature dependence of enzyme rates through the lens of MMRT and evolution using the model enzyme MalL. An in-depth analysis of a previously characterised arginine mutant is described along with four additional arginine mutants. The arginine mutants were designed to mimic urea ligand binding across the enzyme surface. These mutants were characterised kinetically and with biophysical methods. Two were further characterised structurally, with high resolution structures being produced. These mutant enzymes showed significant rate improvements at low temperatures, suggesting two possible mechanisms for evolution towards psychrophily. These arginine mutants showed significant improvement in crystallographic resolution, indicating surface arginine mutations may be a general route for crystallographic improvement.

Acknowledgements

Firstly, thanks to Professor Vic Arcus for your guidance through this project, and being an endless source of expert knowledge, that has always pointed me in the right direction, and I am greatly appreciative of your mentorship. Thanks also to Dr Erica Prentice for your help and support throughout my time in the lab, and for your helpful critiques of this thesis.

Secondly, thanks to Andrew Howard, Bronwyn Crow, Elizabeth Rzoska-Smith, Emily Grout, Jack McGarvie, Keely Oldham, and everyone else in and around C2, past or present. You all have made the lab a wonderful place to work. A special thanks to Dr Judith Burrows for your support and positivity. Without you this lab would not run nearly so well.

Thirdly, I would like to thank Professor Adrian Mulholland for hosting me in Bristol, and giving me the opportunity to learn from you and your group. Thanks also to the members of the CCC who were very welcoming and made my stay very enjoyable.

I would also like to thank the University of Waikato for funding me with a Doctoral Scholarship that has allowed me carry out my research.

This research was undertaken in part using the MX2 beamline at the Australian Synchrotron, part of ANSTO, and made use of the Australian Cancer Research Foundation (ACRF) detector. Access to the Australian Synchrotron was supported by the New Zealand Synchrotron Group Limited.

Finally, I would like to thank my family, especially Mum and Dad. Your unwavering support and encouragement has been invaluable on this journey. Thanks also to Jacob (I am finally finished), Tara and Peter for your continued support.

Table of Contents

List of Figures

List of Tables

List of Equations

List of Abbreviations

AIC:	Akaike information criterion
AICc:	Corrected Akaike information criterion
APS:	Ammonium persulphate
BIC:	Bayesian information criterion
C_P:	Heat capacity at constant pressure
CSM:	Centre of spectral mass
DMSO:	Dimethyl sulphoxide
DNA:	Deoxyribonucleic acid
E_A:	Activation energy
EDTA:	Ethylenediaminetetraacetic acid
ES:	Enzyme substrate complex
FPLC:	Fast protein liquid chromatography
G:	Gibbs free energy
GUI:	Graphical user interface
H:	Enthalpy
IMAC:	Immobilised metal affinity chromatography
IPTG:	Isopropyl β-D-1-thiogalactopyranoside
LB:	Luria broth
m:	Temperature dependence factor of $\Delta C_P^\ddagger$
MQ:	Milli-Q ultrapure water
MD:	Molecular dynamics
MMRT:	Macromolecular rate theory
nm:	Nanometre

OD:	Optical density
PAGE:	Polyacrylamide gel electrophoresis
PCR:	Polymerase chain reaction
PBD:	Protein data bank
PEG:	Polyethylene glycol
PNG:	*p*-nitrophelyl-α- D-glucopyranoside
R:	Ideal gas constant
RDM:	Arginine double mutant
REES:	Red edge excitation shift
RMSD:	Root mean square deviation
RMSE:	Root mean square error
RNA:	Ribonucleic acid
S:	Entropy
SAXS:	Small angle X-ray scattering
SDS:	Sodium dodecyl sulphate
SEC:	Size exclusion chromatography
SOC:	Super optimal broth with catabolite repression
T:	Temperature
T_0:	Reference temperature
TAE:	Tris-Acetate-EDTA
T_c:	Transition temperature
TEMED:	Tetramethylethylenediamine
TIM:	Triose phosphate isomerase
T_{inf}:	Inflection point
TLC:	Transition state like complex

T_{Opt}:	Temperature optimum
Tris:	2-Amino-2-(hydroxymethyl)propane-1,3-diol
TS:	Transition state complex
TST:	Transition state theory
UV:	Ultraviolet
WT:	Wildtype

1 Introduction

Enzymes possess an extraordinary ability to catalyse chemical reactions, resulting in rate accelerations of up to 10^{17} over the uncatalysed reaction (Fersht, 1999). Understanding the precise mechanisms behind these rate increases is therefore of considerable interest. The rate of any chemical reaction is defined by the difference in energy between the reactants and the transition state. In general, enzymes are considered to function by lowering this difference in activation energy, increasing the proportion of reactants able to cross the transition state energy barrier.

An important factor in understanding enzyme function is their response to varied temperature. The complexities of enzyme catalysis with temperature has recently been the focus of much debate. Macromolecular rate theory (MMRT) is an extension of the traditional kinetic understanding of enzyme catalysis (Arcus *et al.*, 2016; Hobbs *et al.*, 2013; Walker *et al.*, 2023). This extension is vital for a full understanding of the temperature dependence of enzyme catalysis, and provides a number of mechanistic insights into biological reaction rates at a number of scales (Alster *et al.*, 2022; Duffy *et al.*, 2021; Hobbs *et al.*, 2013; Liang *et al.*, 2018; Prentice *et al.*, 2020; Schipper *et al.*, 2014).

1.1 Temperature dependence of chemical reaction rates

1.1.1 The Arrhenius model

The first formalisation in the development of modern kinetics and the temperature dependence of reaction rates was done by Svante Arrhenius in the late 19th century (Arrhenius, 1889). His treatment (Equation 1.1) was based entirely on empirical observations of chemical reaction rates.

Equation 1.1. Arrhenius equation

$$k = Ae^{\frac{-E_A}{RT}}$$

Arrhenius' work expanded on the prior work of Jacobus Henricus van't Hoff and Leopold Pfaundler (Laidler & King, 1983). At the time there were many competing models for the temperature dependence of enzyme rates, some of which had better empirical fits to experimental data. The simplicity of the Arrhenius model and the predictive insights into the mechanisms of chemical reactions led to it gaining prominence (Laidler, 1965).

Arrhenius introduced the concept of activation energy (E_A) which described the minimum required energy to cross a barrier associated with the reaction (Justi & Gilbert, 1999). The temperature dependence of the reaction was therefore explained by the fraction of molecules in this higher energy or 'activated' state (Justi & Gilbert, 1999). The pre-exponential term (A) or frequency factor described the frequency of reactant collisions, using kinetic theory in which molecules were treated as hard spheres (Justi & Gilbert, 1999). The activation energy and temperature (T) are exponentially related to the rate (*k*), such that linear increases in temperature result in exponential increases in rate. Likewise small decreases in activation energy yield exponential increases in rate. R is the gas constant, and relates the temperature scale with energy and amount of substance. The Arrhenius equation is a pseudo-equilibrium constant derived from the Van't Hoff equation (Equation 1.2), which relates the equilibrium constant (K_C), to the temperature, and the standard internal energy change (ΔU°).

Equation 1.2. Van't Hoff equation

$$\frac{d\ln(K_C)}{dT} = \frac{\Delta U^{\circ}}{RT^2}$$

For a reaction at equilibrium K_C is the ratio of the forward (k_1) and reverse (k_{-1}) rate constants, and thus Equation 1.2 can be represented as Equation 1.3,which can be decomposed into Equation 1.4 and Equation 1.5, where $\Delta U^{\circ} = E_1 - E_{-1}$.

Equation 1.3. Van't Hoff rate constant equation

$$\frac{d\ln(k_1)}{dT} - \frac{d\ln(k_{-1})}{dT} = \frac{\Delta U^{\circ}}{RT^2}$$

Equation 1.4. Van't Hoff equation for forward rate constant

$$\frac{d\ln(k_1)}{dT} = \frac{E_1}{RT^2} + I$$

Equation 1.5. Van't Hoff equation for reverse rate constant

$$\frac{d\ln(k_{-1})}{dT} = \frac{E_{-1}}{RT^2} + I$$

Experimentally, it has been determined that the constant I can be set equal to zero (Laidler, 1965), and therefore integration of Equation 1.4 (assuming E_1 is independent of temperature), yields Equation 1.6.

Equation 1.6. Derivation of Arrhenius equation from Van't Hoff equation

$$k_1 = A_1 e^{-E_1/RT}$$

1.1.2 Transition state theory

The Arrhenius model was expanded by both Henry Eyring and Meredith Evans and Michael Polanyi, working independently in 1935. These models unified aspects of previous kinetic, thermodynamic and statistical mechanic models into one combined model (Justi & Gilbert, 1999). The new model was termed transition-state theory (TST). TST further developed the activation energy as well as defined and provided a theoretical basis for the pre-exponential factor (Laidler & King, 1983).

Transition state theory considered a chemical reaction as a process in which reactant molecules are distributed across a potential energy surface (Justi & Gilbert, 1999). A reaction is treated as a movement over the potential energy surface in which the system passes through a continuous series of configurations from reactant(s) to product(s) (Justi & Gilbert, 1999). For a reaction to proceed the system must pass through high energy configuration of minimum probability termed the transition state (Justi & Gilbert, 1999). TST incorporated the statistical mechanic and thermodynamic aspects of the system passing through the transition state and combined it with the kinetic contributions from molecular collisions (Justi & Gilbert, 1999).

Transition state theory (Equation 1.7) quantified the reaction rate in terms of the difference in energy between the reactant state and the transition state ($\Delta G^{\ddagger}$) (Arcus & Mulholland, 2020). The pre-exponential factor of the Arrhenius Model was reworked as the barrierless frequency for the reaction ($k_B T/h$) where k_B is the Boltzmann constant and h is the Planck constant. The transmission coefficient (κ) accounts for the proportion of reactants that reach the transition state but do not form products, and is generally set to one for most enzymatic reactions (Arcus *et al.*, 2016; Laidler & King, 1983). The temperature dependence of the transmission coefficient may become relevant for some enzymatic reactions (Pu *et al.*, 2006; Truhlar, 2015)

Equation 1.7. Transition state theory, Eyring-Polanyi equation

$$k = \kappa \left(\frac{k_B T}{h}\right) e^{-\Delta G^{\ddagger}/RT}$$

The Gibbs free energy ($\Delta G^{\ddagger}$) can be considered as temperature dependent according to Equation 1.8. The Eyring-Polanyi equation (Equation 1.7) can thus be expanded into Equation 1.9 and Equation 1.10

Equation 1.8. Temperature dependence of Gibbs free energy

$$\Delta G^{\ddagger} = \Delta H^{\ddagger} - T\Delta S^{\ddagger}$$

Equation 1.9. Expanded Eyring-Polanyi equation

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

Equation 1.10. Log form of the expanded Eyring-Polanyi equation

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

1.2 Models for temperature dependence of enzyme rates

1.2.1 Denaturation

The traditional and most widely cited reason for the temperature dependence of enzyme rates is thermal denaturation of the enzyme. In this model rate increases up to the temperature optimum follow Arrhenius kinetics (Equation 1.1 and Equation 1.7) and any decrease at elevated temperatures is attributable to increasing levels of denaturation (Daniel *et al.*, 2001). The temperature optimum is therefore not an intrinsic property of enzyme catalysis, but rather an artefact of enzyme stability. The level of denaturation is both time and temperature dependent (Daniel *et al.*, 2001).

An increasing amount of experimental evidence shows that thermal denaturation alone is insufficient to fully account for rate deviations with temperature. For the denaturation model to hold true, the optimum temperature should increase with shorter measurement times and the rate should increase exponentially with temperature according to Arrhenius kinetics (Thomas & Scopes, 1998). Both of these features can be demonstrated to be false. Using detailed kinetic assays Thomas and Scopes were able to show that the reduction in activity of phosphoglycerate kinase occurred in advance of thermal denaturation (Thomas & Scopes, 1998). Psychrophilic enzymes often show large discrepancies between their optimum temperatures and their thermal denaturation temperatures. Many psychrophilic enzymes show a decrease in rate up to 20 K before their denaturation temperature, unlike their mesophilic and thermophilic counterparts, whose denaturation temperatures more closely match their optimum (Figure 1.1) (Feller & Gerday, 2003). Denaturation is therefore insufficient to explain this loss in rate as temperature increases. It has been proposed that localised instability of the active site or supporting regions, where these regions are more heat-labile than the bulk of the enzyme, are responsible for the loss of activity at high temperatures (D'Amico *et al.*, 2003).

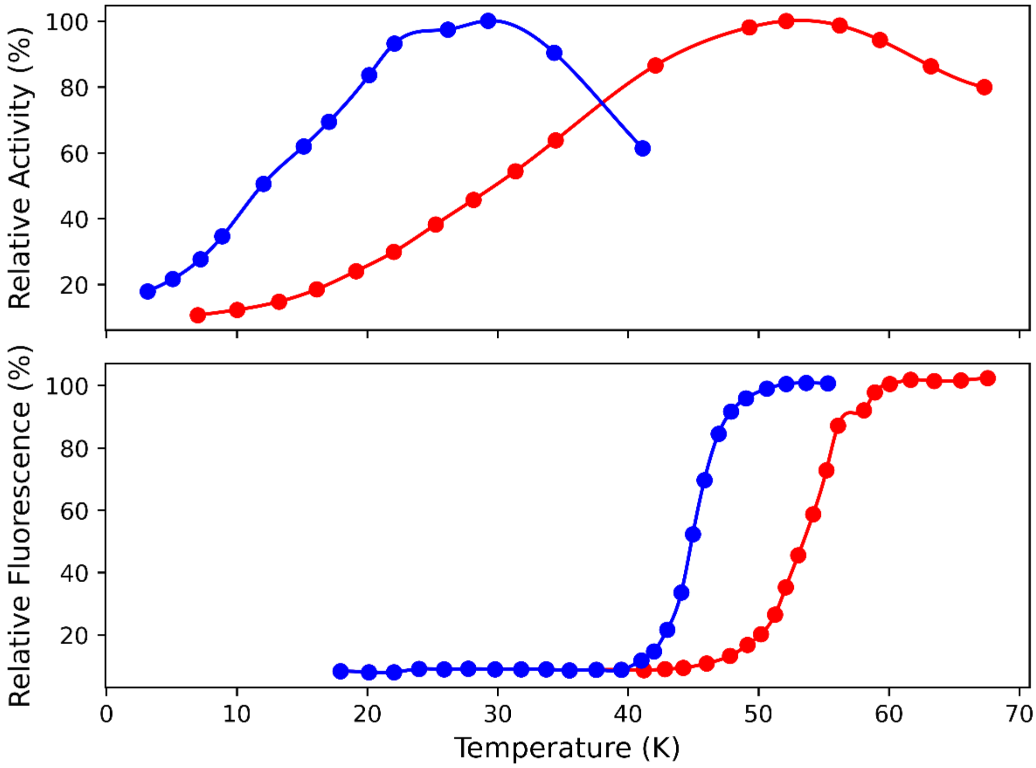


Figure 1.1. Enzyme activity versus thermal denaturation for psychrophilic (blue), mesophilic (red). Psychrophilic enzymes show significant decreases in activity at temperatures where there is no measurable thermal denaturation. Adapted from D'Amico* et al. *(2006).

1.2.2 Equilibrium models

To account for this discrepancy, the equilibrium model introduces an equilibrium between a catalytically active and a reversibly inactive form of the enzyme. The equilibrium between these two states can explain rate deviations in the absence of irreversible denaturation (Daniel & Danson, 2010). In this model the active enzyme (E_{act}) is in rapid equilibrium with the inactive enzyme (E_{inact}), and irreversible thermal denaturation proceeds from this state (Equation 1.11). Temperature dependence of the enzyme rate is thus based on both the k_{cat} (governed by Arrhenius kinetics), and the equilibrium between active and inactive states, which is temperature dependent based on Equation 1.12. ΔH_{eq} is the enthalpy change of the inactivation and T_{eq} is the midpoint of the transition.

Equation 1.11. Equilibrium model

$$E_{act} \overset{K_{eq}}{\rightleftharpoons} E_{inact} \rightarrow X$$

Equation 1.12. Temperature dependence of inactivation

$$ln(K_{eq}) = \frac{\Delta H_{eq}}{R}\left[\frac{1}{T_{eq}} - \frac{1}{T}\right]$$

The equilibrium model does not itself give a physical description of the molecular basis of the enzyme inactivation (Daniel & Danson, 2010). Curvature at fast timescales is a feature of this model (even at t=0), which implies that the conversion between the active and inactive forms is very rapid, and occurs on a much faster timescale than denaturation. In addition, the ΔH_{eq} is an order of magnitude smaller than ΔH_{unfold}. Therefore the inactive enzyme cannot be significantly unfolded and is distinct from denaturation (Daniel & Danson, 2010).

Similarly Åqvist and colleagues have invoked an equilibrium model to describe curvature in enzyme rates (Sočan *et al.*, 2020). In this model the enzyme substrate (ES) complex is in equilibrium with an inactivated form of the ES complex (ES'), with catalysis proceeding from the active form Equation 1.13. The enzyme temperature dependence arises from the equilibrium shift from the ES state to the ES’ state as the temperature is increased (Åqvist *et al.*, 2020).

Equation 1.13. Equilibrium model

$$\begin{tikzcd} E + S \arrow[r, shift left, "k_1"] & ES \arrow[l, shift left] \arrow[r, "k_3"] \arrow[d, shift left, "k_2"] & E + P \\ & ES' \arrow[u, shift left] & \end{tikzcd}$$

1.2.3 Multistate

The multistate model is similar to the equilibrium model in that the active enzyme is in a reversible equilibrium with inactive or partially active forms; however, it postulates that the enzyme exists as a broad conformational ensemble of states (Elias *et al.*, 2014). Each state is structurally similar to the folded state involving only minor perturbations in bond lengths and angles and may have different catalytic efficiencies. The enzyme temperature dependence arises from the broadening of the structural ensemble with increased temperature, which shifts the structural conformation of the enzyme away from those states with optimal catalytic efficiency (Elias *et al.*, 2014).

1.3 Macromolecular rate theory

Illustrations in this section are derived from MalL wildtype (Section 1.9) (Walker *et al.*, 2023), unless otherwise indicated.

The Arrhenius and Eyring-Polanyi equations have also been widely applied to biological systems. It is generally accepted that deviations from these models at high temperatures are caused by irreversible thermal denaturation. However, an increasing body of evidence has shown that denaturation alone is unable to account for these deviations. The rate constant of an enzyme following Arrhenius/Eyring-Polanyi kinetics is expected to increase exponentially with temperature. Enzymes have been shown to have a non-exponential increase in rate with temperature as well as significant decreases in rate at temperatures where the rate of thermal denaturation is insignificant (Thomas & Scopes, 1998). The loss of activity prior to denaturation is particularly relevant for psychrophilic enzymes (enzymes adapted to work at cold temperatures) which often show substantial rate decreases at temperatures well below their thermal denaturation temperature (Figure 1.1) (D'Amico *et al.*, 2006).

Macromolecular rate theory (MMRT) provides an alternate model for describing the temperature dependence of enzyme catalysed rates (Hobbs *et al.*, 2013). MMRT adds an additional parameter for the activation heat capacity ($\Delta C_P^{\ddagger}$) into the rate equation. The heat capacity is the relationship between a system’s ability to partition energy and the change in temperature. At constant pressure the heat capacity is defined by Equation 1.14.

Equation 1.14. Heat capacity at constant pressure

$$C_P = \left(\frac{\partial q}{\partial T}\right)_P = \left(\frac{dH}{dT}\right)_P$$

A protein system can absorb energy into available rotational, translational, vibrational and electronic modes, of which the vibrational modes make the greatest contribution to the heat capacity (Gómez *et al.*, 1995). At biologically relevant temperatures electronic modes are generally not accessible, and the number of vibrational modes (3N-6) greatly outnumber the rotational and translational modes (3 each in the x, y, z coordinates). Proteins are large and flexible and thus have a high number of accessible vibrational modes (Gómez *et al.*, 1995). The activation heat capacity is the difference in heat capacity between the reactant state and the transition state. The large heat capacity of proteins thus contributes to a significant change in heat capacity between the enzyme substrate complex and the enzyme transition state complex. Therefore, for enzymatic reactions the activation heat capacity contributes significantly to the rate. The activation heat capacity defines the temperature dependence of the activation enthalpy ($\Delta H^{\ddagger}$) and activation entropy ($\Delta S^{\ddagger}$), according to Equation 1.15 and Equation 1.16. The reaction rate and Gibbs free energy can thus be calculated according to Equation 1.17 and Equation 1.18 respectively. T_0 is an arbitrary reference temperature.

Equation 1.15. Temperature dependence of activation enthalpy

$$\Delta H^{\ddagger} = \Delta H^{\ddagger}_{T_0} + \int_{T_0}^{T} \Delta C^{\ddagger}_{P}\, dT$$

Equation 1.16. Temperature dependence of activation entropy

$$\Delta S^{\ddagger} = \Delta S^{\ddagger}_{T_0} + \int_{T_0}^{T} \frac{\Delta C_P^{\ddagger}}{T}\, dT$$

Equation 1.17. Temperature dependence of reaction rate

$$ln(k) = ln\left(\frac{k_B T}{h}\right) - \frac{\Delta H^{\ddagger}_{T_0} + \int_{T_0}^{T} \Delta C_P^{\ddagger}\, dT}{RT} + \frac{\Delta S^{\ddagger}_{T_0} + \int_{T_0}^{T} \frac{\Delta C_P^{\ddagger}}{T}\, dT}{R}$$

Equation 1.18. Temperature dependence of Gibbs free energy

$$\Delta G^{\ddagger} = \Delta H^{\ddagger}_{T_0} - T\Delta S^{\ddagger}_{T_0} + \int_{T_0}^{T} \Delta C_P^{\ddagger}\, dT - T\int_{T_0}^{T} \frac{\Delta C_P^{\ddagger}}{T}\, dT$$

Substituting Equation 1.15 and Equation 1.16 into Equation 1.10 and integrating $\Delta C_p^{\ddagger}$ as a constant yields the MMRT 1.0 equation, Equation 1.19. Similarly integrating Equation 1.18 with $\Delta C_p^{\ddagger}$ held constant gives the activation Gibbs free energy form, Equation 1.20. In MMRT 1.0 the activation heat capacity is assumed to be constant across all temperatures. This is likely an approximation.

Equation 1.19. MMRT 1.0, Constant heat capacity

$$\ln(k) = \ln\left(\frac{k_B T}{h}\right) - \frac{\Delta H^{\ddagger}_{T_0} + \Delta C_P^{\ddagger}(T - T_0)}{RT} + \frac{\Delta S^{\ddagger}_{T_0} + \Delta C_P^{\ddagger}(\ln(T/T_0))}{R}$$

Equation 1.20. MMRT 1.0, Constant heat capacity

$$\Delta G^{\ddagger} = \Delta H^{\ddagger}_{T_0} - T\Delta S^{\ddagger}_{T_0} + \Delta C_P^{\ddagger}(T - T_0 - T\ln(T/T_0))$$

The inclusion of activation heat capacity into the rate equation adds a new dimension of understanding to the temperature dependence of enzyme rates. The heat capacity of an enzyme is defined mostly by the number of available vibrational modes, thus the activation heat capacity for an enzymatic reaction is generally large and negative, due to a restriction in conformational freedom as the enzyme tightly binds to the substrate at the transition state on the reaction coordinate (Arcus *et al.*, 2016). A non-zero activation heat capacity results in a curved temperature dependence of the rate, where the degree of curvature is defined by the magnitude of the activation heat capacity (Figure 1.2) (Hobbs *et al.*, 2013). If $\Delta C_P^{\ddagger} = 0$ the function collapses back to an Arrhenius/Eyring function.

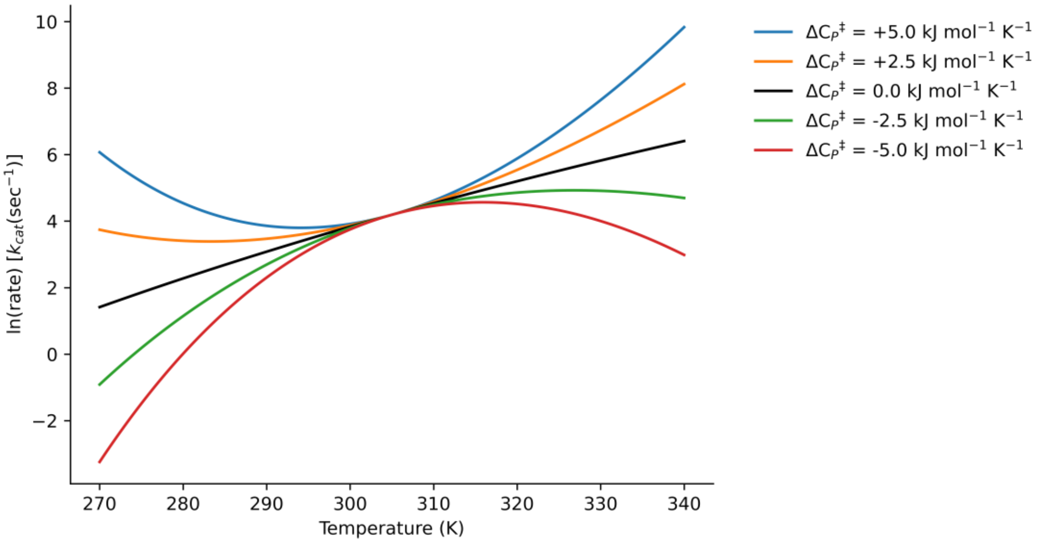


Figure 1.2. Effect of activation heat capacity on temperature dependence of rates. Curves are calculated using Equation 1.19. Direction and degree of curvature is defined by the sign and magnitude of the activation heat capacity, respectively. If the activation heat capacity is zero, the function collapses to an Arrhenius/Eyring function (Black).

The value of the activation heat capacity is also itself temperature dependent. This temperature dependence can be approximated as a linear change with temperature according to Equation 1.21 (Prabhu & Sharp, 2005), where $\Delta C_{P,0}^{\ddagger}$ is the value of the activation heat capacity at the y-intercept (0 K), and m is the slope of the line. This equation may also be represented by Equation 1.22 which places the intercept at the reference temperature, providing a more biologically relevant value for the activation heat capacity.

Equation 1.21. Linear heat capacity equation

$$\Delta C_P^{\ddagger} = \Delta C_{P,0}^{\ddagger} + mT$$

Equation 1.22. Linear heat capacity equation

$$\Delta C_P^{\ddagger} = \Delta C_{P,T_0}^{\ddagger} + m(T - T_0)$$

Substituting Equation 1.21 into Equation 1.17 and Equation 1.18 and integrating gives the MMRT 1.5 equations, Equation 1.23 and Equation 1.24.

Equation 1.23. MMRT 1.5, Linear heat capacity

$$\ln(k) = \ln\left(\frac{k_B T}{h}\right) - \frac{\Delta H_{T_0}^{\ddagger} + \Delta C_{P,0}^{\ddagger}(T - T_0) + \frac{m}{2}(T^2 - T_0^2)}{RT} + \frac{\Delta S_{T_0}^{\ddagger} + \Delta C_{P,0}^{\ddagger}(\ln(T/T_0)) + m(T - T_0)}{R}$$

Equation 1.24. MMRT 1.5, Linear heat capacity

$$\Delta G^{\ddagger} = \Delta H_{T_0}^{\ddagger} - T\Delta S_{T_0}^{\ddagger} + \Delta C_{P,0}^{\ddagger}(T - T_0 - T\ln(T/T_0)) - \frac{m}{2}(T - T_0)^2$$

Note that a linear dependence of $\Delta C_P^{\ddagger}$ results in an extra T^2 term in $\Delta G^{\ddagger}$ and the overall $\Delta G^{\ddagger}$ function behaves like a cubic. Using a linearly temperature dependent activation heat capacity allows for a better fit to many experimental datasets. For example the enzymes glucokinase and fructose bisphosphate aldolase II were found to deviate from Equation 1.19 and were better fit with Equation 1.23 (Prentice *et al.*, 2020). A temperature dependent activation heat capacity allows for the degree of curvature to vary over the temperature range allowing for a more accurate fit (Figure 1.3).

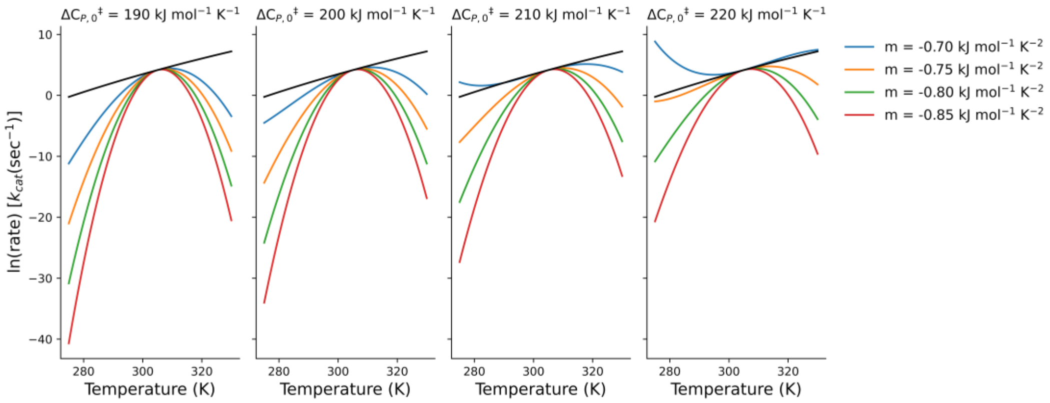


Figure 1.3. Effect of linear activation heat capacity on temperature dependence of rates. Curves are calculated using Equation 1.23. Activation heat capacity is linearly dependent on temperature. If the activation heat capacity is zero, the function collapses to an Arrhenius/Eyring function (Black).

1.3.1 Temperature optimum and inflection point

The temperature optimum of an enzyme is often reported as a measure of thermal adaptation in enzyme characterisations (Prentice *et al.*, 2020). The temperature optimum (T_{opt}) is the temperature at which the rate is maximal, defined where the first derivative of the rate equation is equal to zero ($dk/dT = 0$) (Figure 1.4 A). In general the T_{opt} is only loosely correlated with the growth temperature, with T_{opt} values increasing from psychrophiles to thermophiles. Enzymatic T_{opt} values are consistently higher than environmental and growth temperatures and show a large degree of variance between individual enzymes. This implies that the T_{opt} is not under direct selective pressure. The inflection point (T_{inf}), defined where the second derivative of the rate equation is equal to zero ($d^2k/dT^2 = 0$), is instead hypothesised to be acted on by evolutionary pressure (Figure 1.4 B). The T_{inf} is located in the steepest portion of the temperature curve and enzyme rates are approximately linear about this point. Placing the T_{inf} at the growth temperature allows the activity of multiple enzymes to scale linearly with minor fluctuations in temperature. Thus, metabolic activity remains coordinated across small variations in environmental temperature (Figure 1.4 B). If instead the T_{opt} was placed at the growth temperature minor changes in temperature would result in diverging enzyme activity as small changes in temperature result in steep decreases in enzyme rate away from the T_{opt} (Prentice *et al.*, 2020).

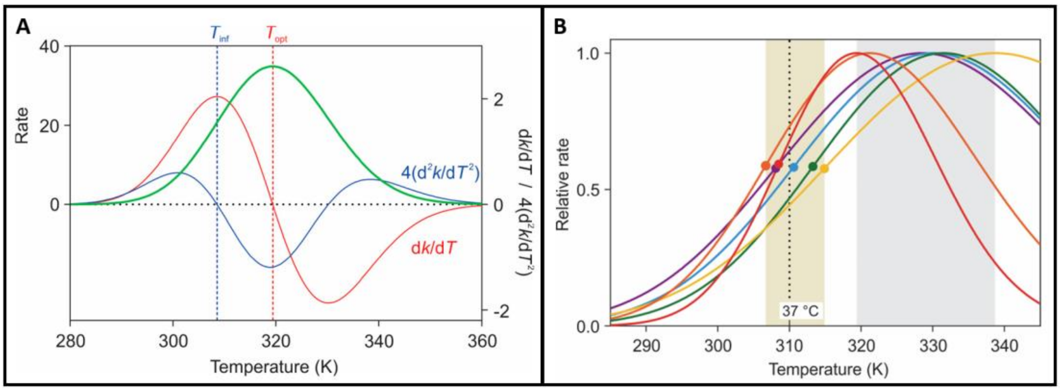


Figure 1.4. The inflection point (T_{inf}) is evolutionarily selected for to enable coordination of metabolism. A) The temperature dependence of enzyme rates (green), with first (red) and second (blue) derivatives. Vertical lines represent where the first and second derivatives equal zero, marking the T_{opt} and T_{inf}, respectively. The second derivative is scaled by a factor of four for clarity. B) Relative temperature dependence of enzyme rates for glycolytic enzymes from E. coli. The T_{inf} values (circles) are narrowly clustered around the organism's growth temperature (beige box). The T_{opt} is widely distributed above the growth temperature (grey box). Reprinted (adapted) with permission from Prentice et al. (2020). Copyright 2020 American Chemical Society.

1.3.2 Cooperative transitions and activation heat capacity

High-resolution temperature-rate data previously collected for the model enzyme MalL shows two arms with a transition at approximately 313 K (Figure 1.5). These two arms can be fit independently using the $\Delta G^{\ddagger}$ form of MMRT 1.0 (Equation 1.20) (Walker *et al.*, 2023). The low temperature arm has a $\Delta C_P^{\ddagger}$ close to zero, while the high temperature arm has a large negative $\Delta C_P^{\ddagger}$. This suggests that a cooperative transition occurs between these two arms. A cooperative transition occurs where each component of a system is not independent and the state of an individual component is influenced by neighbouring components. Thus state changes occur rapidly with only minor changes to input parameters. This gives rise to transitions with sigmoidal character.

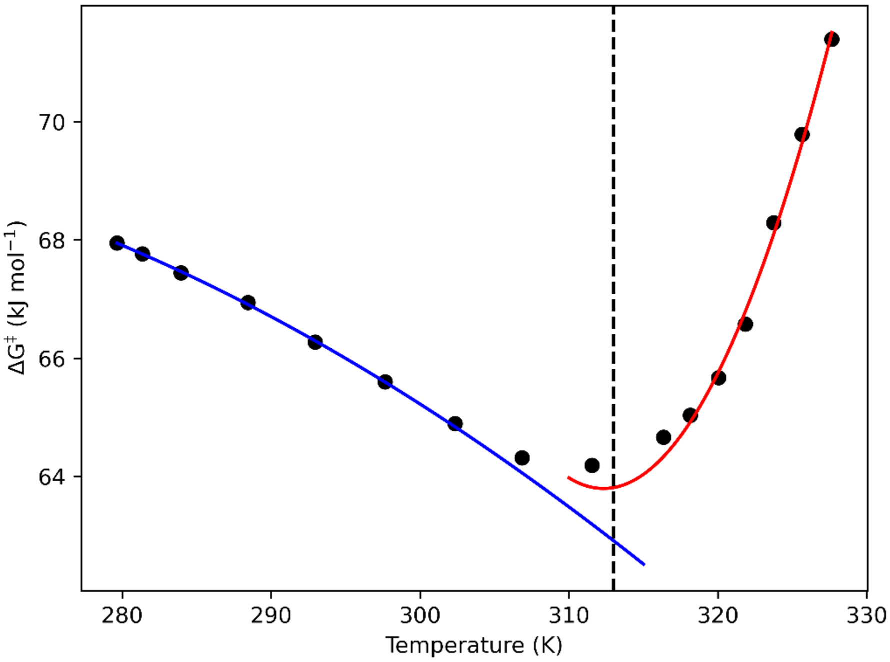


Figure 1.5. Temperature dependence of activation Gibbs free energy. Shows two distinct arms with a cooperative transition between the two arms. Midpoint of transition (T_C) is shown with the dotted line. Curves are the fit of Equation 1.20 to the low temperature portion of the data (blue) and the high temperature portion of the data (red).

The activation heat capacity for a system undergoing a cooperative transition can be modelled using a sigmoid with the general form of Equation 1.25.

Equation 1.25. Sigmoidal activation heat capacity equation

$$\Delta C_P^{\ddagger} = \frac{\left(\Delta C_{P,lowT}^{\ddagger}\right) + \left(\Delta C_{P,highT}^{\ddagger}\right) e^{\frac{-\Delta\Delta H^{\ddagger}(1-T/T_C)}{RT}}}{1 + e^{\frac{-\Delta\Delta H^{\ddagger}(1-T/T_C)}{RT}}}$$

$\Delta C_{P,lowT}^{\ddagger}$ and $\Delta C_{P,highT}^{\ddagger}$ are the values of $\Delta C_P^{\ddagger}$ at low and high temperatures, respectively. $\Delta\Delta H^{\ddagger}$ describes the slope of the transition, with a larger $\Delta\Delta H^{\ddagger}$ resulting in a steeper transition. T_C is the midpoint of the transition (Figure 1.5, Figure 1.6). As with MMRT 1.0 and MMRT 1.5 substituting Equation 1.25 into Equation 1.17 and Equation 1.18 yield the MMRT 2.0 equations, however these do not have a simple solution due to the required integration of Equation 1.25.

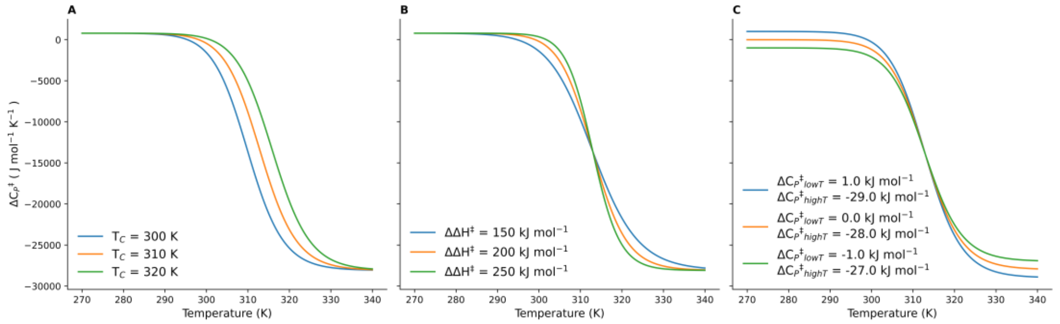


Figure 1.6. Temperature dependence of sigmoidal activation heat capacity. A) T_C shifts the transition of the curve to different temperatures. B) ΔΔH‡ changes the slope of the transition. C) $\Delta C^{\ddagger}_{P,lowT}$ and $\Delta C^{\ddagger}_{P,highT}$ **set the $\Delta C^{\ddagger}_{P}$ values at either side of the transition.**

Cooperative transitions occur in many aspects of biology, for example protein folding, ligand binding, enzyme catalysis and allosteric regulation (Fersht, 1999). The enzyme-substrate (ES) complex has a larger conformational freedom with at least two states, and the transition-state (E-TS) complex is much more tightly constrained (Arcus & Mulholland, 2020). For catalysis to occur the enzyme-substrate complex must first visit a transition-state-like complex (TLC) and this is the bottleneck in phase space for catalysis (Walker *et al.*, 2023). The TLC is a distinct substate of the ES complex from which the transition state can be reached. The $\Delta\Delta H^{\ddagger}$ is defined by the difference in enthalpy between these two states. Substrate binding is followed by a conformational equilibrium between the ES and TLC complexes, and catalysis follows from the TLC complex (Walker *et al.*, 2023) (Equation 1.26). A constriction in the conformational freedom along a coordinate will necessarily give rise to a change in heat capacity.

Equation 1.26. Enzyme catalysis proceeds through a transition-state like complex

$$E + S \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} ES \underset{k_{con\_r}}{\overset{k_{con\_f}}{\rightleftharpoons}} TLC \xrightarrow{k_{chem}} E + P$$

Thus as temperature increases, the wider conformational space sampled in the ES results in a more negative activation heat capacity, and the increasing entropic barrier leads to a decrease in the rate (Arcus *et al.*, 2016) (Figure 1.7). This is best modelled as a sigmoidal $\Delta C_P^\ddagger$ as in MMRT 2.0; however, this leads to an unwieldy equation with six fit parameters that can be difficult to converge. The linear $\Delta C_P^\ddagger$ model is a suitable approximation as this captures both the size of the change in $\Delta C_P^\ddagger$ across the temperature range and the slope of the transition ($\Delta\Delta H^\ddagger$). The form of MMRT used for an individual dataset is often dependent on the specifics of the system under investigation, and the temperature range examined.

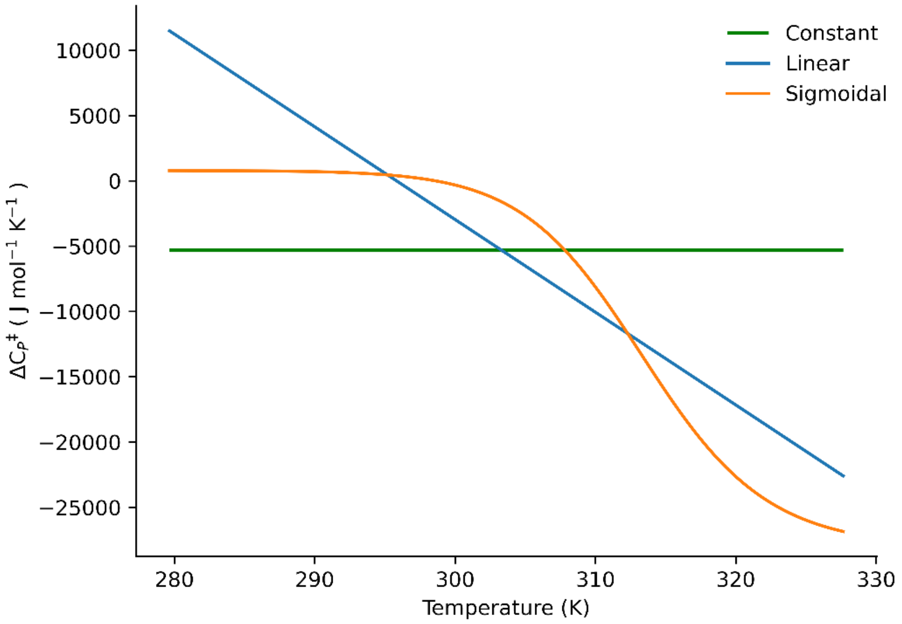


Figure 1.7. Activation heat capacity for different MMRT forms.

Macromolecular rate theory has been used across a wide variety of biological rate data, at various scales. MMRT has been used to explain the anomalous temperature dependence of enzymes from traditional Arrhenius kinetics (Hobbs *et al.*, 2013). It has been shown to be applicable to multiple enzymes acting together such as during metabolism (Prentice *et al.*, 2020), as well as the temperature dependent kinetics of membrane conductance and neuron excitation (Pahlavan *et al.*, 2023). MMRT can also be used to explain the temperature dependence of organism level processes such as respiration in soils, and nitrogen and methane cycling (Alster *et al.*, 2016b; Alster *et al.*, 2022; Alster *et al.*, 2020; Schipper *et al.*, 2014) and respiration in plant leaves (Liang *et al.*, 2018). MMRT has also been applied to ecosystem scales such as photosynthesis and respiration in the terrestrial biosphere (Duffy *et al.*, 2021).

1.4 Evolution of psychrophily

Psychrophiles are extremophilic bacteria capable of growth and activity at extremely low temperatures. These psychrophiles require many adaptations to thrive at low temperatures, including extensive optimisation of their enzymes whose activity is highly temperature sensitive. Psychrophilic enzymes, in general compared to mesophilic (moderate temperatures) and thermophilic (high temperatures) enzymes, display high activity at low temperatures, have a decreased temperature optimum, and have a reduced rate compared to mesophilic homologs at moderate temperatures (Collins & Feller, 2023).

Psychrophilic enzymes have many molecular adaptations to improve activity at low temperatures. The two parameters important for enzyme activity are the k_{cat} and the K_M. Adaptation to low temperatures would ideally require optimisation of both parameters; however, most psychrophilic enzymes optimise k_{cat} at the expense of K_M (Collins & Feller, 2023). Optimisation of k_{cat} requires tuning of the thermodynamic parameters (via Equation 1.8) to decrease the activation energy for the reaction at low temperatures.

Enzymes require a degree of conformational fluctuations to sustain catalysis. Psychrophiles often display an increased flexibility, and associated decrease in stability, as compared to their mesophilic homologs (Collins & Feller, 2023). Psychrophiles achieve this increase in flexibility by a decrease in the number of stabilising interactions or by the introduction of specific flexibility enhancing modifications (Collins & Feller, 2023). Destabilising interactions include modification of hydrogen and aromatic bonding networks, weakening of the hydrophobic core, more hydrophobic surface residues, and less commonly, reduced binding of stabilising ions, weakened charge-dipole interactions in α-helices, and an excess surface charge (Collins & Feller, 2023). Flexibility enhancing interactions include clustering of highly mobile glycine residues, improved loop flexibility by lengthening loops or reducing the number of restrictive proline residues, and increased protein-solvent interactions by increased surface charges or increased solvent cavity content and size (Collins & Feller, 2023).

The energy barrier of a reaction ($\Delta G^{\ddagger}$) is composed of both entropic and enthalpic contributions and can be reduced by decreasing the enthalpy of activation ($\Delta H^{\ddagger}$) and/or increasing the entropic contribution ($\mathrm{T}\Delta S^{\ddagger}$) (Collins & Feller, 2023). Psychrophilic enzymes are generally characterised by having a decreased enthalpy at the expense of a lowered entropic contribution. A lowered enthalpy of activation indicates a decrease in the number of weak enthalpic interactions that are broken during the activation step (Collins & Feller, 2023). These interactions also contribute to protein stability and rigidification. Therefore the increased flexibility exhibited by psychrophilic enzymes results in a decrease of these interactions and therefore a decrease in $\Delta H^{\ddagger}$. A greater flexibility in the enzyme-substrate complex requires greater reorganisation to reach the transition-state complex, and therefore lowering the $\Delta H^{\ddagger}$ consequently results in an increased entropic penalty (lowered $\Delta S^{\ddagger}$) (Collins & Feller, 2023).

1.4.1 Heat capacity in psychrophile evolution

Psychrophilic enzymes tend to have a lower activation heat capacity ($\Delta C_P^{\ddagger}$) than their mesophilic homologs. A more negative $\Delta C_P^{\ddagger}$ will shift the optimum temperature to lower temperatures. This can be achieved by increasing the heat capacity of the enzyme-substrate complex (more low-frequency vibrational modes / greater flexibility) or decreasing the heat capacity of the enzyme-transition state complex (fewer low-frequency vibrational modes / more rigid) (Arcus *et al.*, 2016). An important consequence of a more negative heat capacity, is a steeper temperature dependence and more defined curvature in the rate versus temperature. This places a lower limit on temperature adaptation where a large negative $\Delta C_P^{\ddagger}$ results in an enzyme that is overly sensitive to small changes in temperature (Arcus *et al.*, 2016).

1.5 X-ray crystallography

X-ray crystallography is a method for determining the molecular structure of a crystal. It is widely used for structural determination of small molecules as well as proteins and other biomolecules (DNA, RNA). It utilises the interaction of X-ray radiation with elementary particles and is based on the scattering of X-ray radiation by electrons (Drenth, 1999). X-ray crystallography utilises a monochromatic X-ray source and measures the diffraction of the scattered X-rays. When an X-ray (of a particular wavelength) hits an electron some of the energy is absorbed. This energy is then re-emitted at the same wavelength as a spherical wave (Blow, 2002). These waves can interact with each other and add up constructively or destructively resulting in a diffraction pattern (Figure 1.8). Constructive interference occurs when waves line-up in phase with each other and sum up additively. When waves line up out of phase they add up destructively (Drenth, 1999).

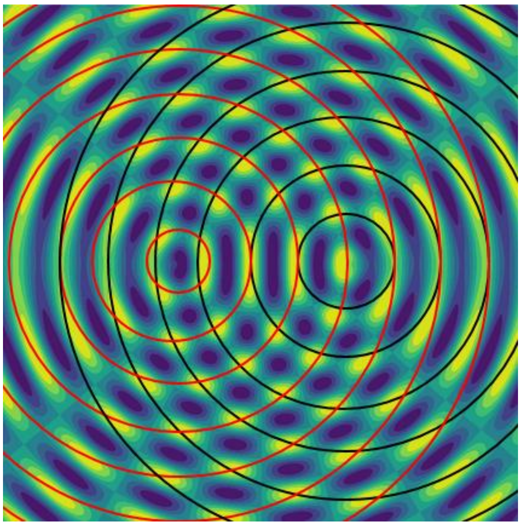


Figure 1.8. Interference pattern of spherical waves. Adapted from Schorsch (2005).

When this process occurs in a regular crystal lattice diffracted X-rays add up constructively at specific angles relative to the angle of incident radiation (Blow, 2002). Diffracted X-rays add up constructively when distances in the crystal lattice correspond to an integer multiple of the wavelength of incident radiation (Drenth, 1999). When this occurs the X-rays add constructively at an angle of 2θ (where θ is the angle of incident radiation) from the incident radiation (Figure 1.9). This is represented mathematically using Bragg's Law (Equation 1.27).

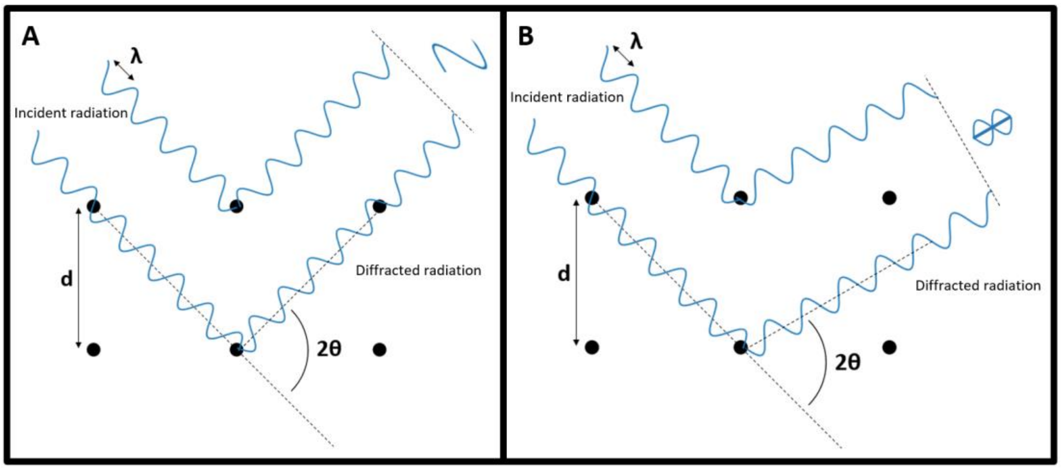


Figure 1.9. Braggs Law. A) Diffracted X-rays are in phase add constructively leading to amplification of the signal. B) Diffracted X-rays are out of phase and add destructively leading to signal loss. Adapted from Dang Ngoc Chan (2011).

Equation 1.27. Bragg's law

$$n\lambda = 2d\sin\theta$$

In a protein crystal the crystal lattice consists of regularly spaced planes of electron density, which divide the unit cell and scatter X-rays proportionally to the average electron density of the plane. When a protein crystal is irradiated a diffraction pattern is produced (Figure 1.10).

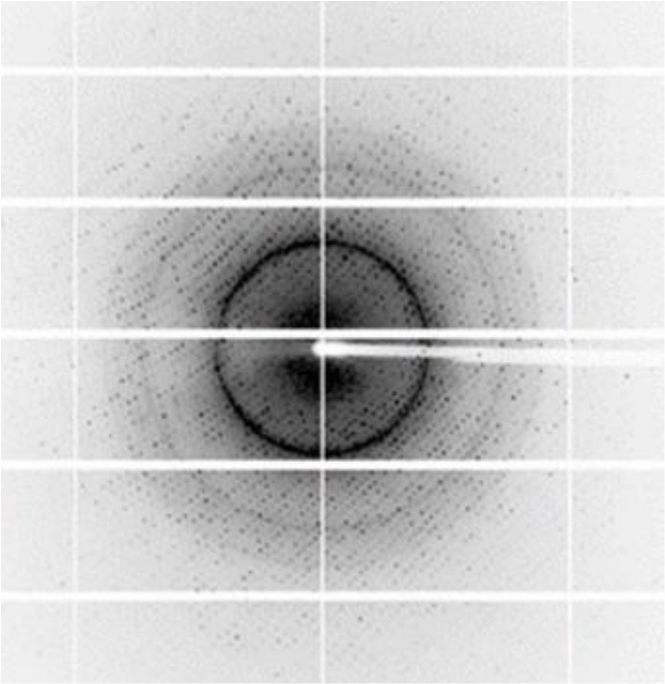


Figure 1.10. Diffraction pattern of a protein crystal (Hamill, 2020).

Each spot, called a reflection, corresponds to a particular set of planes of the crystal in reciprocal space. Planes spaced more closely together (lower *d*) capture finer details and diffract at wider angles, so their reflections appear further from the centre in the diffraction pattern. The position and intensity of the reflection contains information about the electron density of the plane, from which X-rays were diffracted to create it. The intensity is defined by the amplitude and phase of the diffracted X-rays. By rotating a crystal a full 3D map of electron density can be calculated. Since the measured intensity is a sum of all X-rays diffracted from a particular plane the phase information is lost. The phase information must be recreated for electron densities to be calculated directly. This can be done using molecular replacement, where phases are calculated using a homologous structure, or an Artificial Intelligence generated predicted structure, such as AlphaFold (Jumper *et al.*, 2021; Varadi *et al.*, 2022). Other methods such as isomorphous replacement and anomalous dispersion exploit properties of heavy atoms incorporated into the protein crystal to calculate the phases. Once the electron density map has been calculated a protein model can be built into the calculated density. The model is iteratively refined and validated with the aim to create an atomistic interpretation of the observed electron density.

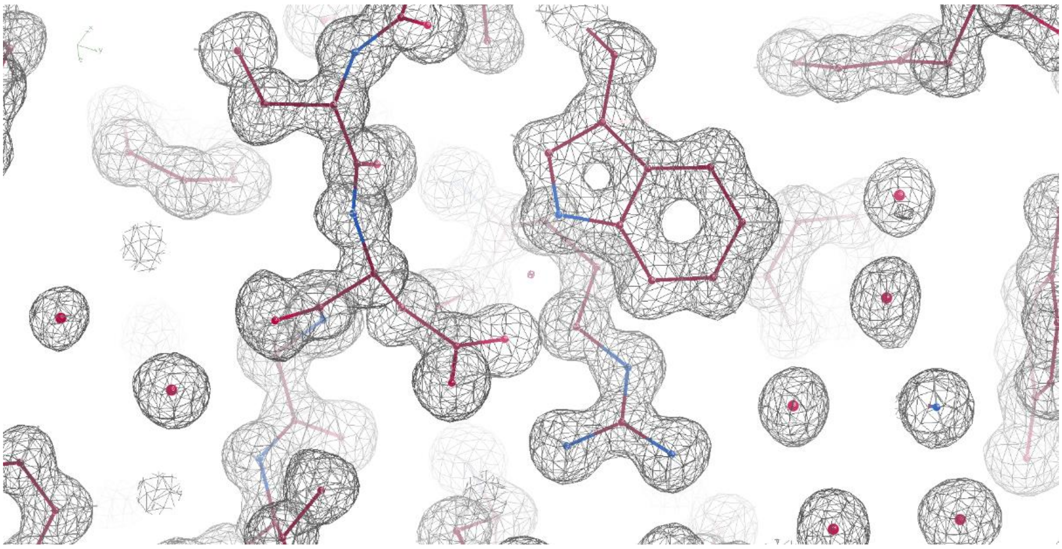


Figure 1.11. Model building. Atom level model is built to explain observed electron density. Data shown is the 2Fo-Fc map and model for MalL D492R (Chapter 5) at a contour level of 1.01 RMSD.

1.5.1 Protein crystallography

A well-formed crystal lattice is essential for X-ray diffraction. The observed intensities in the diffraction pattern are directly proportional to the number of points in the scattering lattice. High-resolution details captured by the closely spaced planes are more susceptible to crystal defects as minor changes in crystal order will cause the signal loss into the baseline. Thus a well-ordered, large crystal will produce a high-quality regular lattice that will diffract optimally, and a significant challenge in X-ray crystallography is the production of protein crystals of suitable quality.

Protein crystals form when highly pure protein becomes supersaturated, without significantly disturbing the natural state of the protein, to prevent the protein from precipitating out of solution (McPherson & Gavira, 2013). Protein crystals are grown in an aqueous solution that contains additives and salts to promote crystallisation. The specific conditions for crystal growth are highly specific to each individual protein, and can often require extensive screening and optimisation to identify conditions that promote crystal formation.

A common method to achieve supersaturation for protein crystallisation is the hanging-drop vapour diffusion method (Figure 1.12). In this method a small drop (1-4 µL) containing a mixture of protein and crystallisation solution is suspended above a reservoir of crystallisation solution and the system is sealed. As the protein drop contains a lower concentration of the crystallisation reagents, the drop undergoes a net loss of water to the reservoir as the system equilibrates, raising the saturation of the protein, and inducing crystallisation.

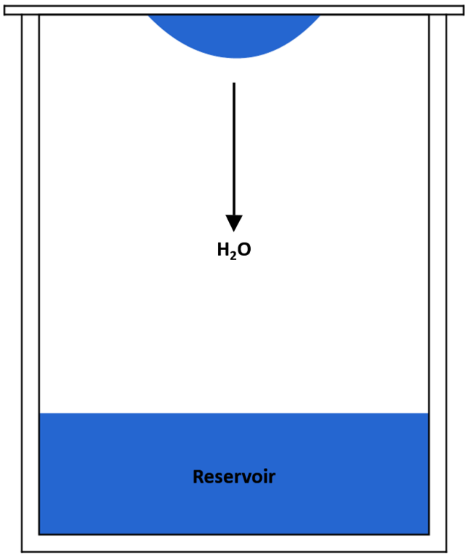


Figure 1.12. Hanging-drop vapour diffusion (Hamill, 2020)

There are two stages in the development of protein crystals, nucleation and growth. The nucleation phase remains poorly understood. The nucleation of crystals is a two-step process, whereby a condensed protein phase forms in solution, followed by nucleation of an ordered crystal from within this phase (Vekilov, 2016). The surface free energy of the phase boundary makes this process unfavourable, thus limiting the number of potential crystallisation sites to the few fluctuations that overcome the free energy barrier (Vekilov, 2016). Once this new phase reaches a certain critical size, crystal growth has a corresponding decrease in free energy and is spontaneous (Vekilov, 2016). Crystal growth from a nucleus is also a two-step process. Crystal growth occurs most readily at step-edges, whereby a crystal plane is extended (McPherson *et al.*, 2000). The formation of new step-edges is the rate-limiting step and requires a nucleation event, to form a new ordered arrangement on top of the crystal growth surface (McPherson *et al.*, 2000). Steps may also be formed by translocations or defects in crystal order (Drenth, 1999).

Crystal formation and growth is dependent on the supersaturation of the protein solution. The supersaturation can be controlled by altering experimental parameters (McPherson *et al.*, 2000). The precise experimental conditions that allow crystals to grow are difficult to predict and often require extensive screening and optimisation (McPherson & Gavira, 2013). After a potential condition is identified during screening, the condition is optimised to maximise crystal quality. Optimisation involves the systematic altering of experiment to conditions to identify changes that optimise crystal growth and quality. The three general factors for a crystallisation condition are the pH, ionic strength, and precipitant (McPherson & Gavira, 2013). The pH affects the ionisation state of surface amino-acid residues, altering the degree of surface interactions between adjacent protein units. Salts alter the ionic strength and compete with the protein for interactions with water, promoting crystallisation. Precipitants interact with the solvent and the protein to alter the saturation, thereby promoting crystallisation (McPherson & Gavira, 2013).

1.6 Small angle X-ray scattering

Small-Angle X-ray Scattering (SAXS) is a powerful method for collecting low resolution structural information about proteins in solution (Kikhney & Svergun, 2015). Similar to X-ray diffraction (Section 1.5) SAXS measures the intensity of scattered X-rays; however, at much smaller angles (Figure 1.13). In a SAXS experiment the sample is irradiated with a collimated (X-rays are parallel and diverge minimally with propagation), monochromatic X-ray source, and intensities of the scattered X-rays are detected (Kikhney & Svergun, 2015). A scattering profile of pure solvent is subtracted from the sample scattering profile leaving only the signal from the particle of interest.

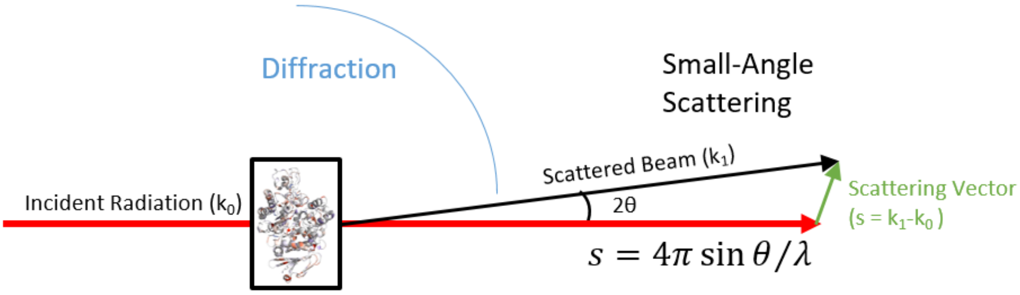


Figure 1.13. Small Angle Scattering. Adapted from Kikhney and Svergun (2015) and Pauw (2013).

Due to the random orientation of particles in solution, only radially averaged information can be obtained (Kikhney & Svergun, 2015). The scattering intensity (I) is generally represented as a function of momentum transfer (s), where λ is the wavelength of the incident radiation and 2θ is the scattering angle (Equation 1.28). The scattering amplitude can be calculated from the intensity (Equation 1.29) and is a Fourier transform of the excess electron density (Equation 1.30) (Kikhney & Svergun, 2015).

Equation 1.28. Momentum transfer

$$s = 4\pi \sin\theta / \lambda$$

Equation 1.29. Scattering intensity.

$$I(s) = \langle I(s) \rangle_{\Omega} = \langle A(s)A*(s) \rangle_{\Omega}$$

Equation 1.30. Scattering amplitude

$$A(s) = \int \Delta\rho(r) e^{isr} \, dr$$

Scattering intensity is proportional to the sample concentration; however, if the sample concentration is too high inter-particle distances become the same order of magnitude as intra-particle distances, and both then contribute to the scattering pattern. Inter-particle interactions can be seen in the scattering intensity at low-angles. A decrease in intensity indicates repulsive interactions, while an increase indicates attractive interactions, such as non-specific aggregation and sample denaturation (Kikhney & Svergun, 2015).

1.6.1 Structural information

The scattering profile of a particle in solution yields many useful structural parameters, such as molecular weight, particle volume, maximum dimension (D_{max}) and radius of gyration (R_g) (Kikhney & Svergun, 2015). For a monodisperse system (all particles are identical) the experimentally obtained parameters directly match the particle of interest. For a polydisperse system experimental parameters represent an average of the particles in the system (Kikhney & Svergun, 2015).

1.6.1.1 Radius of gyration

The radius of gyration is a measure of a particle's overall size. It is the average root-mean-square of distances to the centre of mass, weighted by the scattering length density (measure of scattering power, calculated from electron density) (Kikhney & Svergun, 2015). The radius of gyration is calculated using the Guinier approximation, where scattering at very-low angles can be treated as linear, and is therefore defined by only two parameters (Equation 1.31). Plotting ln(I) vs s^2 gives a linear line where the slope yields the R_g and the Intercept gives the forward scattering (I(0)), which is proportional to the molecular weight and sample concentration (Kikhney & Svergun, 2015).

Equation 1.31. Guinier approximation

$$I(s) = I(0)e^{\frac{s^2R_g^2}{-3}}$$

1.6.1.2 Distance distribution function

The distance distribution function (p(r)) is a histogram of all possible inter-atom distances within a particle. The intensity and probability density functions are Fourier transforms of each other related by Equation 1.32.

Equation 1.32. Probability density function

$$p(r) = \frac{r^2}{2\pi^2} \int_0^\infty \frac{s^2 I(s) \, sin(sr)}{sr} \, ds$$

For a globular compact protein the distance distribution function should be symmetric and bell shaped, with upper and lower limits p(0) and p(D_{max}) being equal to zero (Kikhney & Svergun, 2015).

1.6.1.3 Kratky plot

The Kratky plot (s^2I(s) vs s) visualises particular features of the scattering profile to emphasise those that relate to folding-state and flexibility (Figure 1.14). A folded protein gives a Kratky plot with a bell-shaped curve. An extended chain (such as a denatured protein) gives a Kratky plot with a plateau over a specific range of s (Kikhney & Svergun, 2015). A dimensionless Kratky plot removes information about the protein size, leaving only information about the shape. It is produced by normalising the data, such that I(0)=1 and dividing s by R_g (Kikhney & Svergun, 2015).

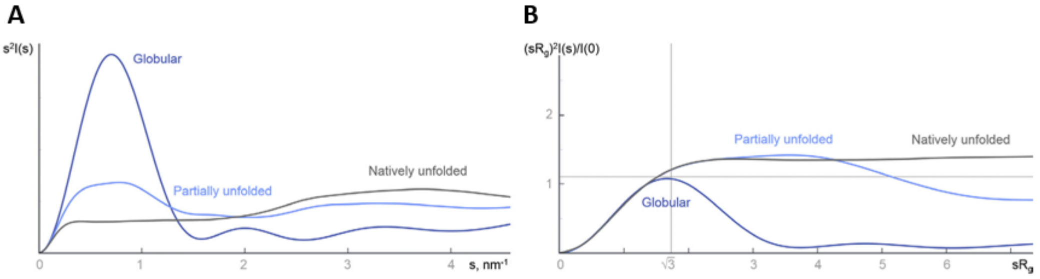


Figure 1.14. Kratky plots of proteins in various states of folding A) Kratky plot B) Dimensionless Kratky plot. Adapted from Kikhney and Svergun (2015). Reproduced with permission.

1.7 Red edge excitation shift spectroscopy

Red edge excitation shift (REES) spectroscopy is a useful and sensitive method for probing the dynamics and flexibility of proteins. The role of small perturbations to the dynamics, and small shifts in the conformational space sampled is increasingly being associated with the function and activity of the protein (Catici *et al.*, 2016). REES is a simple non-destructive and sensitive technique to identify small changes to protein dynamics. REES relies on the intrinsic fluorescence of the amino acid tryptophan which is the fluorophore in a REES experiment. REES occurs where low energy excitation of a fluorophore leads to a red shift in the maxima of the emission spectrum (Figure 1.15) (Catici *et al.*, 2016).

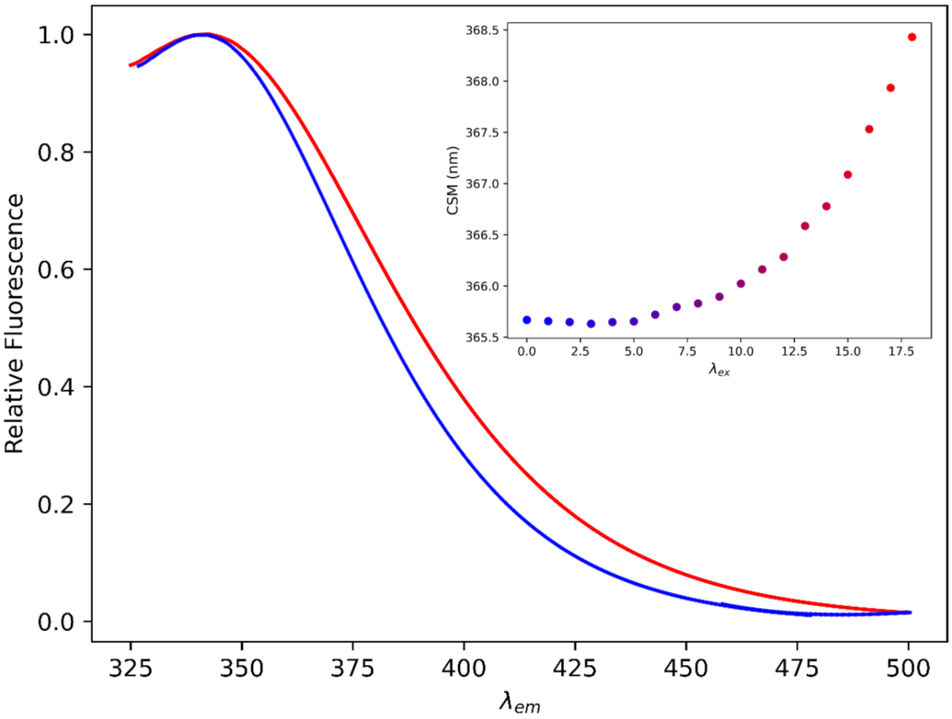


Figure 1.15. REES Effect. Emission maxima shifts to longer wavelengths with lower energy excitation wavelengths (Blue: High Energy, Red: Lower Energy). Inset shows the shift in Emission maxima (represented as CSM – the centre of spectral mass) to higher wavelengths as excitation wavelength is increased (energy decreases). Adapted from Kwok et al. (2021).

In general for most fluorophores emission occurs from the lowest vibrational level of the first electronic excited state (Haldar *et al.*, 2011). Fluorescent emission should therefore be independent of excitation wavelength. This is not true for a polar fluorophore in a motionally restricted environment, where the ability of the bulk solvent's ability to reorganise is also restricted.

Tryptophans are polar and exhibit a dipole moment that is changed upon excitation. This results in a reorganisation of the solvent around the tryptophan fluorophore. The viscosity of the surrounding environment affects the timescale of this reorganisation (τ_S). The REES effect originates from the ability of the solvent to reorganise versus the lifetime of fluorescence (τ_F) (Haldar *et al.*, 2011). In a flexible environment, such as directly in bulk solvent, the reorganisation time is less than the fluorescence lifetime, and the excited fluorophore is able to reach the relaxed, reorganised conformation, from which emission occurs at a longer wavelength (λ_R) (Haldar *et al.*, 2011). In a completely restricted environment the reorganisation time is greater than the fluorescence lifetime and solvent reorganisation does not occur and fluorescence occurs at shorter wavelengths (λ_{FC}) directly from the excited state [termed the Frank-Condon (FC) state], rather than the lower energy relaxed state (Catici *et al.*, 2016).

Using low energy excitation (at the red edge) it is possible to selectively excite tryptophan molecules in different solvation environments (Catici *et al.*, 2016). Under these conditions fluorophores with a solvent environment that minimise the difference between the ground state and excited state (closer to the relaxed state) are selectively excited (Figure 1.16) (Haldar *et al.*, 2011). Thus emission occurs at lower energies and higher wavelengths resulting in the emission maxima shift as excitation wavelength decreases (Figure 1.15) (Haldar *et al.*, 2011). The REES effect requires that the solvent reorganisation time be comparable or longer than the fluorescence lifetime, so fluorescence occurs from various partially relaxed states (Haldar *et al.*, 2011). As mobility increases the ability of the solvent to reorganise becomes less restricted and becomes shorter than the fluorescence lifetime, and thus emission is independent of excitation and the REES effect is lost (Haldar *et al.*, 2011).

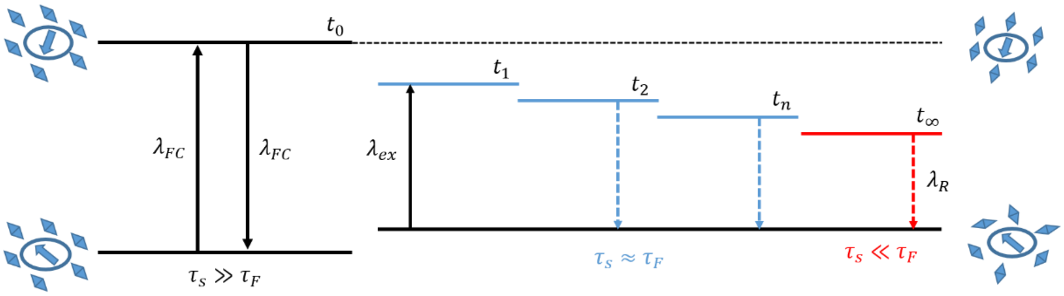


Figure 1.16. Jablonski diagram showing REES effect. Where the solvent reorganisation (*τ_S*) time is greater than the fluorescence lifetime (*τ_F*), both excitation and emission occur with the Frank-Condon (FC) state. Where the solvent reorganisation time is less than the fluorescence lifetime, solvent reorganisation is able to occur and emission occurs from the relaxed state (*λ_R*). Where the fluorescence lifetime and solvent reorganisation time are approximately equal, excitation and emission can occur with partially relaxed states. This occurs selectively with red-shifted (lower energy) excitation wavelengths. Adapted from Haldar* et al. *(2011) and Kwok* et al. *(2021).

Changes in the emission spectra is recorded by the centre of spectral mass (CSM), calculated by Equation 1.33, where f_i is the fluorescence intensity and λ_{Em} is the emission wavelength. For a protein exhibiting the REES effect CSM increases with increasing excitation wavelength. This curve can be fit with Equation 1.34 where CSM_0 is the CSM independent of the excitation wavelength (λ_{Ex}), and A and R give the curvature of the exponential.

Equation 1.33. REES shift: Centre of spectral mass

$$CSM = \frac{\sum(f_i \lambda_{Em})}{\sum f_i}$$

Equation 1.34. REES equation

$$CSM = CSM_0 + Ae^{R\Delta\lambda_{Ex}}$$

The change in fluorescence emission can be considered as a two-state system, whereby at low excitation (blue) wavelengths emission largely occurs from the normal excited state. At higher (red) wavelengths emission occurs from the relaxed state at higher wavelengths (Kwok *et al.*, 2021) (Figure 1.17 A). This can be modelled by Equation 1.35. $CSM(\lambda_{Ex}^{FC})$ is the CSM for normal Frank-Condon emission, and is responsive to changes in the solvent environment. Increasing solvent exposure (such as during unfolding) will cause a red shift (increase) and a decrease in solvent exposure will cause a blue shift (decrease) (Kwok *et al.*, 2021). $CSM(\lambda_{Ex}^{R})$ is the CSM for emission from a fully relaxed state and is fixed for a given system, being uniquely defined by the protein structure and physiochemical environment (Kwok *et al.*, 2021). ΔG_m describes the gradient of the slope describing the free energy change of the solvent reorganisation (Warrender *et al.*, 2023). An increase in ΔG_m represents a decrease in the flexibility of the protein and a sharper transition (Figure 1.16 B) (Kwok *et al.*, 2021).

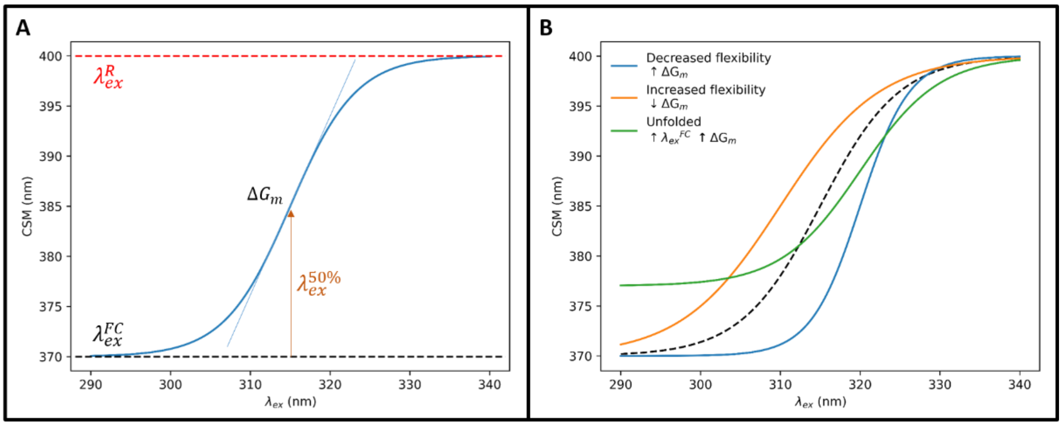


Figure 1.17. REES Two-State Model. A) Graphical representation of Equation 1.35. B) Graphical representation of the relationship between protein conformation and thermodynamic parameters. Adapted from Kwok* et al. *(2021).

Equation 1.35. REES two-state model

$$CSM(\lambda_{Ex}) = \frac{CSM(\lambda_{Ex}^{FC}) + CSM(\lambda_{Ex}^{R})e^{\Delta G_m(\lambda_{Ex}-\lambda_{Ex}^{50\%})/RT}}{1 + e^{\Delta G_m(\lambda_{Ex}-\lambda_{Ex}^{50\%})/RT}}$$

REES is a powerful tool for analysing changes to protein dynamics. It is sensitive and easily accessible. It has been particularly applied to measure antibody stability (Knight *et al.*, 2020; Warrender *et al.*, 2023). REES was used to analyse a variety of antibody allotypes, and was able to identify specific point mutations that confer additional stability (Warrender *et al.*, 2023).

1.8 Stopped-flow kinetics

Stopped-Flow spectrophotometry is a powerful technique for the analysis of enzyme kinetics. Many methods for kinetic analysis follow the change in concentration of a reagent or product as a function of time. This approach requires that the reaction of interest occurs on a timescale longer than the time needed for reagent mixing and instrument activation (Zheng *et al.*, 2015). This 'dead time' may be on the order of seconds for a traditional bench-top setup, however many biological systems may interact at timescales faster than this. Stopped-flow solutions reduce this dead time to only milliseconds (Zheng *et al.*, 2015).

Stopped-flow was developed as a modification of the continuous flow method as a way to conserve reagents (Bagshaw, 2013). During operation a small volume of each reagent is rapidly applied from the reservoir drive syringes, though a mixer and into an observation cell where the flow is stopped, and data acquisition begins (Figure 1.18). The dead time between sample mixing and data acquisition using stopped-flow may be as little as 1-2 ms (Zheng *et al.*, 2015). The reaction in the observation cell may be followed using a number of methods, for example absorbance or fluorescence. The small volumes involved allows for fine control of temperature and rapid temperature equilibrium.

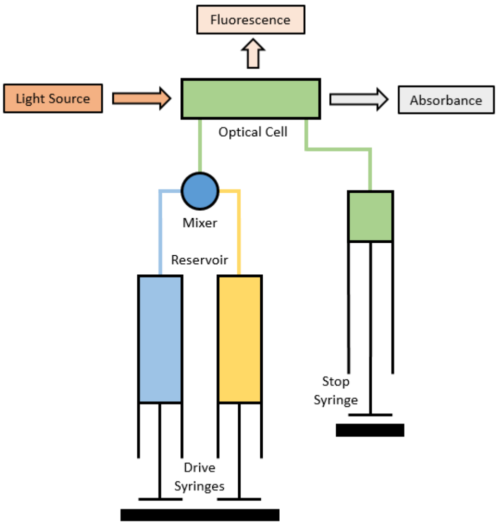


Figure 1.18. Schematic of Stopped-flow instrument. Adapted from Zheng **et al.** ***(2015).***

1.9 Glucosidase MalL from *Bacillus subtilis* as a model enzyme

MalL (E.C 3.2.1.10) is an isomaltase enzyme involved in maltose utilisation in *Bacillus subtilis*. MalL was first characterised by Schönert *et al.* (1998), and was identified as a specific Oligo-1,6-α-glucosidase (sucrase-isomaltase-maltase). The activity of MalL can be followed spectrophotometrically at 405 nm by the cleavage of *p*-nitrophenyl-α-D-glucopyranoside (PNG). MalL is a large and stable monomeric enzyme, which is well suited for studying dynamic effects. The relatively large size of MalL allows for the study of dynamic changes without confounding changes to regions near the active site. MalL has been extensively used as a model for MMRT with many mutants that have been characterised, both kinetically and structurally. Insights from the dynamic mechanisms of MalL have contributed to the development of MMRT.

The catalytic core of MalL is an 8-stranded β/α barrel first described in triosephosphate isomerase (TIM). The catalytic residues consist of an aspartic acid, glutamic acid, and an aspartic acid on the C-terminal end of β_4, β_5, and β_7 respectively. Several loop extensions between β/α pairs form the active site structure and pocket. MalL also has a C-terminal Greek-key β-sheet domain. MalL is typically divided into three domains, the catalytic barrel domain (A), the β_3-α_3 loop extension (B) and the C-terminal auxiliary domain (C). A fourth 'Lid' domain can also be identified as a β_8-α_8 extension. This domain forms part of the active site pocket. Based on the molecular dynamics (van der Kamp *et al.*, 2018; Walker *et al.*, 2023), MalL can be divided into four regions, residues 1-193, 194-321, 322-459, and 460-561 (Figure 1.19).

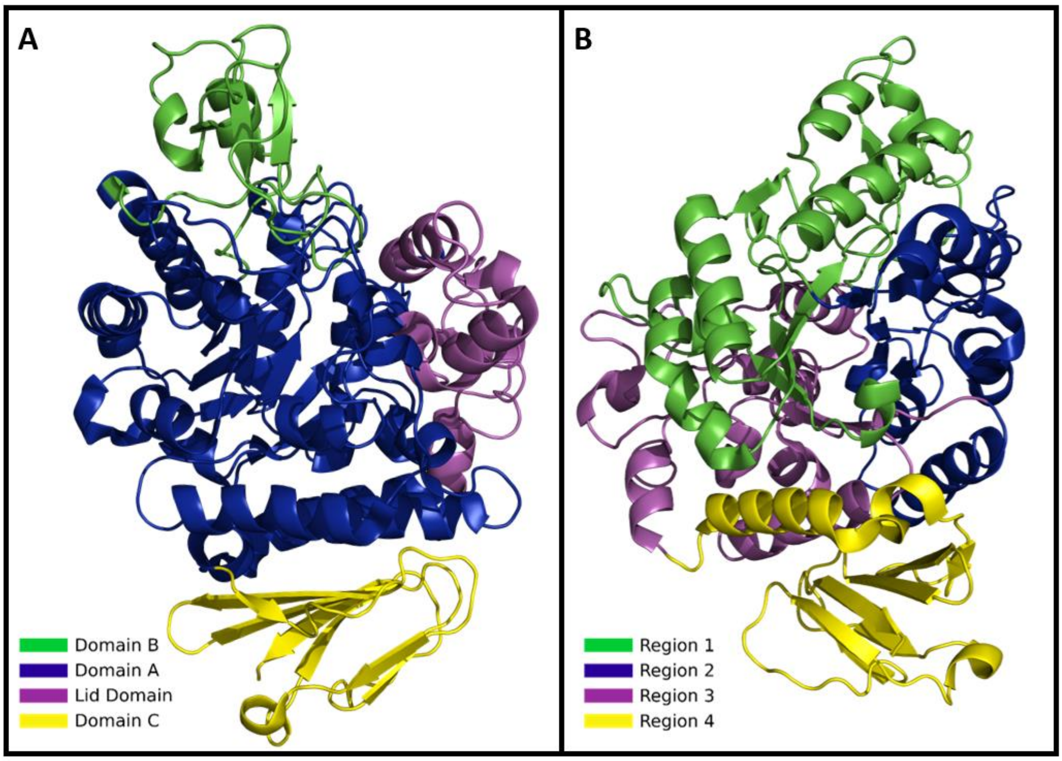


Figure 1.19. Structure of MalL. A) Domain architecture of MalL. Domain A (Blue) is the catalytic domain. Domain B (Green) is a β_3 to α_3 extension. Domain C is a C-terminal auxiliary domain. B) Dynamical regions of MalL identified by molecular dynamics.

The kinetics of MalL and various mutations thereof, have been extensively studied, and characterised using MMRT (Figure 1.20) (Hamill, 2020; Hobbs *et al.*, 2013; Walker *et al.*, 2023). The characterisation of these mutants has provided various insights into enzyme function and dynamics. For example the mutant enzyme MalL V200S has a large increase to both the T_{opt} and the activation heat capacity compared to MalL wildtype (Hobbs *et al.*, 2013), resulting from the stabilisation of the ES complex (Figure 1.21). MalL V200S has also been well characterised by REES spectroscopy (Section 1.7) (Jones *et al.*, 2017; Kwok *et al.*, 2021). Fitting of REES data for MalL V200S to Equation 1.35, gives a ΔG_m of 6.0 ± 0.2 mJ mol^{-1} nm^{-1} compared to 4.0 ± 0.2 mJ mol^{-1} nm^{-1} for MalL wildtype, consistent with a decrease in flexibility (Kwok *et al.*, 2021).

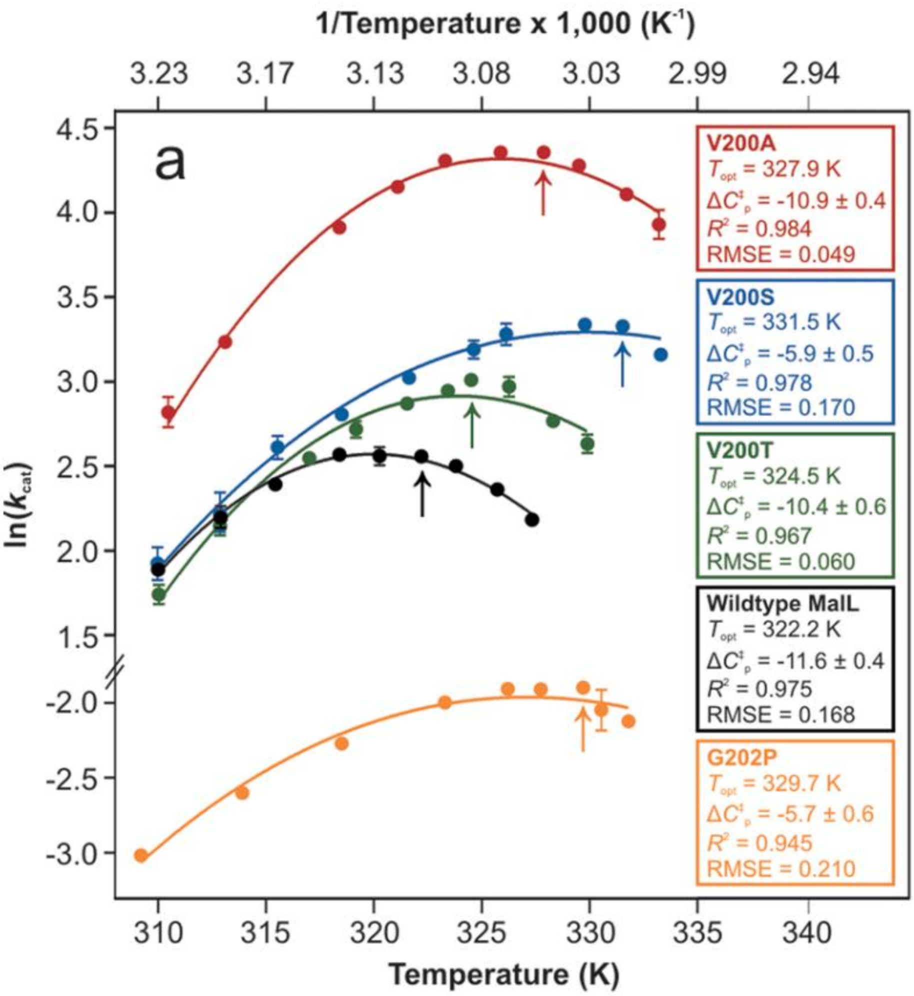


Figure 1.20. Characterisation of a range of individual MalL enzymes and mutations. Data are fit to MMRT 1.0 equation (constant activation heat capacity). Arrows indicate T_{opt} for each enzyme. Reprinted (adapted) with permission from Hobbs et al. ***(2013). Copyright 2013 American Chemical Society.***

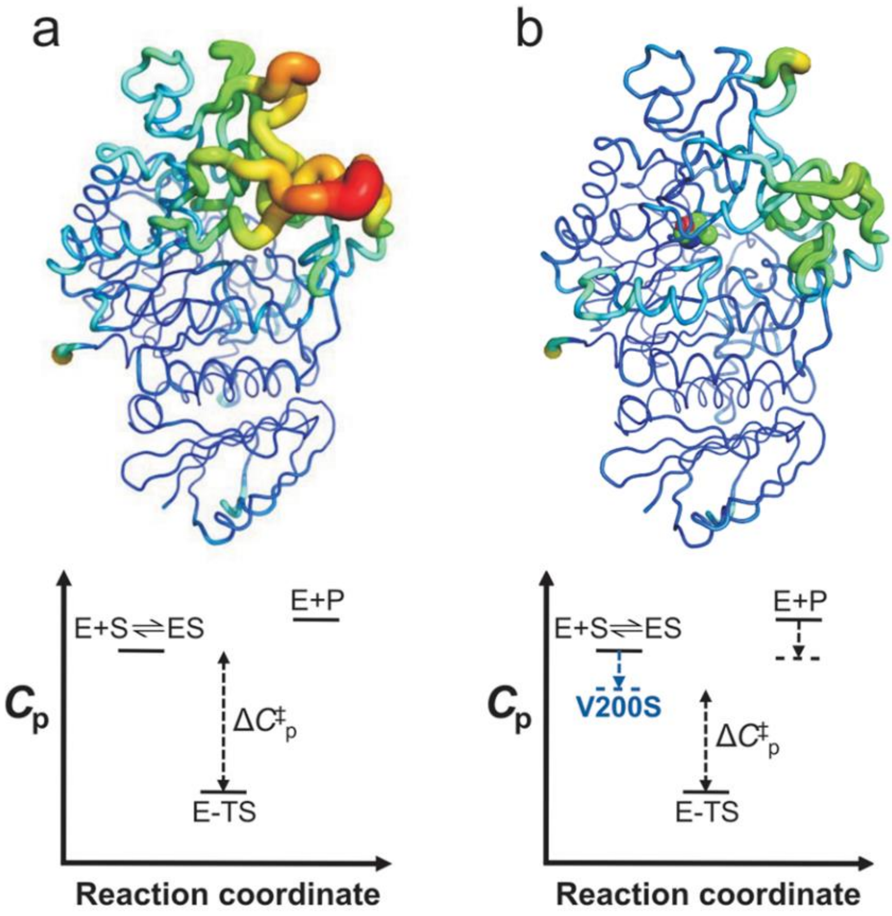


Figure 1.21. Molecular dynamics simulations of MalL wildtype and MalL V200S. Structures are coloured from blue to red, with increasing tube width, for increasing flexibility, calculated from the trajectories. The lower portion shows a reaction scheme of heat capacity for each enzyme. V200S mutation is indicated as spheres. A) MalL wildtype. B) MalL V200S has a reduced activation heat capacity by stabilisation of the ES state, leading to rate increases at high temperature. Reprinted (adapted) with permission from Hobbs* et al. *(2013). Copyright 2013 American Chemical Society.

The role of activation heat capacity and MMRT on enzyme catalysis has been explored using molecular dynamics simulations. This has allowed for the difference in heat capacity between the ES and TS to be calculated, with the calculated value closely matching the experimental value (van der Kamp *et al.*, 2018). This also allowed for the identification of specific residue regions that contribute differently to the change in heat capacity between the ES and TS states (van der Kamp *et al.*, 2018).

In addition to kinetic characterisation, 3D structures (obtained by protein crystallography) of several MalL variants have been obtained and characterised including the wildtype MalL, the V200S, V200A, G202P, and S536R mutants, as well as the wildtype MalL enzyme bound to the transition state analogue 1-deoxynojirimycin (PDB: 4M56, 4MAZ, 4M8U, 4MB1, 7LV6 and 5WCZ, respectively) (Hobbs *et al.*, 2013; van der Kamp *et al.*, 2018; Walker *et al.*, 2023).

1.10 Aims

The previously characterised MalL S536R mutant (Hamill, 2020; Walker *et al.*, 2023) crystallised with a significant resolution improvement (1.1 Å, compared to 2.3 Å for MalL wildtype) without any impact on the overall kinetics of the enzyme. This S536R mutation was designed to place an arginine guanidine group in an identified urea binding site. We therefore, aimed to further explore the utility of designed arginine mutations in light of the improvement in activity afforded by low concentrations of urea. To achieve this several arginine mutants of the model enzyme MalL have been designed and characterised, kinetically, structurally, and by biophysical methods. Macromolecular rate theory (MMRT) was used as a model to gain insight into the kinetics and dynamics of each mutant enzyme.

2 Materials and Methods

2.1 Protein production and purification

2.1.1 Buffers and media

Composition of general buffers and media is given in Table 2.1. Buffer composition were as previously described for MalL (Hobbs et al., 2013; Schönert et al., 1998).

Table 2.1. Buffers and media for MalL proteins

Solution	Components	pH
LB Media	10 g L^{-1} NaCl, 10 g L^{-1} tryptone, 5 g L^{-1} yeast extract	NA
LB Agar	10 g L^{-1} NaCl, 10 g L^{-1} tryptone, 5 g L^{-1} yeast extract, 15 g L^{-1} agar	7.0
Lysis Buffer	10 mM KCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 25 mM imidazole	7.0
Elution Buffer	10 mM KCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 0.5 mM imidazole	7.0
Size Exclusion Buffer	20 mM HEPES	7.0
Assay Buffer	40 mM Na_3PO_4, 150 mM NaCl	7.0

Table 2.2. Antibiotics

Antibiotic	Stock Concentration	Working Concentration
Ampicillin	100 mg mL^{-1}	0.1 mg mL^{-1}
Kanamycin	50 mg mL^{-1}	0.05 mg mL^{-1}

2.1.2 Protein expression

2.1.2.1 Storage of expression strain

Glycerol stocks were made with equal concentrations of 50% v/v glycerol and bacterial culture, and were stored at -80°C.

2.1.2.2 Expression protocol

Glycerol stocks of the relevant strain were used to inoculate 20 mL starter cultures in LB media with the appropriate antibiotic (0.1 mg mL^{-1} ampicillin or 0.05 mg mL^{-1} kanamycin). Starter cultures were grown overnight at 37°C with shaking at 180 rpm. Starter cultures were used to seed 1 L LB media expression cultures, in a 2 L baffled flask with appropriate antibiotic. Expression cultures were grown at 37°C with shaking at 180 rpm until OD_{600} was approximately 0.5-0.9. Protein expression was induced with the addition of Isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.75 mM Expression culture was then moved to 18°C with shaking at 180 rpm overnight. Cells were harvested by centrifugation at 4500 g for 20 min at 4°C. Cells were stored at -20°C until needed.

2.1.2.3 Small scale expression trials

Glycerol stocks of the relevant strain were used to inoculate 10 mL starter cultures with the appropriate antibiotic. Cultures were incubated overnight at 37°C with shaking at 180 rpm. Starter culture (100 µL) was used to seed 20 mL of LB media expression cultures with the appropriate antibiotic. Expression cultures were grown at 37°C with shaking at 180 rpm until OD_{600} was approximately 0.5-0.7. Expression cultures were reduced to 10 mL and induced with IPTG to a final concentration of 0.75 mM. Expression culture was then moved to 18°C with shaking at 180 rpm overnight. Cells were harvested by centrifugation at 4500 g for 20 min at 4°C. Cells were stored at -20°C until needed.

Cells were defrosted on ice and resuspended in 1 mL Lysis buffer (Section 2.1). Cells were lysed on ice or in a cold block by sonication. Cells were sonicated on a QSonica 700 sonicator, for a total of 120 seconds at amplitude 5, in 2 second pulses with 2 seconds of wait time between each pulse. Soluble protein was obtained in the supernatant by centrifugation at 16200 g. The pellet was resuspended in 1 mL lysis buffer. Representative samples of the supernatant and resuspended pellet were taken. Nickel Sepharose resin (100 µL, Cytiva, USA) was equilibrated with Lysis buffer (Section 2.1) and mixed with the supernatant for five minutes. The Nickel resin was separated by centrifugation at minimum speed for 30 seconds and the supernatant discarded. A representative sample of the discarded supernatant was taken. The nickel resin was washed twice with Lysis buffer and separated by centrifugation at minimum speed for 30 seconds. Representative samples were mixed 1:3 with Lysis buffer. The samples and the Nickel resin were analysed by polyacrylamide gel electrophoresis (Section 2.1.4).

2.1.3 Protein purification

All MalL variants used the same protocol for purification. Cells were defrosted on ice and resuspended in approximately 20-30 mL Lysis buffer by vortexing. Cells were lysed on ice or in a cold block by sonication. Cells were sonicated on a QSonica 700 sonicator, for a total of 90 seconds at amplitude 12, in 1 second pulses with 1 second of wait time between each pulse. Soluble protein was obtained in the supernatant by centrifugation at 14000 g for 30 min at 4°C. The supernatant was filtered in series through 1.2, 0.45 and 0.2 µm Minsart syringe filter (Sartorius, Germany) prior to Immobilised Metal Affinity Chromatography.

2.1.3.1 Immobilised metal affinity chromatography (IMAC)

Protein was purified using IMAC based on the affinity of the N-terminal Hexa-His tag for Nickel. Prior to use the HisTrap FF (5mL, Cytiva, US) column was stripped and recharged by flushing the column in series with two column volumes of EDTA (100 mM, pH 8.0), two column volumes of MQ H_2O, one column volume of 100 mM $NiCl_2$, and two column volumes of MQ H_2O. The column was equilibrated with two column volumes of Lysis buffer prior to loading the filtered supernatant.

Protein elution was performed on a Äkta Purifier FPLC. Non-target protein was washed from the column with 4% Elution buffer and 96% Lysis buffer at 1 mL min^{-1}, until absorbance at 280 nm had plateaued. The protein was eluted using an imidazole gradient between 25 mM and 0.5 M (0% Elution buffer / 100% Lysis buffer - 100% Elution buffer / 0% Lysis buffer) over 50 mL at 1 mL min^{-1}. Protein elution was followed by UV trace at 280 nm and the main peak collected for subsequent purification.

2.1.3.2 Size exclusion chromatography (SEC)

SEC was performed on an Äkta Purifier FPLC using a HiLoad 16/60 Superdex 200 pg preparative column. Column was equilibrated with size exclusion buffer (Section 2.1) prior to use. IMAC purified protein was concentrated (Section 2.1.6) to approximately 5 mL and was filtered through a 0.2 µm syringe filter (Pall Corporation, US) and loaded onto the column via a 5 mL injection loop. Elution was performed using size exclusion buffer at 0.5 mL min^{-1} and followed by UV trace at 280 nm. Protein peak was collected for subsequent analysis.

2.1.4 Polyacrylamide gel electrophoresis (PAGE)

Denatured PAGE was performed using Native gels with SDS running buffer. Native gels were made five at a time with a 12% acrylamide resolving layer and a 5% acrylamide stacking layer, according to the scheme in Table 2.3. Gels were run with 1 X Tris Glycine-SDS running buffer. Samples were prepared with 1:3 of 4 X SDS loading dye and denatured at 95°C for 5 minutes. Completed gels were stained with Fairbanks A stain (Wong *et al.*, 2000). Background staining was removed with Fairbanks D stain. Buffer compositions are given in Table 2.4. Gels were imaged using the iBright FL1000 imaging system (Invitrogen, US) and processed using Corel AfterShot 3 (Alludo, Canada).

Table 2.3. PAGE gel components

Component	Resolving Gel (mL)	Stacking Gel (mL)
MQ H_2O	10.35	8.625
30% w/v Acrylamide	12	2.125
Resolving Buffer	7.5	0
Stacking Buffer	0	1.6
10% w/v APS	0.15	0.063
TEMED	0.015	0.0063

Table 2.4. PAGE buffer compositions

Solution	Composition
4 X SDS loading dye	250 mM Tris (pH 6.8), 20% glycerol, 4% SDS, 10% mercaptoethanol, 0.025% w/v bromophenol blue
TG-SDS Running Buffer	25 mM Tris, 250 mM glycine, 0.1% SDS
Resolving Buffer	1.5 M Tris, pH 8.0
Stacking Buffer	1.0 M Tris, pH 6.8
Fairbanks A Stain	0.05% Coomassie blue (R 250), 25% isopropanol, 10% acetic acid
Fairbanks D Stain	10% acetic acid

2.1.5 Measuring protein concentration

Protein concentration was measured by UV at 280 nm using either the Nanodrop 2000 (Thermo Fischer Scientific, USA) or DS-11 FX (DeNovix, USA) spectrophotometer. Protein concentration was corrected with the extinction coefficient as calculated by ProtParam based on the protein sequence (Gasteiger *et al.*, 2005).

2.1.6 Protein concentration

Protein was concentrated to the appropriate volume or concentration using a 10 kDa molecular weight cut-off Amicon Ultra (Merck Millipore, USA) spin concentrator, by centrifugation at 2400 g.

2.2 Kinetic assay procedure

Enzymatic activity of MalL can be followed by the cleavage of *p*-nitrophenyl-α-D-glucopyranoside (PNG) (Sigma Life Sciences; N1377-5G) into glucose and *p*-nitrophenol (Schönert *et al.*, 1998). The production of *p*-nitrophenol can be measured spectrophotometrically at 405 nm. Protein assays were carried out on the stopped-flow spectrophotometer (TgK Scientific, UK). Protein was diluted to 0.026 mg mL^{-1} in Assay buffer (Section 2.1) and mixed 1:1 in the spectrophotometer with substrate [PNG in Assay buffer (Section 2.1)] at the appropriate concentration for a final protein concentration of 0.013 mg mL^{-1}. Temperature control was achieved using a circulating water bath attached to the observation cell on the stopped-flow spectrophotometer. Accurate temperature measurements were taken at the start time of each assay at the reaction cell. Reaction was repeated in quadruplicate with five non-recorded dummy shots (0.2 sec) prior to each measurement. The first three replicates were collected for 45 seconds in 0.088 second intervals, and the final replicate was collected for 20 seconds in 0.039 second intervals. The longer initial two reactions are required to maintain temperature at the reaction cell through multiple shots, and only the initial reaction is taken for rate calculations.

2.2.1 Michaelis-Menten assay

Michaelis-Menten assays were performed as described above, at 25°C on the stopped-flow spectrophotometer, with PNG concentration varied between 0.05 mM and 5 mM.

2.2.2 Temperature optimum assay

Temperature optimum assay was performed as described above with PNG concentration at approximately 10 X K_M (1 mM PNG for most variants). Temperature was varied between 5/10°C to 45/55°C.

2.2.3 Data analysis

Data analysis and fitting was performed using Python (Rossum & Drake, 2009) or GraphPad Prism 9.1.12. Reaction rates were measured by linear regression of the linear portion of the first 10 seconds of rate data. Data were fit to have a maximal r-squared value with at least 5 seconds of rate data (Section 8.4) (Code Snippet 2-1).

Code Snippet 2-1. Linear regression of rate data

```
This program uses linear regression to find the slope of the linear portion of rate data

current_x_value = 10
target_R_squared_value = 0.99
current_r_squared_value = 0

while current_r_squared_value < target_R_squared_value:
  x = time array [0-current_x_value]
  y = absorbance value [for each x]

  linear_regression = linear regression of x vs y
  new_r_squared = r squared score of linear_regression

  if new_r_squared >= target_R_squared_value:
    accept fit

  current_x_value = current_x_value - 0.1

  if current_x_value < 5:
    target_R_squared_value = target_R_squared_value - 0.1
```

2.2.3.1 Fitting of MMRT equations

All MMRT equations were fit using the software as described in Section 6 using a reference temperature (T_0) of 278.15 K.

2.3 Mutagenesis and cloning

MalL wildtype and S536R were obtained as described in Hamill (2020). All other variants were generated as described below.

2.3.1 Generation of mutants

All variants were obtained by gene synthesis (Twist Bioscience, USA) in the pET-28b(+) expression vector. Variants were transformed into *E. coli* DH5α (Section 2.3.2). These stocks were used to generate new plasmid. The plasmid was extracted (Section 2.3.4) and used for transformation into *E. coli* BL21 (DE3).

2.3.2 Transformation of *E. coli* by heat shock

Plasmids were transformed into chemically competent *E. coli* DH5α or BL21 by heat shock. Chemically competent cells were defrosted on ice. Plasmid (100-1000 ng) was added to chemically competent cells (100 µL) and incubated on ice for 30 minutes. Cells were heat shocked at 42°C for 45 seconds then incubated on ice for 2 minutes. One millilitre of sterile SOC media [20 g L^{-1} tryptone, 5 g L^{-1} yeast extract, 585 mg L^{-1} NaCl, 186 mg L^{-1} KCl, 2.46 g L^{-1} $MgSO_4 \cdot 7H_2O$, 2.03 g L^{-1} $MgCl_2 \cdot 6H_2O$, 20 mM glucose (sterile filtered)] was added to the cells. The cells were then incubated at 37°C for 30 minutes. Aliquots of transformed cells were plated on LB Agar plates with appropriate antibiotic (Section 2.1). LB Agar plates were grown overnight at 37°C. Colonies were screened for successful transformation by colony PCR (Section 2.3.3).

2.3.3 Colony PCR

Transformation colonies were screened for presence of the gene insert by PCR flanking the gene insert position using T7 primers (Table 2.5). Single colonies were picked from a plate using a sterile pipette tip, streaked onto an LB Agar plate with appropriate antibiotic for preservation, and the cell remnants were resuspended in the PCR reaction mix for use as template DNA. PCR reactions used HOT FIREPol (SolisBioDyne, Estonia). The composition of the PCR reaction mix is given in Table 2.6 and PCR cycling conditions are given in Table 2.7. Denaturation, Annealing and Extension steps were repeated 29 times. PCR results were assessed by agarose gel electrophoresis (Section 2.3.5).

Table 2.5. PCR primers

Primer	Sequence
T7 Forward	5’-TAATACGACTCACTATAGGG-3’
T7 Reverse	5’-GCTAGTTATTGCTCAGCGG-3’

Table 2.6. PCR reaction composition

Component	Volume	Final Concentration
5 X HOT FIREPol Master Mix	4 µL	1 X
T7 Forward (10 pmol μL^{-1})	0.5 µL	0.25 µM
T7 Reverse (10 pmol μL^{-1})	0.5 µL	0.25 µM
MQ H_2O	15 µL	-

Table 2.7. PCR Cycling parameters for colony PCR

Step	Temperature	Step Length
Pre-Denaturation	95	15 minutes
Denaturation	95	15 seconds
Annealing	60	45 seconds
Extension	72	2 minutes
Post-Extension	72	10 minutes

The LB Agar plates were grown overnight at 37°C. Colonies with a confirmed insert were used to inoculate 10 mL of LB media with appropriate antibiotic. Culture was grown overnight at 37°C with shaking at 180 rpm and used to create storage glycerol stocks (Section 2.1.2.1).

2.3.4 Plasmid purification

E. coli containing the relevant plasmid was grown overnight in 5 mL LB media with the appropriate antibiotic. Plasmid DNA was isolated using the QIAprep Spin Miniprep kit (QIAGEN, Germany) according to the manufacturer's instructions. Plasmid DNA was eluted into 15 µL of elution buffer and stored at -20°C.

2.3.5 Agarose gel electrophoresis

Agarose gels were made as required to between 0.8-2.0% w/v agarose in TAE buffer (40 mM Tris-Acetate, 1 mM EDTA). Thiazole orange (10,000 X, 13 mg mL^{-1} in DMSO) was added to a final concentration of 1 X, to allow for DNA visualisation. Samples were loaded with 5 X DNA loading dye [25% w/v glycerol, 0.2% w/v bromophenol blue]. Gels were imaged using the iBright FL1000 imaging system (Invitrogen, US) and processed using Corel AfterShot 3 (Alludo, Canada).

2.4 Crystallisation

Protein was expressed as described in Section 2.1.2 and concentrated to 4-10 mg mL^{-1}.

2.4.1 Condition screening

Potential crystallisation conditions were identified using sitting-drop vapour diffusion. Pre-prepared conditions (Crsytal-HT-HR2-130, Index-HR2-144, PEGRx-HT-HR2-130, SaltRx-HR2-136) from Hampton Research were used for screening. Aliquots of 100 µL of each condition was pipetted into 96-well Intelli-Plate 96-2 Low Profile crystallisation plates. A Mosquito crystallisation robot (TTP LabTech Ltd, UK) was used to dispense the 200 nL sitting drops of 1:1 protein and crystallisation solution. Each tray was sealed with Crystal Clear Sealing Film (Hampton Research, USA) and stored at 18°C. Trays were checked at regular intervals to monitor crystal development and identify potential conditions.

2.4.2 Condition fine screening

Conditions identified during screening were further optimised by hanging-drop vapour diffusion. Conditions were systematically varied to identify conditions that produced optimal crystal growth. Altered conditions include protein concentration, ionic strength, precipitant concentration, size of the hanging drop, and the ratio of protein to crystallisation solution in the drop. Drop sizes of 1-4 µL were suspended on a siliconized glass coverslip (HR3-515, Hampton Research, USA) above 500 µL of crystallisation solution on a 24 well VDX plate (Hampton Research, USA). Plates were sealed with Glisseal N grease (Borer chemie, Switzerland) and left at 18°C. Trays were checked at regular intervals to monitor crystal development.

2.4.3 Final crystallisation conditions

Crystallisation conditions for each protein is given in Table 2.8.

Table 2.8. Final crystallisation conditions

Protein	Conditions
MalL T150R	0.1 M Tris pH 8.5, and 22% w/v PEG 3350
MalL D492R	0.09 M Tris pH 8.5, 22% w/v PEG 3350, 5%v/v glycerol

2.4.4 Data collection

2.4.4.1 Crystal preparation

Crystals were removed from crystallisation drops with a cryo-loop (Hampton Research, USA) and briefly soaked in cryoprotectant solution before being frozen in liquid nitrogen for data collection. Composition of cryoprotectant solutions for each protein is given in Table 2.9.

Table 2.9. Cryoprotectant solutions

Protein	Cryoprotectant
MalL T150R	0.105 M Tris pH 8.5, 23.5% w/v PEG 3350, 20% v/v glycerol
MalL D492R	0.085 M Tris pH 8.5, 23.5% PEG 3350, 20% v/v glycerol

2.4.4.2 Data collection

X-ray data were collected at the Australian Synchrotron on the MX2 macromolecular crystallography beamline (McPhillips *et al.*, 2002), using an EIGER X 16M pixel detector (Dectris Ltd, Switzerland).

2.4.5 Data processing

2.4.5.1 Indexing and integration

Data were indexed, integrated and scaled in XDS (Kabsch, 2010) using the automated data processing pipeline at the Australian Synchrotron. Unit cell and space group parameters were determined using Pointless (Evans, 2011).

2.4.5.2 Scaling

Data were scaled and merged in AIMLESS (Evans & Murshudov, 2013) within the CCP4 software suite (Winn *et al.*, 2011).

2.4.5.3 Matthews coefficient

The number of molecules in the asymmetric unit was calculated using the Matthews coefficient (Kantardjieff & Rupp, 2003) in the CCP4 software suite (Winn *et al.*, 2011).

2.4.5.4 Molecular replacement

Molecular replacement was performed in PHASER (McCoy *et al.*, 2007) within the PHENIX software suite (Adams *et al.*, 2010), using the previously solved wildtype MalL structure (PDB: 4M56) as a search model (Hobbs *et al.*, 2013).

2.4.6 Model refinement

The model was iteratively refined using rounds of manual refinement in COOT (Emsley *et al.*, 2010) with the $2F_O$-F_C electron density map contoured to 1σ, followed by automated refinement in phenix.refine (Afonine *et al.*, 2012) within the PHENIX software suite (Adams *et al.*, 2010). Further refinement and validation was performed using the PDB-REDO web server (Joosten *et al.*, 2014).

2.4.7 Structure analysis and visualisation

Temperature B-Factors were calculated using Baverage (Dodson, 1991). RMSD values and structural deviations were calculated using proSMART (Nicholls *et al.*, 2014). Structures were visualised using PyMOL version 2.3.3 (Schrödinger, 2000). Hydrogen bond analysis was performed using the FindHBond tool in UCSF Chimera version 1.15 (Pettersen *et al.*, 2004). Unit cell and asymmetric unit volumes were calculated using the MATTPROB web server (Kantardjieff & Rupp, 2003; Matthews, 1968; Weichenberger & Rupp, 2014). Ramachandran analysis was performed using PROCHECK in the CCP4 software suite (Laskowski *et al.*, 1993).

2.4.7.1 Crystal contacts and unit cell analysis

Crystal contacts were identified using UCSF Chimera 1.15 (Pettersen *et al.*, 2004). Adjacent monomers were identified using the Crystal Contacts tool with a contact distance set to 10 Å. Hydrogen bonds were identified using the FindHbond tool, set to find inter-model contacts only. Contents of the unit cell were generated using the PyMOL script supercell (Thomas Holder).

2.5 Red edge excitation shift (REES) spectroscopy

REES was performed on a Hitachi F-7000 fluorescence spectrophotometer using a 3D scan with excitation wavelengths between 290-340 nm, emission wavelengths between 325-500 nm. Protein was at 0.1 mg mL^{-1} and the temperature was maintained at 25°C or 37°C by a circulating water bath. Protein was equilibrated at the experimental temperature for five minutes prior to measurement. The centre of spectral mass was calculated from the measured data using Equation 1.33. The CSM data are then fit with Equation 1.35.

2.6 Small angle X-ray scattering (SAXS)

SEC-SAXS scattering data were collected at the Australian Synchrotron on the SAXS beamline (Ryan *et al.*, 2018), using an EIGER 2M pixel detector (Dectris Ltd, Switzerland). Protein (50 µL) was introduced to the SEC column (3 mL) at 6 mg mL^{-1} in size exclusion buffer (Section 2.1) with 0.1% w/v sodium azide to prevent radiation damage. Primary data reduction was performed in ScatterBrain 2.82 (Stephen Mudie, Australian Synchrotron). Data analysis was performed in the ATSAS software package (Manalastas-Cantos *et al.*, 2021). Buffer subtraction was performed in CHROMIXS (Panjkovich & Svergun, 2018). Data analysis was performed in PRIMUS (Konarev *et al.*, 2003).

2.7 Melting Temperature

Relative melting temperatures were determined using a SYPRO orange thermal shift assay (Huynh & Partch, 2015), with a protein concentration of 0.09 mg mL^{-1} and 2.5 X SYPRO. Melts were performed on a Corbett Research Rotor-Gene 6000. Melt was performed between 25 and 99°C in 0.2°C steps. SYPRO fluorescence was measured with an excitation wavelength of 470 nm and an emission wavelength of 555 nm. Data were fitted with a Boltzmann sigmoid and T_m was determined as the midpoint of the transition.

3 Urea binding to guide design of mutations that influence enzyme dynamics and catalysis

3.1 Introduction

3.1.1 The role of enzyme dynamics in protein activity and crystallisation

Enzyme dynamics refers to internal motions that occur within a protein structure. These movements can be on the order of femtoseconds to seconds (Agarwal *et al.*, 2020). These dynamic motions can be critical for enzyme catalysis (Agarwal *et al.*, 2020). Here the importance of enzyme dynamics in determining enzymatic activity is investigated as well as effects on protein crystallography. This is achieved through the study of the dynamic effects of urea binding, followed by the creation of a point mutations to emulate urea binding. Finally, we report the kinetic and structural characterisation of these mutant enzymes.

3.1.1.1 Influence of enzyme dynamics in protein crystallography

With advances in data processing and refinement, crystallisation has become the rate limiting step in X-ray protein structure determination. As such many protein engineering methods have been developed to improve both the chances of a protein crystallising and the quality of generated crystals.

Historically, one of the first methods for improving crystallisation success was to work with a homologous enzyme more amenable to crystallisation, especially thermophilic homologues. For example, one of the first structures of the 30S ribosome was derived from the bacterium *Thermus thermophilus* (Wimberly *et al.*, 2000). Alternately cofactors, ligands and inhibitors can be included in the crystallisation conditions. Many proteins become more ordered and are stabilised through the binding of ligands and cofactors, and are thus more likely to crystallise and improve the resolution of diffracting crystals (Derewenda, 2010). In a similar manner antibody fragments have been used to promote crystallisation (Davis *et al.*, 1990). Another approach to improve crystallisation success is to remove or separate disordered regions of the protein of interest and focus on individual stable domains. This can involve the removal of N- and C-terminal flexible regions, domain linkers and shortening of flexible surface loop regions (Dale *et al.*, 2003). This is especially useful to improve the resolution of poorly diffracting crystals, where low resolution structures can identify these features.

Many efforts to date have focused on the removal of specific structural features impeding crystallisation in the protein of interest. This approach involves a detailed prior knowledge of the protein in order to develop a unique system-specific strategy. Little progress has been made in the development of general methods to rationally improve the crystallisation of a wide variety of proteins. The main aims of protein engineering for crystallisation are to create proteins that are more stable and less prone to aggregation, are more ordered and rigid or have more favourable crystallisation interactions between adjacent monomers in the crystal matrix.

Residues involved in crystal contacts tend to have lower temperature factors than other residues (Eyal *et al.*, 2005). This is indicative of the decrease in conformational freedom of these residues during the formation of the crystalline state. The properties of the amino acids at these crystal interfaces may interfere with the crystallisation process (Cooper *et al.*, 2007). It has been suggested that surface residues with high conformational entropy such as lysine, glutamine and glutamate increase the entropy barrier required to form crystal contacts and thus restrict the ability of proteins to crystallise (Cooper *et al.*, 2007). It was thus proposed that crystallisation success could be improved by mutating surface patches with high entropy residues to lower entropy residues, such as alanine, which have a higher propensity to form crystal contacts (Longenecker *et al.*, 2001). This process has been termed surface entropy reduction.

3.1.1.2 MMRT

Macromolecular rate theory (MMRT) describes the temperature dependence of enzyme rates, incorporating an additional parameter, the activation heat capacity ($\Delta C_P^\ddagger$) (Section 1.3). An enzyme's heat capacity is mainly determined by the number and energy of the vibrational modes of the enzyme. The range of conformational states experienced by the enzyme is reflected in the vibrational modes of the enzyme, and thus in turn may relate to the enzyme rate through changes in $\Delta C_P^{\ddagger}$ (Jones *et al.*, 2018). The different forms of MMRT vary in the complexity of the temperature dependence of the activation heat capacity, either constant, linear, or sigmoidal. MMRT 2.0 postulates the existence of an equilibrium between the enzyme-substrate (ES) complex and a transition-state-like conformation (TLC) (Section 1.3.2).

3.2 Use of urea to modulate protein dynamics

Urea at sub-denaturing concentrations has been used in crystallography to study the mechanisms of protein folding and urea denaturation. Urea has been observed as a general modulator of conformational freedom, especially for surface residues. It has been noted that at low concentrations, denaturants such as urea and guanidinium hydrochloride, exhibit stabilising effects on protein conformations. This has been posited to be due to cross-linking hydrogen bonding interactions causing a loss in conformational freedom (Dunbar *et al.*, 1997). Urea binds primarily to the protein surface and therefore acts analogously to surface entropy reduction (Liepinsh & Otting, 1994; Pike & Acharya, 1994). This has also been shown to be associated with a decrease in crystallographic temperature factors (Pike & Acharya, 1994). In contrast, urea has also been associated with an increase in conformational flexibility with urea binding, where at low temperatures, low concentrations of urea increased the activity of a thermophilic NADH oxidase by up to 250% (Žoldák *et al.*, 2003). Urea at higher concentrations has also been associated with increased temperature factors in crystal structures (Ratnaparkhi & Varadarajan, 1999).

Given the observed diverse effects of urea on the dynamics of side chains in crystal structures, the effect of this compound on the enzyme catalysis is of interest. Urea has been shown to influence the behaviour and dynamics of proteins even at relatively low concentrations (i.e. in the absence of denaturing conditions). Urea therefore can be considered to act as a general modulator of protein activity. MMRT is uniquely able to investigate the effects of urea, as it provides a mechanism to quantify changes to the enzyme dynamics along the reaction coordinate. Consequently we were interested in studying the effects of urea binding on protein activity and if these effects could be replicated by mutagenesis. The model enzyme MalL was assayed with varying concentrations of urea across a range of temperatures (Warrender, unpublished). Varying urea concentrations altered the activity of MalL across the temperature range (Figure 3.1).

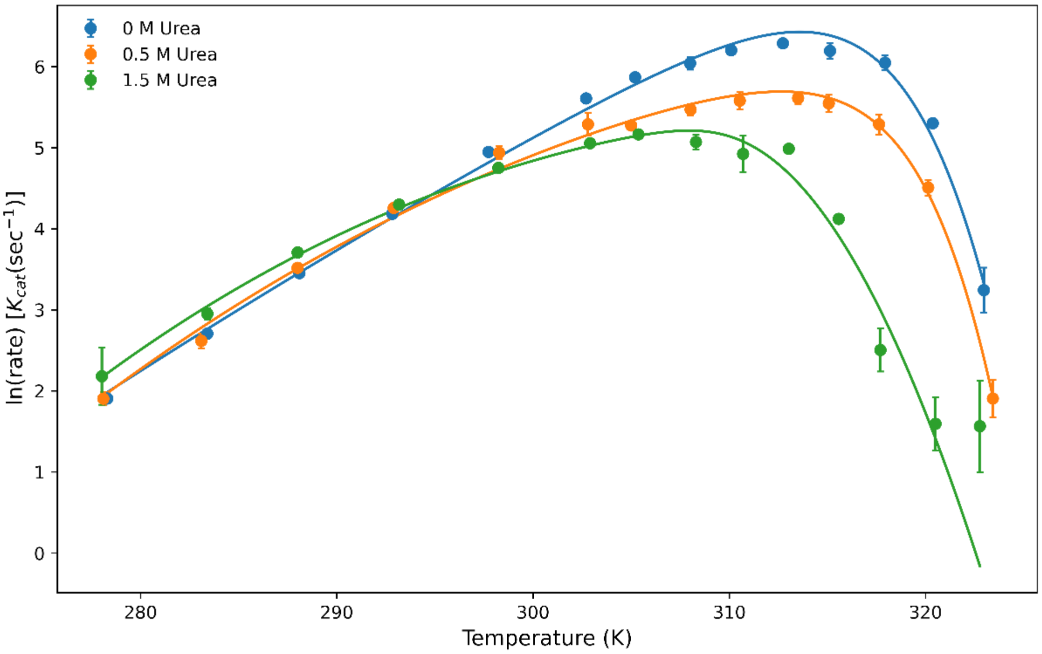


Figure 3.1. Changing activity of MalL with urea bound (Warrender, unpublished). Substrate is para-nitrophenol-α-D-glucopyranoside. Points are the average of at least two replicates, and error bars where visible are the standard deviation of the replicates. Data curves are the fit to MMRT 2.0.

MalL exhibits a varied temperature dependence with increasing urea concentration. Increasing urea concentration resulted in a decreased temperature optimum from approximately 313 K in the 0 M urea dataset to 305 K at 1.5 M urea. Interestingly for the 1.5 M urea dataset, at low temperatures urea is acting as an activator, where the rate increases by ~40% at 278 K. This is in contrast to its action at high temperatures where it acts as either an inhibitor or destabilises the enzyme. This is in agreement with urea effecting a greater conformational flexibility observed by Žoldák *et al.* (2003), where a similar rate increase at low temperatures was observed. The decrease in temperature optimum and increase in rate at low temperatures is a characteristic often observed in psychrophilic enzymes. The mechanism of urea action may therefore be analogous to the evolution towards psychrophily.

3.3 MalL S536R

MalL S536R has been previously characterised in Hamill (2020). This thesis presents an extended structural and kinetic analysis for this mutant. These data have recently been published in Walker *et al.* (2023).

3.3.1 Rational mutation design

The structure of MalL was solved in our laboratory co-crystallised with urea at 2.5 M (Prentice, unpublished). The urea bound structure was solved with an improved resolution of 1.7 Å when compared to MalL wildtype. This is a significant improvement to the MalL wildtype structure which was solved at a resolution of 2.3 Å. The structure had a root mean square deviation (RMSD) of 0.20 Å compared to the wildtype structure. Analysis of this and other urea structures identified a total of 12 urea binding sites (Section 4). Analysis of urea binding within this structure identified a binding site that was of particular interest due to its position buried in the auxiliary C-terminal domain. A site in the C-terminal domain allowed for the study of dynamic changes originating from an auxiliary domain, far removed from the active site. This site was subsequently selected for further study.

We then aimed to imitate the intermolecular interactions of urea binding by introducing a mutation that could replicate these interactions. The amino acid arginine contains a guanidine group that is chemically similar to urea and was therefore chosen to mimic the effects of urea binding. Thus the intermolecular interactions of urea could potentially be replicated as stable intramolecular interactions in the mutant enzyme. The design of this mutant enzyme was completed in collaboration with the Institute for Protein Innovation (Boston, USA). As a result of this work a residue (Ser536) located approximately 3 Å from the urea binding site was identified for mutation to arginine. This serine to arginine mutation was designed to place the guanidine group of urea into the urea binding site.

Mutations may cause dynamic effects to enzyme activity from sites distal to the active site, or from alternative domains. In this sense, they are analogous to allosteric modulators.

3.3.2 Crystallography improvement

MalL S536R was crystallised and the structure solved to atomic resolution at 1.10 Å. The S536R is available in the PDB (PDB: 7LV6). This represents a 1.2 Å resolution improvement over the wildtype enzyme which was solved at 2.3 Å resolution. For comparison there are currently over 100,000 structures in the PDB database solved to a resolution of 2.3 Å or better, while there are less than 4,000 solved to 1.1 Å or better. Data collection and refinement statistics for this structure are provided in Table 3.1. The overall structural architecture is very similar to wildtype, with a RMSD between C_α atoms of just 0.325 and 0.381 Å (555 and 556 atoms compared, respectively) between MalL S536R and each monomer in the wild-type enzyme (PDB: 4M56). A potentially important difference between these two structures is the contents of the asymmetric unit with one molecule in the asymmetric unit for S536R and two molecules for wildtype. See discussion below.

Table 3.1. Data collection and refinement statistics for MalL S536R

Statistic	MalL S536R
Wavelength (Å)	0.953735
Space group	P 1 2_1 1
Unit cell lengths (Å)	a = 48.75 b = 101.00 c = 61.75
Unit cell angles (°)	α = 90.00 β = 113.06 γ = 90.00
Resolution (Å)	1.1 - 44.85 (1.1 - 1.12)
R_{merge}	0.107 (0.548)
Completeness (%)	94.2 (87.1)
Redundancy	10.9 (6.9)
No. of observations	2270037 (65265)
No. of unique reflections	208774 (9509)
Mean I/σI	12.7 (2.9)
$CC_{1/2}$	0.998 (0.884)
R factor	0.126
R_{free}	0.145
Protein atoms	9633
Solvent atoms	785
Solvent content (%)	38.9
Average temperature factor ($Å^2$)	16.13
RMSD bond lengths (°)	0.01
RMSD bond angles (Å)	1.2
Ramachandran analysis:	
- Percentage in favoured regions	98.0
- Percentage in allowed regions	2.0
- Percentage in disallowed regions	0

* Values in brackets correspond to the outer resolution shell.

There was a drastic improvement in the observed resolution and electron density maps for MalL S536R compared to the wildtype enzyme. The mutated arginine residue (Arg536) was well defined in the electron density and occupies approximately the same binding site as urea (Figure 3.2 A); however, the residues involved in guanidine binding differ between urea and Arg536.

The mutation to arginine introduces additional hydrogen bonds that stabilise a mobile loop on the surface of the C-terminal domain containing the residues 536-547 (Figure 3.2 B). Arg536 interacts with the backbone of Arg542 which creates a link across the loop by hydrogen bond interactions to the side chain of Asp544 which in turn links to Ser547 via backbone-backbone interactions (Figure 3.2 B). This results in a significant rigidification of the entire C-terminal domain. In the wildtype structure these residues are poorly defined or missing in the density suggesting that they are disordered. Moreover, the bond distance between Asp544 and Ser547 decreases from 3.4 and 4.3 Å in each chain of the wildtype structure to 2.6 Å in the mutant S536R structure. In the wildtype structure, in chain B Arg542 is shifted 4.8 Å to interact with the backbone atoms of Asp544 and Ser547 (Figure 3.2 C).

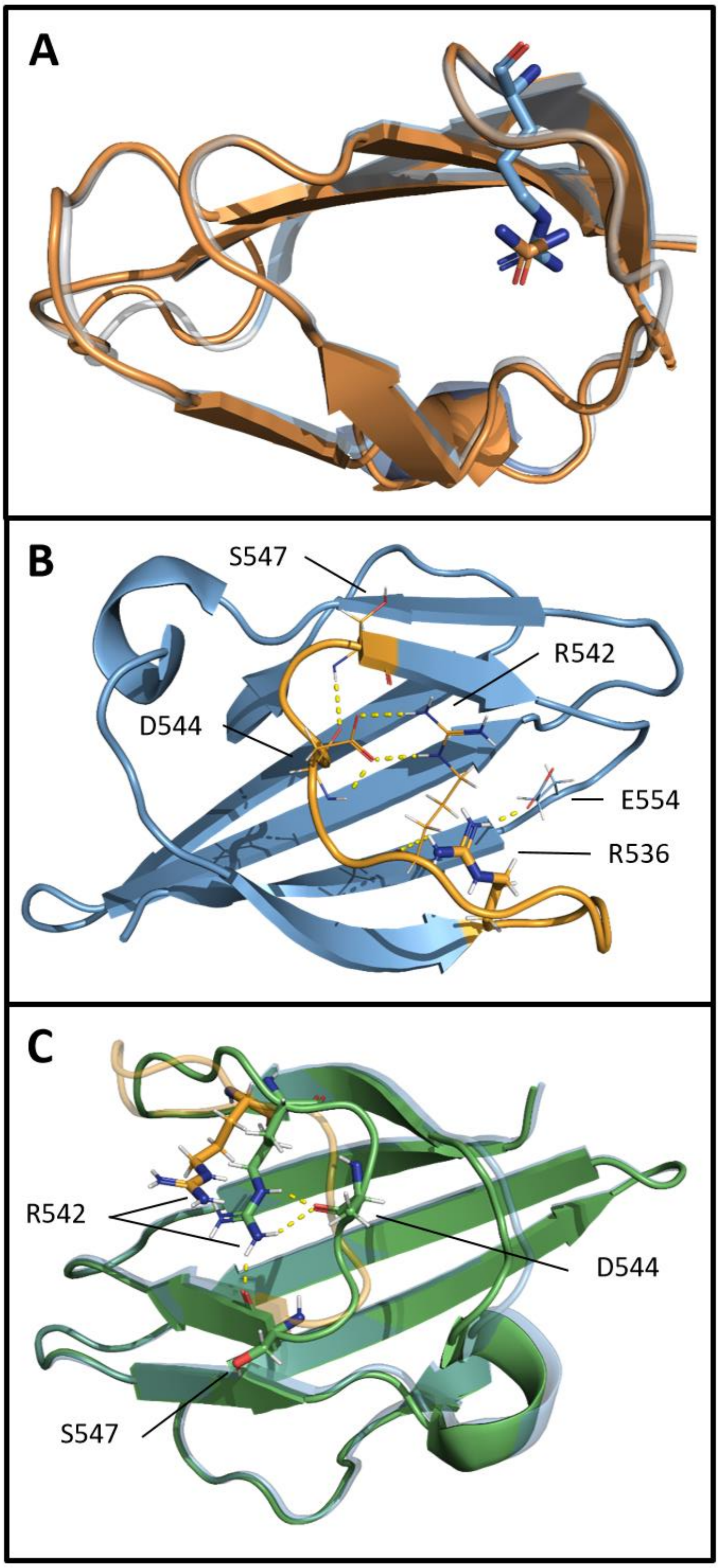


Figure 3.2. A) Overlay of bound urea in the wildtype MalL structure (orange) with the arginine-536 mutation (blue; PDB: 7LV6). B) Mutation of serine-536 to arginine (R536) induces a hydrogen bonding network in the C-terminal domain of MalL causing ordering of a surface loop (coloured yellow). C) Arg542 shifts in the wildtype MalL Chain B (coloured green, PDB: 4M56) structure to bond with the backbone of residues serine-547 and aspartate-544.

MalL S536R shows a significant drop in temperature factors compared to the wild type structure. This is consistent with the improved resolution and suggests an overall rigidification of the structure. In the wildtype structure MalL consists of a rigid β-barrel core. The outer α-helix barrel is stabilised on one side by the C-terminal domain packing against the barrel. The C-terminal domain as well as the B domain is somewhat mobile, relative to the rest of the structure, with increased temperature factors in these regions (Figure 3.3 B, C). In the S536R mutant the entire core is rigidified compared to the wildtype enzyme (Figure 3.3). The areas of most significant flexibility occur in the surface regions of the protein, particularly in the B domain and the lower face of the C-terminal domain; however, these regions are still relatively more rigid than the wildtype enzyme.

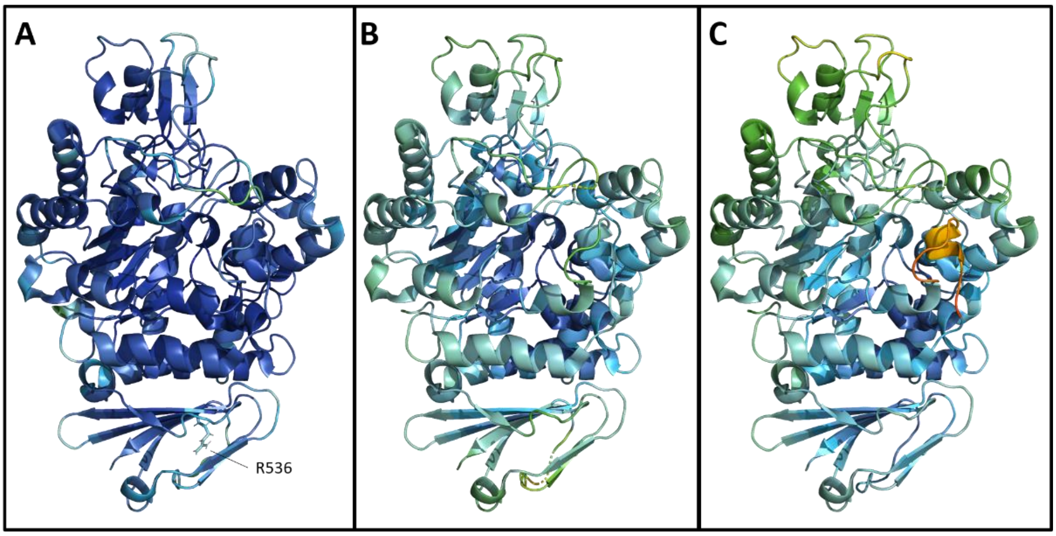


Figure 3.3. Temperature factor analysis shows rigidification of MalL S536R (A) across the entire structure. Mutation site in C-terminal domain is labelled. B-C) Temperature factor analysis of wildtype MalL Chain A (B) and wildtype MalL Chain B (C). Wildtype and S536R structures have been depicted based on the same temperature factor colouration scale. Temperature factors are shown in scale from low to high (blue-green-red).

The comparison between these two structures raises some intriguing questions. Of course, there is the possibility that the mutation S536R is simply fortuitous resulting in only one molecule in the asymmetric unit, fortuitous ordering of a surface loop, and a very high resolution crystal structure. However, the successful design of an arginine mutant that fills a pocket that binds urea, and gave a very high resolution structure gave impetus to further study this effect on both structure and function.

3.3.3 MMRT kinetics for MalL S536R

Kinetic assays presented in this section were collected by Emma Walker based on established protocols for MalL (Section 2.2) with the substrate *p*-nitrophenyl-α-D-glucopyranoside (PNG). Kinetic assays of MalL wildtype (WT) are also presented here for comparison. Data analysis and fitting was conducted separately. Model quality is assessed using the Akaike Information Criterion (AIC). The AIC is a measure of the amount of information lost by a model and selects for a model that best explains the data with the fewest parameters, with a lower AIC indicating a better model (Portet, 2020). The corrected AIC (AICc) accounts for small sample sizes with an additional penalty term for the number of parameters.

3.3.3.1 Temperature characterisation of MalL S536R

The effect of temperature on the rate of MalL S536R was assayed, and the data has been fit with both MMRT 1.5 (linear heat capacity) (Figure 3.4) and MMRT 2.0 (Two-State, sigmoidal heat capacity) (Figure 3.5).

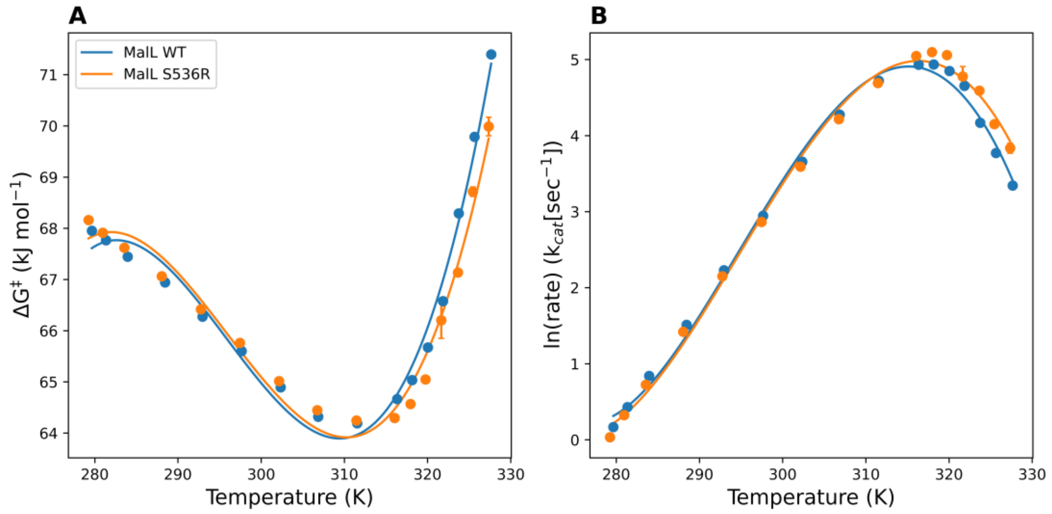


Figure 3.4. Results of temperature characterisation of MalL S536R and wildtype fit to the MMRT 1.5 (linear heat capacity) equation. Points are the average of three replicates, and where visible error bars are the standard deviation of the three replicates. Data collected by Emma Walker. A) Fit of MMRT 1.5 to Temperature vs $\Delta G^{\ddagger}$. B) Fit of MMRT 1.5 to Temperature versus ln(Rate) [$k_{cat}(sec^{-1})$].

Results from the fit to the data are given in Table 3.2. Results for MalL wildtype and MalL S536R are consistent with a fit to MMRT 1.5 (R^2 = 0.9724 and 0.9653, respectively). The reference temperature (T_0) for both datasets is 278.15 K.

Table 3.2. MMRT 1.5 fit parameters with standard error

Parameter	MalL WT	MalL S536R
$\Delta H^{\ddagger}_{T_0}\ (kJ\ mol^{-1})$	19.0 ± 16.0	29.8 ± 15.9
$\Delta S^{\ddagger}_{T_0}\ (J\ mol^{-1}K^{-1})$	-174.0 ± 56.5	-136.2 ± 56.3
$\Delta C^{\ddagger}_{P,0}\ (kJ\ mol^{-1}K^{-1})$	210 ± 18.0	186.4 ± 18.5
$\Delta C^{\ddagger}_{P,T_0}\ (kJ\ mol^{-1}K^{-1})$	12.5 ± 1.5	11.2 ± 1.5
$m\ (J\ mol^{-1}K^{-2})$	-709.9 ± 59.3	-629.9 ± 60.9
$T_{opt}\ (K)$	315.1	316.3
$T_{inf}\ (K)$	307.5	308.4
AICc	191.2	192.3

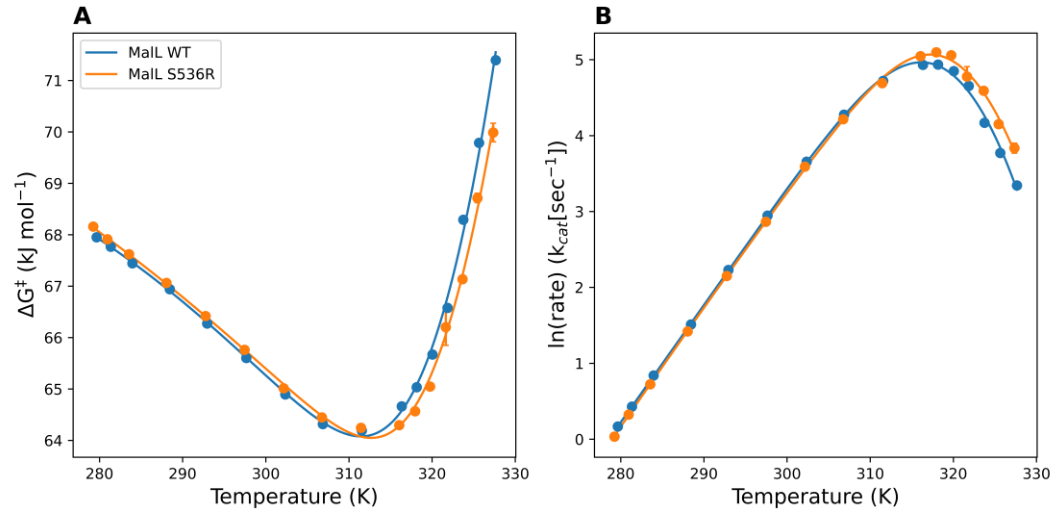


Figure 3.5. Results of temperature characterisation of MalL S536R and wildtype fit to the MMRT 2.0 (Two-state, sigmoidal heat capacity) equation. Points are the average of three replicates, and where visible error bars are the standard deviation of the three replicates. Data collected by Emma Walker. A) Fit of MMRT 2.0 to Temperature vs $\Delta G^{\ddagger}$. B) Fit of MMRT 2.0 to Temperature versus ln(Rate) [$k_{cat}(sec^{-1})$].

Results from the MMRT 2 fit to the data are given in Table 3.3. Results for MalL WT and MalL S536R are consistent with a fit to MMRT 2 ($R^2 = 0.9949$ and 0.9945, respectively). The reference temperature (T_0) for both datasets is 278.15 K.

Table 3.3. MMRT 2.0 fit parameters with standard error

Parameter	MalL WT	MalL S536R
$\Delta H^{\ddagger}_{T_0}\ (kJ\ mol^{-1})$	96.6 ± 2.3	100.5 ± 1.8
$\Delta S^{\ddagger}_{T_0}\ (J\ mol^{-1}K^{-1})$	102.3 ± 8.2	116.0 ± 6.1
$\Delta\Delta H^{\ddagger}\ (kJ\ mol^{-1})$	186.7 ± 45.5	227.8 ± 63.6
$\Delta C^{\ddagger}_{P,lowT}\ (kJ\ mol^{-1}K^{-1})$	0.8 ± 0.2	0.4 ± 0.2
$\Delta C^{\ddagger}_{P,highT}\ (kJ\ mol^{-1}K^{-1})$	-28.1 ± 6.3	-25.5 ± 5.5
$T_C\ (K)$	313.5 ± 2.7	313.7 ± 2.3
AICc	167.6	164.4

The MMRT 2.0 fits of MalL wildtype and MalL S536R are very similar, being within the margin of error. The slope of the heat capacity transition ($\Delta\Delta H^{\ddagger}$) is slightly increased for MalL S536R compared to wildtype.

3.3.4 Hydrogen bond analysis

Differences in hydrogen bonding were analysed between the wildtype structure and the S536R structure. All hydrogen bonds were found using the FindHBond tool in Chimera 1.15 (Section 2.4.7). Hydrogen bond criteria are described in Mills & Dean (1996), and hydrogen bonds were found with the relax H-bond constraints option enabled (criteria relaxed by 0.4 Å and 20°). Existing explicit hydrogens were removed from the structure prior to analysis. The lists of bonds found were then compared against each other and any bonds without an equivalent bond in the other list, or bonds that were at least 0.3 Å shorter than their equivalent were extracted (Code Snippet 3.1). To discount the effects of the resolution (wildtype: 2.3 Å, S536R: 1.1 Å) the process was repeated with the S536R structure cut back to 2.3 Å. In addition, bonds involving a secondary rotomer or in regions not modelled in the other structure were discounted. The analysis was repeated for both chain A and chain B of the MalL wildtype structure. This does not include structural waters or ligands. A full table of hydrogen bond differences can be found in Section 8.5.

Code Snippet 3.1

```
Bond_list_query = [list of hydrogen bonds to find unique bonds in]
Bond_list_match = [list of hydrogen bonds to match against]


def check_if_bonds_match(query_bond, match_bond):
   query_bond = [
      query_donor_name,
      query_donor_number,
      query_acceptor_name,
      query_acceptor_number,
      query_bond_length
   ]
   match_bond = [
      match_donor_name,
      match_donor_number,
      match_acceptor_name,
      match_acceptor_number,
      match_bond_length
   ]

   if query_donor_name == match_donor_name:
      if query_donor_number == match_donor_number:
         if query_acceptor_name == match_acceptor_name:
            if query_acceptor_number == match_acceptor_number:
               if match_bond_length – query_bond_length > 0.3:
                  return bond shortened
               else:
                  return bonds match
   return bonds do not match

for each query_bond in Bond_list_query:
   for each match_bond in Bond_list_match:
      if check_if_bonds_match(query_bond, match_bond) == bonds match:
         continue to next query bond
   found unique bond
```

In total there were 28 significant hydrogen bonds in S536R, consisting of 21 extra bonds not found in MalL wildtype chain A, and seven shortened bonds (>0.3 A). This is compared to 11 additional hydrogen bonds (nine extra and three shortened) in the wildtype structure (Figure 3.6). There were eight additional hydrogen bonds (10 at 2.3 Å) in the S536R structure that involved residues that were missing or not resolved in the wildtype structure, compared to only two additional bonds in the wildtype structure. With the S536R structure resolution cut to 2.3 Å there were 26 significant hydrogen bonds (16 extra and 11 shortened), while the number of significant hydrogen bonds in wildtype increased to 17 (12 extra and five shortened).

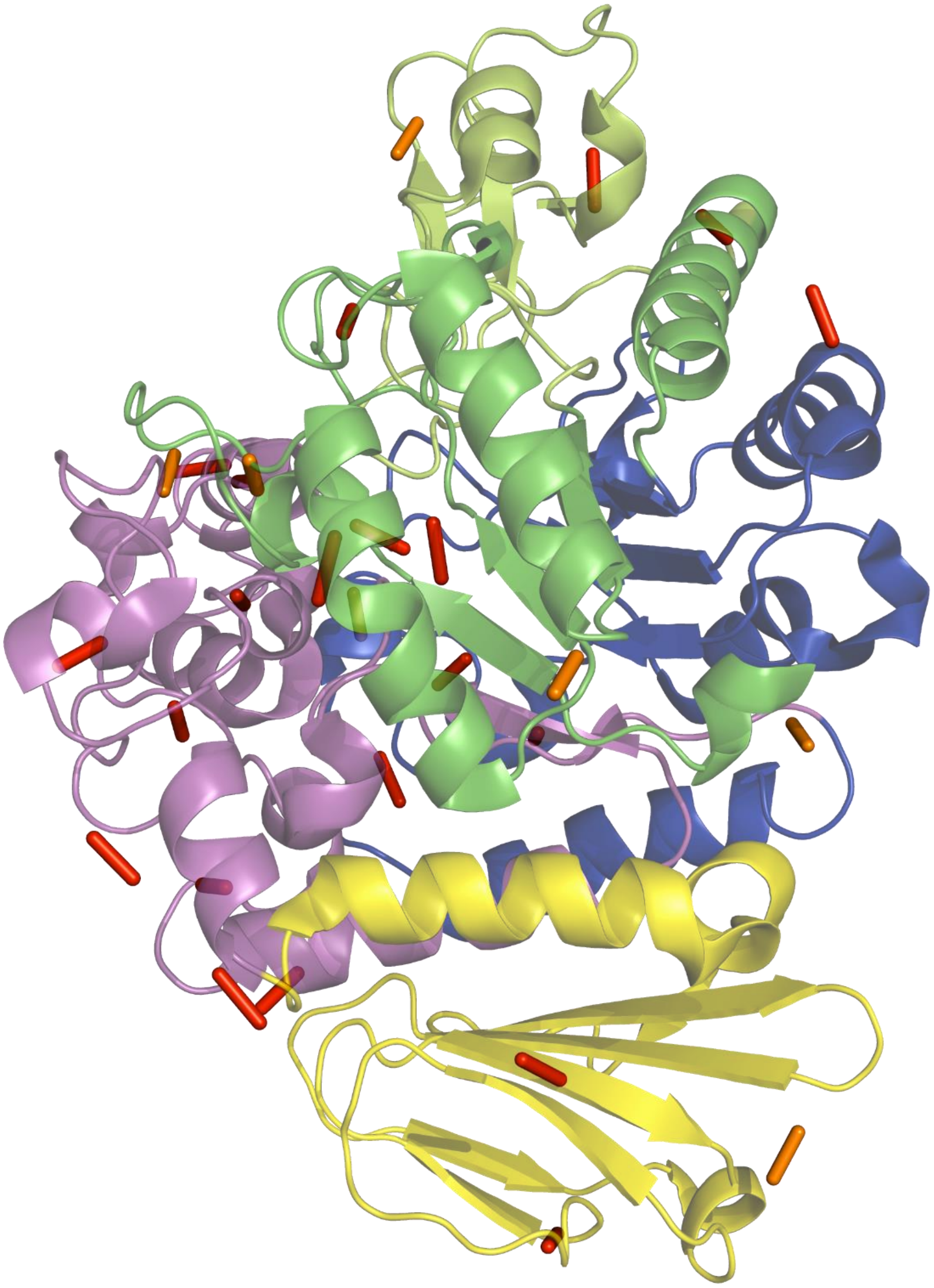


Figure 3.6. Additional hydrogen bonding in S536R, indicating bonds not found in MalL wildtype (Chain A) (Red) and bonds significantly shortened in S536R (>0.3 Å) (Orange).

There were a total of 25 significant hydrogen bonds in the MalL S536R structure compared to the MalL wildtype chain B structure, consisting of 18 extra hydrogen bonds and seven shortened bonds. In the MalL wildtype chain B structure there were 13 extra hydrogen bonds and six shortened bonds. In addition there were six hydrogen bonds in the MalL S536R structure involving residues that were missing or not resolved in the MalL wildtype chain B structure compared to only one in the MalL wildtype chain B structure (Figure 3.7). See discussion for further analysis.

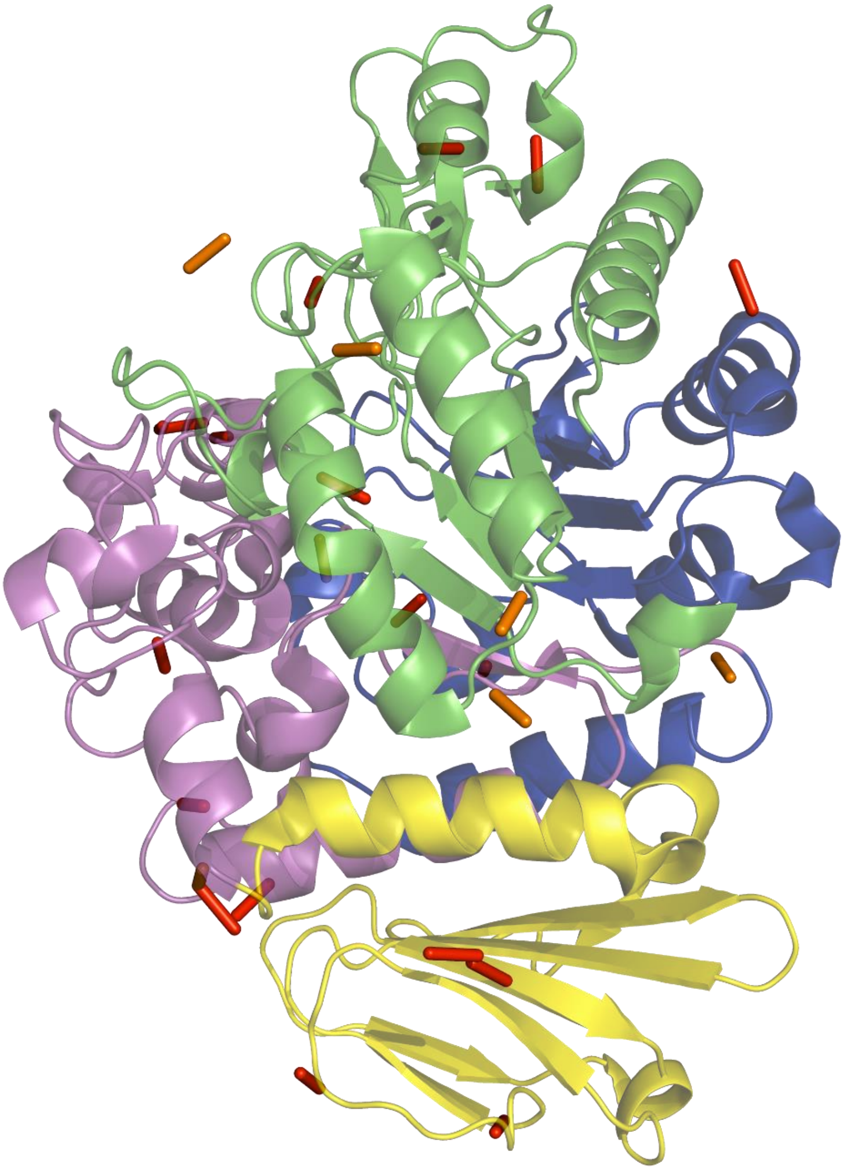


Figure 3.7. Additional hydrogen bonding in S536R, indicating the bonds not found in MalL wildtype (Chain B) (Red), and the bonds significantly shortened (>0.3 Å) (Orange).

3.3.5 REES

The dynamics of MalL S536R was assessed using Red Edge Excitation Shift (REES) spectroscopy (Figure 3.8). Data were fit to the REES equation (Equation 1.35) and results are given in Table 3.4.

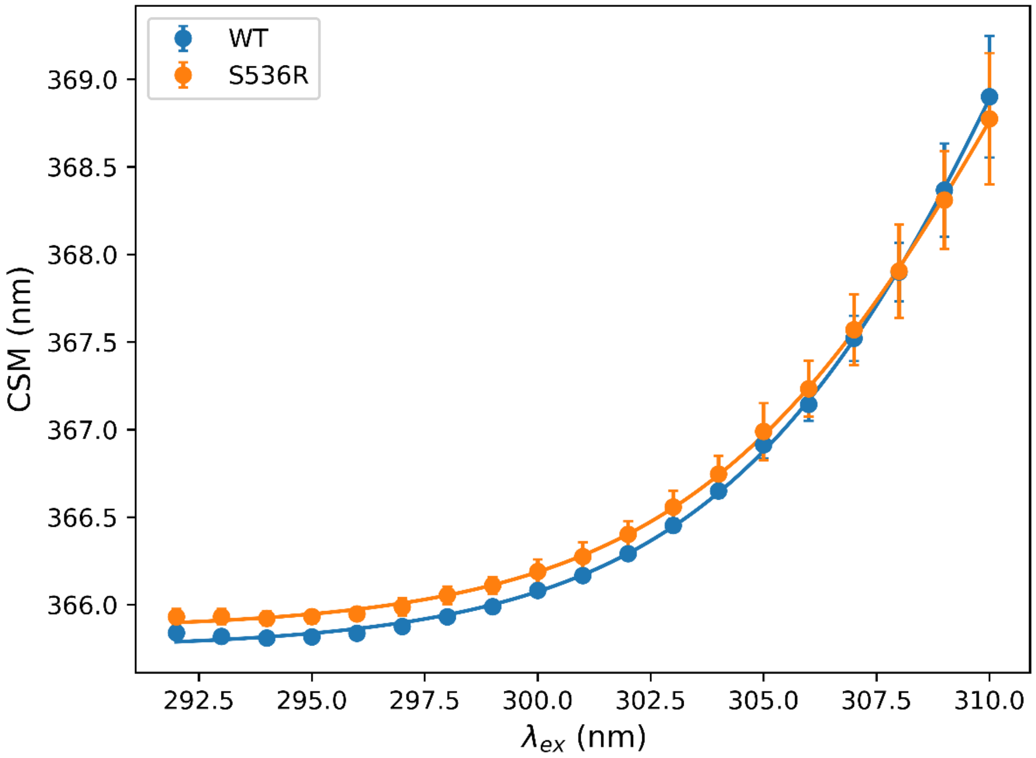


Figure 3.8. REES spectroscopy of MalL S536R and MalL wildtype at 37°C. Error bars where visible are the standard error of three replicates. Data are fit to the REES equation (Equation 1.35).

Table 3.4. Fit of REES equation with standard error

Parameter	**MalL WT**	**S536R**
$CSM(\lambda_{Ex}^{FC})\ (nm)$	365.75 ± 0.02	365.86 ± 0.01
$CSM(\lambda_{Ex}^{R})\ (nm)$	374.2 ± 1.1	372.6 ± 0.6
$\Delta G_m\ (mJ\ mol^{-1} nm^{-1})$	7.1 ± 0.3	7.1 ± 0.3
$\lambda_{Ex}^{50\%}$ **(nm)**	311.9 ± 0.8	311.0 ± 0.6

3.4 Discussion

3.4.1 MalL S536R on the transition-state pathway

The kinetics of MalL wildtype and MalL S536R are very similar, with the fit parameters of both the MMRT 1.5 and MMRT 2 fits being within standard error of each other (Table 3.2, Table 3.3). Thus, the general dynamics of these two enzymes in solution is unchanged with the introduction of the S536R mutation. This is also reflected in the REES results which show only minor differences in fitting parameters, with only the value of $CSM(\lambda_{Ex}^{\mathrm{R}})$ (Section 1.7) being significantly different.

However, upon crystallisation the S536R mutation has a significant effect, resulting in a much more stable and restricted structure, with an improvement in crystallographic resolution from 2.3 Å to 1.1 Å. Support for this restricted structure is further provided by the crystallographic temperature factors, which show a reduction in the variance of the temperature factors across the entire enzyme, including at sites distal to the mutation site.

Based on this evidence, the S536R structure is hypothesised to represent a conformation close to the enzyme transition state complex (Walker *et al.*, 2023). The process of crystallisation has “trapped” this highly ordered conformation. The improvement in crystallographic resolution is a representation of the higher degree of order, and reduced conformational flexibility characteristic of a conformation that favours the transition state on the reaction pathway. This increased order is a factor of the crystallisation conditions, and in solution the MalL S536R enzyme samples the same distribution of conformational states as the wildtype enzyme, involving both the enzyme-substrate complex and TLC complex.

There is a significant difference in the hydrogen bonding between the wildtype structure and that of MalL S536R in the crystallised form. MalL S536R has a greater number and strength of hydrogen bonds, consistent with the greater order of the structure. The majority of these extra hydrogen bonds in the MalL S536R structure (20 at 1.1 Å, 18 at 2.3 Å) can be found in the regions 1-193 and 322-459. These two regions have been found to be important in the transition between the enzyme-substrate and transition-state complexes in dynamics investigations of the enzyme (van der Kamp *et al.*, 2018; Walker *et al.*, 2023). These regions are the most mobile in simulations of MalL wildtype in the reactant state. These simulations also show that these regions are significantly constrained in the transition-state complex and make the greatest contribution to the calculated change in heat capacity ($\Delta C_P^\ddagger$), which has been shown to match the experimentally derived value (van der Kamp *et al.*, 2018).

The origin of the activation heat capacity is in the large conformational flux of the enzyme-substrate complex, primarily in the regions mentioned above, which form the lid and loop regions above the active site. The conformational freedom of these regions facilitates substrate entry and binding. The enzyme-substrate complex is able to sample the transition-state complex, which is highly constrained compared to the enzyme-substrate complex. Formation of the transition state is accompanied by the formation of a large number of extra hydrogen bonds and the formation of other additional weak-bonding interactions. The formation of these extra bonds is consistent with the difference in $\Delta H^\ddagger$ between the two states (van der Kamp *et al.*, 2018; Walker *et al.*, 2023).

3.4.2 Comparison to equilibrium model

The equilibrium model postulates an equilibrium between an active and inactive state (E_{active}, $E_{inactive}$) (Daniel & Danson, 2010). This is consistent with the ES ($E_{inactive}$) and TLC (E_{active}) states under MMRT 2.0 with a non-zero activation heat capacity (Walker *et al.*, 2023). Under this model the enzyme rate is determined by both the chemical step, as well as the conformational equilibrium between the active and inactive state. This can be expressed as Equation 3.1, where the first term in square brackets refers to the chemical step ($\Delta \mathrm{G}^\ddagger_{\mathrm{chem}}$), and the second term refers to the conformational step ($\Delta \mathrm{G}^\ddagger_{\mathrm{conf}}$).

Equation 3.1. Equilibrium model equation

$$\Delta G^{\ddagger} = \left[\Delta H^{\ddagger}_{T_0} - T\Delta S^{\ddagger}_{T_0} + \Delta C^{\ddagger}_P\left(T - T_0 - T\ln\left(\frac{T}{T_0}\right)\right)\right]_{chem} + RT\ln\left(1 + e^{\frac{\left[\Delta H^{\ddagger}_{T_0} - T\Delta S^{\ddagger}_{T_0} + \Delta C^{\ddagger}_P\left(T - T_0 - T\ln\left(\frac{T}{T_0}\right)\right)\right]_{conf}}{RT}}\right)$$

Here, the temperature rate data (Section 3.3.3.1) of MalL wildtype and MalL S536R is fitted with the equilibrium model, and compared with MMRT 2.0.

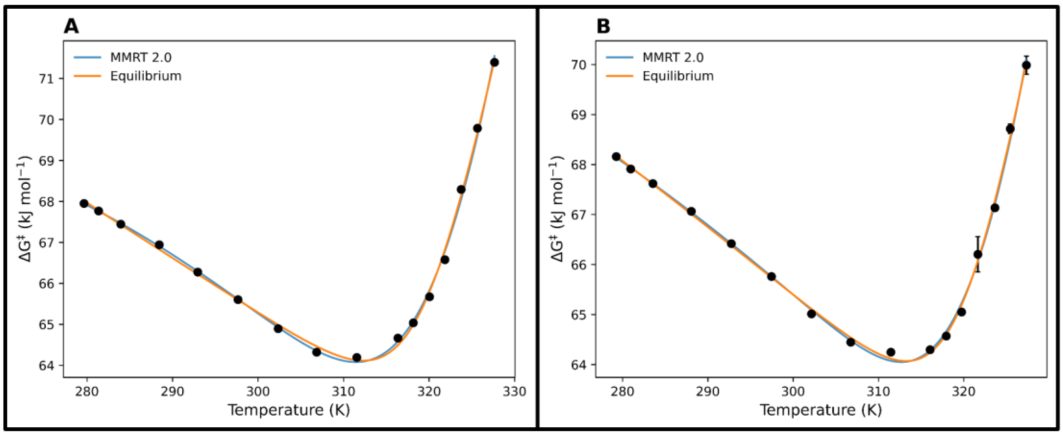


Figure 3.9. Comparison of fits to equilibrium model and MMRT 2.0. A) MalL wildtype. B) MalL S536R.

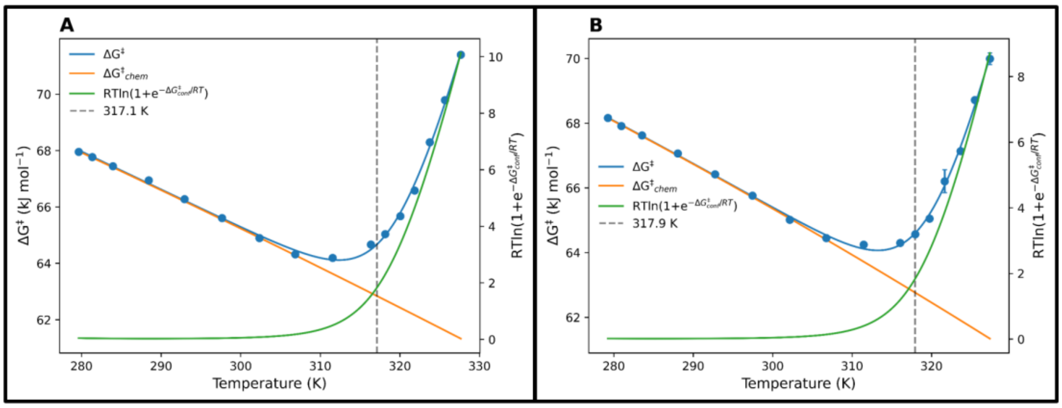


Figure 3.10. Relative contributions of $\Delta G^{\ddagger}_{chem}$ and $\Delta G^{\ddagger}_{conf}$ to $\Delta G^{\ddagger}$. Vertical line indicates where $\Delta G^{\ddagger}_{conf}$ is zero. A) MalL wildtype. B) MalL S536R.

The data are fitted well by both the equilibrium model and MMRT 2.0, with all fits having an $R^2 > 0.99$. The AICc for the equilibrium model was 172.1 and 165.9 for MalL wildtype and MalL S536R, respectively, compared to 167.6 and 164.4 for MMRT 2.0. Fit parameters for MalL wildtype and MalL S536R are given in Table 3.5. Both the equilibrium model and MMRT 2.0 are fitted with six parameters. Interestingly, $\Delta G^{\ddagger}_{\mathrm{conf}}$ crossed the temperature axis at 317.1 K and 317.9 K for MalL wildtype and MalL S536R, respectively, which is similar to the value of T_C for each enzyme (313.5, 313.7, respectively), reminiscent of an unfolding transition with large values of $\Delta H^{\ddagger}_{\mathrm{conf}}$ and $\Delta S^{\ddagger}_{\mathrm{conf}}$. Denaturation is ruled out under the experimental conditions.

Table 3.5. Fit parameters for equilibrium model with standard error

	MalL WT	**S536R**
$\boldsymbol{\Delta H^{\ddagger}_{T_0}(chem)\ (kJ\ mol^{-1}K^{-1})}$	104.6 ± 20.1	104.1 ± 8.7
$\boldsymbol{\Delta S^{\ddagger}_{T_0}(chem)\ (J\ mol^{-1}K^{-1})}$	131.0 ± 69.8	128.8 ± 30.4
$\boldsymbol{\Delta C^{\ddagger}_{P}(chem)\ (kJ\ mol^{-1}K^{-1})}$	0.085 ± 1.215	0.150 ± 0.645
$\boldsymbol{\Delta H^{\ddagger}_{T_0}(conf)\ (kJ\ mol^{-1}K^{-1})}$	190.9 ± 139.3	191.8 ± 128.7
$\boldsymbol{\Delta S^{\ddagger}_{T_0}(conf)\ (J\ mol^{-1}K^{-1})}$	599.6 ± 434.6	600.7 ± 402.5
$\boldsymbol{\Delta C^{\ddagger}_{P}(conf)\ (kJ\ mol^{-1}K^{-1})}$	9.1 ± 11.0	7.8 ± 9.5
$\boldsymbol{T_0(chem)\ (K)}$	278.15	278.15
$\boldsymbol{T_0(conf)\ (K)}$	310	310
AICc	172.1	165.9

The equilibrium model is difficult to fit, with many parameters having large errors, particularly for the parameters associated with the conformational step. This is likely due to the relatively small size of the final $\Delta G^{\ddagger}_{\mathrm{conf}}$ term, compared to its individual components, as well as a large degree of correlation between individual parameters.

3.4.3 Urea for rational mutation design

Urea at low concentration has been used to alter the dynamics of proteins in a systematic way. There are currently few general methods to improve the chance of crystallisation and resolution of protein diffraction. The addition of urea can serve as a general method to guide rational mutagenesis to produce better diffracting crystals. This method has been successfully used to engineer the enzyme MalL. The introduced mutation had little impact on the function of the enzyme and successfully improved diffraction by 1.2 Å. The mutation appears to act as a stabilising influence, with an associated decrease in temperature factors. This stabilisation in general allows for better resolution of surface loops and side chains, which are able to be captured in the structure.

4 Rational enzyme engineering using arginine mutations

4.1 Introduction

Arginine has long been identified as a stabilising element in proteins, with arginine mutations been used as a general method to create stabilised proteins (Mrabet *et al.*, 1992; Sokalingam *et al.*, 2012; Strub *et al.*, 2004). Arginine is able to adopt a large number of conformations due to its size and number of atoms available for bonding interactions (Sokalingam *et al.*, 2012). Arginine is capable of forming five hydrogen bonds with the three nitrogens of the guanidinium group. It also contains two basic sites capable of forming a salt bridge (Aghajari *et al.*, 1998; Wyttenbach *et al.*, 1999). Thus arginine residues are readily able to form stabilising interactions (Mrabet *et al.*, 1992). Surface engineering of lysine into arginine has been identified as a route to increase protein stability (Sokalingam *et al.*, 2012).

Changes in the prevalence of arginine residues is associated with enzyme adaptation to different temperature environments. Psychrophilic enzymes have changes to their amino acid content that relate to protein structural flexibility, including fewer charged amino acids forming fewer salt bridges and hydrogen bonds (Huang *et al.*, 2023). A reduction in arginine content is frequently associated with psychrophilic proteins (Aghajari *et al.*, 1998; Laurell *et al.*, 2000). Decreases in arginine content reduces the number of internal electrostatic interactions that maintain overall fold structure, as well as reducing the strength of inter-domain interactions (Aghajari *et al.*, 1998). These changes contribute to the generally observed increase in conformational flexibility of psychrophilic enzymes.

This increase in flexibility is achieved by the reduction of, or tuning of many stabilising interactions (Section 1.4). This increase in flexibility is often associated with an enthalpy-entropy trade-off. The decreased enthalpy in the ground state results in a more favourable activation enthalpy. This comes at the expense of a greater degree of reorganisation to reach the transition state, and thus an increased entropy cost (Collins & Feller, 2023).

Psychrophilic enzymes often show changes in hydrophobic interactions. The size and relative hydrophobicity of non-polar residue clusters is generally decreased (Struvay & Feller, 2012). This weakens the hydrophobic effect, and lowers the compactness of the enzyme. Psychrophilic enzymes also tend to have an increased number of hydrophobic residues interacting with the solvent, resulting in a destabilising effect caused by an entropically unfavourable solvent reorganisation (Struvay & Feller, 2012). Hydrophobic interactions generally increase with temperature and are primarily entropic at 22°C and enthalpic at high temperatures (Baldwin, 1986). These interactions and thus the hydrophobic effect is weakened at low temperatures. The folded state of a protein is governed by the free energy difference (ΔG) between the folded and unfolded (denatured) states. If the heat capacity difference between these states is non-zero, ΔG is curved, resulting in denaturation at both high and low temperatures (Sanfelice & Temussi, 2016).

4.2 Rational mutation design

Based on the promising results from the MalL S536R mutation, we sought to develop and characterise further arginine point mutations. Collaborators from the Institute for Protein Innovation (Boston, USA) evaluated the *in silico* stability of all possible single point mutations in MalL, and assigned each an energy score (Park *et al.*, 2016). The data for the wildtype enzyme (where each residue is mutated into itself) vary with a maximum energy score of -1144.5 kcal mol^{-1} and a minimum of -1195.7 kcal mol^{-1} with an average of -1167.2 ± 9.4 kcal mol^{-1}. For the mutant enzymes, the most stabilising predicted mutation is Y14F with an energy score of -1196.1 kcal mol^{-1} and the most destabilising mutation is M189I with an energy score of -1080.5 kcal mol^{-1}. On average leucine is the most stabilising mutation with an average energy score of -1164.5 kcal mol^{-1} and proline is the most destabilising with an average energy score of -1158.0 kcal mol^{-1}. On average, polar uncharged amino acids are the most stabilising, followed by hydrophobic amino acids, and then charged amino acids.

Mutations that removed cysteine, phenylalanine, histidine, methionine, tryptophan, and tyrosine were the most stabilising (being more stable than wildtype), with between 41.4-48.0% of such mutations being stabilising, while removing glutamic acid, glycine, lysine, and arginine were the most destabilising, with only 15.8-18.4% of such mutations being stabilising. Mutations to isoleucine, leucine, methionine, asparagine, glutamine, serine, threonine, and valine were the most stabilising, with between 34.2-36.5% of these mutations being stabilising. Proline was the most destabilising mutation with only 17.4% being stabilising, followed by tryptophan at 20.3%, and glycine, lysine, and arginine between 24.6-26.6%. These trends are summarised in Figure 4.1.

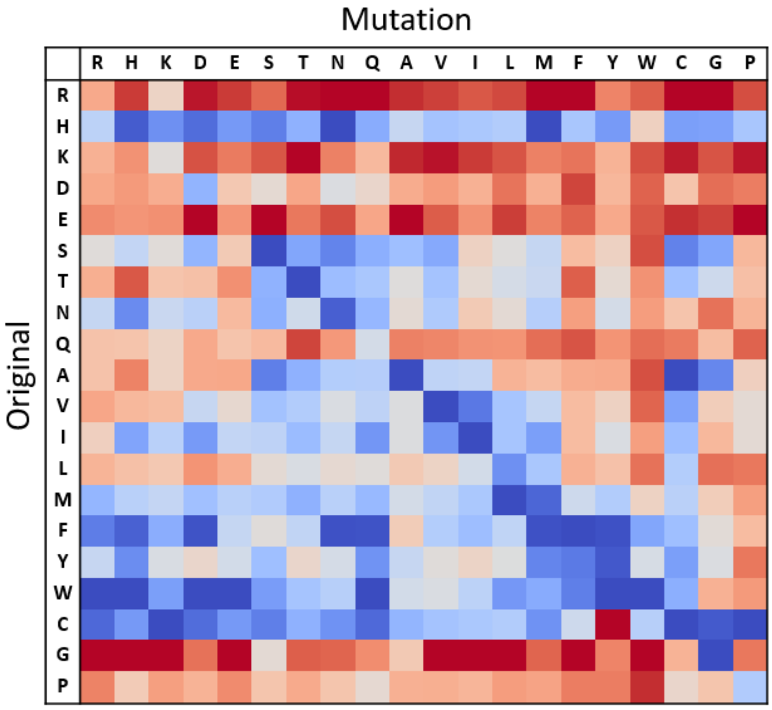


Figure 4.1. Mutation stability matrix. Stability score is calculated as the fraction of mutations with an energy score less than the average for all wildtype mutations. Stabilising mutations are indicated by a blue colour, while destabilising mutations are indicated by a red colour.

Arginine tends to be both destabilising when removed as well as when added in the *in silico* data. Only 15.8% of mutations away from, and 26.6% of mutations to arginine are stabilising. This indicates that arginine residues are in general difficult to place, but when placed well are vital to the surrounding structure.

Potential new arginine mutations were identified by screening all possible arginine mutations for mutants that had a predicted energy score of less than the predicted energy score of the S536R mutant. This initial screen identified 475 potential mutations. This list was further refined by proximity to known urea binding sites from previously determined crystal structures in the presence of urea (Prentice et al., unpublished) (Figure 4.2). There were a total of 12 urea binding sites observed over the seven crystal structures determined in the presence of urea. Any potential arginine mutation that was not within five Ångstroms of a bound urea was discarded. This left 78 potential arginine mutants. These 78 potential mutations were individually screened in PyMOL for overlap of the mutant rotamers with the bound urea. This left a final shortlist of 18 potential mutations (Table 4.1).

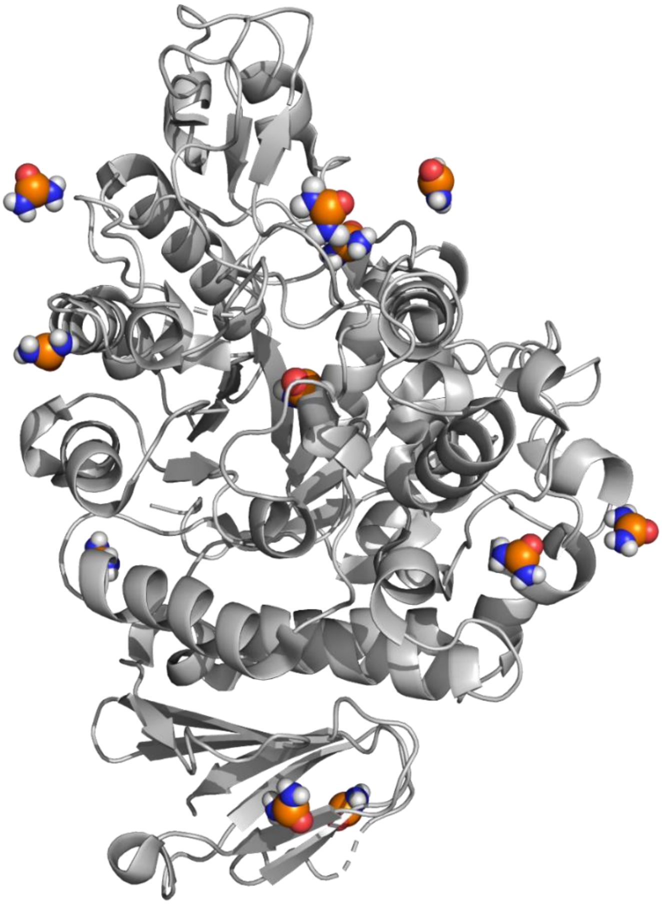


Figure 4.2. Urea binding sites overlaid on MalL wildtype structure. Urea is shown as spheres.

Table 4.1. Potential arginine mutations identified via in silico screening. Variants taken through to experimental characterisation are in bold.

Mutation	Energy (kcal mol^{-1})
S142R	-1178
V376R	-1173
N392R	-1171
S338R	-1169
N445R	-1169
D58R	-1169
T150R	-1168
Q494R	-1167
K165R	-1166
D343R	-1166
G146R	-1162
S446R	-1159
Q540R	-1159
D492R	-1159
T208R	-1159
A437R	-1157
E173R	-1155
D379R	-1154
S536R	-1152

Of the 18 potential mutations four were chosen to bring forward and characterise experimentally. These were chosen to occupy four distinct urea binding sites. The chosen urea sites were selected based on their position, where the urea binds into a pocket, rather than only surface binding. T150R and S142R were chosen to occupy the urea binding sites between the B-domain and the lid domain (residues 374-459, β_8-α_8 extension) which was found to be highly mobile in molecular dynamics simulations (van der Kamp *et al.*, 2018), and the constriction of this domain at the transition state contributes greatly to the activation heat capacity (Figure 4.3 A). V376R was chosen as it occupies a deep pocket that connects the lid domain to the TIM barrel structure in the loop that connects α_7 and β_8 (Figure

4.3 B). D492R was chosen as it occupies the second urea binding site (from MalL S536R) in the C-terminal domain (Figure 4.3 C).

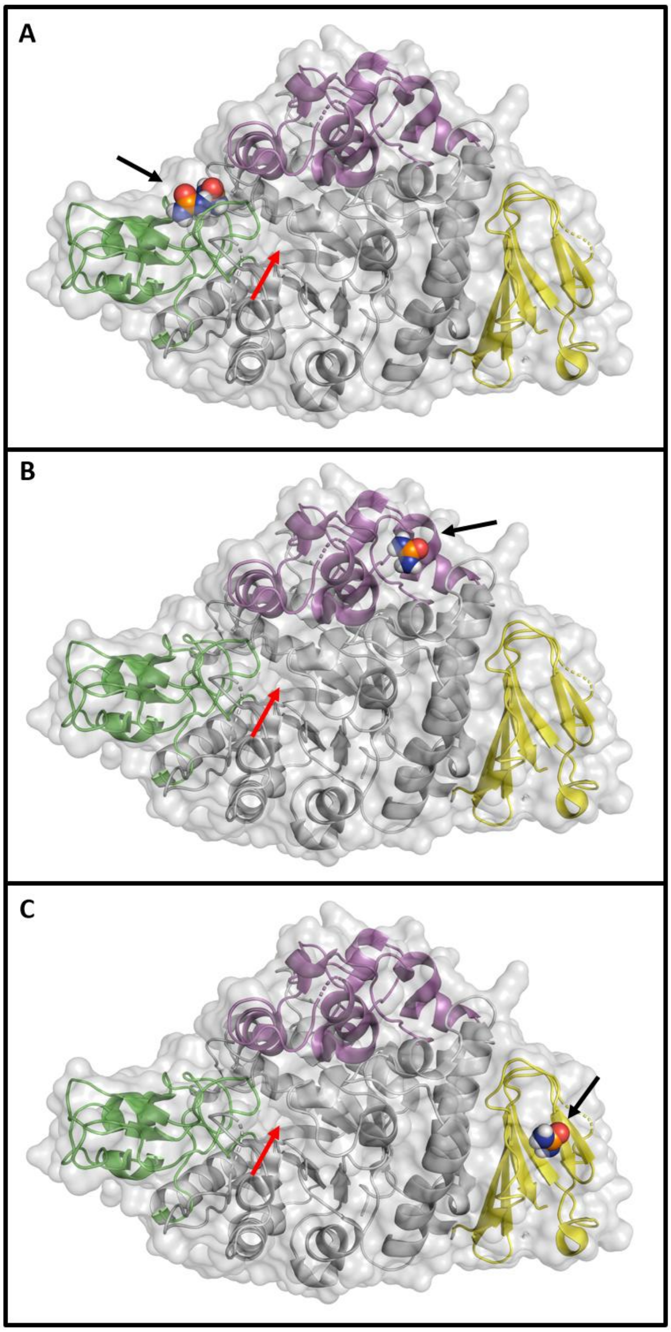


Figure 4.3. Urea binding sites targeted for arginine mutations. Bound urea is shown as spheres. Domain B (residues 99-174) is green. The Lid domain (residues 374-459) is magenta. The C-terminal domain (Residues 485-561) is yellow. Domains B and the Lid domain form part of the active site pocket. The active site pocket is indicated with a red arrow. Mutation sites are indicated with a black arrow. A) Mutations S142R and T150R target urea binding sites between Domain B and the lid domain. B) The V376R mutation targets a urea binding site between the Lid domain and the TIM barrel structure. C) The D492R mutation targets a binding site in the auxiliary C-terminal domain.

Of the four mutants (S142R, T150R, V376R, and D492R), three were successfully characterised. S142R was discarded as mutation at this site interfered with substrate binding (K_M was approximately 25 times greater than the K_M of MalL wildtype). Data here are presented alongside MalL wildtype as a baseline for comparison (Section 3).

4.3 Generation of mutant expression strains

DNA encoding mutant proteins were obtained from Twist Biosciences (USA) as inserts in a pET-28a(+) plasmid expression vector. Plasmids were transformed into *E. coli* DH5α for working with the plasmid and *E. coli* BL21 DE3 for protein expression.

4.4 Characterisation of MalL T150R

4.4.1 Expression of MalL T150R

MalL T150R was expressed and purified using established protocols for MalL variants (Section 2.1), adapted from the methods presented in Schönert et al. (1998). MalL T150R was expressed under the control of the lac operon on the pET-28a(+) expression vector in *E. coli* BL21 DE3. Pure soluble protein was obtained by immobilised metal affinity chromatography (IMAC), followed by size exclusion chromatography (SEC) (Section 2.1.3) (Figure 4.4). The purity of obtained protein was confirmed by polyacrylamide gel electrophoresis (Section 2.1.4) (Figure 4.5).

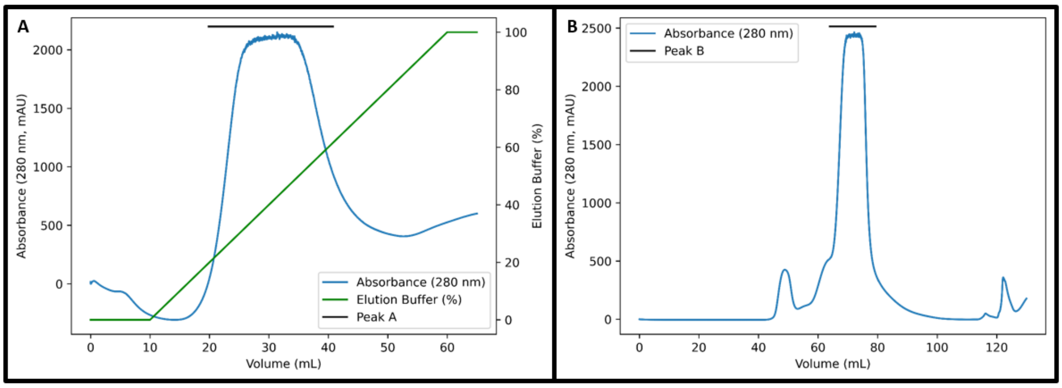


Figure 4.4. Purification of MalL T150R. A) Elution chromatogram (Absorbance, 280 nm) of MalL T150R from IMAC column. Black bar represents eluted protein. B) Elution chromatogram (Absorbance, 280 nm) of MalL T150R from SEC column. Black bar represents eluted protein.

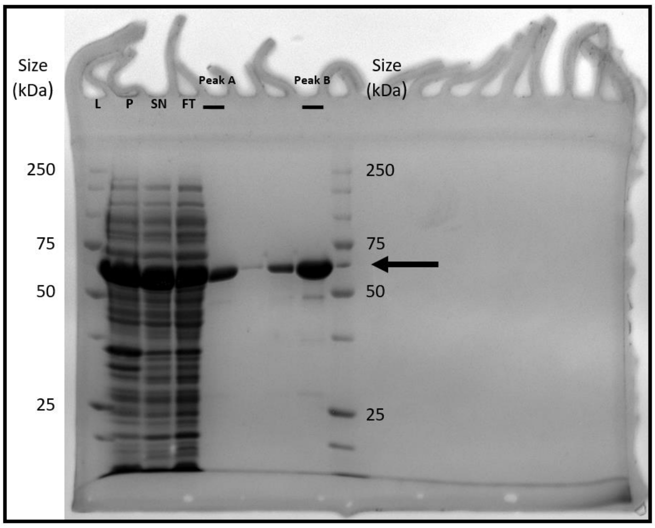


Figure 4.5. PAGE of MalL T150R purification. L = protein ladder, P = Insoluble pellet, SN = Supernatant containing soluble protein loaded onto IMAC column, FT = Supernatant eluted from IMAC column minus bound target protein. Peak A and peak B represent peaks IMAC/SEC chromatograms. The other 2 lanes are the small peaks at approximately 48 mL and 62 mL in Figure 4.4 B above. Arrow indicates target protein.

MalL T150R was purified by IMAC, eluting between 15-80% elution buffer (Section 2.1.1). It was further purified by SEC and eluted between 65-85 mL. Purity was confirmed by SDS-PAGE. An intense band was observed between 50 and 70 kDa, which is consistent with the expected size of 69.5 kDa.

4.4.2 Melting temperature

The melting temperature (T_m) of MalL T150R was assessed by a SYPRO orange thermal shift assay at 48.2 ± 0.1°C, compared to MalL wildtype at 50.6 ± 0.1°C.

4.4.3 Kinetic characterisation of MalL T150R

Kinetic characterisation of MalL T150R was performed using established protocols for MalL variants, using the substrate *p*-nitrophenyl-α-D-glucopyranoside (PNG), as described in Section 2.2.

4.4.3.1 Michaelis-Menten

MalL T150R was characterised using a Michaelis-Menten model to find the parameters k_{cat}, K_M, and K_i (Figure 4.6). The Michaelis-Menten assay was performed at 25°C.

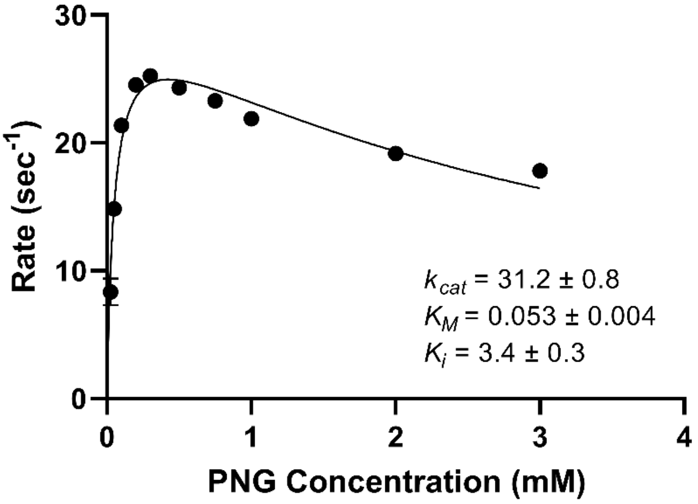


Figure 4.6. Michaelis-Menten kinetics of MalL T150R. Assay was performed at 25°C with the substrate p-nitrophenyl-α-D-glucopyranoside (PNG). Data are fit to a substrate-inhibition model. Data are plotted as the average of at least three replicates, with error bars, where visible, being the standard deviation of these replicates.

Results of the Michaelis-Menten fit of MalL T150R is given in Table 4.2. MalL T150R has a k_{cat} of 31.2 ± 0.8 sec^{-1} and a K_M of 0.053 ± 0.004 mM. MalL T150R was fit with a substrate inhibition model with a K_i = 3.4 ± 0.3 mM. Subsequent assays were performed at 0.5 mM, approximately 10 X K_M.

Table 4.2. Michaelis-Menten parameters for MalL T150R at 25°C with standard error

Parameter	MalL T150R
k_{cat} (sec^{-1})	31.2 ± 0.8
K_M (mM)	0.053 ± 0.004
K_i (mM)	3.4 ± 0.3

4.4.3.2 Temperature characterisation

The temperature dependence of MalL T150R was characterised as described in Section 2.2.2 and fit with both MMRT 1.5 (Figure 4.7) and MMRT 2.0 (Figure 4.8).

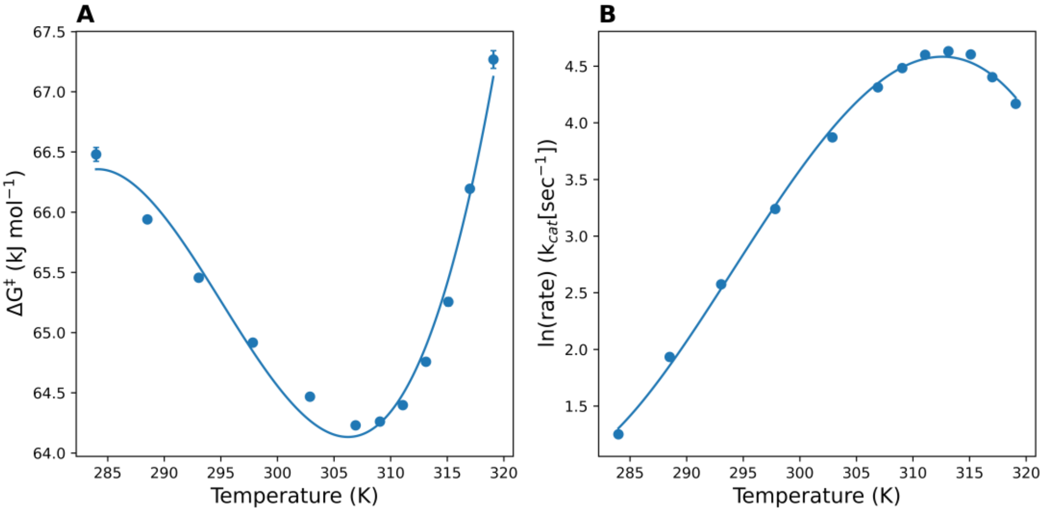


Figure 4.7. Result of temperature characterisation of MalL T150R fit to the MMRT 1.5 (linear activation heat capacity) equation. Points are the average of at least three replicates, and error bars where visible are the standard deviation of the replicates. A) Fit of MMRT 1.5 to Temperature vs ΔG‡. B) Fit of MMRT 1.5 to Temperature vs ln(Rate).

Results of the MMRT 1.5 fit to MalL T150R is given in Table 4.3. Results are consistent with an MMRT 1.5 fit, with an R^2 = 0.9747.

Table 4.3. MMRT 1.5 fit parameters of MalL T150R with standard error

Parameter	MalL WT	MalL T150R
$\Delta H^{\ddagger}_{T_0}\ (kJ\ mol^{-1})$	19.0 ± 16.0	5.7 ± 20.5
$\Delta S^{\ddagger}_{T_0}\ (J\ mol^{-1}K^{-1})$	-174.0 ± 56.5	-216.0 ± 72.3
$\Delta C^{\ddagger}_{P,0}\ (kJ\ mol^{-1}K^{-1})$	210 ± 18.0	215.2 ± 24.5
$\Delta C^{\ddagger}_{P,T_0}\ (kJ\ mol^{-1}K^{-1})$	12.5 ± 1.5	12.3 ± 1.9
$m\ (J\ mol^{-1}K^{-2})$	-709.9 ± 59.3	-729.5 ± 81.0
$T_0\ (K)$	278.15	278.15
$T_{opt}\ (K)$	315.1	312.6
$T_{inf}\ (K)$	307.5	304.9
AICc	191.2	130.1

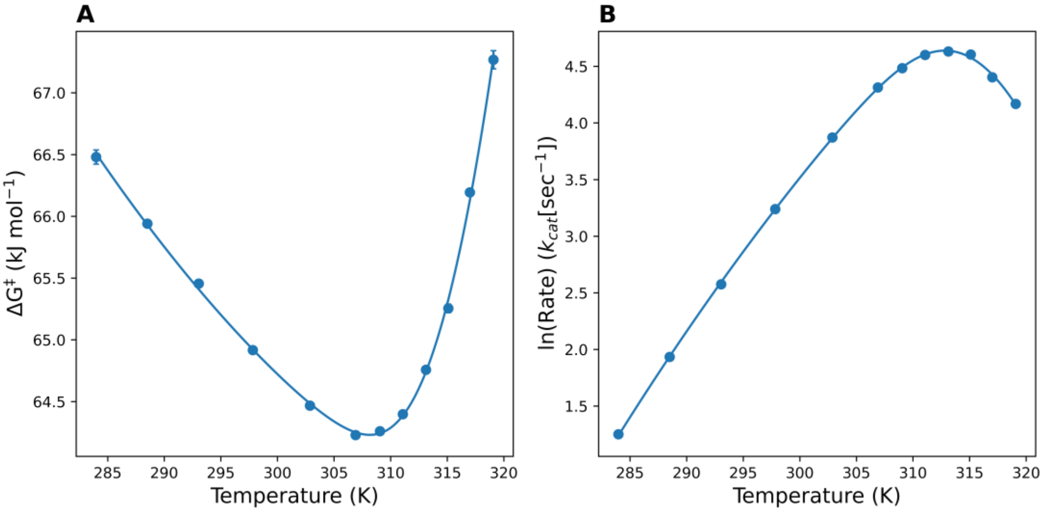


Figure 4.8. Results of temperature characterisation of MalL T150R to the MMRT 2.0 (Two-state, sigmoidal heat capacity) equation. Points are the average of three replicates, and where visible error bars are the standard deviation of the three replicates. A) Fit of MMRT 2.0 to Temperature vs ΔG‡. B) Fit of MMRT 2.0 to Temperature vs ln(Rate).

The results of the fit of MalL T150R to MMRT 2.0 is given in Table 4.4. Results are consistent with a fit to MMRT 2.0 with $R^2 = 0.9976$.

Table 4.4. MMRT 2.0 fit parameters of MalL T150R with standard error

Parameter	MalL WT	MalL T150R
$\Delta H^{\ddagger}_{T_0}\ (kJ\ mol^{-1})$	96.6 ± 2.3	109.2 ± 0.9
$\Delta S^{\ddagger}_{T_0}\ (J\ mol^{-1} K^{-1})$	102.3 ± 8.2	150.5 ± 3.2
$\Delta\Delta H^{\ddagger}\ (kJ\ mol^{-1})$	186.7 ± 45.5	380.7 ± 105.4
$\Delta C^{\ddagger}_{P,lowT}\ (kJ\ mol^{-1} K^{-1})$	0.8 ± 0.2	-0.8 ± 0.2
$\Delta C^{\ddagger}_{P,highT}\ (kJ\ mol^{-1} K^{-1})$	-28.1 ± 6.3	-21.9 ± 3.2
$T_C\ (K)$	313.5 ± 2.7	309.1 ± 1.1
$\Delta C^{\ddagger}_{P,lowT}$ Fit Range	1-6	1-6
$T_0\ (K)$	278.15	278.15
AICc	167.6	104.8

4.5 Characterisation of MalL V376R

4.5.1 Expression of MalL V376R

Expression of MalL V376R was as described above in Section 4.4.1. Pure soluble protein was obtained by immobilised metal affinity chromatography (IMAC), followed by size exclusion chromatography (SEC) (Section 2.1.3) (Figure 4.9). The purity of obtained protein was confirmed by polyacrylamide gel electrophoresis (Section 2.1.4) (Figure 4.10).

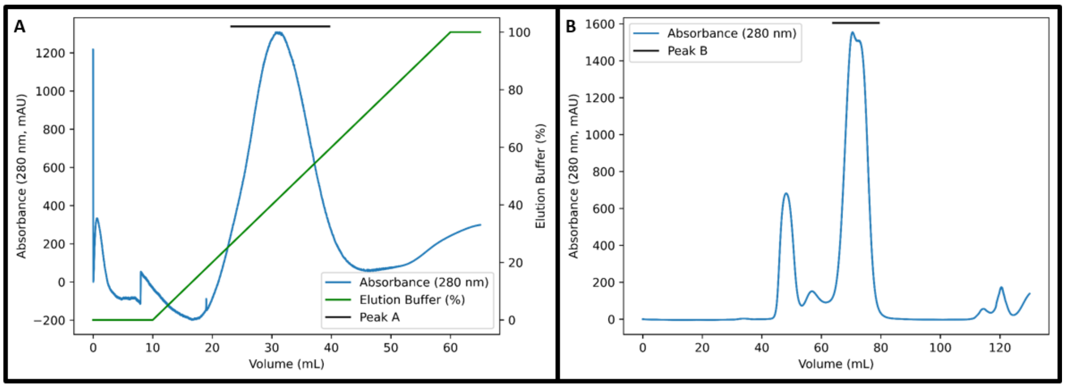


Figure 4.9. Purification of MalL V376R. A) Elution chromatogram (Absorbance, 280 nm) of MalL D492R from IMAC column. Black bar represents eluted protein. B) Elution chromatogram (Absorbance, 280 nm) of MalL D492R from SEC column. Black bar represents eluted protein.

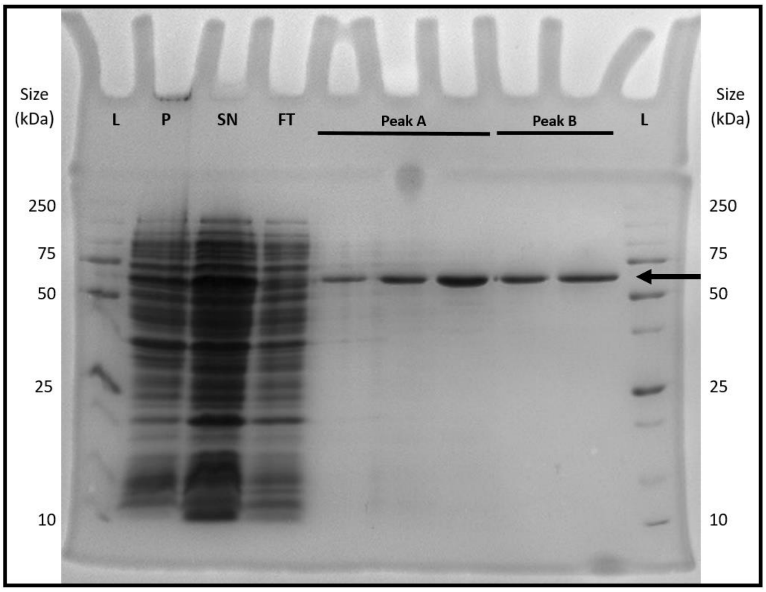


Figure 4.10. PAGE of MalL V376R purification. L = protein ladder, P = Insoluble pellet, SN = Supernatant containing soluble protein loaded onto IMAC column, FT = Supernatant eluted from IMAC column minus bound target protein. Peak A and peak B represent peaks from IMAC and SEC chromatograms, respectively. Arrow indicates target protein.

MalL V376R was purified with Ni IMAC and eluted between approximately 20-70% elution buffer (Section 2.1.1). MalL V376R was further purified by SEC and eluted between 60-80 mL. Purity was confirmed by SDS-PAGE. An intense band between 50 and 75 kDa is consistent with the expected size of 69.5 kDa.

4.5.2 Melting temperature

The melting temperature (T_m) of MalL V376R was assessed by a SYPRO orange thermal shift assay at 47.1 ± 0.2°C, compared to MalL wildtype at 50.6 ± 0.1°C.

4.5.3 Kinetic characterisation of MalL V376R

MalL V376R was characterised using established protocols for MalL variants with the substrate *p*-nitrophenyl-α-D-glucopyranoside (PNG) (Section 2.2).

4.5.3.1 Michaelis-Menten

MalL V376R was characterised using a Michaelis-Menten model to find the parameters k_{cat}, K_M, and K_i (Figure 4.6). The Michaelis-Menten assay was performed at 25°C.

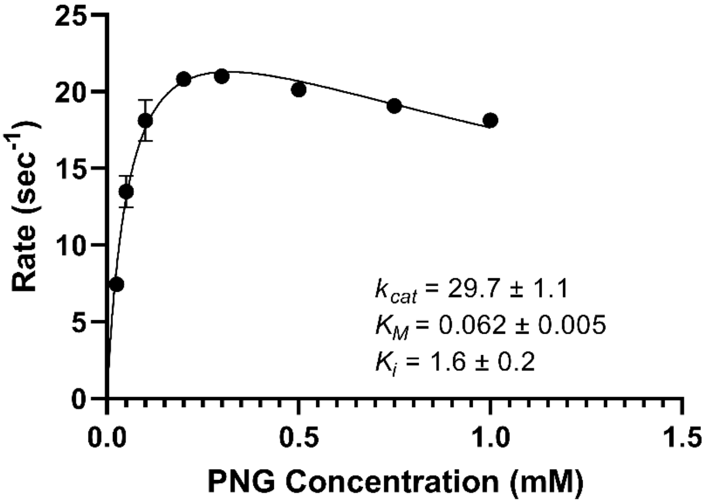


Figure 4.11. Michaelis-Menten kinetics of MalL V376R. Assay was performed at 25°C with the substrate p-nitrophenyl-α-D-glucopyranoside (PNG). Data are fit to a substrate-inhibition model. Data are plotted as the average of at least three replicates, with error bars, where visible, being the standard deviation of these replicates.

The results of the Michaelis-Menten fit are given in Table 4.5. MalL V376R has a k_{cat} of 29.7 ± 1.1 sec^{-1} and a K_M of 0.062 ± 0.005 mM. MalL V376R was fit with a substrate inhibition model with a K_i = 1.6 ± 0.2 mM. Subsequent assays were performed at 0.5 mM, approximately 10 X K_M.

Table 4.5. Michaelis-Menten parameters for MalL V376R at 25°C with standard error.

Parameter	MalL V376R
k_{cat} (sec^{-1})	29.7 ± 1.1
K_M (mM)	0.062 ± 0.005
K_i (mM)	1.6 ± 0.2

4.5.3.2 Temperature characterisation

The temperature dependence of MalL V376R was characterised and fit with both MMRT 1.5 (Figure 4.12) and MMRT 2.0 (Figure 4.13).

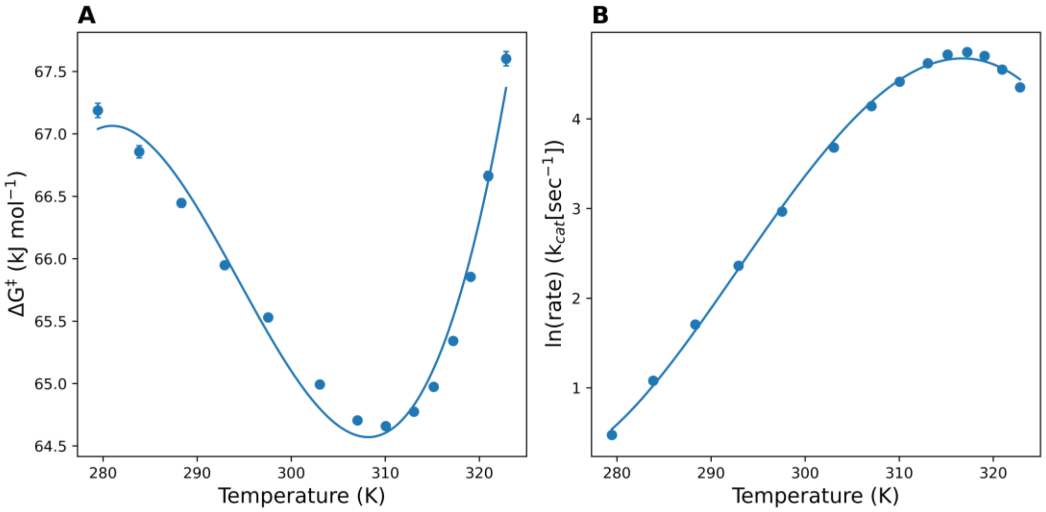


Figure 4.12. Result of temperature characterisation of MalL V376R fit to the MMRT 1.5 (linear activation heat capacity) equation. Points are the average of at least three replicates, and error bars where visible are the standard deviation of the replicates. A) Fit of MMRT 1.5 to Temperature vs ΔG‡. B) Fit of MMRT 1.5 to Temperature vs ln(Rate).

The results of the fit of MalL V376R to MMRT 1.5 is given in Table 4.6. Results are consistent with a MMRT 1.5 fit, with $R^2 = 0.9706$.

Table 4.6. MMRT 1.5 fit parameters of MalL V376R with standard error

Parameter	MalL WT	MalL V376R
$\Delta H^{\ddagger}_{T_0}\ (kJ\ mol^{-1})$	19.0 ± 16.0	48.9 ± 10.2
$\Delta S^{\ddagger}_{T_0}\ (J\ mol^{-1}K^{-1})$	-174.0 ± 56.5	-65.0 ± 36.2
$\Delta C^{\ddagger}_{P,0}\ (kJ\ mol^{-1}K^{-1})$	210 ± 18.0	128.4 ± 13.3
$\Delta C^{\ddagger}_{P,T_0}\ (kJ\ mol^{-1}K^{-1})$	12.5 ± 1.5	7.1 ± 1.0
$m\ (J\ mol^{-1}K^{-2})$	-709.9 ± 59.3	-436.4 ± 44.0
$T_0\ (K)$	278.15	278.15
$T_{opt}\ (K)$	315.1	316.7
$T_{inf}\ (K)$	307.5	307.3
AICc	191.2	151.4

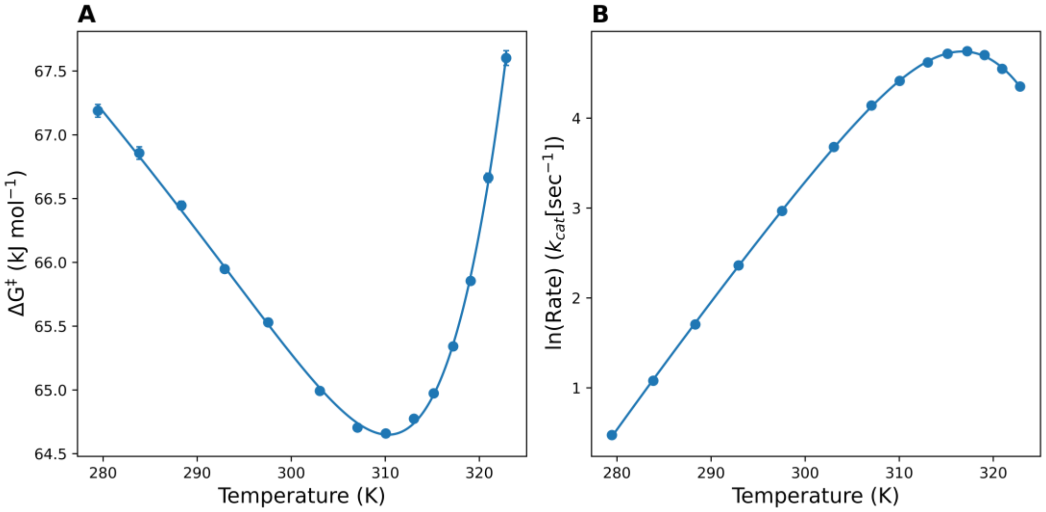


Figure 4.13. Results of temperature characterisation of MalL V376R to the MMRT 2.0 (Two-state, sigmoidal heat capacity) equation. Points are the average of three replicates, and where visible error bars are the standard deviation of the three replicates. A) Fit of MMRT 2.0 to Temperature vs ΔG‡. B) Fit of MMRT 2.0 to Temperature vs ln(Rate).

The results of the fit of MalL V376R to MMRT 2.0 is given in Table 4.7. Results are consistent with a fit to MMRT 2.0 with $R^2 = 0.9987$.

Table 4.7. MMRT 2.0 fit parameters of MalL V376R with standard error

Parameter	MalL WT	MalL V376R
$\Delta H^{\ddagger}_{T_0}\ (kJ\ mol^{-1})$	96.6 ± 2.3	92.0 ± 0.7
$\Delta S^{\ddagger}_{T_0}\ (J\ mol^{-1}K^{-1})$	102.3 ± 8.2	88.7 ± 2.5
$\Delta\Delta H^{\ddagger}\ (kJ\ mol^{-1})$	186.7 ± 45.5	170.0 ± 19.4
$\Delta C^{\ddagger}_{P,lowT}\ (kJ\ mol^{-1}K^{-1})$	0.8 ± 0.2	0.2 ± 0.2
$\Delta C^{\ddagger}_{P,highT}\ (kJ\ mol^{-1}K^{-1})$	-28.1 ± 6.3	-28.4 ± 7.4
$T_C\ (K)$	313.5 ± 2.7	316.1 ± 2.6
$\Delta C^{\ddagger}_{P,lowT}$ **Fit Range**	1-6	1-6
$T_0\ (K)$	278.15	278.15
AICc	167.6	110.1

4.6 Characterisation of MalL D492R

4.6.1 Expression of MalL D492R

Expression of MalL D492R was as described above in Section 4.4.1. Pure soluble protein was obtained by immobilised metal affinity chromatography (IMAC), followed by size exclusion chromatography (SEC) (Section 2.1.3) (Figure 4.14). The purity of obtained protein was confirmed by polyacrylamide gel electrophoresis (Section 2.1.4) (Figure 4.15).

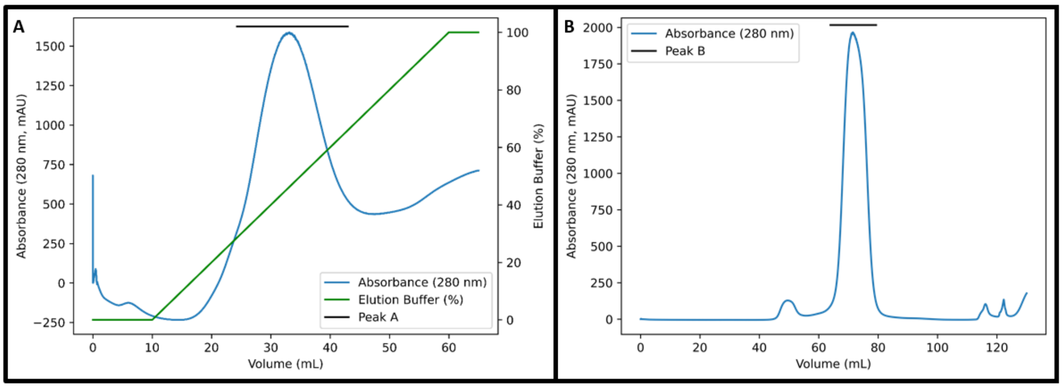


Figure 4.14. Purification of MalL D492R. A) Elution chromatogram (Absorbance, 280 nm) of MalL D492R from IMAC. Black bar represents eluted protein. B) Elution chromatogram (Absorbance, 280 nm) of MalL D492R from SEC. Black bar represents eluted protein.

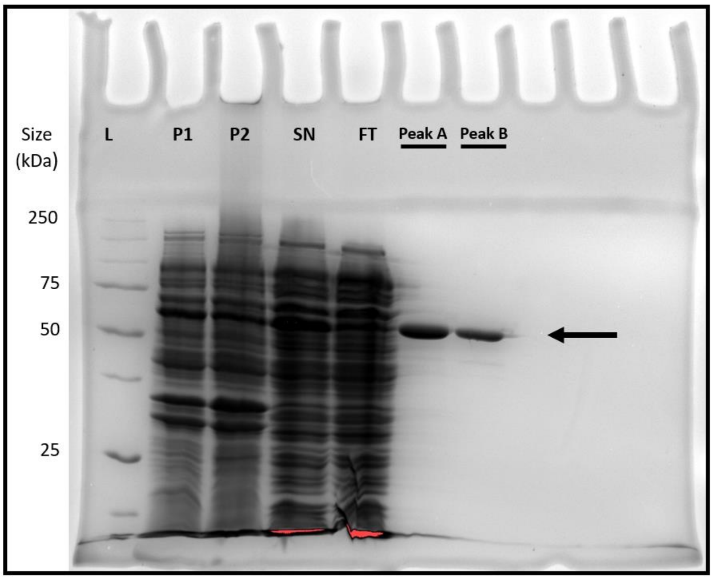


Figure 4.15. SDS-PAGE of MalL D492R purification. L = protein ladder, P1/P2 = Insoluble pellet, SN = Supernatant containing soluble protein loaded onto IMAC column, FT = Supernatant eluted from IMAC column minus bound target protein. Peak A and peak B represent peaks from IMAC and SEC chromatograms, respectively. Arrow indicates target protein.

MalL D492R eluted from the IMAC column between approximately 20-60% elution buffer (Section 2.1.1). MalL D492R was further purified by SEC and eluted between 60-85 mL. SDS-PAGE indicates a high degree of purity with minimal contamination. A band between 50 and 75 kDa is consistent with the expected size of MalL being 69.5 kDa.

4.6.2 Melting temperature

The melting temperature (T_m) of MalL D492R was assessed by a SYPRO orange thermal shift assay at 46.9 ± 0.3°C, compared to MalL wildtype at 50.6 ± 0.1°C.

4.6.3 Kinetic characterisation of MalL D492R

Kinetic assays were performed using established protocols for MalL variants using the substrate *p*-nitrophenyl-α-D-glucopyranoside (PNG), as described in Section 2.2.

4.6.3.1 Michaelis-Menten

MalL D492R was characterised using a Michaelis-Menten model to find the parameters k_{cat}, K_M, and K_i (Figure 4.16). The Michaelis-Menten assay was performed at 25°C.

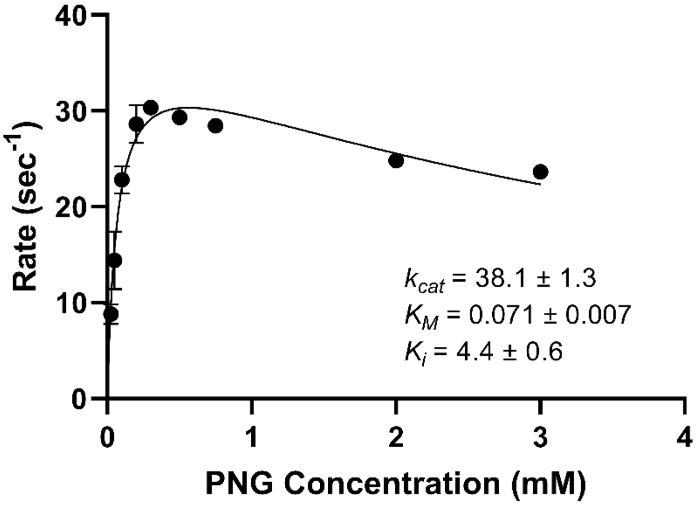


Figure 4.16. Michaelis-Menten kinetics of MalL D492R. Assay was performed at 25°C with the substrate p-nitrophenyl-α-D-glucopyranoside (PNG). Data are fit to a substrate-inhibition model. Data are plotted as the average of at least three replicates, with error bars, where visible, being the standard deviation of these replicates.

The fit parameters for the Michaelis-Menten fit is given in Table 4.8. Data were fit with a substrate inhibition model with a K_i = 4.4 ± 0.6 mM. MalL D492R has a K_M of 0.071 ± 0.007 mM and a k_{cat} of 38.1 ± 1.3 sec^{-1}. Subsequent assays were performed at 0.6 mM, approximately 10 X K_M, while avoiding substrate inhibition effects.

Table 4.8. Michaelis-Menten parameters for MalL D492R at 25°C with standard error

Parameter	MalL D492R
k_{cat} (sec^{-1})	38.1 ± 1.3
K_M (mM)	0.071 ± 0.007
K_i (mM)	4.4 ± 0.6

4.6.3.2 Temperature characterisation

The temperature dependence of MalL D492R was characterised and fit with both MMRT 1.5 (Linear activation heat capacity) and MMRT 2.0 (Sigmoidal activation heat capacity).

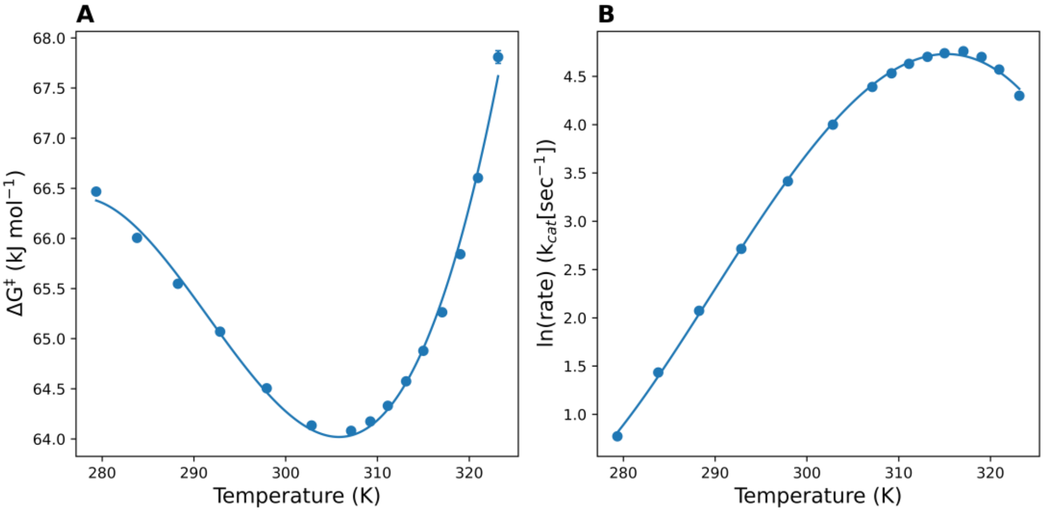


Figure 4.17. Result of temperature characterisation of MalL D492R fit to the MMRT 1.5 (linear activation heat capacity) equation. Points are the average of at least three replicates, and error bars where visible are the standard deviation of the replicates. A) Fit of MMRT 1.5 to Temperature vs ΔG‡. B) Fit of MMRT 1.5 to Temperature vs ln(Rate).

Results from the fit of MMRT1.5 to the data for MalL D492R is given in Table 4.9, alongside MalL wildtype for comparison. The fit is consistent with MMRT 1.5 with R^2 = 0.9917. The reference temperature (T_0) is 278.15 K for both datasets.

Table 4.9. MMRT 1.5 fit parameters of MalL D492R with standard error

Parameter	MalL WT	MalL D492R
$\Delta H^{\ddagger}_{T_0}\ (kJ\ mol^{-1})$	19.0 ± 16.0	69.2 ± 5.9
$\Delta S^{\ddagger}_{T_0}\ (J\ mol^{-1}K^{-1})$	-174.0 ± 56.5	10.0 ± 20.7
$\Delta C^{\ddagger}_{P,0}\ (kJ\ mol^{-1}K^{-1})$	210 ± 18.0	108.4 ± 7.4
$\Delta C^{\ddagger}_{P,T_0}\ (kJ\ mol^{-1}K^{-1})$	12.5 ± 1.5	5.0 ± 0.6
$m\ (J\ mol^{-1}K^{-2})$	-709.9 ± 59.3	-371.9 ± 24.6
$T_0\ (K)$	278.15	278.15
$T_{opt}\ (K)$	315.1	315.2
$T_{inf}\ (K)$	307.5	305.7
AICc	191.2	145.4

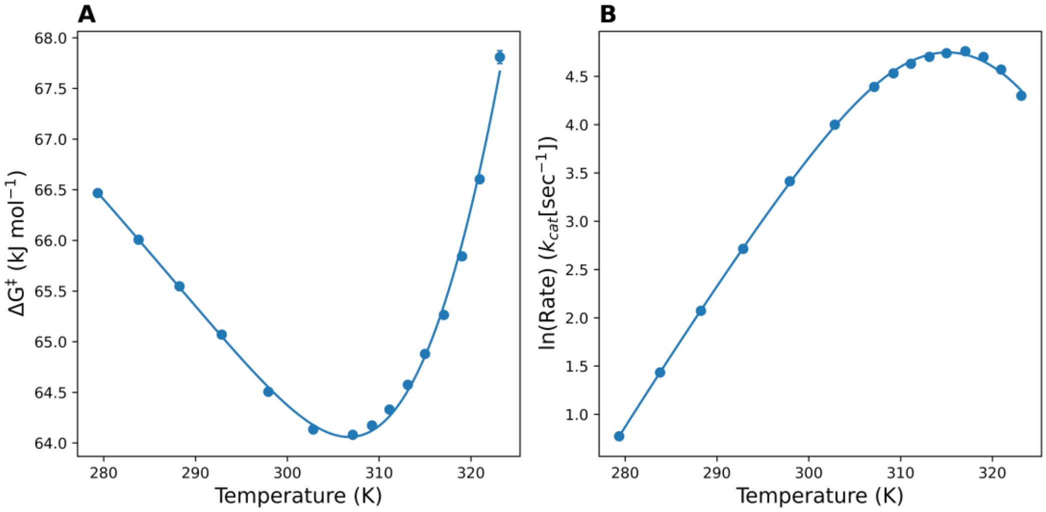


Figure 4.18. Results of temperature characterisation of MalL D492R to the MMRT 2.0 (Two-state, sigmoidal heat capacity) equation. Points are the average of three replicates, and where visible error bars are the standard deviation of the three replicates. A) Fit of MMRT 2.0 to Temperature vs ΔG‡. B) Fit of MMRT 2.0 to Temperature vs ln(Rate).

Fit parameters for the fit of the data to the MMRT 2.0 equation is given in Table 4.10. Results are consistent with a fit to MMRT 2.0 with $R^2 = 0.9937$. The result for T_C did not converge and was fixed at 307 K.

Table 4.10. MMRT 2.0 fit parameters of MalL D492R with standard error

Parameter	MalL WT	MalL D492R
$\Delta H^{\ddagger}_{T_0}\ (kJ\ mol^{-1})$	96.6 ± 2.3	95.2 ± 1.8
$\Delta S^{\ddagger}_{T_0}\ (J\ mol^{-1}K^{-1})$	102.3 ± 8.2	102.9 ± 6.2
$\Delta\Delta H^{\ddagger}\ (kJ\ mol^{-1})$	186.7 ± 45.5	167.3 ± 16.9
$\Delta C^{\ddagger}_{P,lowT}\ (kJ\ mol^{-1}K^{-1})$	0.8 ± 0.2	0.1 ± 0.1
$\Delta C^{\ddagger}_{P,highT}\ (kJ\ mol^{-1}K^{-1})$	-28.1 ± 6.3	-11.7 ± 0.2
$T_C\ (K)$	313.5 ± 2.7	307
$\Delta C^{\ddagger}_{P,lowT}$ **Fit Range**	1-6	1-5
$T_0\ (K)$	278.15	278.15
AICc	167.6	138.1

4.6.4 Dynamic characterisation of MalL D492R

MalL D492R was further characterised by SAXS and REES spectroscopy. D492R was chosen for further characterisation as the kinetics indicated it had the greatest rate acceleration at low temperatures.

4.6.4.1 SAXS

SAXS profiles were collected and analysed for MalL wildtype and MalL D492R. Kratky and Guinier plots are given in Figure 4.19 and Figure 4.20, respectively. The distance distribution function is given in Figure 4.21. Results for the Guinier, distance distribution and Porod volume analyses are given in Table 4.11 and Table 4.12.

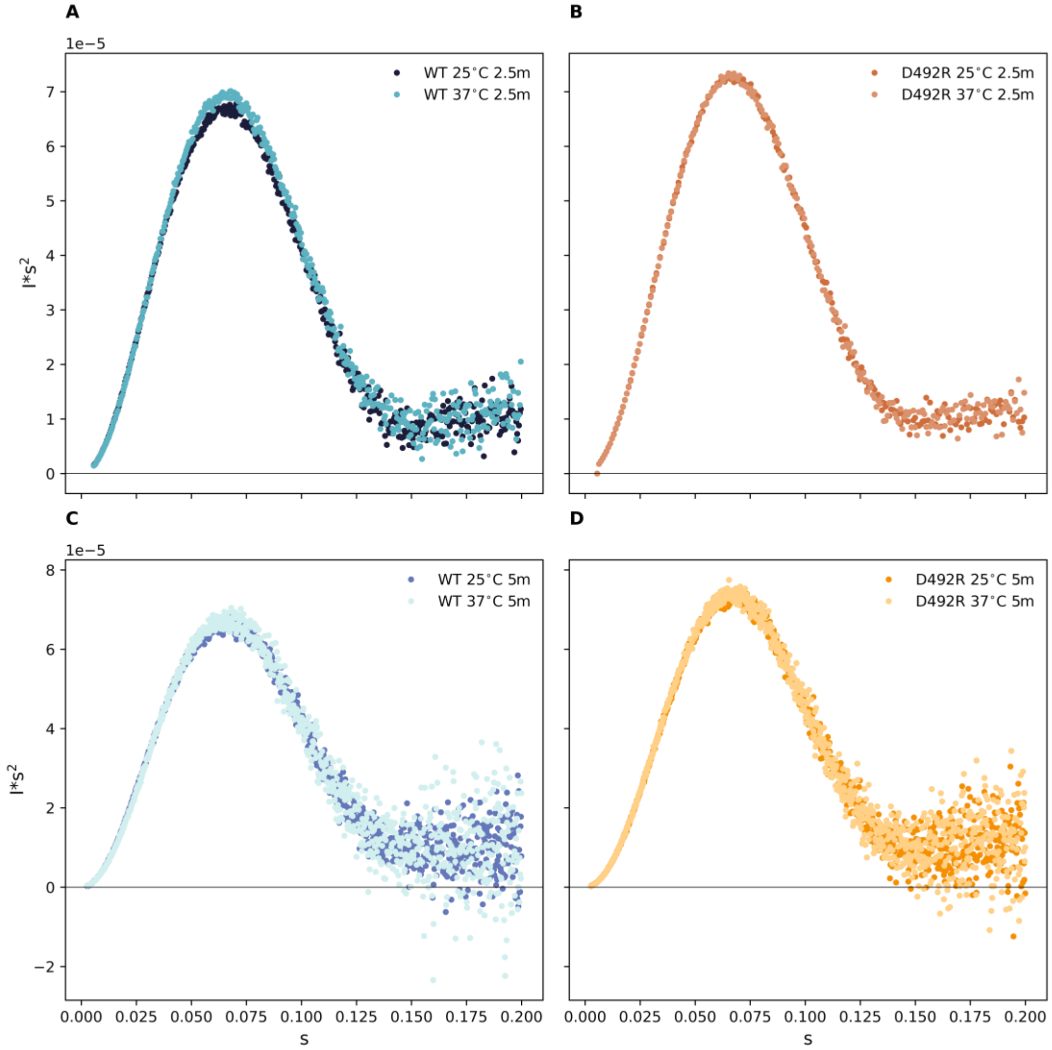


Figure 4.19. Kratky plots of MalL wildtype and MalL D492R at 25°C and 37°C and camera distances of 2.5 m and 5 m.

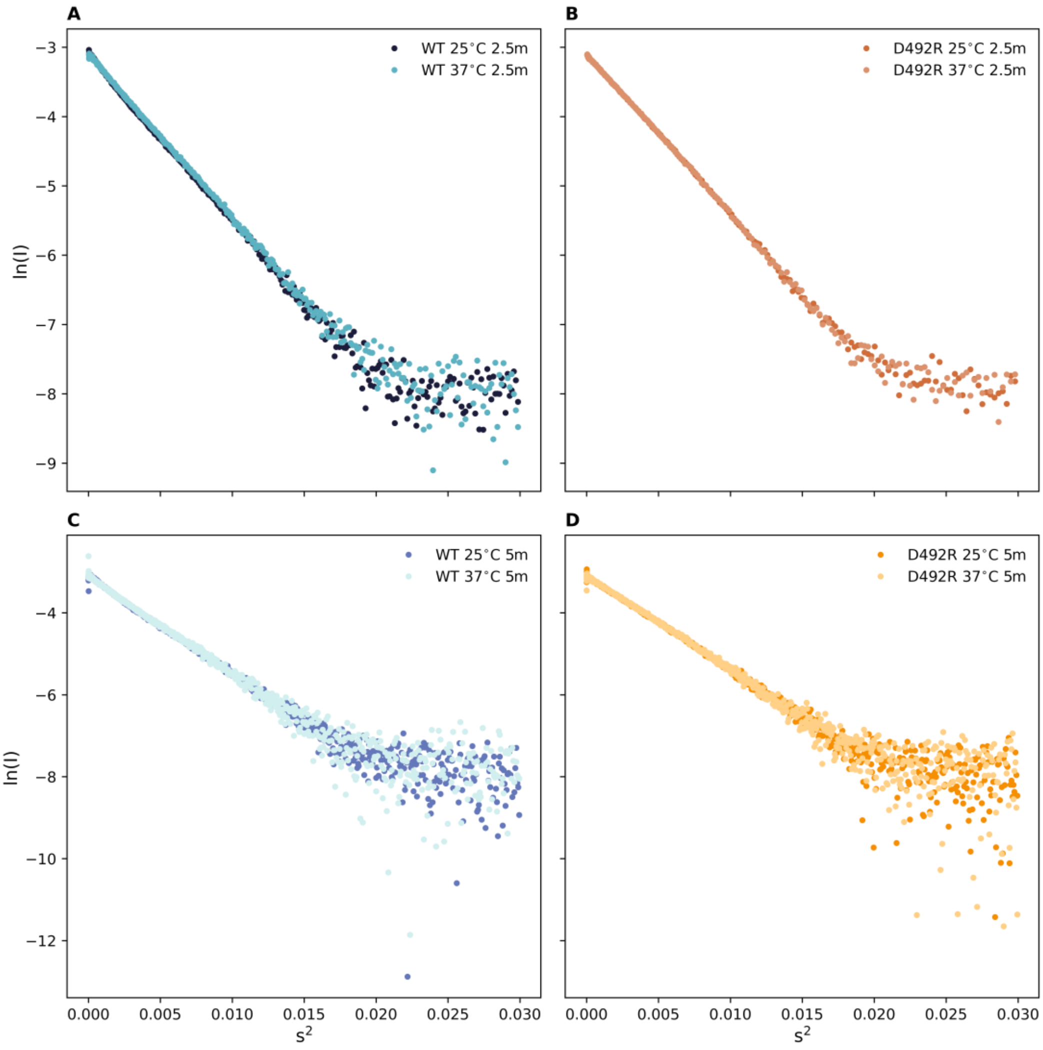


Figure 4.20. Guinier plots of MalL wildtype and MalL D492R at 25°C and 37°C and camera distances of 2.5 m and 5 m.

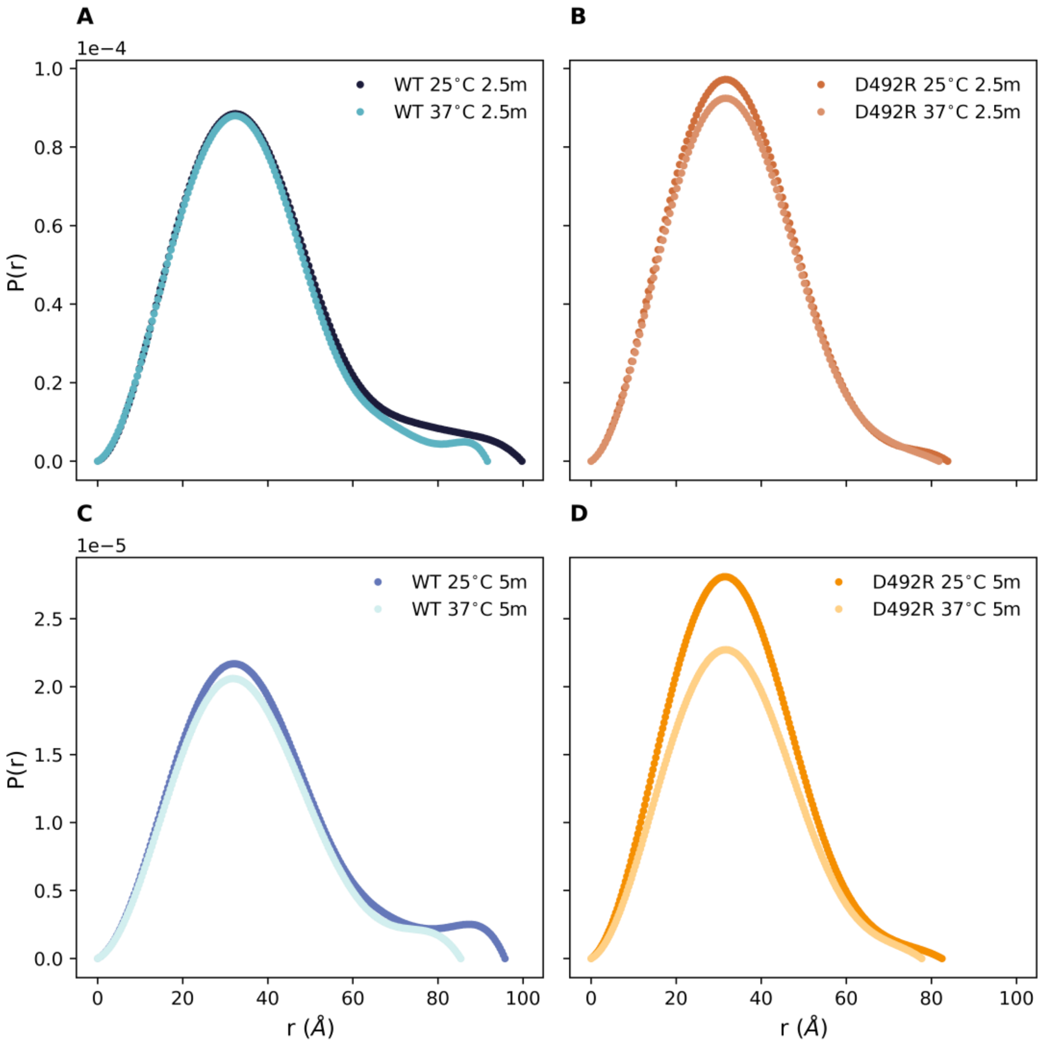


Figure 4.21. Distance distribution plots of MalL wildtype and MalL D492R at 25°C and 37°C and camera distances of 2.5 m and 5 m.

Table 4.11. SAXS results for MalL wildtype

	Protein	**MalL WT**			
	Distance (m)	**2.5**		**5**	
	Temperature (°C)	**25**	**37**	**25**	**37**
Guinier Analysis	**I**	0.044	0.042	0.011	0.01
	R_g	28.61	27.36	28.98	27.78
Distance Distribution Analysis	**R0/I0**	29.17	27.48	29.48	27.57
	p(r) R0/I0	29.18	27.48	29.49	27.57
	Porod Volume	114696	110918	114381	105062
	D_{max}	99.73	91.59	95.72	85.33
Porod Volume Analysis	**Porod Volume**	90659	81660	62614	36938

Table 4.12. SAXS results for MalL D492R

	Protein	**MalL D492R**			
	Distance (m)	**2.5**		**5**	
	Temperature (°C)	**25**	**37**	**25**	**37**
Guinier Analysis	**I**	0.044	0.042	0.013	0.01
	R_g	25.98	25.97	25.91	25.91
Distance Distribution Analysis	**R0/I0**	25.8	25.79	25.7	25.58
	p(r) R0/I0	25.8	25.79	25.7	25.58
	Porod Volume	97412	96322	99264	97370
	D_{max}	83.82	81.76	82.45	77.72
Porod Volume Analysis	**Porod Volume**	84263	75268	41839	31537

4.6.4.2 REES

REES curves were collected and analysed for MalL wildtype and D492R at 25°C and 37°C (Figure 4.22). Data were fit with the REES thermodynamic model (Section 1.7) (Table 4.13).

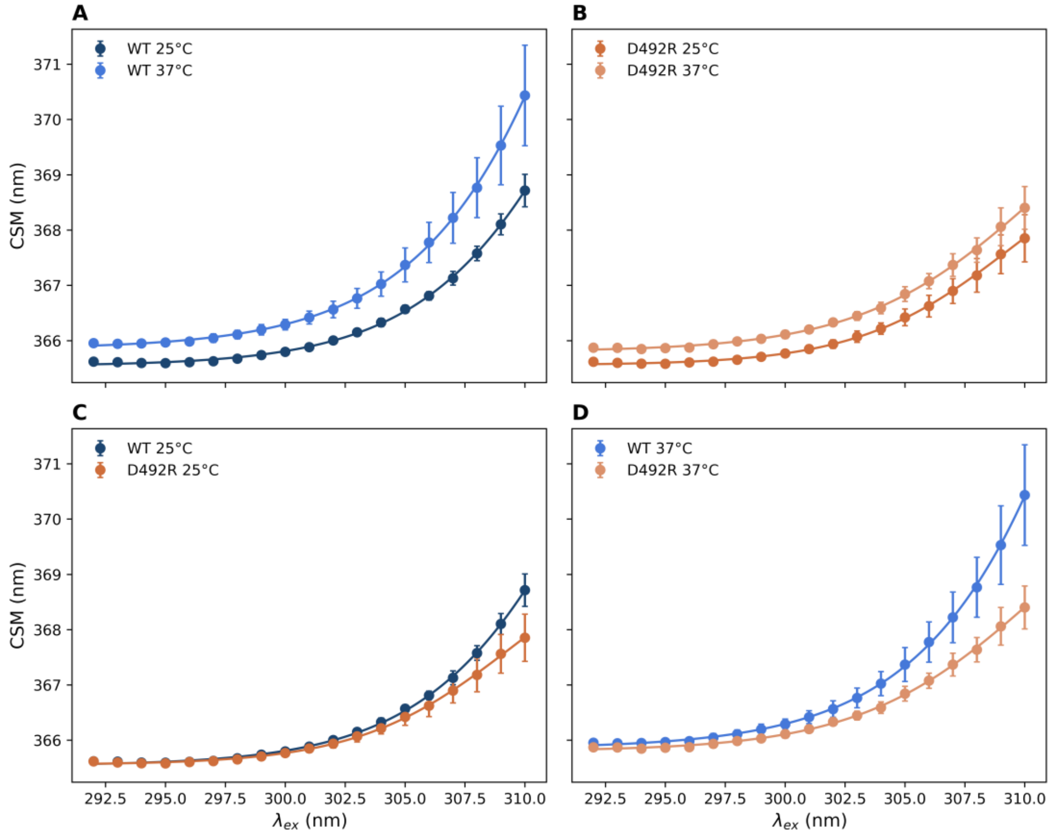


Figure 4.22. REES fit of MalL wildtype versus D492R at 25°C and 37°C. Error bars where visible are the standard error of three replicates. Data were fit with the REES equation (Equation 1.35).

Table 4.13. REES fit of MalL wildtype versus D492R with standard error

Parameter	WT 25°C	WT 37°C	D492R 25°C	D492R 37°C
$CSM(\lambda_{Ex}^{FC})$ (nm)	365.5 ± 0.02	365.8 ± 0.02	365.6 ± 0.01	365.8 ± 0.02
$CSM(\lambda_{Ex}^{\mathrm{R}})$ (nm)	378.7 ± 3.5	394.3 ± 9.2	369.8 ± 0.3	371.5 ± 0.6
ΔG_m $(mJ\ mol^{-1}nm^{-1})$	7.6 ± 0.4	6.5 ± 0.3	8.7 ± .04	7.1 ± 0.4
$\lambda_{Ex}^{50\%}$ (nm)	314.2 ± 1.4	316.7 ± 1.8	309.5 ± 0.5	310.6 ± 0.7

4.7 Characterisation of MalL arginine double mutant (RDM)

Following characterisation of the initial set of mutations, a further double mutant was constructed. MalL arginine double mutant (RDM) consists of two mutations: T150R and D492R. This mutant was designed to test the potential additive effects of arginine mutations in MalL.

4.7.1 Expression of MalL RDM

MalL RDM was expressed as described in Section 4.4.1. MalL RDM was purified by nickel immobilised metal affinity chromatography (IMAC) and size exclusion chromatography (SEC) (Figure 4.23). Purity was confirmed by polyacrylamide gel electrophoresis (PAGE) (Figure 4.24).

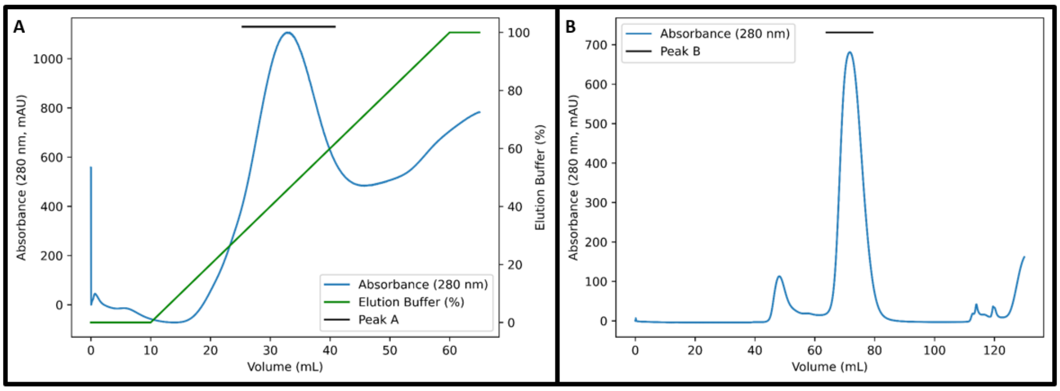


Figure 4.23. Purification of MalL RDM. A) Elution chromatogram (Absorbance, 280 nm) of MalL D492R from IMAC column. Black bar represents eluted protein. B) Elution chromatogram (Absorbance, 280 nm) of MalL D492R from SEC column. Black bar represents eluted protein.

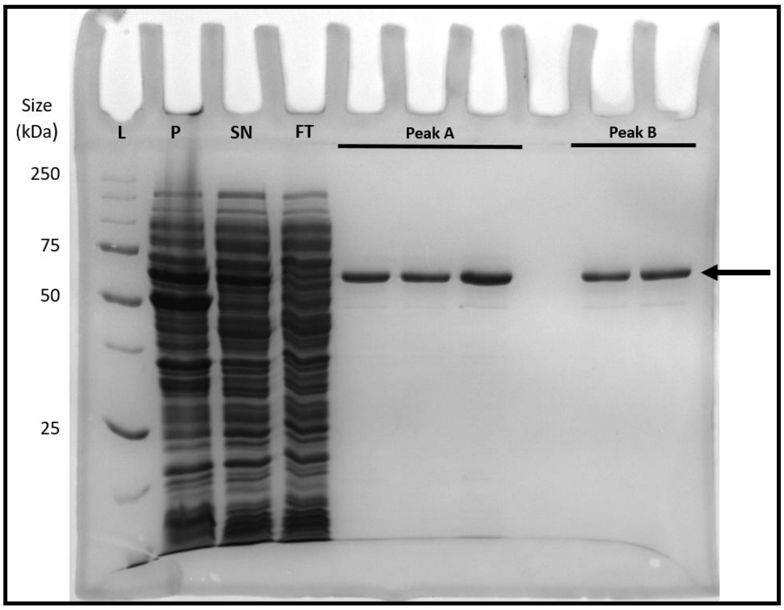


Figure 4.24. PAGE of MalL RDM purification. L = protein ladder, P = Insoluble pellet, SN = Supernatant containing soluble protein loaded onto IMAC column, FT = Supernatant eluted from IMAC column minus bound target protein. Peak A and peak B represent peaks from IMAC SEC chromatograms, respectively. Arrow indicates target protein.

MalL RDM was purified by IMAC and eluted between 20-60% elution buffer (Section 2.1.1). It was further purified by SEC and eluted between 60-80 mL. Purity was confirmed by SDS-PAGE. An intense band between 50 and 75 kDa is consistent with the expected size of 69.5 kDa.

4.7.2 Melting temperature

The melting temperature (T_m) of MalL RDM was assessed by a SYPRO orange thermal shift assay at 42.3 ± 0.5°C, compared to MalL wildtype at 50.6 ± 0.1°C.

4.7.3 Kinetic characterisation of MalL RDM

MalL RDM was characterised using established protocols for MalL variants with the substrate *p*-nitrophenyl-α-D-glucopyranoside (PNG) (Section 2.2).

4.7.3.1 Michaelis-Menten

MalL RDM was characterised using a Michaelis-Menten model to find the parameters k_{cat}, K_M, and K_i (Figure 4.6). The Michaelis-Menten assay was performed at 25°C.

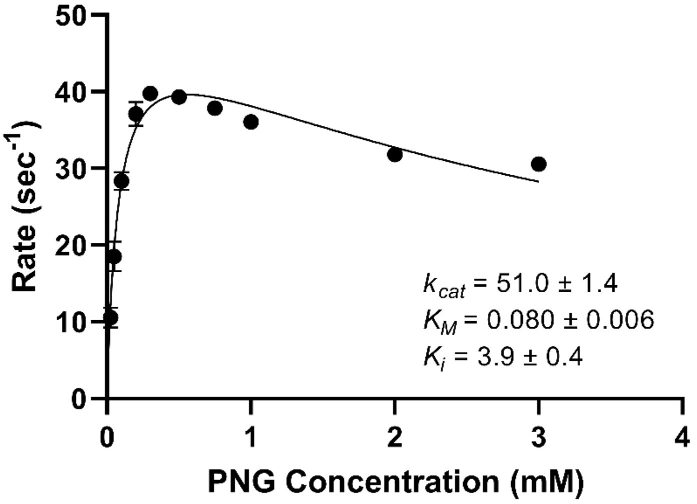


Figure 4.25. Michaelis-Menten kinetics of MalL RDM. Assay was performed at 25°C with the substrate p-nitrophenyl-α-D-glucopyranoside (PNG). Data are fit to a substrate-inhibition model. Data are plotted as the average of at least three replicates, with error bars, where visible, being the standard deviation of these replicates.

Results for the Michaelis-Menten fit of MalL RDM is given in Table 4.14. MalL RDM has a k_{cat} of 51.0 ± 1.4 sec^{-1} and a K_M of 0.080 ± 0.006 mM. MalL RDM was fit with a substrate inhibition model with a K_i = 3.9 ± 0.4 mM. Subsequent assays were performed at 0.75 mM, approximately 10 X K_M.

Table 4.14. Michaelis-Menten parameters for MalL RDM at 25°C with standard error

Parameter	MalL RDM
k_{cat} (sec^{-1})	51.0 ± 1.4
K_M (mM)	0.080 ± 0.006
K_i (mM)	3.9 ± 0.4

4.7.3.2 Temperature characterisation

The temperature dependence of MalL RDM was characterised and fit with both MMRT 1.5 (Figure 4.26) and MMRT 2.0 (Figure 4.27).

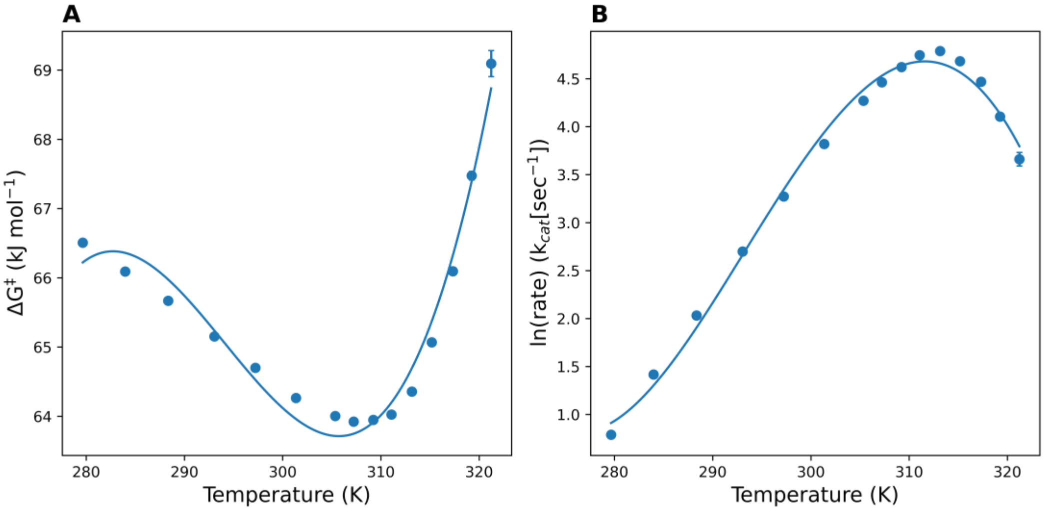


Figure 4.26. Result of temperature characterisation of MalL RDM fit to the MMRT 1.5 (linear activation heat capacity) equation. Points are the average of at least three replicates, and error bars where visible are the standard deviation of the replicates. A) Fit of MMRT 1.5 to Temperature vs ΔG‡. B) Fit of MMRT 1.5 to Temperature vs ln(Rate).

The results of the fit of MalL RDM to MMRT 1.5 are given in Table 4.15. Results are consistent with a MMRT 1.5 fit, with $R^2 = 0.9630$.

Table 4.15. MMRT 1.5 fit parameters for MalL RDM with standard error

Parameter	**MalL WT**	**MalL RDM**
$\Delta H^{\ddagger}_{T_0}$ $(kJ\ mol^{-1})$	19.0 ± 16.0	18.1 ± 18.4
$\Delta S^{\ddagger}_{T_0}$ $(J\ mol^{-1}K^{-1})$	-174.0 ± 56.5	-172.4 ± 65.2
$\Delta C^{\ddagger}_{P,0}$ $(kJ\ mol^{-1}K^{-1})$	210 ± 18.0	227.8 ± 24.9
$\Delta C^{\ddagger}_{P,T_0}$ $(kJ\ mol^{-1}K^{-1})$	12.5 ± 1.5	12.3 ± 1.9
m $(J\ mol^{-1}K^{-2})$	-709.9 ± 59.3	-774.5 ± 82.9
T_0 (K)	278.15	278.15
T_{opt} (K)	315.1	311.6
T_{inf} (K)	307.5	304.1
AICc	191.2	177.2

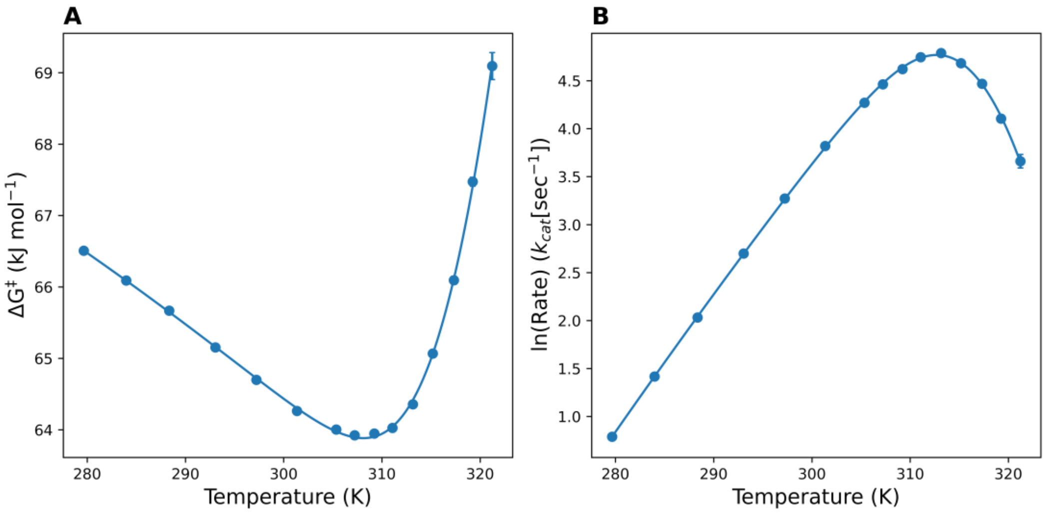


Figure 4.27. Results of temperature characterisation of MalL RDM to the MMRT 2.0 (Two-state, sigmoidal heat capacity) equation. Points are the average of three replicates, and where visible error bars are the standard deviation of the three replicates. A) Fit of MMRT 2.0 to Temperature vs ΔG‡. B) Fit of MMRT 2.0 to Temperature vs ln(Rate).

The results of the fit of MalL V376R to MMRT 2.0 is given in Table 4.16. Results are consistent with a fit to MMRT 2.0 with $R^2 = 0.9993$.

Table 4.16. MMRT 2.0 fit parameters for MalL RDM with standard error

Parameter	MalL WT	MalL RDM
$\Delta H^{\ddagger}_{T_0}\ (kJ\ mol^{-1})$	96.6 ± 2.3	92.9 ± 0.7
$\Delta S^{\ddagger}_{T_0}\ (J\ mol^{-1}K^{-1})$	102.3 ± 8.2	94.3 ± 2.4
$\Delta\Delta H^{\ddagger}\ (kJ\ mol^{-1})$	186.7 ± 45.5	235.4 ± 21.4
$\Delta C^{\ddagger}_{P,lowT}\ (kJ\ mol^{-1}K^{-1})$	0.8 ± 0.2	0.2 ± 0.1
$\Delta C^{\ddagger}_{P,highT}\ (kJ\ mol^{-1}K^{-1})$	-28.1 ± 6.3	-34.9 ± 3.7
$T_C\ (K)$	313.5 ± 2.7	311.5 ± 0.9
$\Delta C^{\ddagger}_{P,lowT}$ **Fit Range**	1-6	1-6
$T_0\ (K)$	278.15	278.15
AICc	167.6	120.2

4.8 Discussion

Overall the characterisation of these mutants reveal significant deviations from the behaviour of MalL wildtype enzyme, in contrast with MalL S536R which has very similar behaviour to wildtype. All mutants were expressed, soluble and active. All mutants had a melting temperature within 8°C of MalL wildtype (all but MalL RDM was within 4°C). MalL RDM had a melting temperature of 42.3 ± 0.5°C, compared to the wildtype enzyme at 50.6 ± 0.1°C. The arginine mutations may be additive in terms of thermal stability, as evidenced by the increased drop in the melting temperature of MalL RDM. This indicates that all mutants maintained their thermal stability at elevated temperatures. The temperature dependence of all the arginine mutants is well captured by fitting with both the MMRT 1.5 and MMRT 2.0 models, with all model fits for MMRT 1.5 having $R^2 \geq 0.96$ and MMRT 2.0 having $R^2 \geq 0.99$. The AICc for the fit of MMRT 1.5 to MalL wildtype is 191.2, and arginine mutants vary between 130.1 and 177.2. For MMRT 2.0 the AICc is 167.6 for MalL wildtype and the AICc values for the arginine mutants vary between 110.1 and 138.1. All MalL variants had a lower AICc for the fit of MMRT 2.0 compared to MMRT 1.5. There is some difference in the values of the k_{cat}, with MalL RDM having a k_{cat} of 51.0 ± 1.4 sec^{-1}, which is significantly higher than for MalL D492R (38.1 ± 1.3 sec^{-1}), which has a higher rate than MalL RDM in the full temperature characterisation (Figure 4.28). Overall, all of the arginine mutants (excluding S536R) show an increase in rate at low temperatures. This may be likened to the evolution towards psychrophily insofar as these mutants display increasing rates at low temperatures and decreasing rates at high temperatures. It is also consistent with these mutants having slightly decreased melting temperature values. (Figure 4.28). MalL D492R and RDM show a nearly two-fold increase in the rate constant at 278 K (5°C) (Figure 4.28). It is noteworthy that position 492 is ~30 Ångstroms from the active site and so this mutation is an allosteric effect.

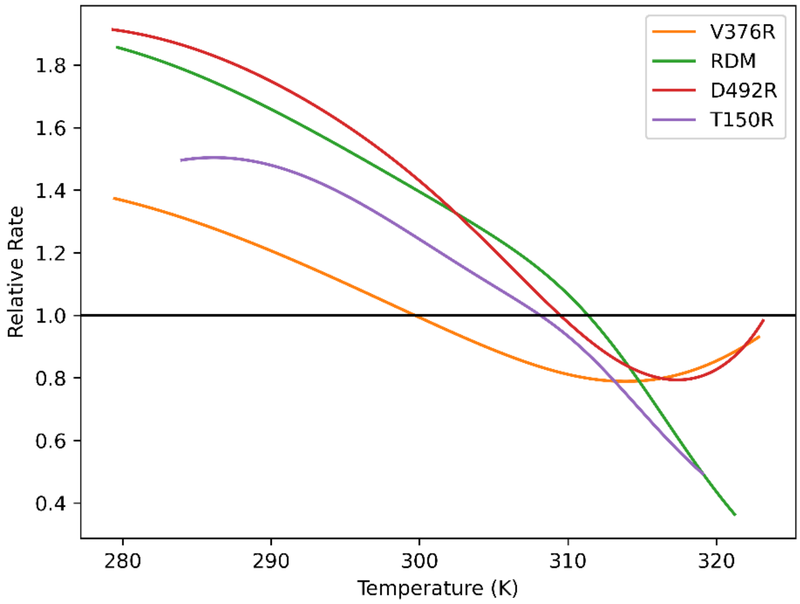


Figure 4.28. Relative rate of arginine mutants compared to MalL wildtype. Curves are the ratio of the MMRT 2.0 fits. Two distinct profiles are observed based on curvature above 310 K; MalL V376R and MalL D492R, and MalL RDM and MalL T150R.

Based on the relative rate profiles, as well as the activation Gibbs free energy and ln(Rate) plots (Figure 4.29) the four arginine mutants can be classified into two groups; group **A**, consisting of MalL V376R and D492R, and group **B**, consisting of MalL RDM and T150R. These are easily distinguished from each other by their differing curvature about 310 °C.

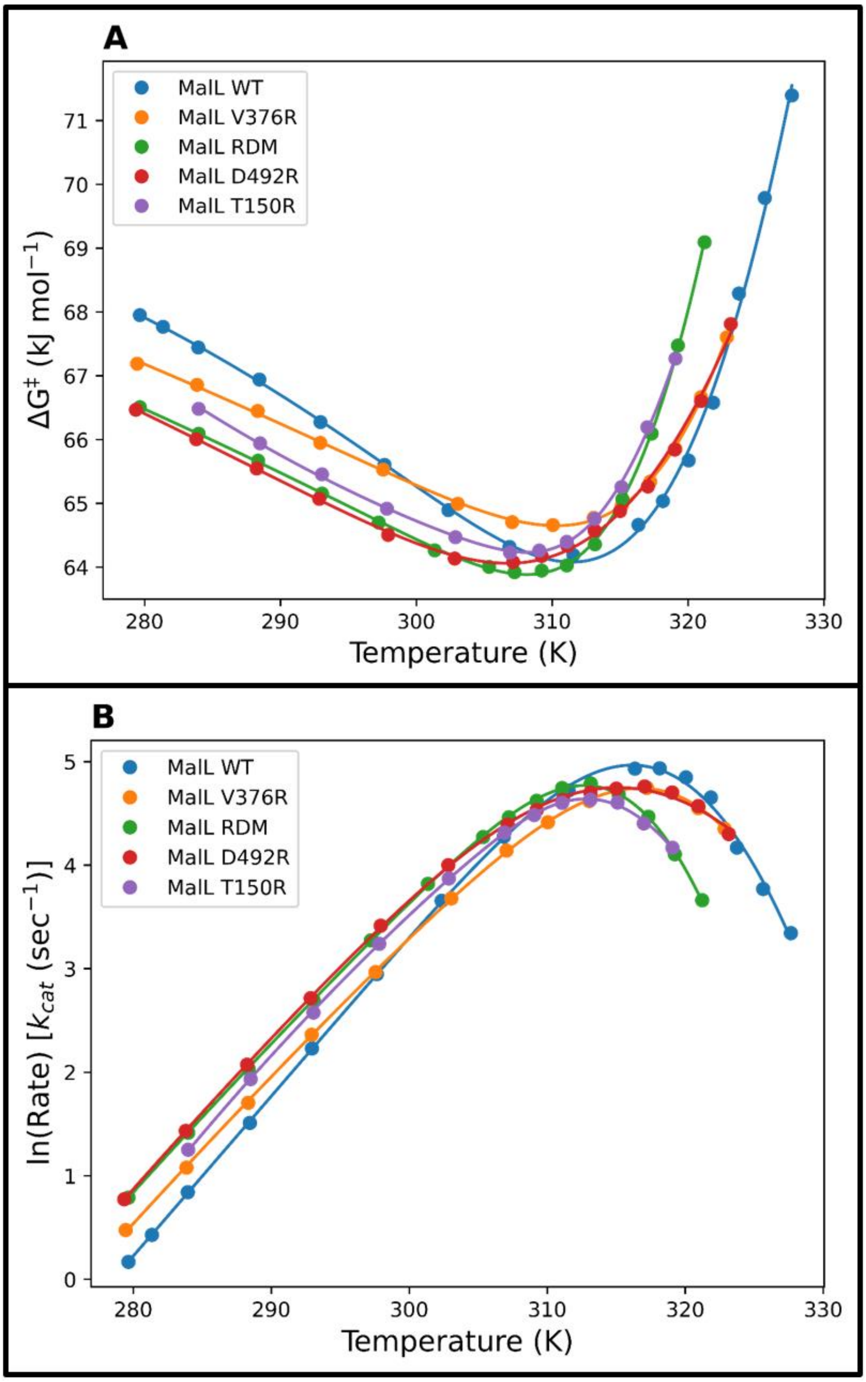


Figure 4.29. Activation Gibbs free energy (A) and Ln(Rate) (B) profiles of arginine mutants. Data are fit to the MMRT 2.0 equation.

The activation entropy is the negative of the derivative of the activation Gibbs free energy. Thus, when the slope of the activation Gibbs free energy is negative, the activation entropy is positive and favourable, and when the slope of the activation Gibbs free energy is positive, the entropy is negative and unfavourable. In the low temperature portion (approximately 278-305 K) MalL wildtype has the steepest slope and thus a more positive activation entropy. MalL T150R is also slightly steeper than the other three mutants, becoming steeper at lower temperatures. This can be seen in the temperature profile of the activation entropy (Figure 4.30). At 293 K the activation entropy for MalL wildtype is 140.9 ± 6.3 J mol^{-1} K^{-1}. The activation entropy of all of the arginine mutants is significantly decreased, with MalL V376R, MalL RDM, MalL D492R, and MalL T150R having activation entropies of 97.4 ± 1.7, 105.0 ± 2.0, 103.5 ± 3.8, and 108.7 ± 3.2 J mol^{-1} K^{-1}, respectively.

The activation entropy of the arginine mutants remains lower than MalL wildtype until around the T_{opt}, where the mutants diverge. MalL V376R and D492R (group **A**) are not as temperature dependent, and thus the entropy decreases at a reduced rate, eventually crossing above MalL wildtype. MalL RDM and T150R (group **B**) decreases earlier than MalL wildtype. The increasing entropy barrier is what results in the temperature optimum. The earlier decrease in entropy is reflected in the lower T_{opt}'s for MalL RDM and T150R, at 311.6 and 312.6 K, respectively, compared to 315.1 K for MalL wildtype. The T_{opt}'s for MalL V376R and D492R remain similar to MalL wildtype at 316.7 and 315.2 K respectively.

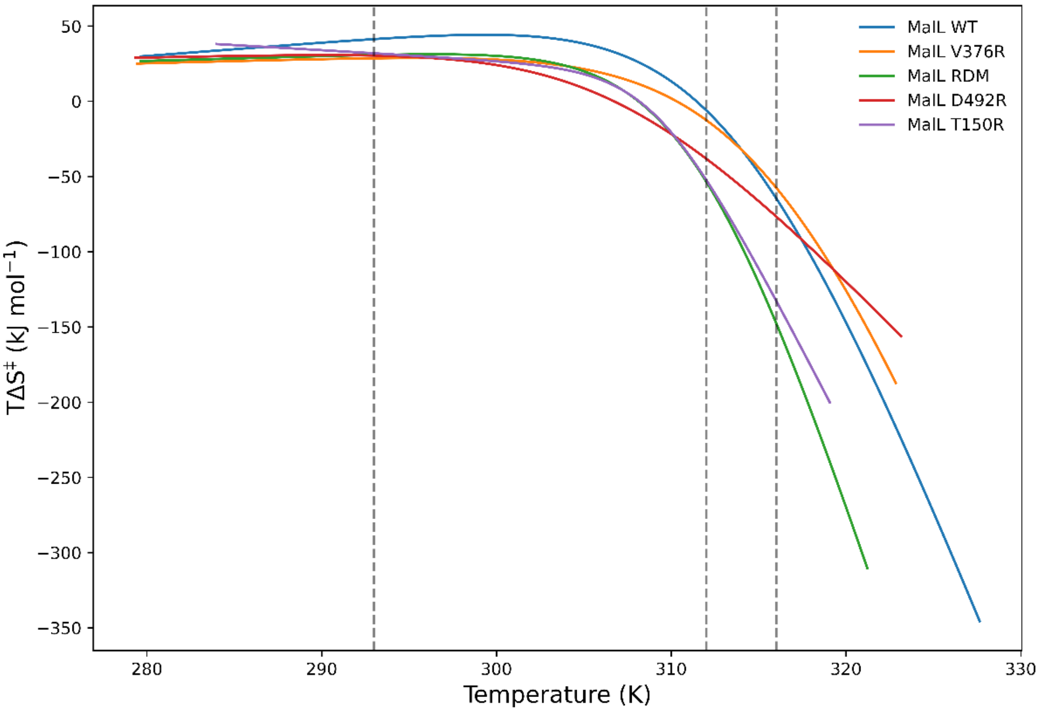


Figure 4.30. Activation entropy. A reduction in activation entropy is observed at low temperatures compared to MalL wildtype. At higher temperatures the less temperature dependent V376R and D492R decrease less steeply and cross MalL wildtype. MalL RDM and T150R decrease earlier with a corresponding lower T_{opt}. Vertical dashed lines represent 293, 312, and 316 K. Activation entropy calculated from the fit of MMRT 2.0 using Equation 1.16.

This reduction in activation entropy for the arginine mutants is unfavourable. This is offset by a corresponding decrease in the activation enthalpy at low temperatures (Figure 4.31). Such enthalpy-entropy trade-offs are common in psychrophilic enzymes (Collins & Feller, 2023). At 293 K the activation enthalpy for MalL wildtype is 107.6 ± 0.1 kJ mol^{-1}. For MalL V376R, RDM, D492R, and T150R, the activation enthalpies are 94.5 ± 0.5, 95.9 ± 0.6, 95.4 ± 1.1, and 97.3 ± 0.9 kJ mol^{-1}, respectively. As with the activation entropy, at high temperatures the arginine mutants diverge. MalL V376R and D492R (group **A**) are less temperature dependent overall with a flattened profile. MalL RDM and T150R (group **B**) have a lower T_{opt}, as well as a lowered T_C, and the activation enthalpy mirrors MalL wildtype at a slightly lower temperature. The enthalpy-entropy trade-off is maintained for MalL RDM and T150R at all temperatures, while for MalL D492R and V376R the trade-off decreases with increasing temperature.

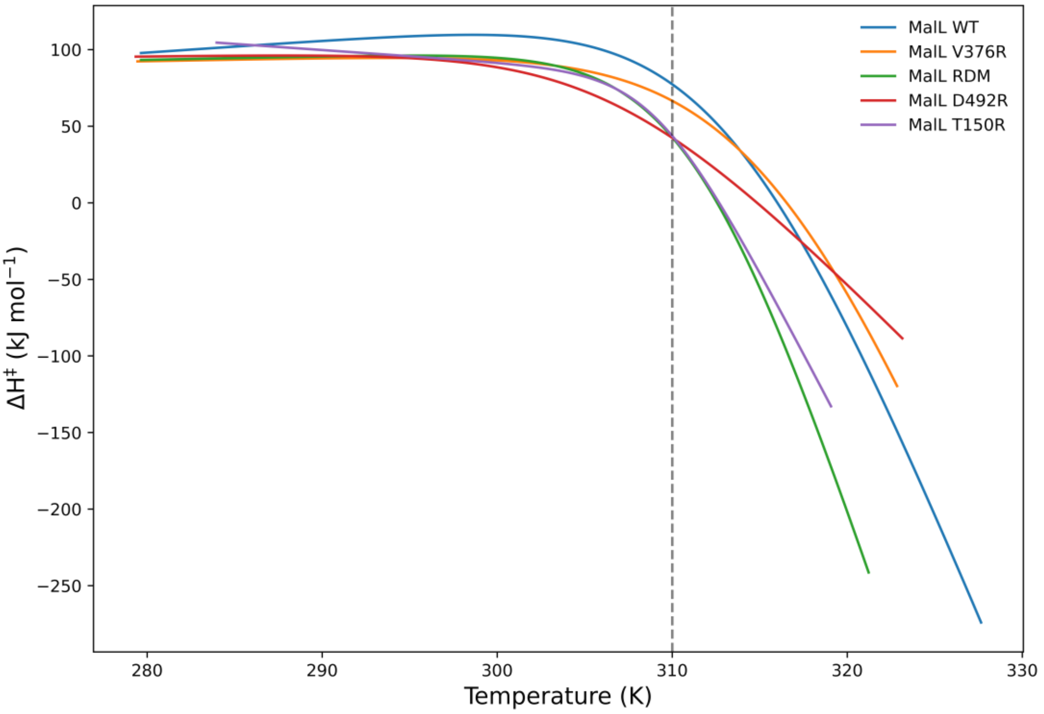


Figure 4.31. Activation enthalpy. Arginine mutants have lowered enthalpy at low-moderate temperatures (< 310 K). At high temperatures arginine mutants bifurcate into two distinct groups. Vertical dashed line represents 310 K. Activation enthalpy is calculated from the fit of MMRT 2.0 using Equation 1.15.

The activation heat capacity represents the temperature dependence of the activation entropy and activation enthalpy. While MMRT 2.0 produces a more accurate fit, it often produces high uncertainties in the values of the fit parameters (Figure 4.33). This is most probably due to at least two parameters being correlated ($\Delta C^{\ddagger}_{P,highT}$ and T_C). MMRT 1.5 is therefore useful to capture the general trends in the data (Figure 4.32).

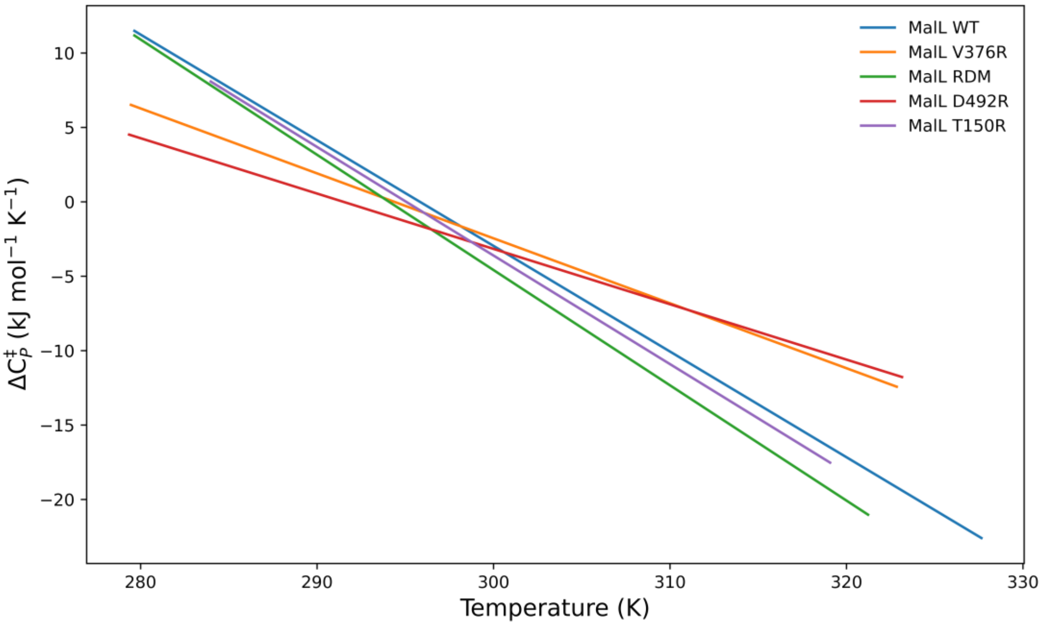


Figure 4.32. MMRT 1.5 activation heat capacity. MalL RDM and T150R have a similar activation heat capacity to MalL wildtype. MalL V376R and D492R have a less temperature dependent activation heat capacity. Activation heat capacity was calculated from the fit of MMRT 1.5 using Equation 1.21.

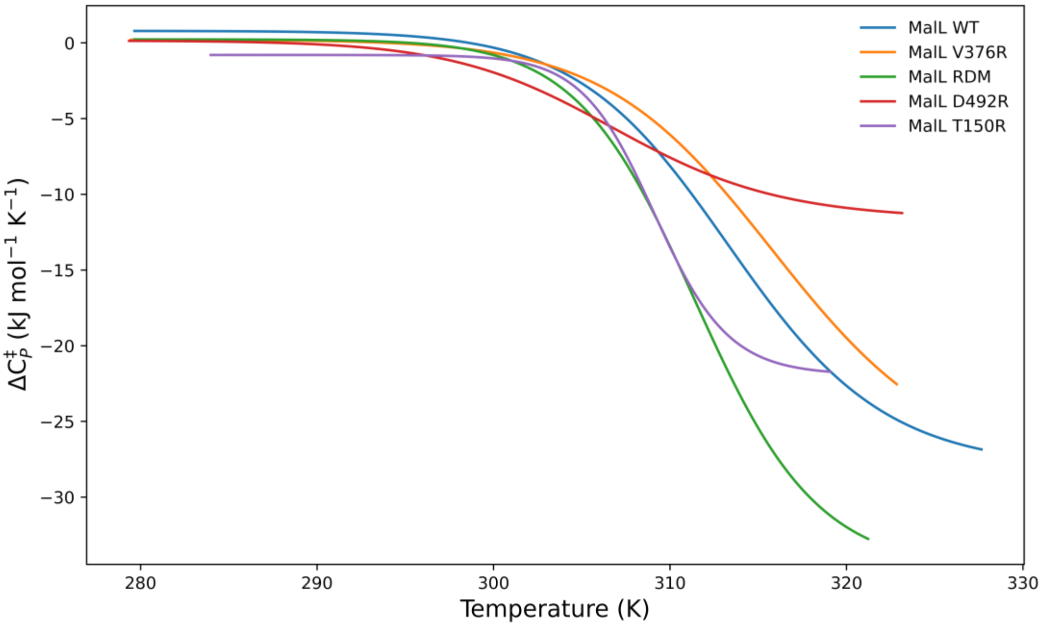


Figure 4.33. MMRT 2.0 activation heat capacity. Arginine mutants have a lower activation heat capacity at low temperatures. At higher temperatures the arginine mutants split into two groups. MalL V376R and D492R (group A) have a less temperature dependent activation, and therefore less negative heat capacity. MalL RDM and T150R (group B) have a more negative activation heat capacity. Activation heat capacity was calculated from the fit of MMRT 2.0 using Equation 1.25.

At low temperatures all of the arginine mutants have a reduction in the activation heat capacity (Figure 4.33). The two groups are again evident in the activation heat capacity profile compared to MalL wildtype. MalL V376R and D492R (group **A**) have a much reduced temperature dependence with the slope of the MMRT 1.5 fit being -371.9 ± 24.6 and -436.4 ± 44.0 J mol^{-1} K^{-1}, respectively, compared to -709.95 ± 9.3 J mol^{-1} K^{-1} for MalL wildtype. MalL RDM and T150R remain similar to MalL wildtype, with slopes of -774.5 ± 82.9 and -729.5 ± 81.0 J mol^{-1} K^{-1}, respectively. This results in a clear bifurcation at high temperatures where at 315 K for example the MMRT 2.0 activation heat capacity for MalL wildtype is -16.1 kJ mol^{-1} K^{-1}. MalL V376R and D492R (group **A**) have a less negative activation heat capacity at -12.6 and -9.8 kJ mol^{-1} K^{-1}, respectively. MalL RDM and T150R have a more negative activation heat capacity at -25.4 and -20.7 kJ mol^{-1} K^{-1}, respectively.

Arginine mutations in general are stabilising due to arginines ability to add enthalpy to the system via the formation of additional non-covalent bonding interactions. The hypothesis that serves to explain these observations is that the additional bonding interactions provided by the addition of arginine increases the preorganization energy into the TLC at the cost of an increased entropy penalty. Traditionally, psychrophily has been defined by an increase in entropy (consistent with our result) but also a decrease in stabilising (enthalpic) interactions, which is not consistent with these data (Collins & Feller, 2023). Although, destabilisation of the ground state does not provide any clues as to the important interactions at the transition state.

MalL RDM and T150R (group **B**) act like traditional psychrophiles, where the kinetics are defined by a decrease in the activation heat capacity, increasing the scale of the temperature dependence of the rate, allowing for an entropy-enthalpy trade-off to raise the rate at low temperatures (Arcus *et al.*, 2016). The more negative activation heat capacity stems from a greater constriction at the transition state complex, indicating an increase in flexibility at the ground state. The transition temperature for these mutants is also decreased slightly to 311.5 ± 0.9 and 309.1 ± 1.1 K for MalL RDM and T150R, respectively, compared to 313.5 ± 2.7 K for MalL wildtype.

MalL V376R and D492R in contrast have a different mechanism of psychrophilic activation. These mutants are instead characterised by a decrease in the flexibility of the enzyme substrate complex, which is unusual for psychrophilic adapted enzymes. The hypothesis is that these mutations stabilise the enzyme in the ground state, bring the ES and TLC states closer together and lower the free energy of the TS state. As the conformational search for the TLC is the part of the rate limiting step, a lowered barrier into this conformation results in a rate acceleration overall (Figure 4.34).

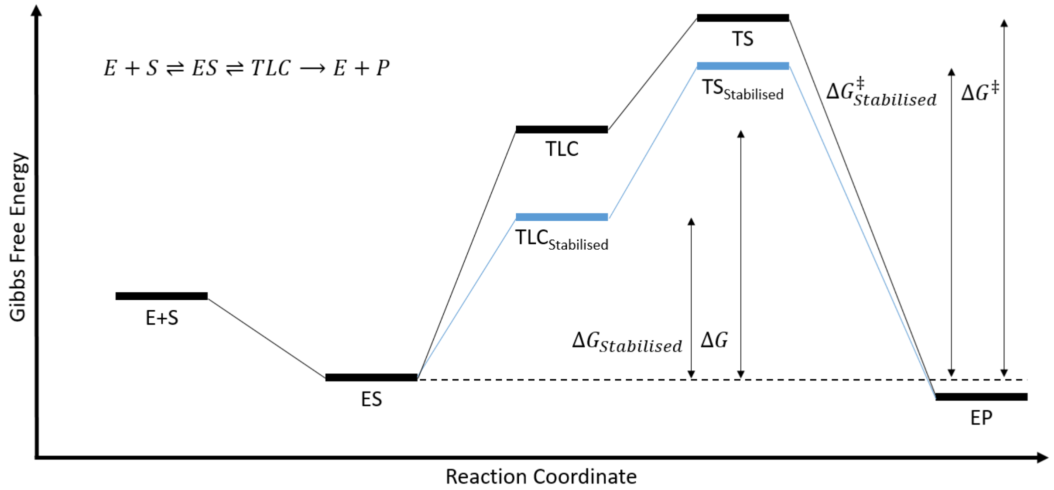


Figure 4.34. Gibbs free energy change along an enzyme-substrate reaction coordinate. MalL D492R and V376R are stabilised relative to MalL wildtype (represented here as a decrease in the Gibbs free energy of the TLC and TS). This reduced $\Delta G_{Stabilised}$ into the TLC shifts the ES ⇌ TLC equilibrium towards TLC at low temperatures, promoting the rate.

This can also been seen in the SAXS results, where the Porod volume, D_{max} and radius of gyration all decrease substantially for MalL wildtype between 25°C and 37°C (Table 4.11), while MalL D492R is more compact than MalL wildtype across both temperatures, and the difference in the Porod volume, D_{max} and radius of gyration is reduced (Table 4.12). A decrease in the particle radius with increasing temperature may be a feature of disordered proteins (Langridge *et al.*, 2014). This will then indicate that MalL D492R has been stabilised, adopting the constrained conformation at all temperatures.

In addition, the REES results show that MalL D492R is less flexible than MalL wildtype. The value of λ_{Ex}^{FC} was unchanged between MalL wildtype and D492R (within margin of error). The increase in the ΔG_m between MalL wildtype and D492R is indicative of a decreased conformational flexibility (Kwok *et al.*, 2021).

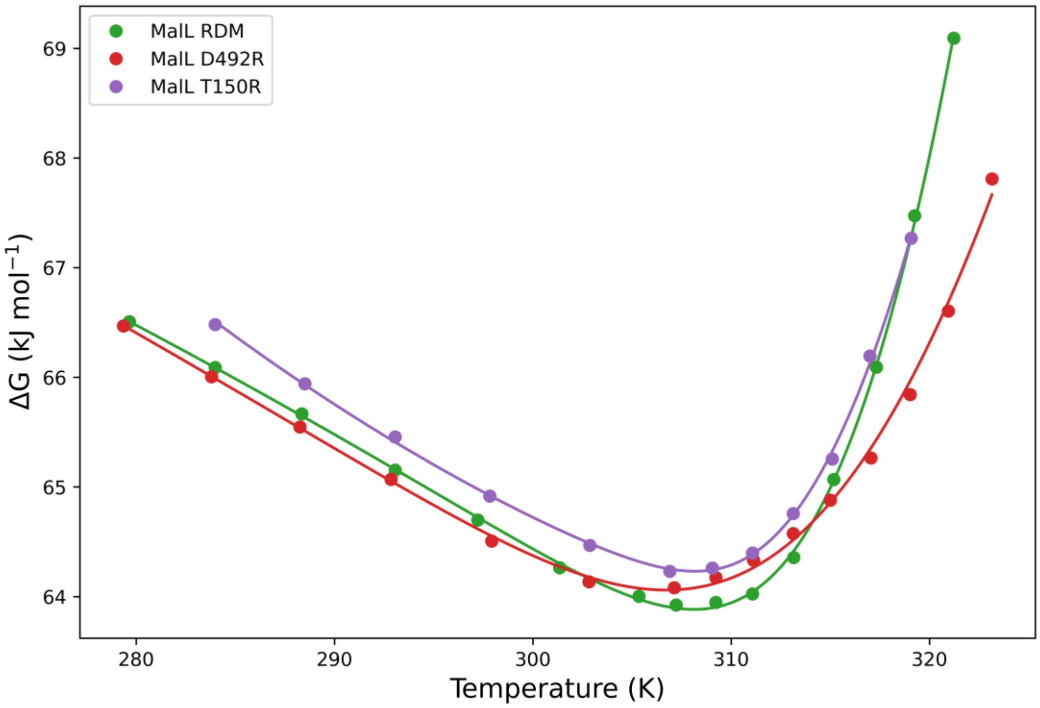


Figure 4.35. MalL RDM (T150R, D492R) shows behaviour consistent with MalL D492R at low temperatures, and MalL T150R at high temperatures. Data are fit with the MMRT 2.0 equation.

MalL RDM is a double mutant composed of the mutations D492R and T150R (Figure 4.35). Interestingly, MalL RDM transitions between the activities of both constituent mutations with temperature. At low temperatures, MalL RDM closely matches MalL D492R. At higher temperatures, MalL RDM more closely matches the activity of MalL T150R. The D492R mutation of MalL RDM is stabilising the ES at low temperatures; however, this effect is lost at high temperatures. The arginine mutants therefore appear to be only partially additive, with each mutation conveying differing effects across the temperature range.

5 Crystal structures of MalL mutants D492R and T150R

5.1 Introduction

The MalL D492R and T150R mutations were designed to be stabilising mutations that may change the catalytic rate by mimicking urea binding in sites distal to the active site. These mutants were characterised, revealing significant kinetic and thermodynamic changes to their activity (Chapter 4). Following the kinetic and thermodynamic characterisation, we sought to crystallise and structurally characterise these mutant enzymes.

5.2 MalL D492R crystallisation and structure determination

5.2.1 Crystallisation

Initial crystallisation conditions were identified by sitting-drop screening a library of 384 common crystallisation conditions (Hampton Research, USA). Identified crystallisation conditions were further optimised using hanging-drop vapour diffusion, by varying ionic strength, pH, precipitant concentration, drop size, and protein concentration. The final crystallisation condition for MalL D492R consisted of 0.09 M Tris pH 8.5, 22% w/v PEG 3350, 5% glycerol. Protein at 8 mg mL^{-1} was mixed 1:1 with precipitant solution in 4 µL drops.

5.2.2 Data collection

Crystals were frozen in liquid nitrogen following a brief suspension in cryoprotectant solution consisting of 0.085 M Tris pH 8.5, 23.5% PEG 3350, 20% v/v glycerol. Crystals were taken to the MX2 beamline at the Australian Synchrotron for data collection. Data were collected for 3600 frames in a full 360° rotation, at 90% attenuation, with a 72 second exposure and a detector distance of 110 mm. Representative crystals and diffraction are shown in Figure 5.1.

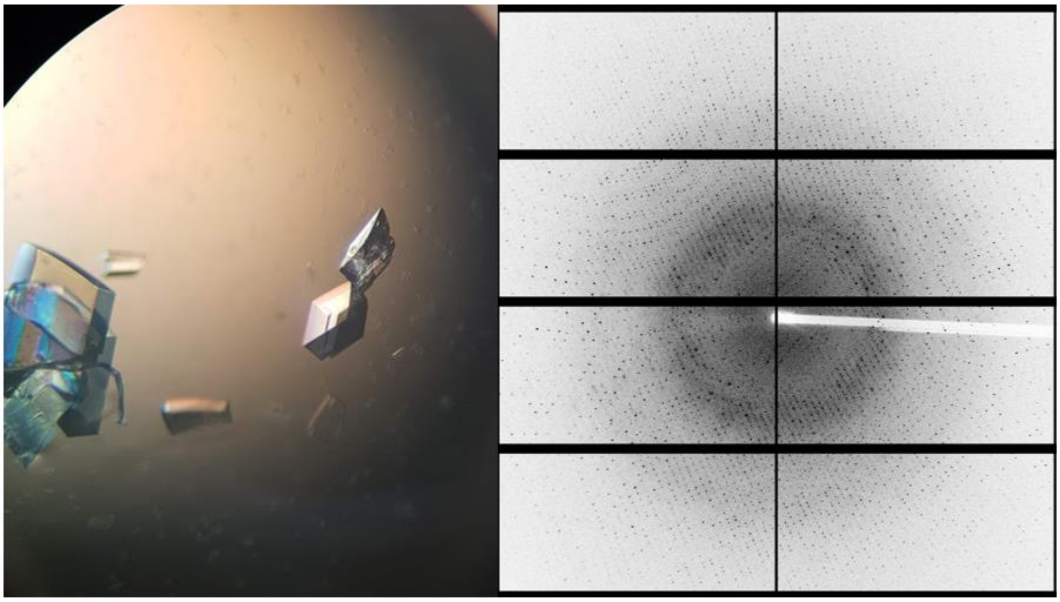


Figure 5.1. Representative crystals and X-ray diffraction for MalL D492R. X-ray diffraction recorded to 1.04 Å.

5.2.3 Data processing

Initial data processing was performed in XDS (Kabsch, 2010). Subsequent processing was performed in the CCP4 software suite (Winn *et al.*, 2011). Space group determination was performed in Pointless (Evans, 2011). Data were scaled and merged in AIMLESS (Evans & Murshudov, 2013). Matthews's coefficient analysis was used to find the number of molecules in the asymmetric unit (Kantardjieff & Rupp, 2003; Matthews, 1968). Statistics from these processes are given in Table 5.1.

Table 5.1. Data collection statistics for MalL D492R.

Data Statistic	MalL D492R
Space group	P 1 2_1 1
Wavelength (Å)	0.9537
Cell Dimensions: **a/b/c (Å)** **α/β/γ (°)**	 60.47/91.84/96.78 90/104.63/90
Mosaicity (°)	0.05
Monomers in the asymmetric unit	2
Resolution Range	48.08-1.04 (1.06-1.04)
Number of observed reflections	3267981 (142395)
Number of unique reflections	486755 (22922)
R_{merge}	0.06 (0.56)
Mean I/σI	14.0 (2.7)
$CC_{1/2}$	1.0 (0.82)
Completeness	99.5 (94.9)
Multiplicity	6.7 (6.2)
Wilson B-factor (Å^2)	7.857

* Values in brackets correspond to the outer resolution shell.

Molecular replacement was performed in PHASER (McCoy *et al.*, 2007) in the PHENIX software suite (Adams *et al.*, 2010). The previously solved MalL wildtype (PDB: 4M56) was used as a search model. Models were iteratively refined by manual model building in COOT (Emsley *et al.*, 2010), and automatic refinement in phenix.refine (Afonine *et al.*, 2012). Further model refinement was completed in PDB-REDO (Joosten *et al.*, 2014). A representative electron density map and model is shown in Figure 5.2.

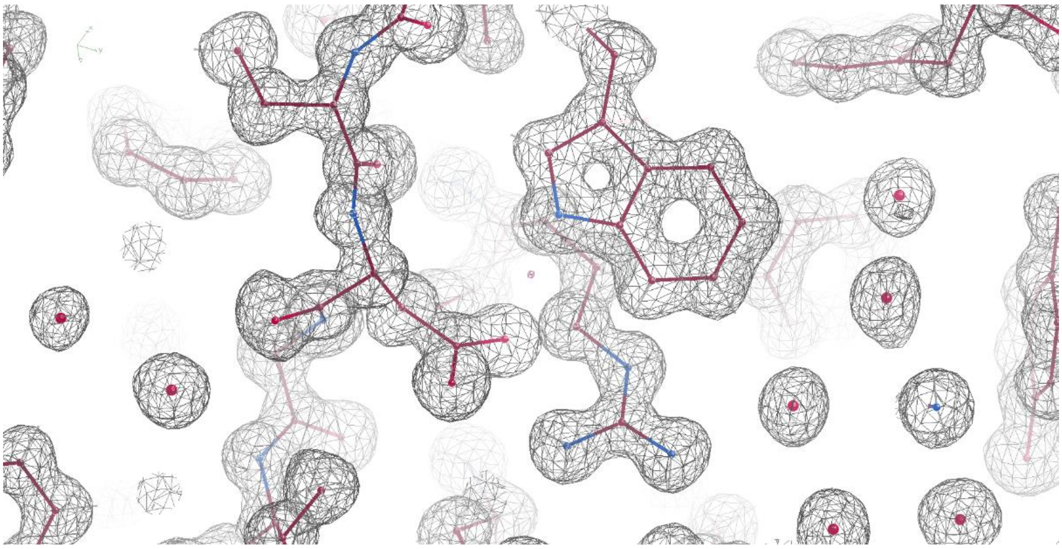


Figure 5.2. Representative 2Fo-Fc electron density map and model fit of MalL D492R at 1.04 Å resolution. Contour level is set to 1.01 RMSD.

There were 23 amino-acids of the 582 long sequence, including the hexa-His tag, missing from the N-terminus of both chain A and chain B, which were not modelled in the final structure (Full length protein sequence is included in Section 8.1). A small number of sidechains were unable to be modelled in the final density. There were 13 additional small molecules identified in the structure. These were identified as Tris and glycerol which are components of the crystallisation and cryoprotectant solutions. A metal ion was found bound to Asp20, Asn45, Phe26 and Asp28 in both chain A and chain B. This site has been identified as a calcium binding site in other glycoside hydrolase enzymes (Kobayashi *et al.*, 2011; Møller Marie *et al.*, 2012; Yamamoto *et al.*, 2010). In both cases the bound metal was identified as calcium. The calcium ions were well defined by the observed density, and bound with octahedral coordination, with distances of approximately 2.3 Å, consistent with calcium binding (Harding *et al.*, 2010). Refinement statistics are given in Table 5.2.

Table 5.2. Refinement statistics for MalL D492R

Refinement Statistic	MalL D492R
R-Factor	0.11
R-Free	0.13
Total number of atoms	20157
Total number of protein atoms	18757
Other molecules:	
- Number	15
- Atoms	152
Number of waters	1248
Solvent content (%)	34.3
RMSD:	
- Bond length (Å)	0.009
- Bond angles (°)	1.14
Average B-Factors (Å^2):	
- Protein monomer:	
Chain A	13.06
Chain B	12.46
- Waters	24.7
Ramachandran analysis:	
- Percentage in favoured regions	89.5
- Percentage in allowed regions	9.5
- Percentage in disallowed regions	0.4

5.3 MalL T150R crystallisation and structure determination

5.3.1 Crystallisation

Initial crystallisation conditions were identified by sitting-drop screening using a library of 384 common crystallisation conditions (Hampton Research, USA). Identified crystallisation conditions were further optimised using hanging-drop vapour diffusion, by varying ionic strength, pH, precipitant concentration, drop size, and protein concentration. The final crystallisation condition for MalL T150R consisted of 0.1 M Tris pH 8.5, and 22% w/v PEG 3350. Protein at 4 mg mL^{-1} was mixed 1:1 with precipitant solution in 2 µL drops.

5.3.2 Data collection

Crystals were frozen in liquid nitrogen following a brief suspension in cryoprotectant solution consisting of 0.105 M Tris pH 8.5, 23.5% PEG 3350, 20% v/v glycerol. Crystals were taken to the MX2 beamline at the Australian Synchrotron for data collection. Data were collected for 3600 frames in a full 360° rotation, at 85% attenuation, with a 36 second exposure and a detector distance of 200 mm. Representative crystals and diffraction is shown in Figure 5.3.

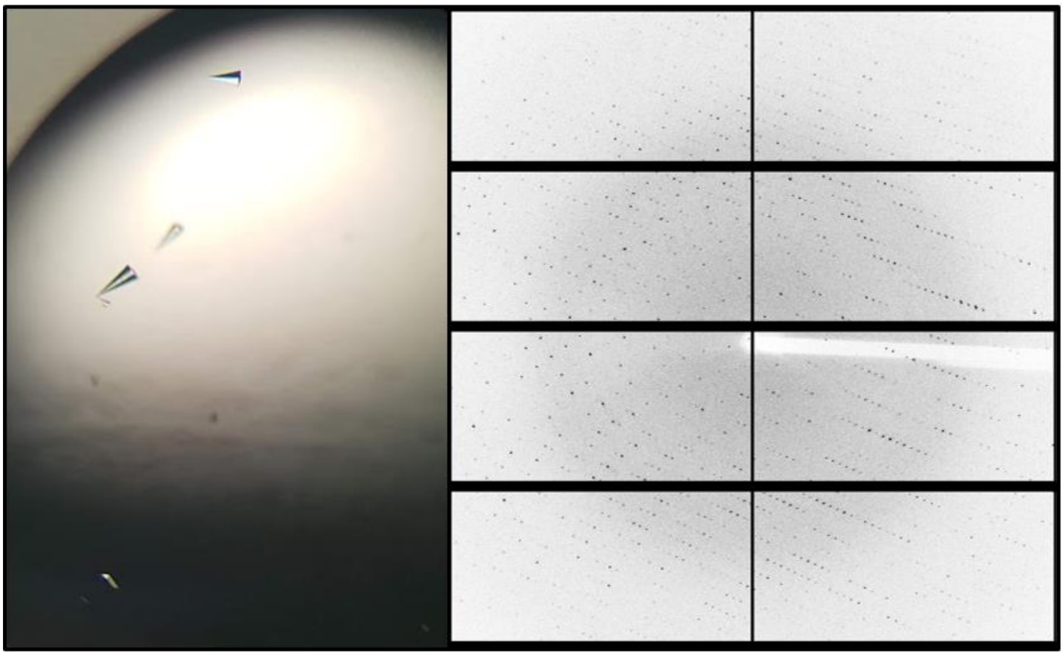


Figure 5.3. Representative crystals and X-ray diffraction for MalL T150R. X-ray diffraction recorded to 1.46 Å.

5.3.3 Data processing

Initial data processing was performed in XDS (Kabsch, 2010). Subsequent processing was performed in the CCP4 software suite (Winn *et al.*, 2011). Space group determination was performed in Pointless (Evans, 2011). Data were scaled and merged in AIMLESS (Evans & Murshudov, 2013). Matthews's coefficient analysis was used to find the number of molecules in the asymmetric unit (Kantardjieff & Rupp, 2003; Matthews, 1968). Statistics from these processes is given in Table 5.3.

Table 5.3. Data collection statistics for MalL T150R.

Data Statistic	MalL T150R
Space group	P 1 2_1 1
Wavelength (Å)	0.9537
Cell Dimensions: **a/b/c (Å)** **α/β/γ (°)**	 49.718/93.061/61.652 90/110.44/90
Mosaicity (°)	0.15
Monomers in the asymmetric unit	1
Resolution Range	48.95-1.46 (1.49-1.46)
Number of observed reflections	615982 (26183)
Number of unique reflections	89153 (3977)
R_{merge}	0.072 (0.921)
Mean I/σI	12.9 (2.0)
$CC_{1/2}$	1.0 (0.76)
Completeness	99.5 (90.5)
Multiplicity	6.9 (6.6)
Wilson B-factor (Å^2)	14.7

Molecular replacement was performed in PHASER (McCoy *et al.*, 2007) in the PHENIX software suite (Adams *et al.*, 2010). The previously solved MalL wildtype (PDB: 4M56) was used as a search model. Models were iteratively refined by manual model building in COOT (Emsley *et al.*, 2010), and automatic refinement in phenix.refine (Afonine *et al.*, 2012). Further model refinement was completed in PDB-REDO (Joosten *et al.*, 2014). A representative electron density map and model is shown in Figure 5.4.

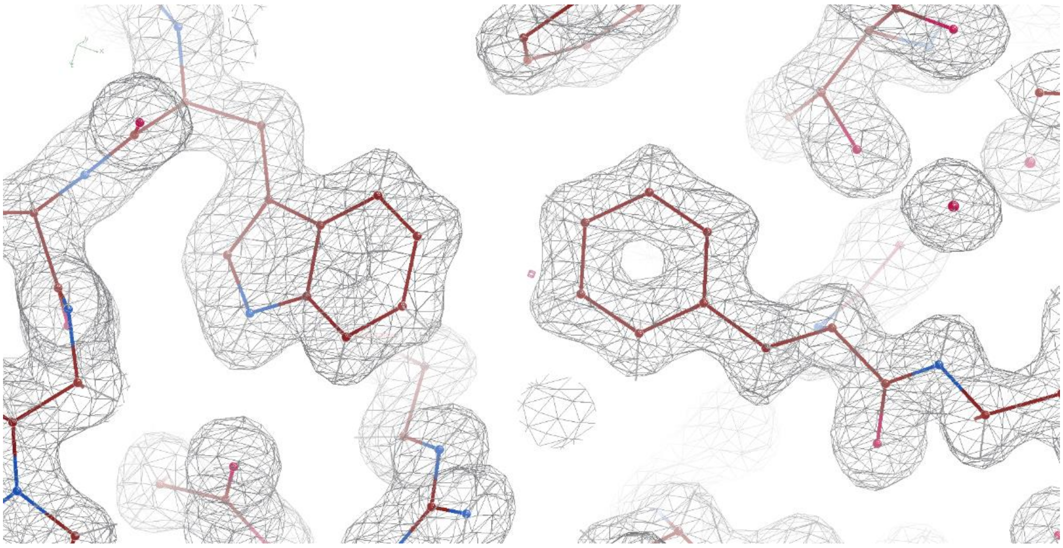


Figure 5.4. Representative 2Fo-Fc electron density map and model fit of MalL T150R at 1.46 Å resolution. Contour level is set to 1.01 RMSD.

There were 23 amino-acids, including the hexa-His tag, missing from the N-terminus which were not modelled in the final structure (Full length 586 amino acid sequence is included in Section 8.1). A small number of sidechains were unable to be modelled in the final density. There were 4 additional small molecules identified in the structure. These were identified as Tris and glycerol which are components of the crystallisation and cryoprotectant solutions. A calcium ion was found bound to Asp20, Asn45, Phe26 and Asp28 which is consistent with calcium binding in other glycoside hydrolase enzymes (Kobayashi *et al.*, 2011; Møller Marie *et al.*, 2012; Yamamoto *et al.*, 2010). The calcium ion was bound with octahedral coordination at a distance of approximately 2.3 Å, consistent with calcium binding. The calcium ion was well fitted into the experimental density. Refinement statistics are given in Table 5.4.

Table 5.4. Refinement statistics for MalL T150R

Refinement Statistic	MalL T150R
R-Factor	0.13
R-Free	0.17
Total number of atoms	9723
Total number of protein atoms	9032
Other molecules: **- Number** **- Atoms**	 5 53
Number of waters	638
Solvent content (%)	35.7
RMSD:	
- Bond length (Å)	0.011
- Bond angles (°)	1.149
Average B-Factors (Å^2):	
- Protein monomer	20.12
- Waters	31.52
Ramachandran analysis:	
- Percentage in favoured regions	88.5
- Percentage in allowed regions	10.5
- Percentage in disallowed regions	0.2

5.4 Discussion

5.4.1 Structural analysis

MalL D492R crystallised with two monomers in the asymmetric unit, as did MalL wildtype. Therefore, the dramatically increased resolution for this structure is not a result of the reduction in the contents of the asymmetric unit (as might be argued for S536R). MalL T150R crystallised with only one monomer in the asymmetric unit. The unit cell of MalL D492R is approximately 15% smaller than the unit cell for MalL wildtype, with a volume of 5.20 x 10^5 Å^3 compared to 6.10 x 10^5 Å^3 for MalL wildtype. Similarly, the size of the asymmetric unit decreased from 3.05 x 10^5 Å^3 in MalL wildtype to 2.60 x 10^5 Å^3 in MalL D492R. The structures of MalL T150R and MalL D492R show the same structural architecture as MalL wildtype. The RMSD (calculated from C$_\alpha$ atoms) between the MalL variants and MalL wildtype varies between 0.369 and 0.421 Å, which indicates a very high degree of structural similarity (Table 5.5). Thus the single point mutations D492R and T150R do not significantly alter the overall structure of the enzyme. An overview of the structural similarity is shown in Figure 5.5.

Table 5.5. C_α RMSD (Å) between MalL variants

	Wildtype Chain A	**Wildtype Chain B**	**T150R**
T150R	0.373 (555)	0.402 (556)	-
D492R Chain A	0.382 (555)	0.393 (556)	0.399 (556)
D492R Chain B	0.369 (555)	0.383 (556)	0.421 (556)

* Number in brackets is number of atoms compared

The observed RMSD between wildtype chain A and chain B was 0.257 Å, and 0.369 Å for between D492R chain A and chain B. Structural variation between chains within the same structure are on a similar scale to the variation between mutants.

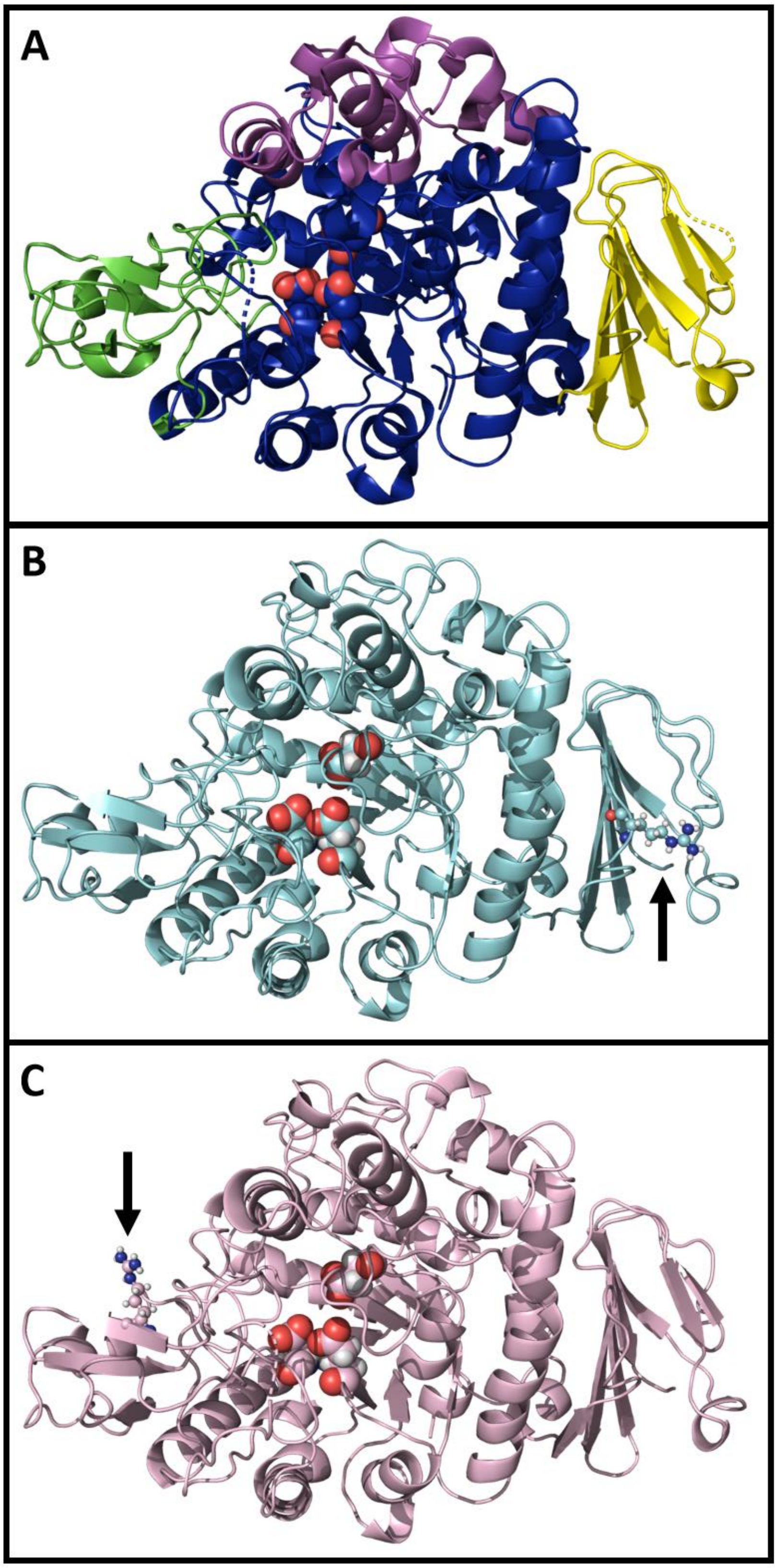


Figure 5.5. Structural comparison of MalL variants. A) Chain A of MalL wildtype (PDB: 4M56). Domain A is coloured blue, domain B is green, 'lid' domain is magenta, and domain C is yellow. B) Chain A of MalL D492R. C) MalL T150R. Catalytic residues are shown as spheres. Mutations are shown as ball-and-sticks, and indicated with an arrow.

The D492R mutation is located approximately 30 Å away from the active site, and theT150R mutation is located approximately 20 Å away. This, combined with the high degree of structural similarity, signifies a lack of an obvious structural basis for variations in enzyme activity. Thus, any changes to the rate must be based on the dynamics of the enzyme, or related to conformations that are not present in the crystal. The three domains of MalL (Section 1.9) are easily identified in the structures above. The catalytic TIM barrel A domain is in the centre (blue), with domain B (green) at the top, the lid domain to the right (magenta), and the auxiliary C domain at the bottom (yellow). The D492R mutation is located in the C domain, and the T150R mutation is located in the B domain, at the interface with the A domain.

5.4.1.1 Crystal packing

Both MalL wildtype and MalL D492R crystallised with two monomers in the asymmetric unit. MalL D492R is significantly more densely packed with a reduced unit cell volume (5.2 x 10^5 Å^3, compared to 6.1 x 10^5 Å^3)¸ as well as increased density, with MalL D492R having a non-solvent, non-hydrogen density of 27.0 Å^3 atom^{-1}, compared to 33.2 Å^3 atom^{-1} (Figure 5.6). There were also significantly more hydrogen bonds between adjacent monomers in MalL D492R compared to MalL wildtype (Figure 5.6).

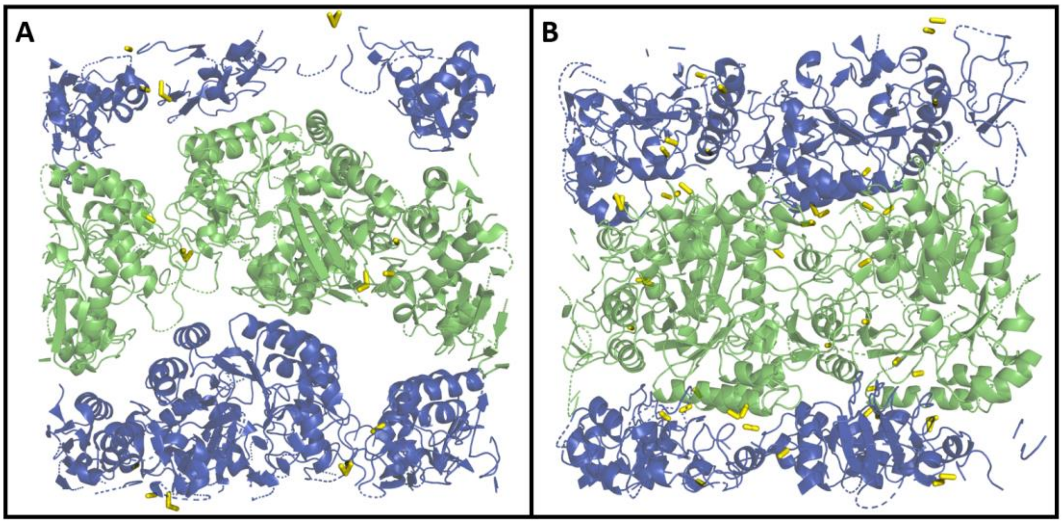


Figure 5.6. Crystal packing in unit cell. Hydrogen bonds between adjacent chains are coloured yellow. A) MalL wildtype. B) MalL D4592R is significantly more densely packed.

The two monomers in the asymmetric unit are related by a pseudo rotational axis, with the two monomers packing face to face (Figure 5.7 A). In contrast, the two monomers in MalL D492R are related by a pseudo translational axis with back to front packing (Figure 5.7 C). Additionally, hydrogen bonds to adjacent monomers from the asymmetric unit were identified (Figure 5.7). There were 10 identified hydrogen bonds between either monomer in the asymmetric unit of MalL wildtype and an adjacent monomer. In contrast, there were 44 identified hydrogen bonds identified in MalL D492R (of which 16 involved residues with alternate conformations). Four of these identified hydrogen bonds involved the D492R mutation, with each arginine residue (in either monomer of the asymmetric unit) forming two hydrogen bonds. Similarly there were 18 hydrogen bonds identified in the MalL T150R structure. A full list of identified bonds can be found in Section 8.7.

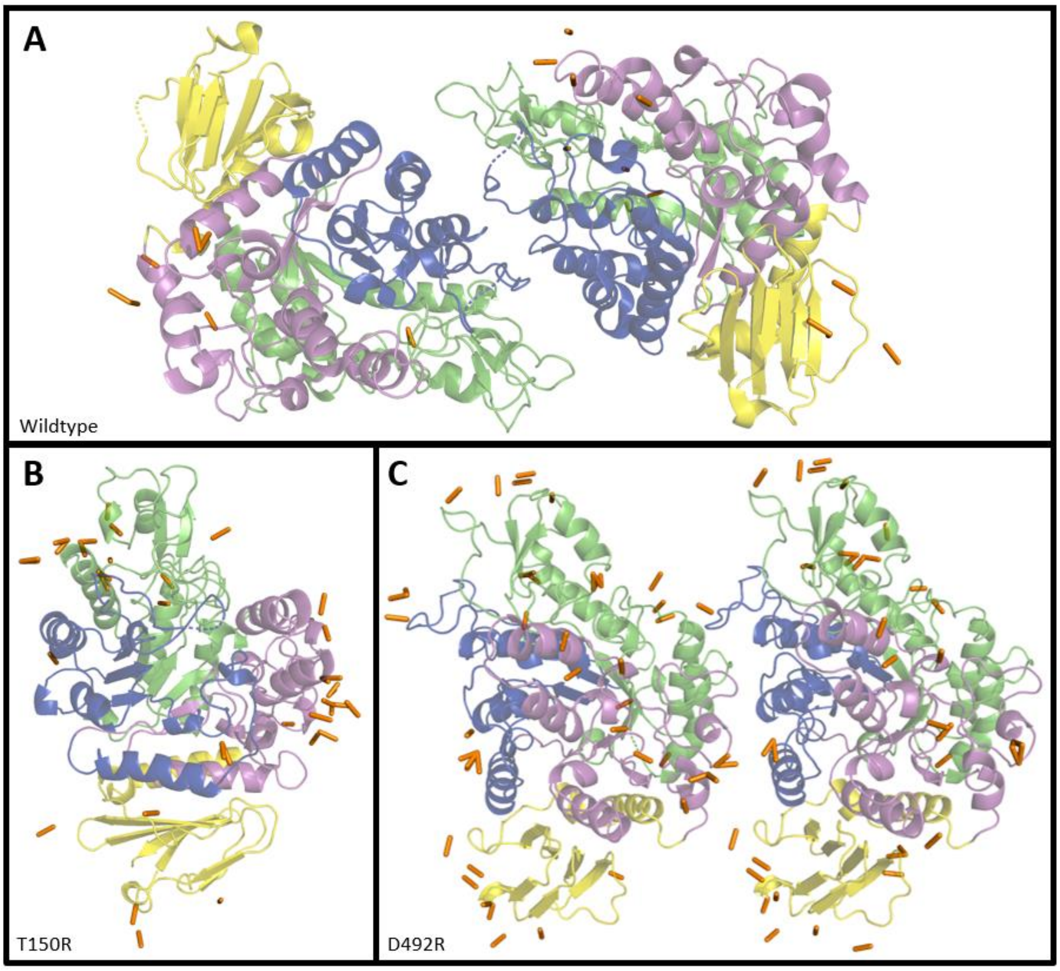


Figure 5.7. Hydrogen bonding to crystal contacts from the asymmetric unit. Structures are coloured by region. Region 1 (residues 1-193) is green, region 2 (residues 194-321) is blue, region 3 (residues 322-459) is magenta, region 4 (residues 460-561) is yellow. A) MalL wildtype has 20 hydrogen bonds. B) MalL T150R has 36 hydrogen bonds. C) MalL D492R has 88 hydrogen bonds.

MalL wildtype and MalL T150R bind to crystal contacts at well-defined contact points, with the main contact occurring between region one (green) and region three (magenta) (Figure 5.7 A, B), although MalL T150R in general has a greater number of bonding interactions at each interface. MalL D492R in contrast has tighter crystal packing, and more distributed crystal interfaces, however as with both MalL wildtype and MalL T150R, the interface between regions one and three remains important (Figure 5.7 C).

5.4.1.2 ProSMART structural comparison

The ProSMART analysis allows for conformation-independent structural alignment and scoring by comparing local regions of structure between protein chains (Nicholls *et al.*, 2014). This also allows for the identification of rigid substructures that are conserved between the two compared structures. This analysis finds two such substructures in MalL (Figure 5.8). The first substructure (shown in red) consists of the majority of the core of the protein, extending into both the B domain and the C domain. This indicates that the core structure of the protein is well conserved between all variants, and the majority of structural differences occur in the surface regions. The effect of the D492R and T150R mutations on the local structure are plainly visible. The mutations limit the extension of the rigid core structure into the C domain and B domain, respectively. The second substructure (shown in green) is part of the lid domain, which forms part of the active site binding pocket. This region also contributes significantly to the activation heat capacity in molecular dynamics simulations of MalL wildtype (van der Kamp *et al.*, 2018).

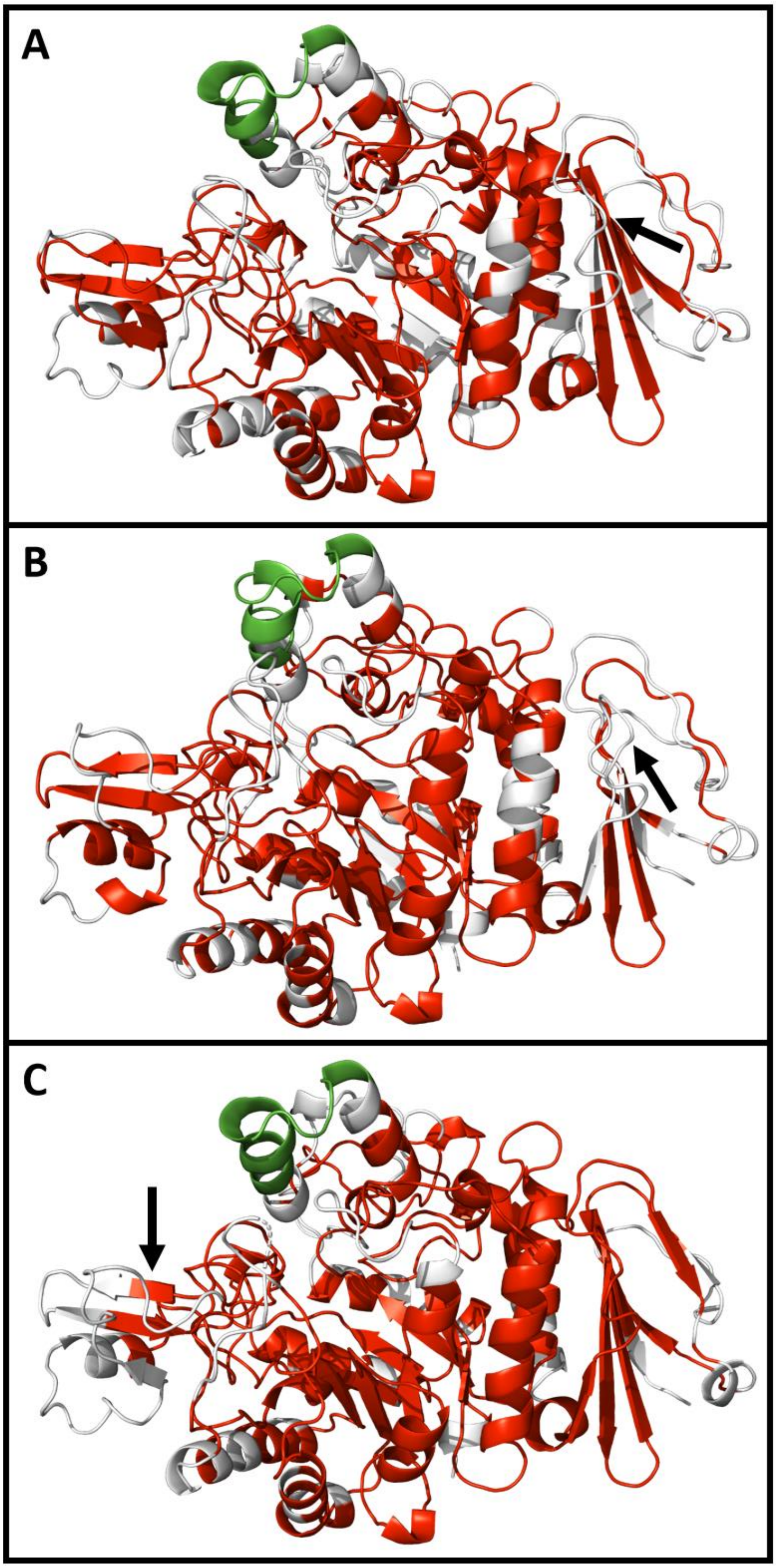


Figure 5.8. Substructure analysis of MalL variants versus chain A of MalL wildtype. A) Chain A of MalL D492R. B) Chain B of MalL D492R. C) MalL T150R. Reveals a stable core structure (red), and lid domain (green) as stable structural elements. Mutation positions are located with an arrow.

The Procrustes score is a measure of the raw structural dissimilarity of the immediate local backbone environment, with a redder colour indicating greater structural dissimilarity (Nicholls *et al.*, 2014). This indicates a high degree of structural similarity, except in the regions immediately surrounding the mutation locations, as well as a few surface loops (Figure 5.9). These surface loops are highly mobile and poorly defined in the electron density of each MalL structure. The C domain of MalL D492R (Figure 5.9 A, Figure 5.9 B) shows significant (although minor) backbone changes throughout the C domain, with the greatest change occurring in the mutation location and extending into the lower portion of the domain. Similarly, the effect of the T150R mutation extends through the B domain and does not appear to significantly interact with the A domain (Figure 5.9 C).

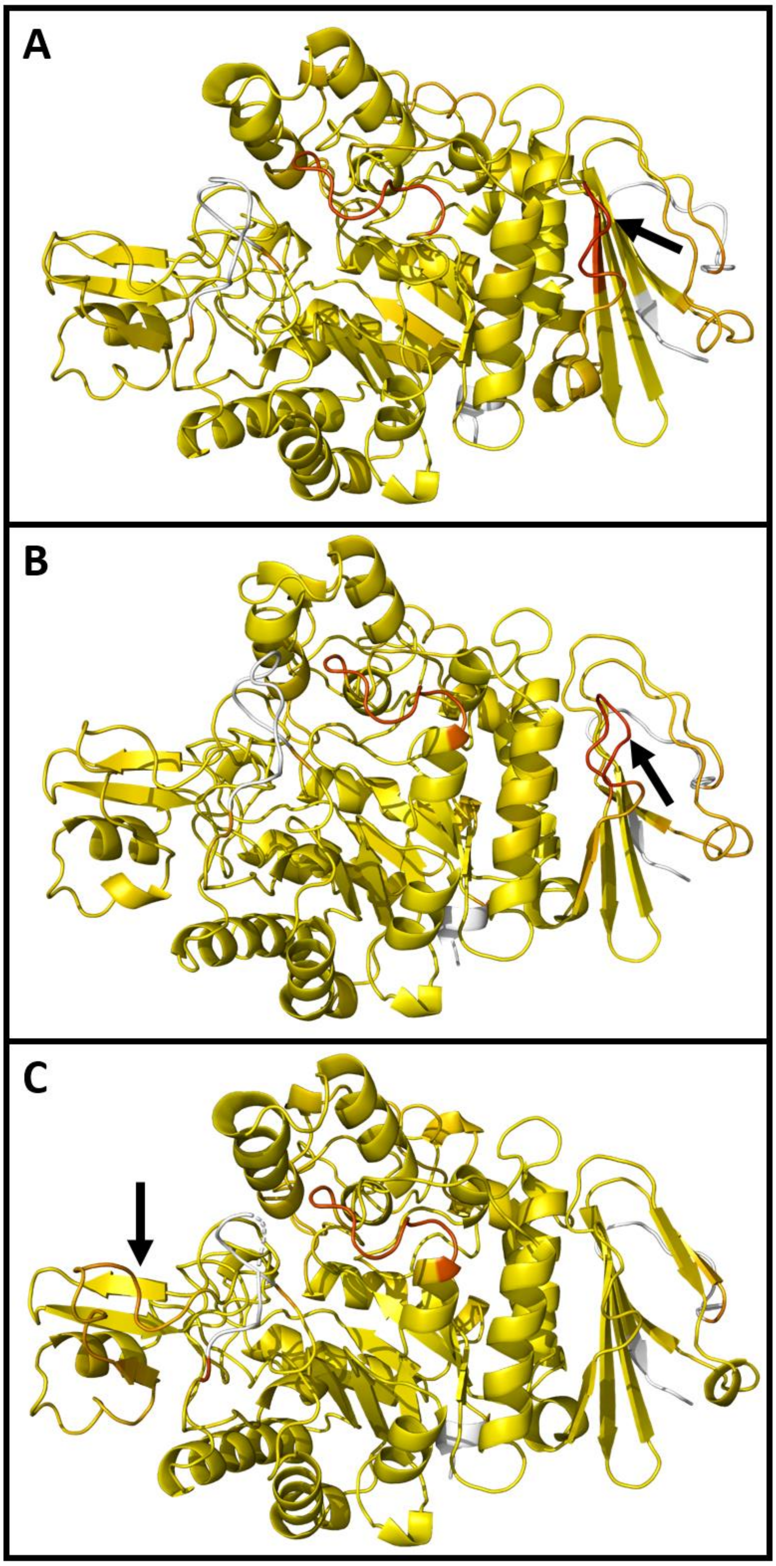


Figure 5.9. Backbone Procrustes analysis of MalL variants versus chain A of MalL wildtype. A) Chain A of MalL D492R. B) Chain B of MalL D492R. C) MalL T150R. Coloured by structural similarity. Yellow indicates structural similarity. Red indicates structural dissimilarity. White indicates regions that could not be compared. Reveals major backbone deviations in poorly defined surface loops and in regions surrounding mutations. Mutation positions are located with an arrow.

The side-chain RMSD is a measure of the average distance between corresponding side-chains after local superposition (Nicholls *et al.*, 2014). This shows that the side chain positioning in the protein core is largely conserved between mutants, with the bulk of side chain deviations occurring in surface regions.

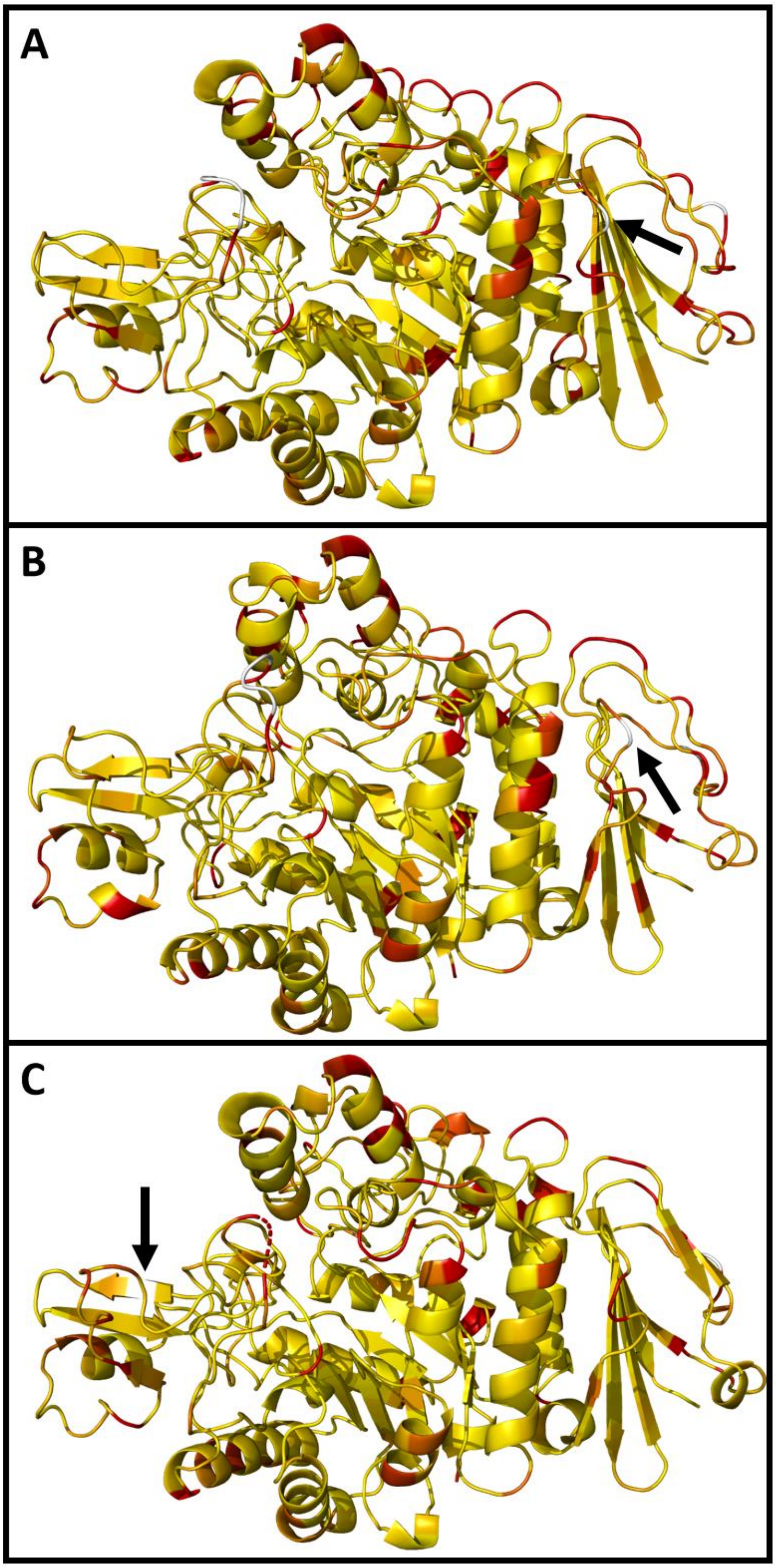


Figure 5.10. Sidechain RMSD analysis of MalL variants versus chain A of MalL wildtype. A) Chain A of MalL D492R. B) Chain B of MalL D492R. C) MalL T150R. Coloured by structural similarity. Yellow indicates structural similarity. Red indicates structural dissimilarity. White indicates regions that could not be compared. Mutation positions are located with an arrow.

5.4.1.3 Comparison to urea binding

The T150R and D492R mutations were designed based on their overlap with urea binding sites. In each case an arginine was introduced such that the guanidinium group might occupy the urea binding site. The crystal structures of MalL D492R and MalL T150R reveal that neither mutation occupies the urea binding site.

The D492R mutation was designed to mimic urea binding at a site in domain C of MalL (Figure 5.11). In the wildtype urea bound structure (not published), urea bound to the backbone of Leu519, the amide carbonyl of Gln519, and the carboxyl group of Asp592 (Figure 5.11 A). This aspartic acid was mutated into arginine with the intention that the guanidine group of arginine would occupy the urea binding site. Arg492 instead binds down to carboxyl group of Glu521 on the bottom loop of domain C (Figure 5.11 B). The backbone interactions with Gln489 and Lys308 are maintained. The interactions of the carboxyl group of Asp492 in the urea wildtype structure with the backbone of Gln494 and Val495 are lost upon the mutation to arginine.

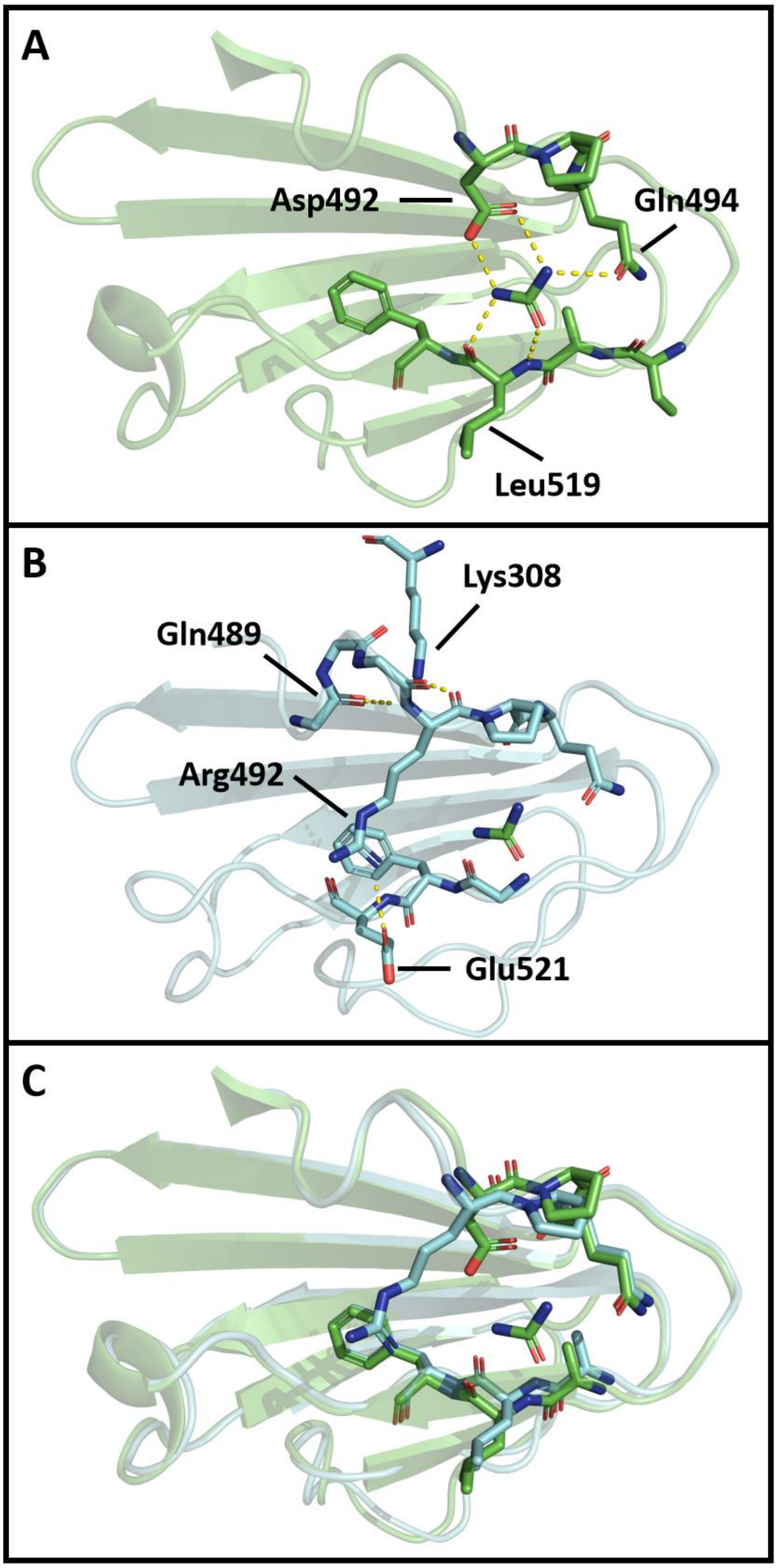


Figure 5.11. D492R mutation at urea binding site. Domain C is shown as cartoon. A) Urea binds to Asp492, Gln494, and the backbone of Leu519. B) MalL D492R with bound urea (green) overlaid. The D492R mutation does not result in binding in the urea binding pocket. Arg492 binds to Glu521 in the lower loop of domain C. C) Overlay of wildtype structure with urea bound and MalL D492R structure.

The urea that the T150R mutation was designed to mimic binds to the backbone of Ser145, Ser147 and Trp149 as well as the hydroxyl group of Thr150 (Figure 5.12). The T150R mutation does not result in occupation of the urea binding site. The introduced arginine instead binds to Tyr159 and Lys165 on the adjacent β-strand and loop, through carbon-hydrogen mediated hydrogen bonds (Figure 5.12). Carbon-hydrogen mediated hydrogen bonds occur when a carbon-hydrogen pair acts as a donor to form a hydrogen bond with another electronegative atom (Derewenda *et al.*, 1995). These bonds are usually of lesser strength than a typical hydrogen bond; however, these bonds may still contribute significantly to the overall structure and function of a protein (Derewenda *et al.*, 1995). The β-strand backbone interactions to Tyr159 is maintained with either Tyr150 or Arg150.

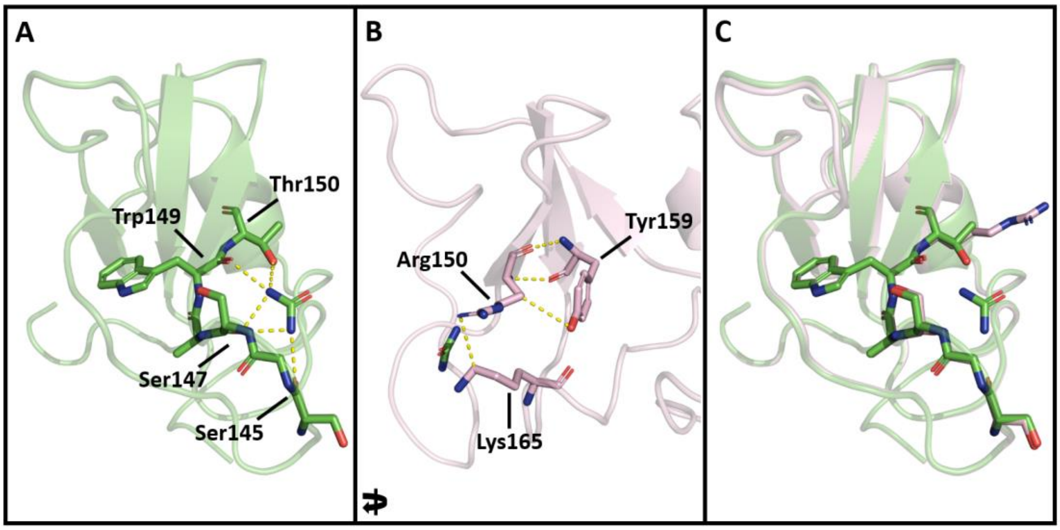


Figure 5.12. T150R mutation at the urea binding site. Domain B is shown as cartoon. A) Urea binds to the backbone of Ser145, Ser147 and Trp149, and the sidechain of Tyr150. B) MalL T150R with bound urea (green) overlaid. The mutant residue R150 does not bind in the urea binding pocket. Rotated 45° left compared to A and C. C) Overlay of wildtype structure with urea bound and MalL T150R structure.

Neither mutation here bound in the urea binding site, in contrast to the S536R mutation. This is likely due to the lack of a structure validation step to check that the introduced mutation occupied the binding site. These mutations were chosen due to their *in silico* energy score, and if a rotamer existed that could, in principle occupy the binding site.

5.4.2 Temperature factor analysis of MalL variants

Both MalL D492R and MalL T150R crystallised with significantly improved resolution, at 1.04 and 1.46 Å, respectively, compared to 2.3 Å for MalL wildtype. This is indicative of a significant change in the structural dynamics of these enzymes (along with improved crystal packing as described previously). Temperature factors describe the decrease in scattering intensity due to thermal motion. The temperature factors in a protein structure are associated with the relative flexibility of the protein in the crystal, with higher temperature factors indicating a higher level of flexibility, and low temperature factors indicating rigidity (Sun *et al.*, 2019). The values of observed temperature factors are the result of many competing factors, including the structure resolution, crystal contacts, and refinement method. To allow for a more accurate structural comparison, temperature factors can be normalised using Equation 5.1, where B is the temperature factor, $\bar{B}$ is the average temperature factor, and σ is the standard deviation. Temperature factors analysis for the arginine mutants are shown in Figure 5.13, with the temperature factors for MalL wildtype give in Figure 5.14, for reference.

Equation 5.1. Normalised temperature factors

$$B_{norm} = \frac{B - \bar{B}}{\sigma}$$

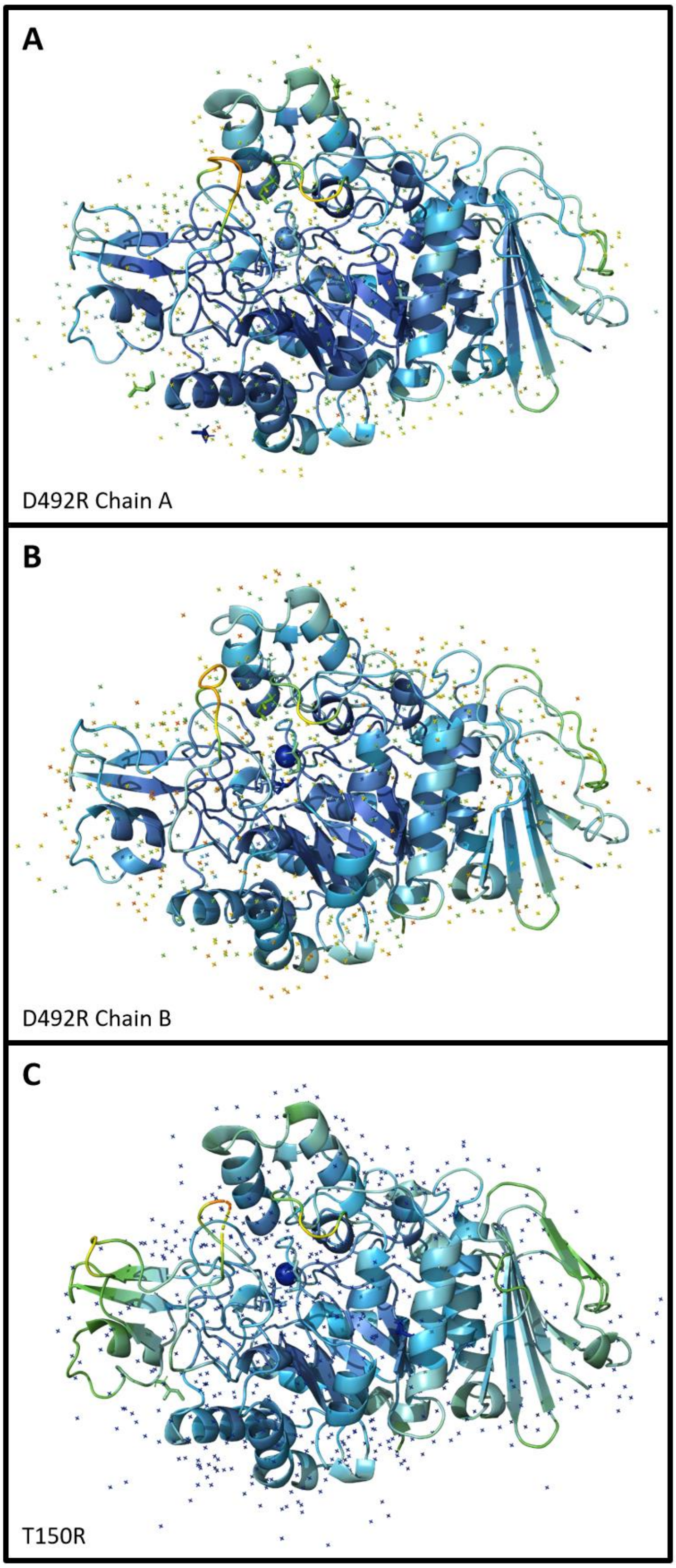


Figure 5.13. Structures of MalL variants coloured by average temperature factor per residue. Temperature factors are shown in scale from low to high (blue-green-red). Colours are scaled to individual structures only. A) MalL D492R Chain A, 1.04 Å, and temperature-factor range 6.7 - 48.0 Å^2. B) MalL D492R Chain B, 1.04 Å, and temperature-factor range 6.4 - 39.8 Å^2. C) MalL T150R, 1.46 Å, temperature-factor range 11.1 - 59.7 Å^2.

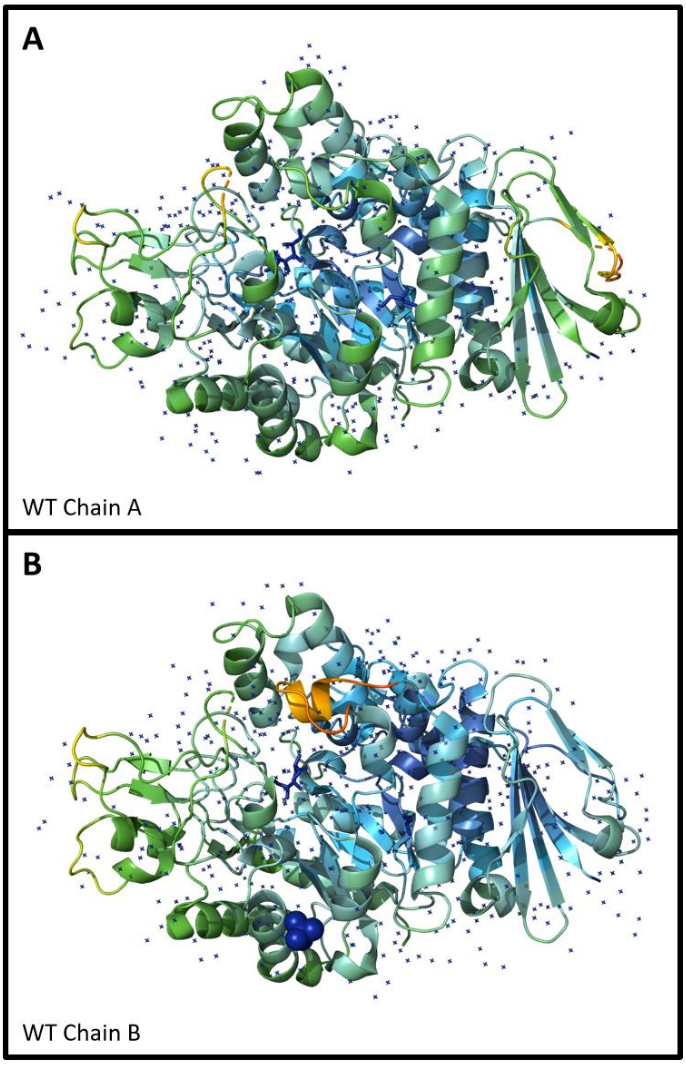


Figure 5.14. Structures of MalL variants coloured by average temperature factor per residue. Temperature factors are shown in scale from low to high (blue-green-red). Colours are scaled to individual structures only. A) MalL wildtype Chain A, 2.3 Å, and temperature-factor range 9.5 – 62.9 Å^2. B) MalL wildtype Chain B, 2.3 Å, and temperature-factor range 11.2 - 89.8 Å^2.

Enzyme temperature factors are generally low in the region surrounding the active site (Sun *et al.*, 2019). This is particularly evident in the MalL wildtype structures (Figure 5.14), where the core region is dark blue, and the surrounding surface regions are more flexible and represented in green and orange. Overall the arginine mutants (particularly MalL D492R) show a decreased range of temperature factors, more intermolecular contacts and increased rigidity across the whole structure, consistent with the higher order and improved resolution of the structures. The average main chain temperature factor per residue for each arginine mutant versus MalL wildtype is presented in Figure 5.15 and Figure 5.16. The overall temperature factors of MalL D492R and MalL T150R are lower than MalL wildtype, consistent with the improved resolution (Figure 5.15 A, Figure 5.16 A). The MalL T150R temperature factors are overall much closer to MalL wildtype than MalL D492R.

Compared to MalL wildtype MalL D492R is significantly rigidified in the region of residues 100-180, which consists of domain B (Figure 5.15 B). This is contrasted by region 400-485, which has an increased flexibility and higher temperature factors (Figure 5.15 B). This region consists of part of the 'lid' domain, which forms part of the active site pocket, as well part of the TIM barrel structure at the interface to domain C. The C domain of MalL D492R, in general, is relatively flexible, with increased temperature factors compared to the rest of the structure. This is somewhat surprising, given that there is an overall decrease in temperature factors across the entire structure, and that the introduced mutation is in this domain. MalL T150R is rigidified in the region of residues 150-250, which forms part of the B domain and the left side of the TIM barrel (as viewed in Figure 5.13 C). It is also similarly more flexible in domain C (Figure 5.16 B).

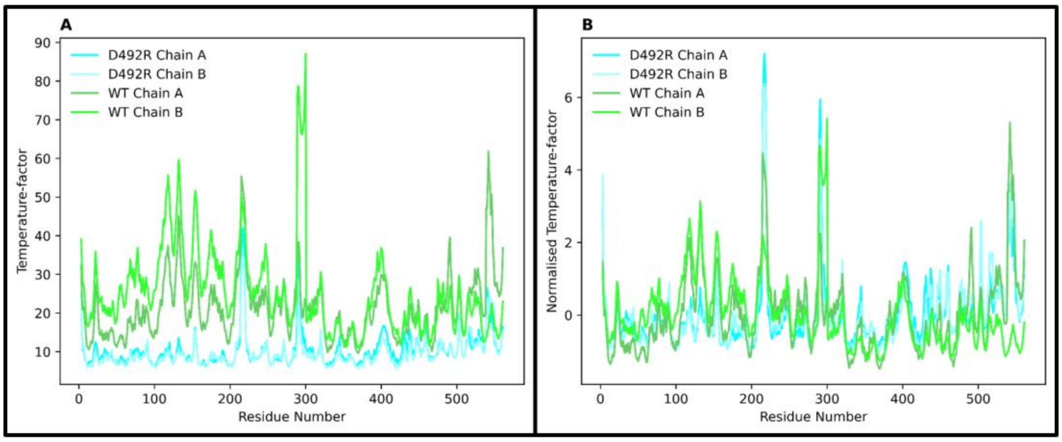


Figure 5.15. Temperature factors of MalL D492R versus MalL wildtype. Temperature factors are the average of main chain atoms per residue. A) Temperature factors. B) Normalised temperature factors.

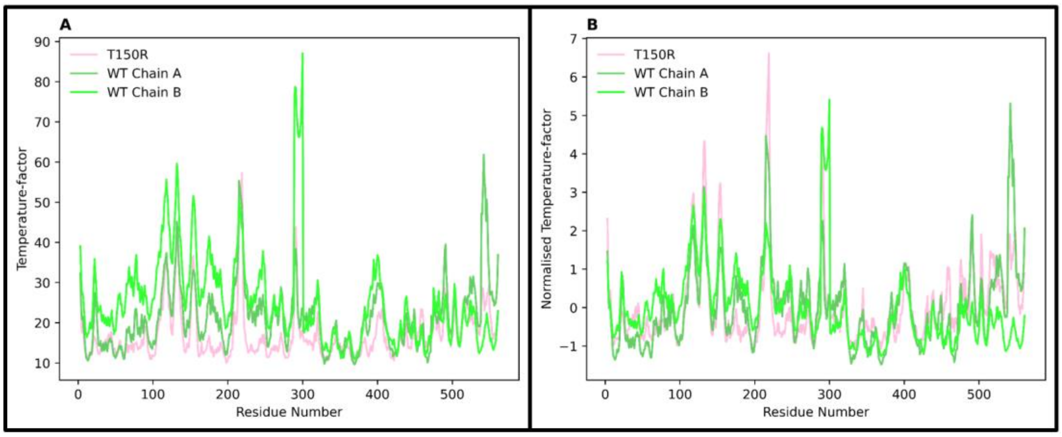


Figure 5.16. Temperature factors of MalL T150R versus MalL wildtype. Temperature factors are the average of main chain atoms per residue. A) Temperature factors. B) Normalised temperature factors

5.4.3 Hydrogen bond analysis

Hydrogen bond differences were analysed between the MalL arginine variants and MalL wildtype as described in Section 3.3.4. Results of this analysis are presented in Table 5.6. The difference in resolution between the structures may potentially be a confounding factor; however, analysis presented in Section 3.3.4 indicates that general trends are consistent when high resolution structures are cut-off to lower resolution. A full table of hydrogen bond differences can be found in Section 8.5.

Table 5.6. Hydrogen bond comparison. First number is the number of hydrogen bonds in each structure not found in the comparison structure. Number in brackets is the number of bonds in each structure that is significantly shorter (>0.3 Å) than in the comparison structure.

	Comparison Structure				
Structure	**D492R Chain A**	**D492R Chain B**	**T150R**	**WT Chain A**	**WT Chain B**
D492R Chain A	-	-	-	15 (7)	16 (7)
D492R Chain B	-	-	-	15 (7)	10 (8)
T150R	-	-	-	15 (7)	12 (8)
WT Chain A	12 (1)	9 (0)	8 (2)	-	-
WT Chain B	9 (2)	5 (2)	7 (2)	-	-

There were significantly more hydrogen bonds in the arginine mutant structures as compared to MalL wildtype. Extra hydrogen bonds (sum of unique and shortened hydrogen bonds) ranged between 23 and 18 across the arginine mutants, while extra bonds in MalL wildtype ranged between only 13 and 7. These bonds are visualised in Figure 5.17 and Figure 5.18. Overall the distribution of extra hydrogen bonds is relatively spread out across the structures. There is a higher proportion of extra hydrogen bonds in region 1 (green) in MalL D492R, which correlates with the decrease in temperature factors observed above, and indicates this region is significantly stabilised relative to MalL wildtype. This is counteracted by a reduced number of extra hydrogen bonds in region 3. MalL T150R has relatively few extra hydrogen bonds observed in domain B and domain C, despite there being several extra hydrogen bonds in the MalL wildtype structure when compared against MalL T150R. This indicates that the T150R has disrupted the existing bonding network, without forming additional hydrogen bonding interactions in this region.

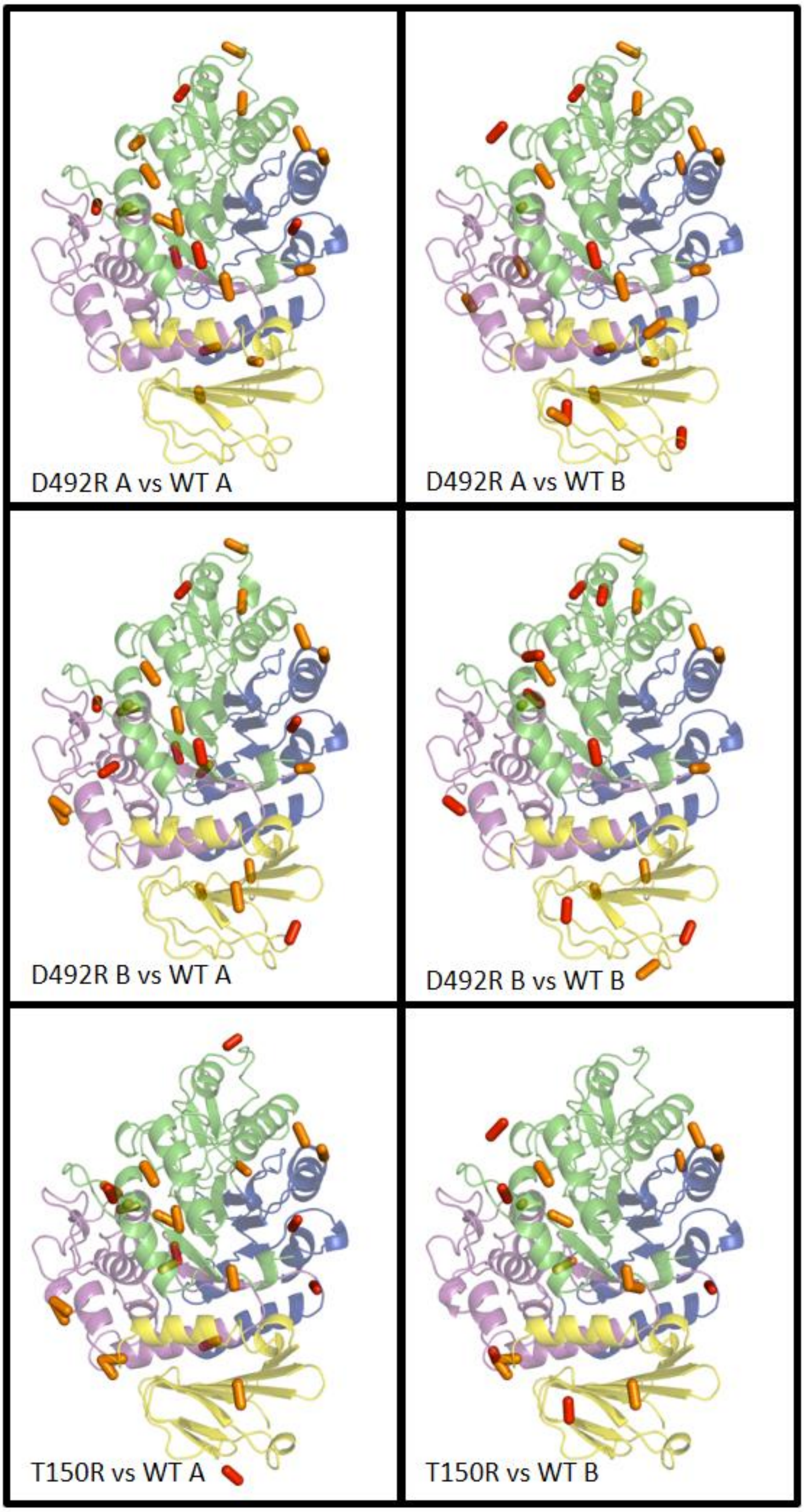


Figure 5.17. Additional hydrogen bonding in arginine mutants, indicating bonds not found in MalL wildtype (Yellow) and bonds significantly shortened (>0.3 Å) in each arginine mutant (Red). Chain identifiers are noted where appropriate.

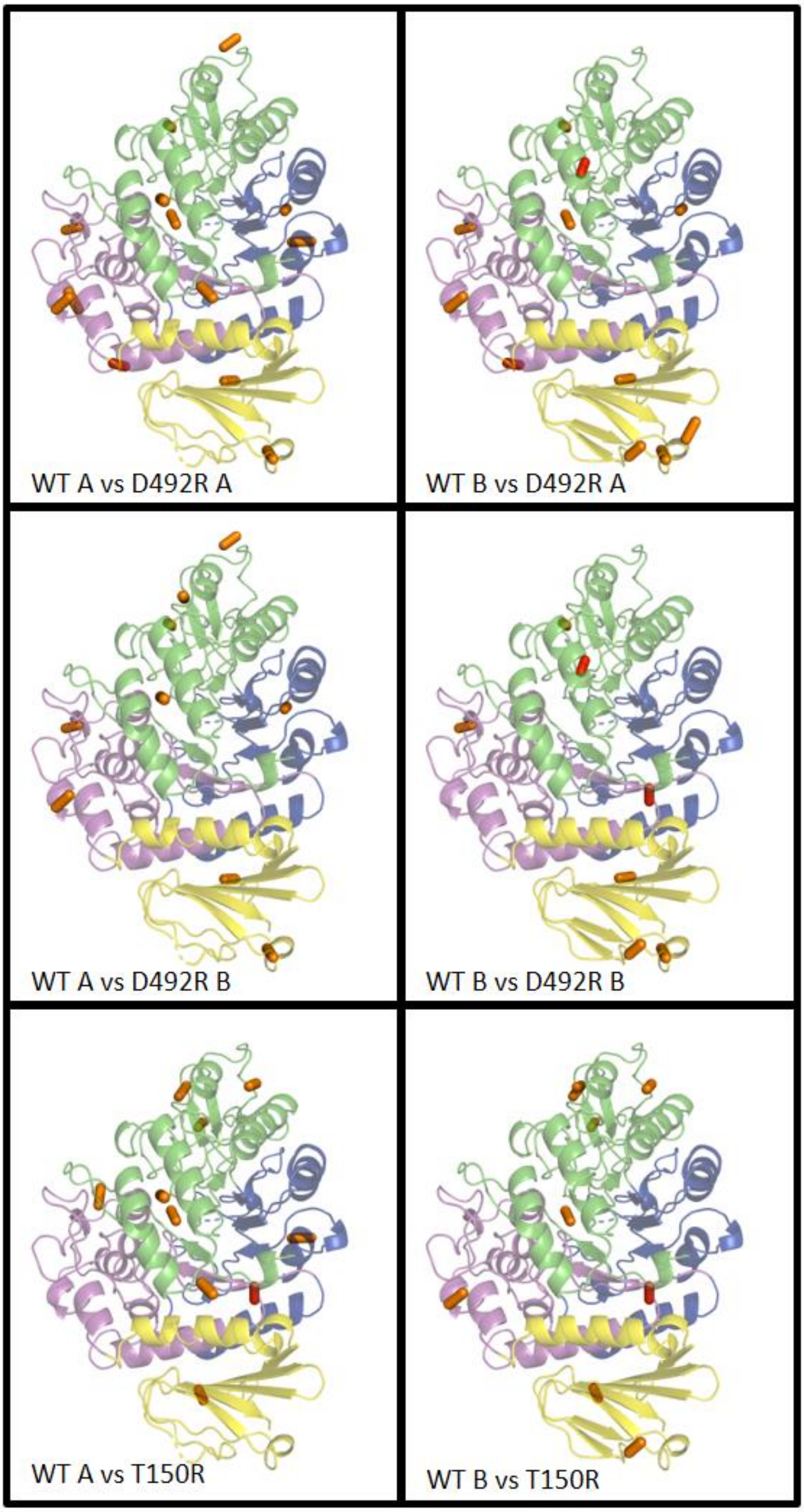


Figure 5.18. Additional hydrogen bonding in MalL wildtype, indicating bonds not found in each arginine mutant (Yellow) and bonds significantly shortened (>0.3 Å) in MalL wildtype (Red). Chain identifiers are noted where appropriate.

5.4.3.1 Hydrogen bonds contributing to structure rigidification

To identify hydrogen bonds that contribute most to the stabilisation of the MalL variants, the hydrogen bonds of all variants were compared. This analysis identified five additional hydrogen bonds present in MalL D492R (Chain A and Chain B), MalL T150R, and MalL S536R, that were not present in MalL wildtype (Chain A and Chain B) (Figure 5.19, Orange). A further five bonds were identified that were present in at least three of the stabilised mutant structures (Figure 5.19, Red). One bond was identified that was significantly shortened in all mutant structures (Figure 5.19, Cyan). A full list of bonds is available in Section 8.6.

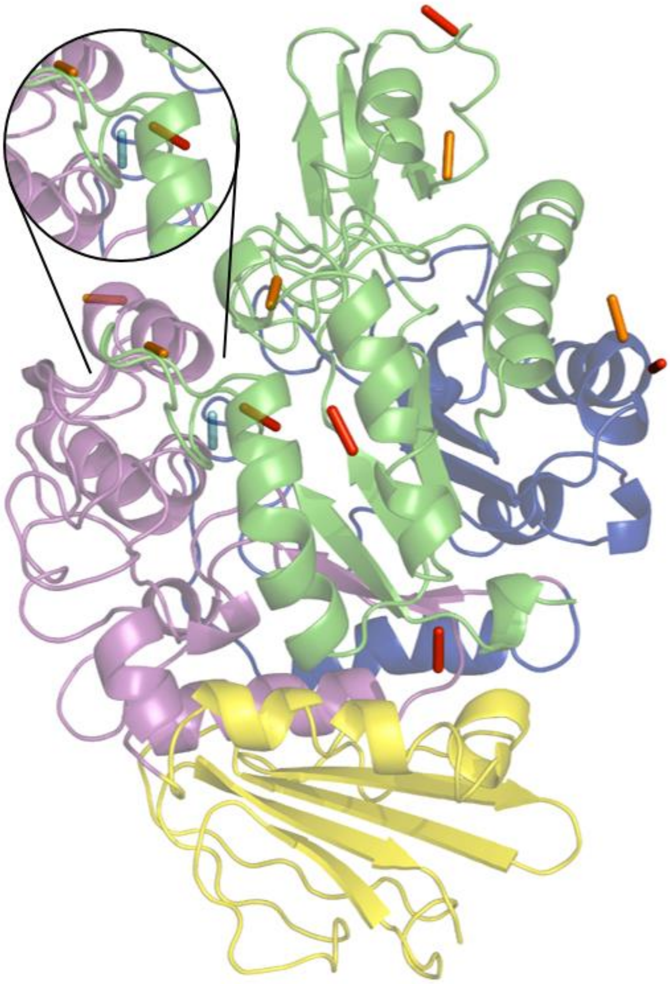


Figure 5.19. Additional hydrogen bonds contributing to stabilisation of MalL variants. Orange bonds are present in all (D492R Chain A and Chain B, T150R, S536R) structures. Orange are present in at least 3. Cyan bond is significantly shortened (>0.3 Å) in all structures compared to MalL wildtype. Inset shows position of shortened bond.

All but one of these additional hydrogen bonds are located on the surface of the structure, primarily in regions one (Figure 5.19, Green) and three (Figure 5.19, Magenta). Hydrogen bonds in these locations are consistent with a structure that is moved towards the TLC, as we have shown with MalL S536R (Section 3.4.1). Notably, there were no common hydrogen bonds identified in the C-terminal domain.

Three bonds of particular interest were identified. The hydrogen bonds between Arg186 and Glu243, and Lys475 and Asp45 connect region 1 (green) with regions 4 (yellow) and 2 (blue), respectively (Figure 5.20 A). A third hydrogen bond between Arg422 and Asp332 connects the lid domain (magenta) with the catalytic A domain (blue) (Figure 5.20 B). This bond is also particularly of note as Asp332 is one of the catalytic residues. Asp332 is pulled slightly away (~0.3 Å) towards Arg422, compared to MalL wildtype. As these additional hydrogen bonds which join domains are common to each structure, they are potentially crucial in the effects seen in catalysis and the increased resolution of these structures.

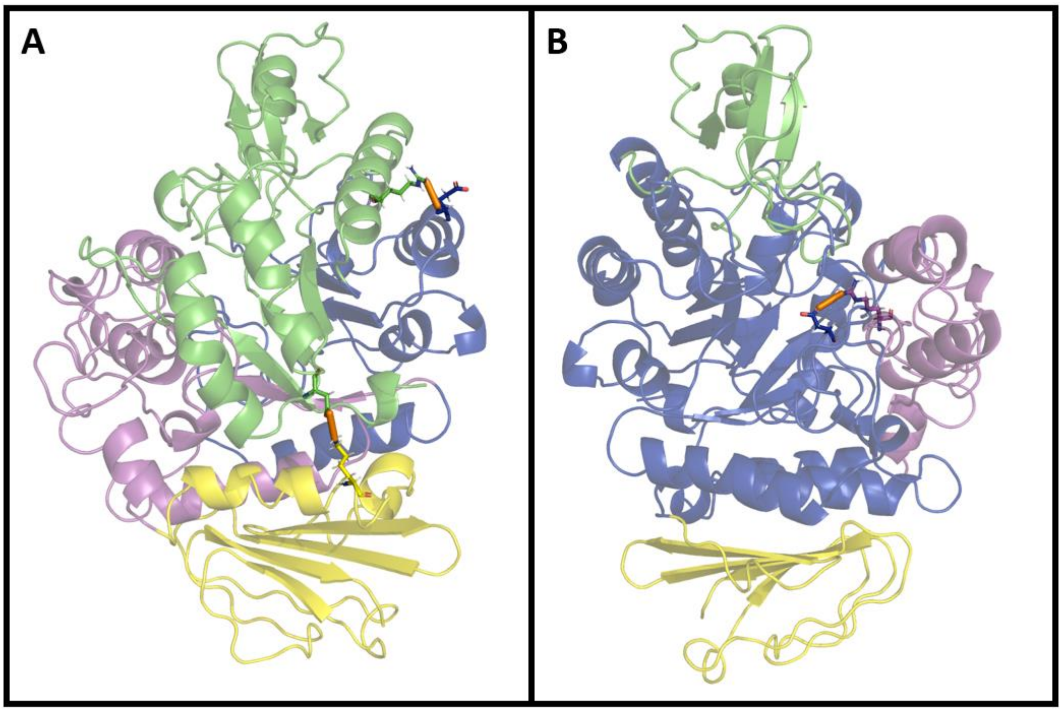


Figure 5.20. Additional hydrogen bonds contributing to stabilisation of MalL variants. A) Hydrogen bonds connect region 1 (green) with regions 2 (blue) and 4 (yellow). B) Hydrogen bond connects domain A (blue) with the 'Lid' domain.

5.4.4 Structural analysis and MMRT

The MalL T150R kinetics suggest that it behaves as a traditional psychrophile with an increase in flexibility in the ground state, and an increased activation heat capacity (Section 4.8) (Figure 5.21). The lack of excess hydrogen bonds in domains B and C, coupled with the increase in temperature factors in domain C, indicate that flexibility in distal regions play an important role in adaptation to temperature (Feller & Gerday, 2003). Tuning the flexibility of surface residues has been shown to mediate a transition between psychrophilic and mesophilic activity (Isaksen *et al.*, 2016). The MalL T150R mutation does not overlay the urea binding site, and instead packs against Lys165 on an adjacent loop, which is much more ordered in the MalL T150R structure, compared to MalL wildtype where it is disordered and missing in the electron density. This results in a stabilising interaction, with the burying of some hydrophobic surface area, and may contribute to the decrease in temperature factors observed in this region (residues 150-250). This mutation does not have a clear mechanism of action for its effect of the enzyme catalytic rate and may simply act by adding entropy to the system, destabilising the ground state.

MalL D492R is characterised by a constrained conformation consistent with a structure more like the TLC, as was seen for MalL S536R which was trapped in the TLC (Walker *et al.*, 2023) (Figure 5.21). This can be seen in the structure in the increased number of hydrogen bonds, especially in region one, whose rigidification is a large contributor into the transition state (van der Kamp *et al.*, 2018). This can also be seen in the decreased temperature factors in this region. MalL D492R also shows an increase in temperature factors in the lid domain, as well as a decreased number of hydrogen bonds in region 3 which contains the lid domain. This region is the greatest contributor to the activation heat capacity, and the greater flexibility in this region may contribute to the lowered activation heat capacity of MalL D492R at low temperatures.

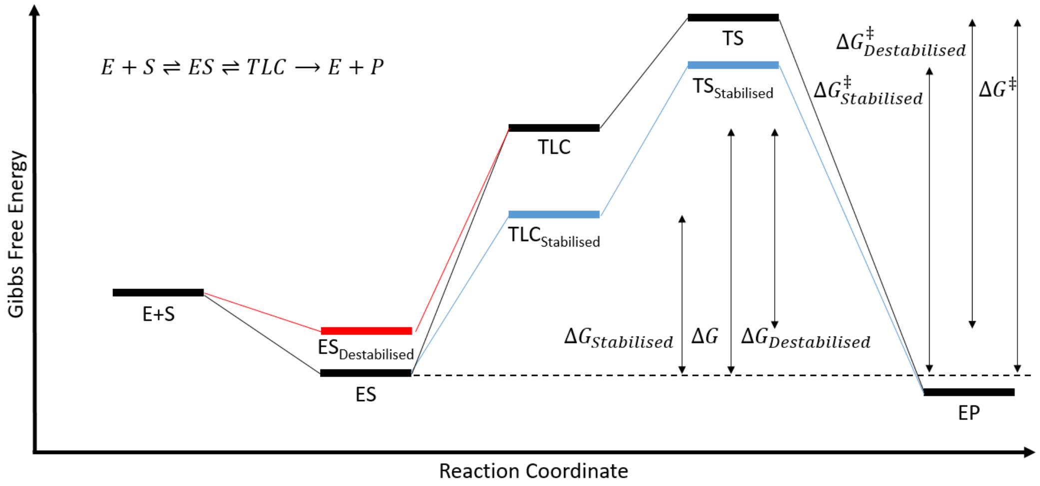


Figure 5.21. Gibbs free energy change along the reaction coordinate. MalL D492R (blue pathway) is characterised by a constrained conformation in the TLC and TS, lowering the barrier into the TLC. MalL T150R (red pathway) is characterised by increased flexibility in the ground state, leading to a similarly reduced barrier into the TLC.

MalL T150R (red pathway) is characterised by a destabilised ground state. This serves to increase the Gibbs free energy of the ES complex resulting in a lowered barrier into TLC and a lowered activation Gibbs free energy. In contrast MalL D492R (blue pathway) is characterised by stabilising interactions, with the TLC and TS having a lower Gibbs free energy, resulting in a lowered barrier into the TLC and lower overall activation Gibbs free energy.

5.4.5 Urea for rational mutation design

As with the production of the S536R mutant, the D492R and T150R mutations both produced improved protein crystals that diffracted with a resolution improvement of 1.26 and 0.84 Å, respectively. However, these mutations did not bind into the urea binding sites as expected. The mutated arginine residue in MalL D492R is involved in the formation of additional interactions to the bottom loop of the C-terminal domain, contributing to the stabilisation of this domain. In addition, the D492R mutation is also involved in the formation of additional crystal contacts not present in the wildtype enzyme. In contrast, the arginine in MalL T150R forms only one additional weak interaction with Lys165 on an adjacent loop, and is not involved in any crystal contacts. The origin of the stabilisation of T150R cannot be conclusively determined, and may instead be based in the dynamics of the solvent interaction. However both arginine mutations significantly improved crystal quality, without any changes to the general structure of the enzyme. This indicates that surface engineering of arginine residues may be a viable general method for the improvement of protein crystallisation and diffraction.

6 Graphical User Interface for Interactive Fitting of Macromolecular Rate Theory

6.1 Introduction

The application of the macromolecular rate theory (MMRT) equations to biological temperature rate data can be non-intuitive to those unfamiliar with the equations. Here we present a general purpose graphical user interface to aid in the fitting of MMRT equations.

6.1.1 Fitting of MMRT equations to data

The general application of curve fitting algorithms carries many implicit assumptions, where decisions need to be made prior to fitting, such as the initial values and constraints on any fitting parameters. This is particularly true for the MMRT equations, where several parameters are temperature dependent, and whose value depends on the choice of reference temperature. The value of the fitted parameters can change significantly with temperature. Thus, for full understanding it is important to be able to see how the value of the fit parameters vary at low, moderate and high temperatures.

The use of MMRT across multiple disciplines is becoming more common, and there is no standardised method for fitting MMRT equations. Traditional data analysis tools such as Microsoft Excel or GraphPad Prism do not have inbuilt support for MMRT equations. This particularly makes the fitting of MMRT equations difficult, as all fits require that a reference temperature be explicitly chosen. MMRT 2.0 is unable to be fit by traditional tools as it requires the evaluation of an integral function. Thus it requires a fitting procedure using a more powerful algorithm, such as those provided by programming languages, such as Python, R, or Matlab. The use of programming languages to analyse data has a high barrier of entry, and requires a reasonable amount of technical skill to set-up, run and maintain.

Here, two software packages are developed and presented to aid in the fitting and analysis of MMRT equations to biological rate data (Section 1.3). The software package 'MFit1' is designed to fit and analyse fits for the MMRT 1.0 (constant activation heat capacity) and MMRT 1.5 (linear activation heat capacity) equations. The package 'Mfit2' is designed for fitting and analysis of the MMRT 2.0 (complex activation heat capacity) equation. All versions of MMRT are fitted using Equation 6.1, and vary with different models for the activation heat capacity. MMRT 1.0 has a constant activation heat capacity and integration of Equation 6.1 yields the MMRT 1.0 equation (Equation 6.2). MMRT 1.5 has a linearly temperature dependent activation heat capacity (Equation 6.3) and integration of Equation 6.1 yields the MMRT 1.5 equation (Equation 6.4). MMRT 2.0 has a sigmoidal activation heat capacity (Equation 6.5), and cannot be fully integrated for a definite solution.

Equation 6.1. Temperature dependence of the activation Gibbs free energy

$$\Delta G^{\ddagger} = \Delta H^{\ddagger}_{T_0} - T\Delta S^{\ddagger}_{T_0} + \int_{T_0}^{T} \Delta C_P^{\ddagger}\, dT - T\int_{T_0}^{T} \frac{\Delta C_P^{\ddagger}}{T}\, dT$$

Equation 6.2. MMRT 1.0 equation (constant activation heat capacity)

$$\Delta G^{\ddagger} = \Delta H^{\ddagger}_{T_0} - T\Delta S^{\ddagger}_{T_0} + \Delta C_P^{\ddagger}(T - T_0 - T\, ln(T/T_0))$$

Equation 6.3. Linear activation heat capacity

$$\Delta C_P^{\ddagger} = \Delta C_{P,0}^{\ddagger} + mT$$

Equation 6.4. MMRT 1.5 equation (linear activation heat capacity)

$$\Delta G^{\ddagger} = \Delta H^{\ddagger}_{T_0} - T\Delta S^{\ddagger}_{T_0} + \Delta C_{P,0}^{\ddagger}(T - T_0 - T\ln(T/T_0)) - \frac{m}{2}(T - T_0)^2$$

Equation 6.5. Sigmoidal activation heat capacity

$$\Delta C_P^{\ddagger} = \frac{\left(\Delta C_{P,lowT}^{\ddagger}\right) + \left(\Delta C_{P,highT}^{\ddagger}\right)e^{\frac{-\Delta\Delta H^{\ddagger}(1-T/T_C)}{RT}}}{1 + e^{\frac{-\Delta\Delta H^{\ddagger}(1-T/T_C)}{RT}}}$$

6.1.2 Non-linear least squares curve fitting

Non-linear least squares (NLLS) is a mathematical method that aims to estimate the value of a set of parameters of a given model, that produces the 'best fit' to a given set of experimentally observed data. The objective of the 'best fit' is to find a set of optimal parameters for the fitting function, such that the fitting function will correctly describe the data and average out experimental uncertainties (Johnson, 1992). During a least squared fit, the 'best fit' is that the sum of the squared distances between the model fit and observed data is at a minimum when evaluated at the optimal parameter values. For this reason, the procedure may be referred to as an optimisation (parameter values are optimised) or a minimisation (residuals are minimised).

All non-linear least squares procedures require that the user provides an initial estimate of the parameter values. The quality of the parameter values can be scored using the residuals. The initial parameters are used as inputs to find a better estimate of the parameters. The new values of the parameters are then used to iteratively find even better estimates of the parameters, until the parameter values have stabilised within some specified limit (Johnson, 1992) (Figure 6.1).

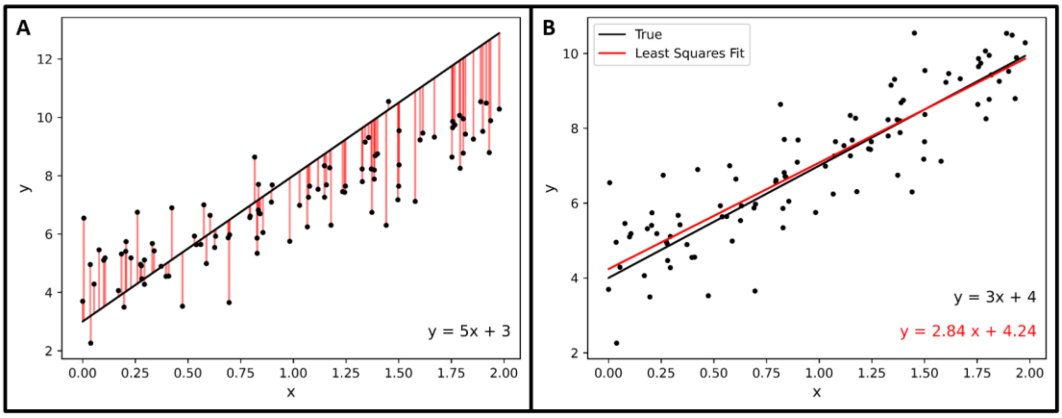


Figure 6.1. Non-linear least squared fitting of a simple linear function. Data x values are evenly distributed. Data y values are randomly distributed around y=3x+4. A) The non-linear least squares optimisation uses the sum of squared residuals as an objective function to score the quality of estimated parameter values. Residuals for parameter values slope=5 and intercept=3 (away from the minimum) are shown in red. B) The least squares fit is obtained by minimising the sum of square residuals. Obtained parameter values are only an estimation of their true value.

Two common methods for solving non-linear least squares problems are gradient descent and the Gauss-Newton algorithm. The Gauss-Newton algorithm treats the local region of parameter space as linear. Linear least squares (where all parameters are linearly related) can be solved in a single iteration. Thus locally optimised parameter values can be found. By iteratively finding improved parameter values the parameters will converge at a local minimum where no further improvements can be made. Gradient descent works by calculating the gradient surrounding a set of parameter values. The parameter values are then iteratively optimised by stepping in the direction opposite to the steepest gradient (the steepest gradient points towards a maxima). The Levenberg-Marquardt algorithm is a combination of these two methods and uses a scaling factor to adaptively switch between both methods. Further away from a minimum it acts more like a gradient descent algorithm, and closer to a minimum it acts more like the Gauss-Newton algorithm.

The parameter space for a simple linear function of the form $y = ax + b$ can be represented in three dimensions (Figure 6.2). The X and Y dimensions represent different potential values of a and b. The Z dimension is the root mean square error (RMSE) calculated as the square root of the mean of the sum of squared residuals (Equation 6.6). Non-linear least squares algorithms act to find a local minima from a given set of starting parameter values. Where the data are noisy, or the parameters are not strictly independent of each other (are correlated), this may be a non-trivial task, and many such local minima may exist (de Levie, 2000). Therefore an accurate estimation of the parameter values is critical for a true ‘best fit’.

Equation 6.6. RMSE equation

$$RMSE = \sqrt{\frac{1}{n}\sum\left(\hat{Y}_i - Y_i\right)^2}$$

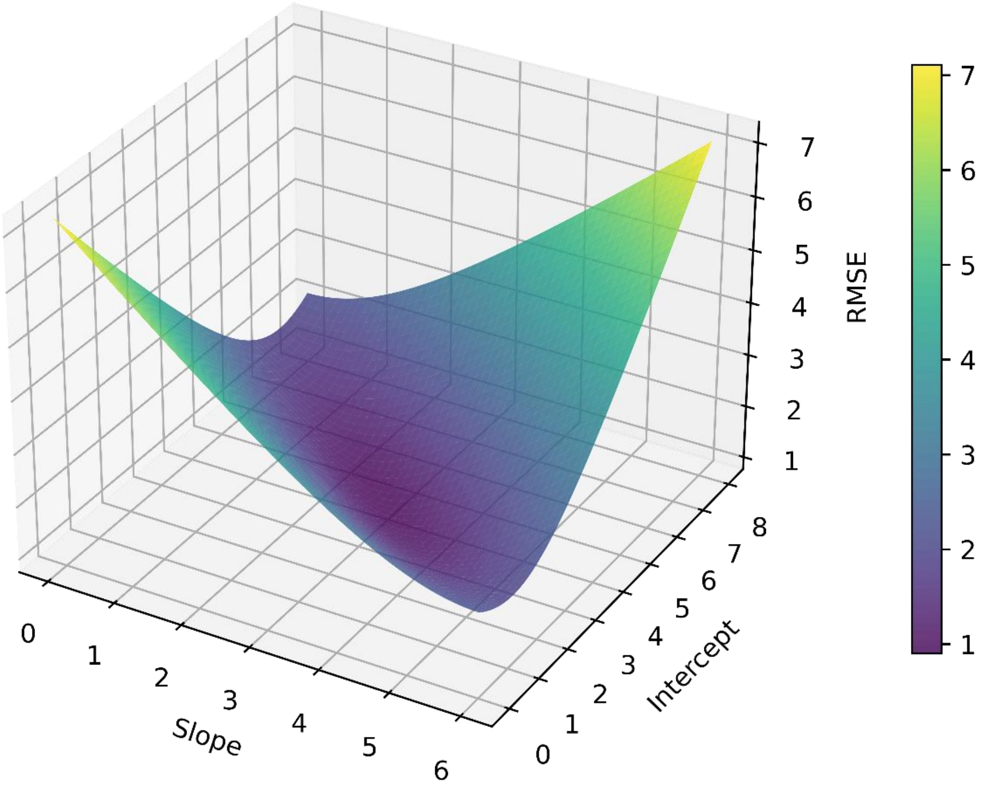


Figure 6.2. General principle of non-Linear Least-Squares regression. Z-axis is the square root of the mean of the sum of squared residuals (RMSE). RMSE is at a minimum when parameters are at an optimal value. Here, the true parameter values are slope=3, Intercept=4. The optimised parameter values are slope=2.84, Intercept=4.24. Broad minimum hinders convergence on exact true parameter values.

6.2 General overview

Both MFit1 and MFit2 can accept temperature rate data in a wide variety of formats as input. These data are then automatically fit with the MMRT equations using default parameters. From there the fit of individual datasets can be adjusted, and the temperature profiles of fit parameters inspected. When a fit has been completed, plots and fit statistics can be exported from the program.

6.3 Use of MFit software

6.3.1 Program start-up

When the program is started the user is presented with the home screen (Figure 6.3). From here the user can close or restart the program using the options in the file menu, access information about the program, open the user manual in the help menu, or click on the start button to proceed with data input. This will launch the data input screen (Figure 6.4).

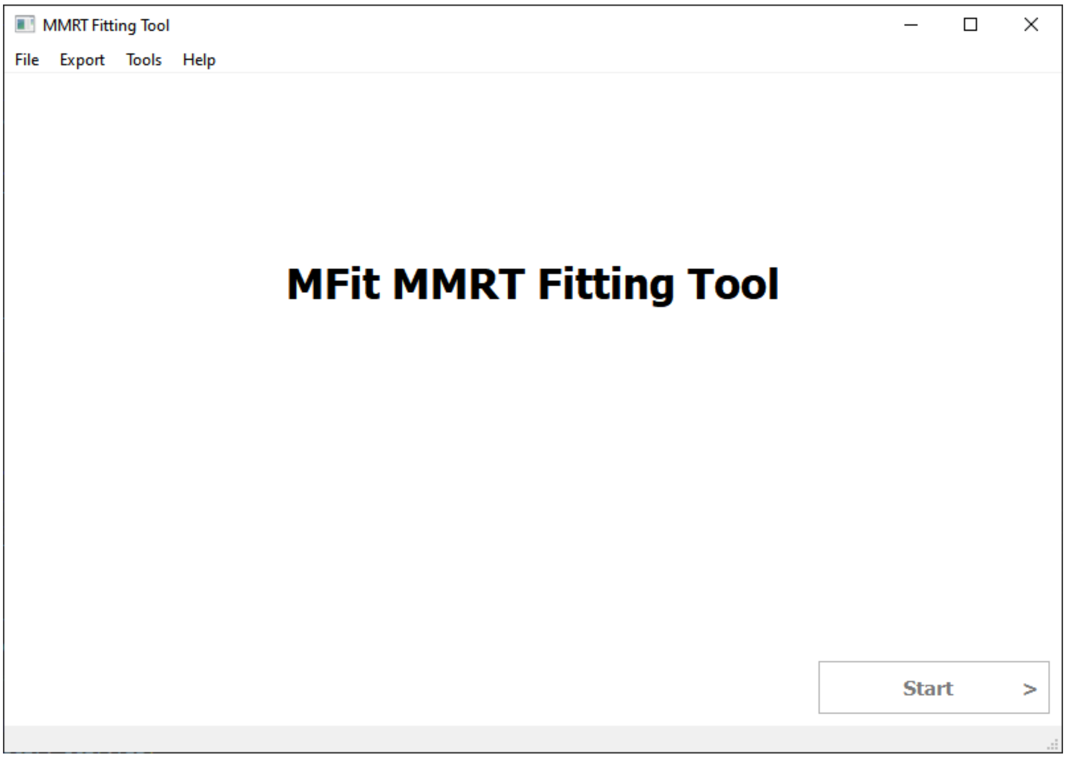
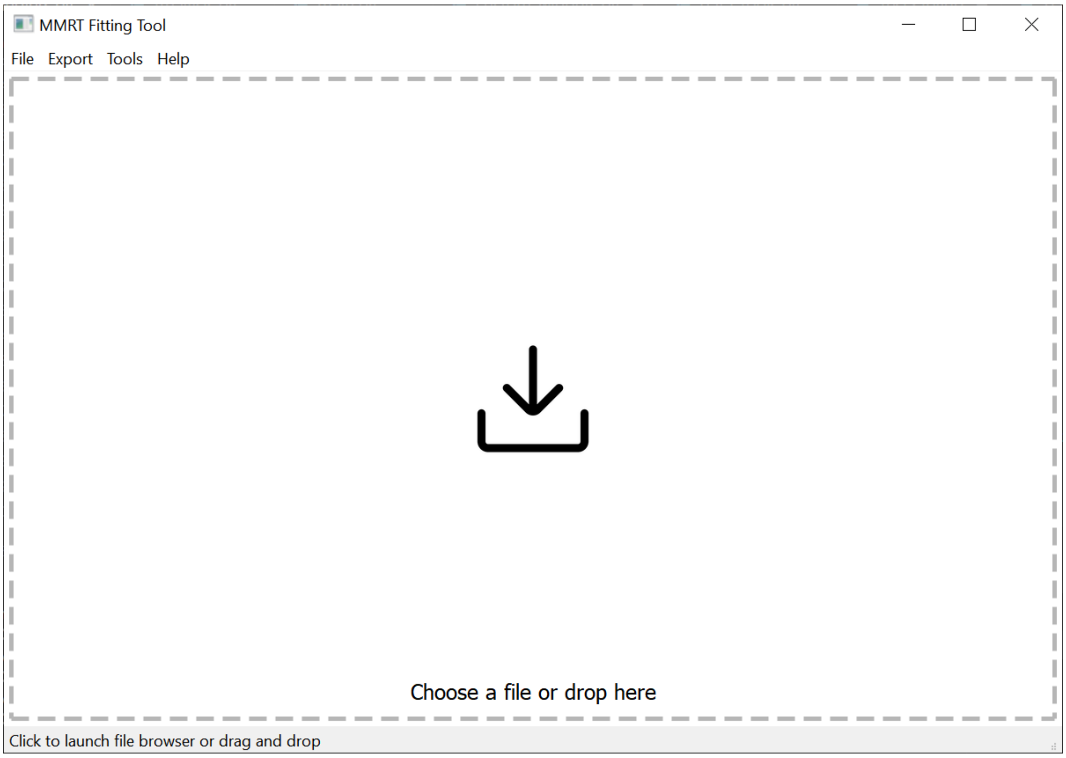


Figure 6.3. Home screen of MFit MMRT Fitting Tool

6.3.2 File input and data processing

To import data into MFit, the user can click anywhere in the program window to launch the file browser, or drag and drop files into the program window. MFit accepts any number of files, thus data from multiple files can be analysed in one session.



Choose a file or drop here

Click to launch file browser or drag and drop

Figure 6.4. Data input screen of MFit MMRT Fitting Tool

MFit accepts a wide variety of tabulated data as input. It can accept excel formatted files (.xlsx, .xls, .xlsm), comma-separated values (.csv), tab-delimited files (.tab, .txt, .tsv), space-delimited files (.txt), or MFit session files (Section 6.3.10) (.mmrt). Files should be formatted in columns with temperature data in the first column, and replicate rate data in subsequent columns. A header row containing titles may be optionally present. Temperature may be inputted as either Celsius or Kelvin. If the temperature values are less than 150 they will be automatically converted to Kelvin, which is required for fitting.

Rate data can be accepted directly in terms of the rate constant (k), as the natural log of the rate constant ($\ln(k)$), or as activation Gibbs free energy values ($\Delta G^{\ddagger}$). The activation Gibbs free energy can be calculated from the rate data using transition state theory, as Equation 6.7.

Equation 6.7. Eyring-Polanyi equation rearranged for ΔG‡

$$\Delta G^{\ddagger} = RT\left[\ln\left(\frac{k_B T}{h}\right) - \ln(k)\right]$$

As the data are imported they will be sorted in order of temperature. When data are imported, the MFit software will attempt to automatically determine the rate type (k, $\ln(k)$, $\Delta G^{\ddagger}$) (Code Snippet 6.1). This can be changed later (Section 6.3.7) if the automatic determination is incorrect. The first four data points are analysed, and if the slope of temperature versus rate is negative, the data will be imported as activation Gibbs free energy. If the rate is positive then the software attempts to distinguish between rate and log(rate) data. It does this by analysing the variance in the gradient of the rate data up to the maximum rate measured. This method relies on the assumption that log(rate) data will vary less than rate data. The gradients of all pairs of non-adjacent (at least one data point in-between) data points are measured and the standard deviations calculated. If the standard deviations of all slopes are less than the cut-off (which is set at 0.5) then the data are imported as log(rate) data. Finally, if the standard deviations are greater than the cut-off, the data are imported as rate data. Rate and log(rate) data will be converted to activation Gibbs free energy using Equation 6.7. This places a practical limit on the number of data points a dataset may contain, as the complexity of the rate determination step will rise exponentially as the number of data points increases. Large datasets (250 data points) take approximately 40 seconds to import. Datasets imported as activation Gibbs free energy (and having a negative initial slope) will be imported in constant time, regardless of dataset size.

Code Snippet 6.1. Pseudocode representation of rate type determination

```
Initial_x = array of first four temperature points
Initial_y = array of first four unknown rate data points

initial_slope = get_slope(initial_x, initial_y)
if initial_slope < 0:
    return rate is activation Gibbs free energy

truncated_x = array of temperatures truncated at max rate
truncated_y = array of rate data truncated at max rate

slopes = []
for pair in get_pairs(truncated_x, truncated_y):
    slope = get_slope(pair)
    slopes.append(slope)

standard_deviation_slopes = std(slopes)

if standard_deviation_slopes < 0.5:
    return rate is ln(rate)
else:
    return rate is rate
```

6.3.3 Data fitting

Once the data has been imported the MFit software will attempt to fit the data to the MMRT equations, using the default parameters (which can be altered later). MFit1 will fit both MMRT 1.0 and MMRT 1.5, while MFit2 will fit MMRT 2.0 (Section1.3, Section 6.1.1). The default parameters for fitting of MMRT 1.0 and MMRT 1.5 are given in Table 6.1, and Table 6.2. Different fit parameters can be provided if required (Section 6.3.6). The MMRT equations are fit by minimisation using the Levenberg-Marquardt algorithm. The program will output the parameter fit values, as well as parameter errors and a variety of fitting statistics.

Table 6.1. Default initial parameters for MMRT 1.0 fitting

Parameter	Value
$\Delta H^{\ddagger}_{T_0}\ (kJ\ mol^{-1})$	10
$\Delta S^{\ddagger}_{T_0}\ (J\ mol^{-1}K^{-1})$	-150
$\Delta C^{\ddagger}_{P,T_0}\ (kJ\ mol^{-1}K^{-1})$	-1
$T_0\ (K)$	278.15

Table 6.2. Default initial parameters for MMRT 1.5 fitting

Parameter	Value
$\Delta H^{\ddagger}_{T_0}\ (kJ\ mol^{-1})$	10
$\Delta S^{\ddagger}_{T_0}\ (J\ mol^{-1}K^{-1})$	-150
$\Delta C^{\ddagger}_{P,0}\ (kJ\ mol^{-1}K^{-1})$	-1
$m\ (J\ mol^{-1}K^{-2})$	0
$T_0\ (K)$	278.15

MMRT 2.0 is fit in two stages. In the first stage MMRT 1.0 (Equation 1.20) is fit to the low temperature section of the input rate data (Figure 6.5). The activation heat capacity is assumed to be relatively constant over this range (Figure 6.6), so is well fit by the MMRT 1.0 equation. The initial parameter values used for the fit are given in Table 6.3. The only user changeable parameter is the range of data points which define the low temperature portion of the overall dataset (Section 6.3.6.1).

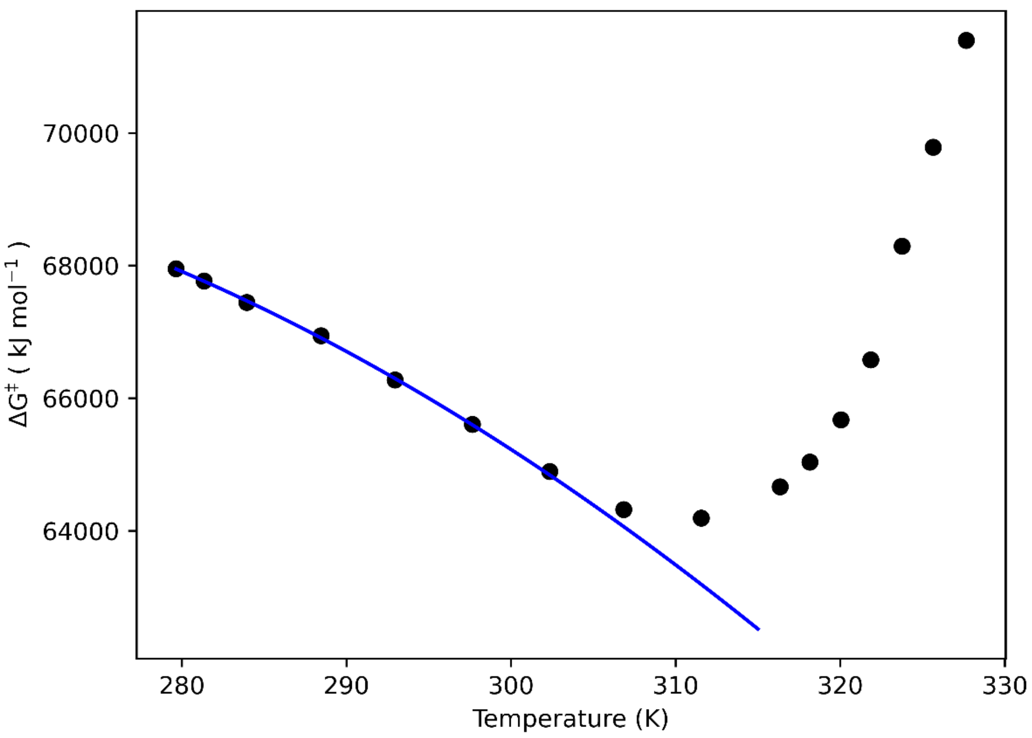


Figure 6.5. Fit of MMRT 1.0 to the low temperature arm of MalL wildtype

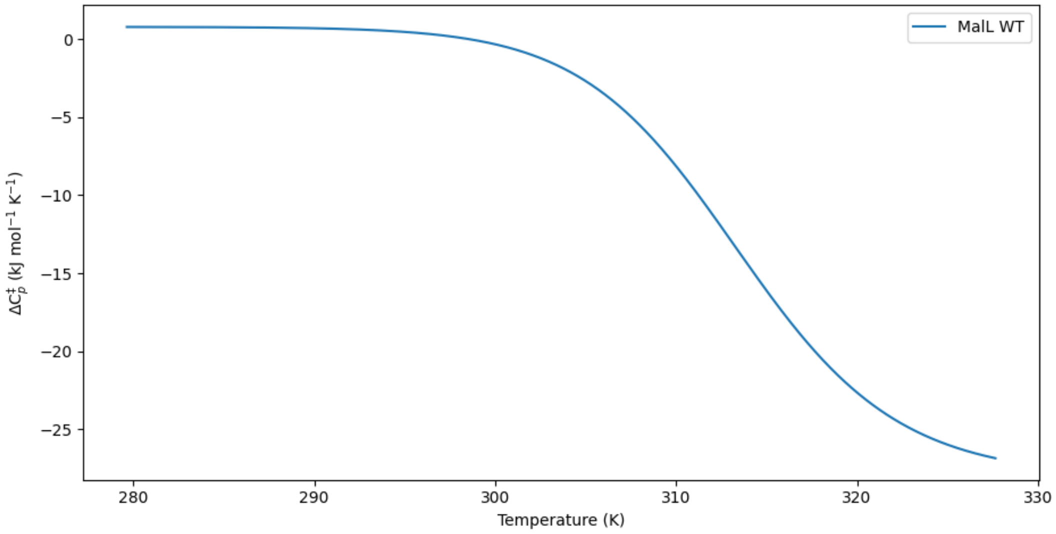


Figure 6.6. Temperature dependence of the activation heat capacity for MalL wildtype.

Table 6.3. Initial parameters for MMRT 1.0 fit of the low temperature arm

Parameter	Value
$\Delta H^{\ddagger}_{T_0}$ $(kJ\ mol^{-1})$	45
$\Delta S^{\ddagger}_{T_0}$ $(J\ mol^{-1}K^{-1})$	100
$\Delta C^{\ddagger}_{P,T_0}$ $(\Delta C^{\ddagger}_{P,lowT})$ $(kJ\ mol^{-1}K^{-1})$	0.5
$\Delta C^{\ddagger}_{P,lowT}$ **Fit Range**	1-6
T_0 (K)	288

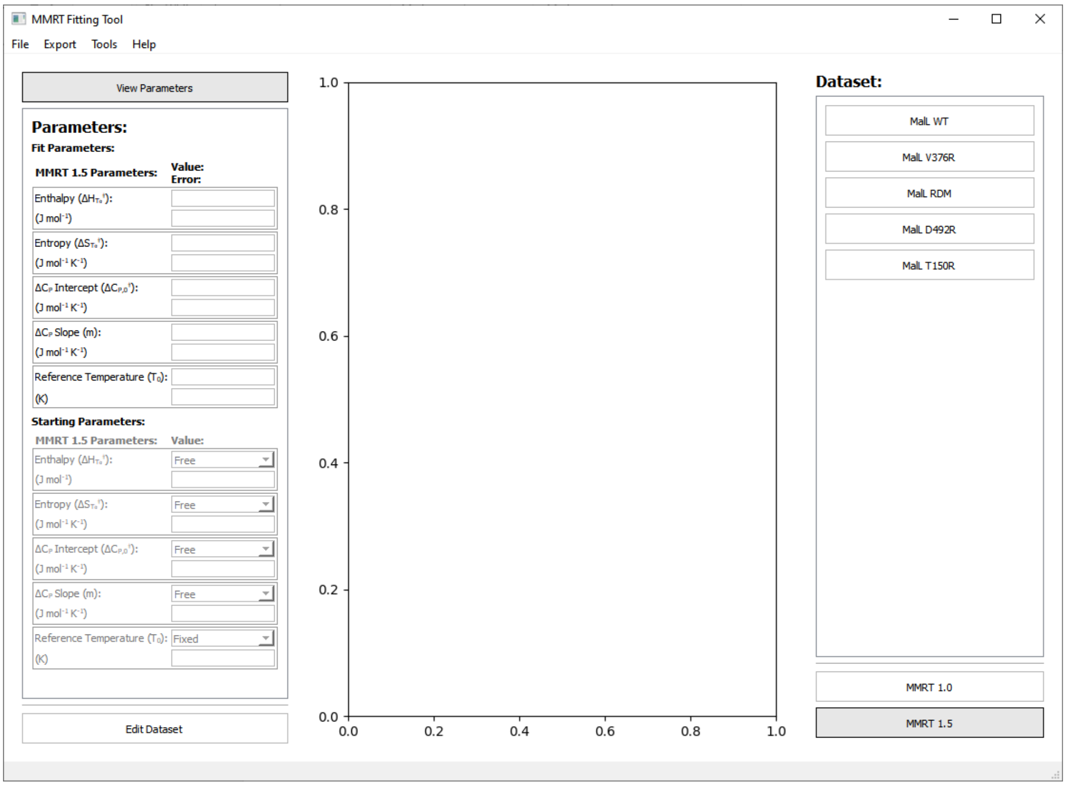
In the next stage the MMRT 2.0 equation (Section 1.3) is fit to the data with the activation heat capacity (low temperature) being held constant at the value found in the previous stage. The midpoint of the transition (T_C) is initially set to the temperature value at the data point with the minimum value of activation Gibbs free energy, and is constrained to be within the temperature range of the dataset. Initial values for the other MMRT 2.0 parameters are given in Table 6.4.

Table 6.4. Default initial parameters for MMRT 2.0 fitting

Parameter	Value
$\Delta H^{\ddagger}_{T_0}$ $(kJ\ mol^{-1})$	45
$\Delta S^{\ddagger}_{T_0}$ $(J\ mol^{-1}K^{-1})$	100
$\Delta\Delta H^{\ddagger}$ $(kJ\ mol^{-1})$	90
$\Delta C^{\ddagger}_{P,highT}$ $(kJ\ mol^{-1}K^{-1})$	20
T_0 (K)	278.15

6.3.4 Data display and plotting

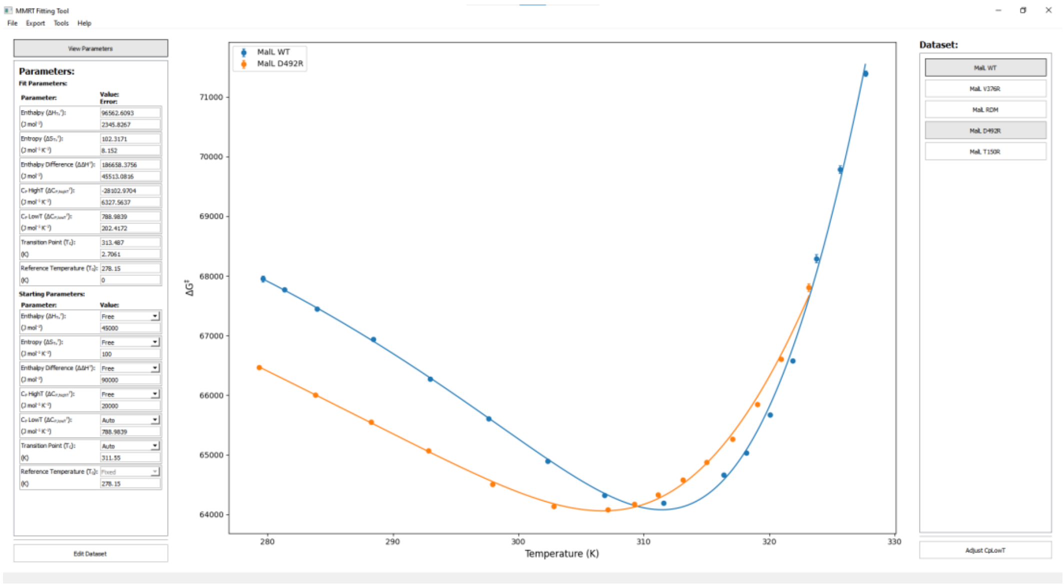
From here the user is presented with the main screen of the program. The MFit1 program screen is split into three sections (Figure 6.7). On the left is the parameter viewer and editor, as well as dataset edit options (shown by clicking ‘Edit Dataset’). In the centre is the plot window, where datasets and their fits are displayed. On the right is the dataset selector panel, and buttons to select the fit type displayed (MMRT 1.0 or MMRT 1.5).



Enthalpy (ΔH$_{T_a}$‡): Free
Entropy (ΔS$_{T_a}$‡): Free
ΔC$_P$ Intercept (ΔC$_{P,0}$‡): Free
ΔC$_P$ Slope (m): Free
Reference Temperature (T$_0$): Fixed

Figure 6.7. Main screen of the MFit1 program.

MFit2 is similarly configured, with controls for adjusting the range of the fit of activation heat capacity at low temperatures, in place of the fit type selection panel (Figure 6.8). Clicking on a dataset button will plot the dataset and associated fit curve. Data are plotted with error bars alongside a curve calculated from the fit values. Holding ‘control’ and clicking a dataset will overlay that dataset onto the plot area. The selected dataset is indicated by a black border (Figure 6.8).



Parameters:

Fit Parameters:

Parameter:	**Value:** **Error:**
Enthalpy ($\Delta H_{T_0}^{\ddagger}$): (J mol^{-1})	96562.6093 2345.8267
Entropy ($\Delta S_{T_0}^{\ddagger}$): (J mol^{-1} K^{-1})	102.3171 8.152
Enthalpy Difference ($\Delta\Delta H^{\ddagger}$): (J mol^{-1})	186658.3756 45513.0816
C_P HighT ($\Delta C_{P,HighT}^{\ddagger}$): (J mol^{-1} K^{-1})	-28102.9704 6327.5637
C_P LowT ($\Delta C_{P,LowT}^{\ddagger}$): (J mol^{-1} K^{-1})	788.9839 202.4172
Transition Point (T_C): (K)	313.487 2.7061
Reference Temperature (T_0): (K)	278.15 0

Starting Parameters:

Parameter:	**Value:**
Enthalpy ($\Delta H_{T_0}^{\ddagger}$): (J mol^{-1})	Free 45000
Entropy ($\Delta S_{T_0}^{\ddagger}$): (J mol^{-1} K^{-1})	Free 100
Enthalpy Difference ($\Delta\Delta H^{\ddagger}$): (J mol^{-1})	Free 90000
C_P HighT ($\Delta C_{P,HighT}^{\ddagger}$): (J mol^{-1} K^{-1})	Free 20000
C_P LowT ($\Delta C_{P,LowT}^{\ddagger}$): (J mol^{-1} K^{-1})	Auto 788.9839
Transition Point (T_C): (K)	Auto 311.55
Reference Temperature (T_0): (K)	Fixed 278.15

Figure 6.8. Main screen with dataset selected

6.3.5 Dataset fit success

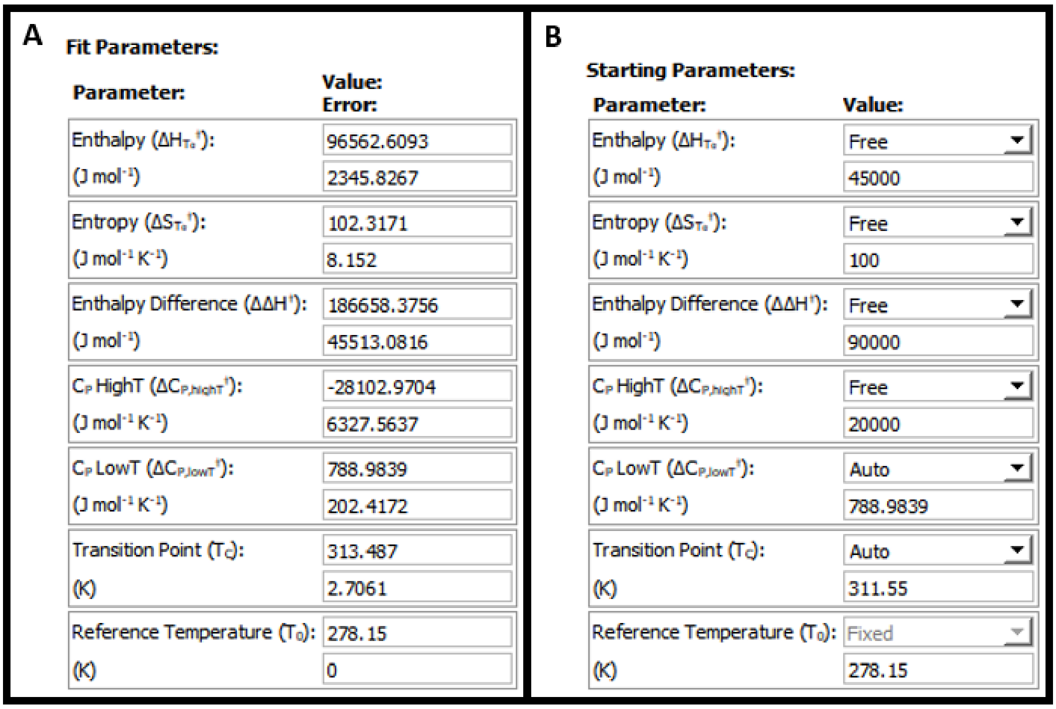
The dataset selector buttons will display information about the success of the fit for each dataset (right hand side, Figure 6.8). If the fit was unable to converge on a solution the text will turn red. This can occur if the values of the initial parameters are too far away from the true values, or the fit times out after 5 seconds. If a solution is only partly found the text will turn yellow. This most commonly occurs for fits of MMRT 2.0 where the transition point (T_C) does not converge and is close to the minimum or maximum value.

6.3.6 Viewing and editing parameters

Selecting a dataset will also populate the parameter panel with the details of the individual dataset fit (left hand side, Figure 6.8). The parameter panel is split into two components. The first shows the result and standard error of the fit to the data (Figure 6.9 A), and the second shows the starting values and constraints of the parameters used to initiate the fit (Figure 6.9 B).

The initial parameters used for the fitting procedure can be altered by the user. This may be needed if the default parameters do not converge on an appropriate solution for the input dataset. The values in the text fields can be directly edited.

Parameters can be fixed at the initial value by changing the combo box for each parameter to ‘Fixed’. The reference temperature (T_0) is always fixed at the initial value. The low temperature activation heat capacity is fixed by default at the pre-defined value (Section 6.3.3). This is represented in the parameter combo box as ‘Auto’. This parameter can be set to vary from an entered value by changing the combo box to ‘Custom’ or fixed at an entered value by changing the combo box to ‘’Fixed”. The initial value of the transition temperature is also automatically determined from the data (as the temperature at which the activation Gibbs free energy is at a minimum), however is allowed by default to vary during the fit. This is represented in the parameter combo box as ‘Auto’. As with the low temperature activation heat capacity, the transition temperature can be either fixed or allowed to vary from an entered value by changing the combo box for this parameter to ‘Custom’ or ‘Fixed’. Editing any parameter initial value or constraint will rerun the fit and the updated plot and final parameter values will be displayed.



Fit Parameters:

Parameter:	Value: Error:
Enthalpy ($\Delta H_{T_0}^{\ddagger}$): ($J\ mol^{-1}$)	96562.6093 2345.8267
Entropy ($\Delta S_{T_0}^{\ddagger}$): ($J\ mol^{-1}\ K^{-1}$)	102.3171 8.152
Enthalpy Difference ($\Delta\Delta H^{\ddagger}$): ($J\ mol^{-1}$)	186658.3756 45513.0816
C_P HighT ($\Delta C_{P,highT}^{\ddagger}$): ($J\ mol^{-1}\ K^{-1}$)	-28102.9704 6327.5637
C_P LowT ($\Delta C_{P,lowT}^{\ddagger}$): ($J\ mol^{-1}\ K^{-1}$)	788.9839 202.4172
Transition Point (T_C): (K)	313.487 2.7061
Reference Temperature (T_0): (K)	278.15 0

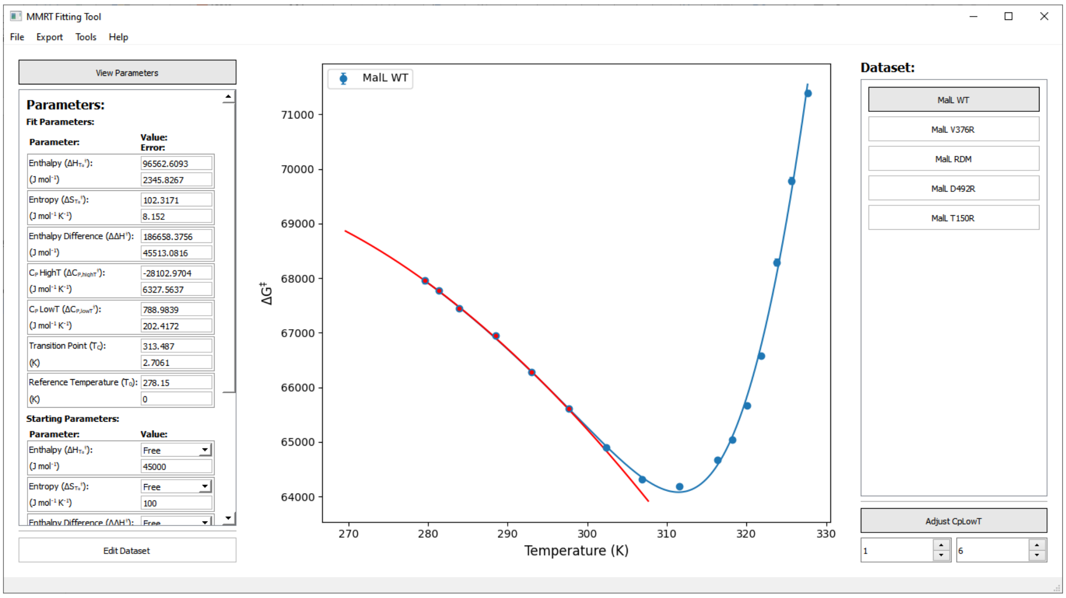
Starting Parameters:

Parameter:	Value:
Enthalpy ($\Delta H_{T_0}^{\ddagger}$): ($J\ mol^{-1}$)	Free 45000
Entropy ($\Delta S_{T_0}^{\ddagger}$): ($J\ mol^{-1}\ K^{-1}$)	Free 100
Enthalpy Difference ($\Delta\Delta H^{\ddagger}$): ($J\ mol^{-1}$)	Free 90000
C_P HighT ($\Delta C_{P,highT}^{\ddagger}$): ($J\ mol^{-1}\ K^{-1}$)	Free 20000
C_P LowT ($\Delta C_{P,lowT}^{\ddagger}$): ($J\ mol^{-1}\ K^{-1}$)	Auto 788.9839
Transition Point (T_C): (K)	Auto 311.55
Reference Temperature (T_0): (K)	Fixed 278.15

Figure 6.9. MFit2 Parameter Panel A) Values of fitted parameters with standard error. B) Initial values and constraints used to fit the dataset.

6.3.6.1 Editing low temperature activation heat capacity

The range of the low temperature activation heat capacity fit can be changed by the user. Clicking on ‘Adjust CpLowT’ will overlay the MMRT 1.0 fit to the limited data range (by default the first six data points) and show the control panel for editing the range (Figure 6.10). The range is controlled by editing which data points are included in the fit. The data points in the current fit are indicated by red dots on the plot. Changing the values in the spin boxes will change which data points that are included, and the value of the low temperature activation heat capacity will be re-determined, and this new value will be used for a new fit of the full MMRT 2.0 equation. In selecting the number of data points to fit here, the user should assess the quality of the fit to the chosen data points, and that the chosen data points are representative of the low temperature, constant heat capacity range (Figure 6.5, Figure 6.6). Choosing an appropriate range here is required for accuracy in fit parameters (when reported at low temperatures), as well as the enthalpy difference and transition temperature which report on the transition between low and high temperatures.



Parameters:

Fit Parameters:

Parameter:	Value:	Error:
Enthalpy ($\Delta H_{T_0}^{\ddagger}$): (J mol^{-1})	96562.6093	2345.8267
Entropy ($\Delta S_{T_0}^{\ddagger}$): (J mol^{-1} K^{-1})	102.3171	8.152
Enthalpy Difference ($\Delta\Delta H^{\ddagger}$): (J mol^{-1})	186658.3756	45513.0816
C_P HighT ($\Delta C_{P,highT}^{\ddagger}$): (J mol^{-1} K^{-1})	-28102.9704	6327.5637
C_P LowT ($\Delta C_{P,lowT}^{\ddagger}$): (J mol^{-1} K^{-1})	788.9839	202.4172
Transition Point (T_C): (K)	313.487	2.7061
Reference Temperature (T_0): (K)	278.15	0

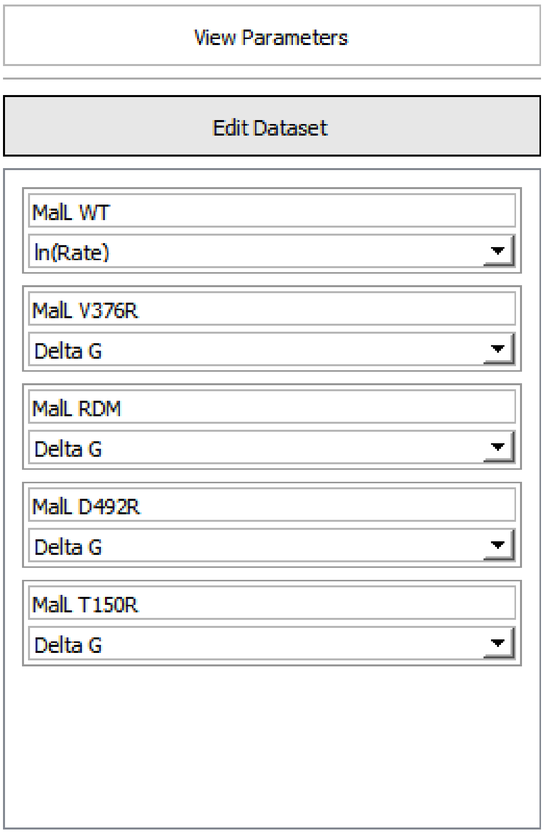
Starting Parameters:

Parameter:	Value:	
Enthalpy ($\Delta H_{T_0}^{\ddagger}$): (J mol^{-1})	Free	45000
Entropy ($\Delta S_{T_0}^{\ddagger}$): (J mol^{-1} K^{-1})	Free	100
Enthalpy Difference ($\Delta\Delta H^{\ddagger}$):	Free	

Figure 6.10. Low temperature activation heat capacity fit controls. Blue curve is the fit of MMRT 2.0 to the data. The red curve is the limited fit of MMRT 1.0 to fix $\Delta C_{P,lowT}^{\ddagger}$.

6.3.7 Editing dataset information

The dataset edit panel can be displayed by clicking ‘Edit Dataset’. From here the dataset name can be changed. The type of rate data can also be changed using the combo box, if the automatic determination was not accurate (Figure 6.11).



MalL WT
ln(Rate)

MalL V376R
Delta G

MalL RDM
Delta G

MalL D492R
Delta G

MalL T150R
Delta G

Figure 6.11. Dataset editor panel

6.3.8 Excluding data points from fit

In the ‘Tools’ menu the ‘Edit Data Points’ option can be turned on (Figure 6.12). This allows individual data points to be excluded from the fit. Excluded data points are shown in red. To toggle exclusion of a data point, the point on the plot can be clicked, or by holding left click and dragging, multiple data points can be excluded at once (Figure 6.13).

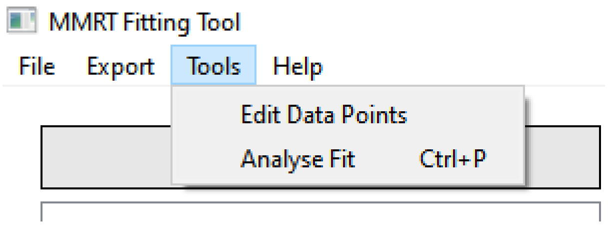


Figure 6.12. Edit Data points setting.

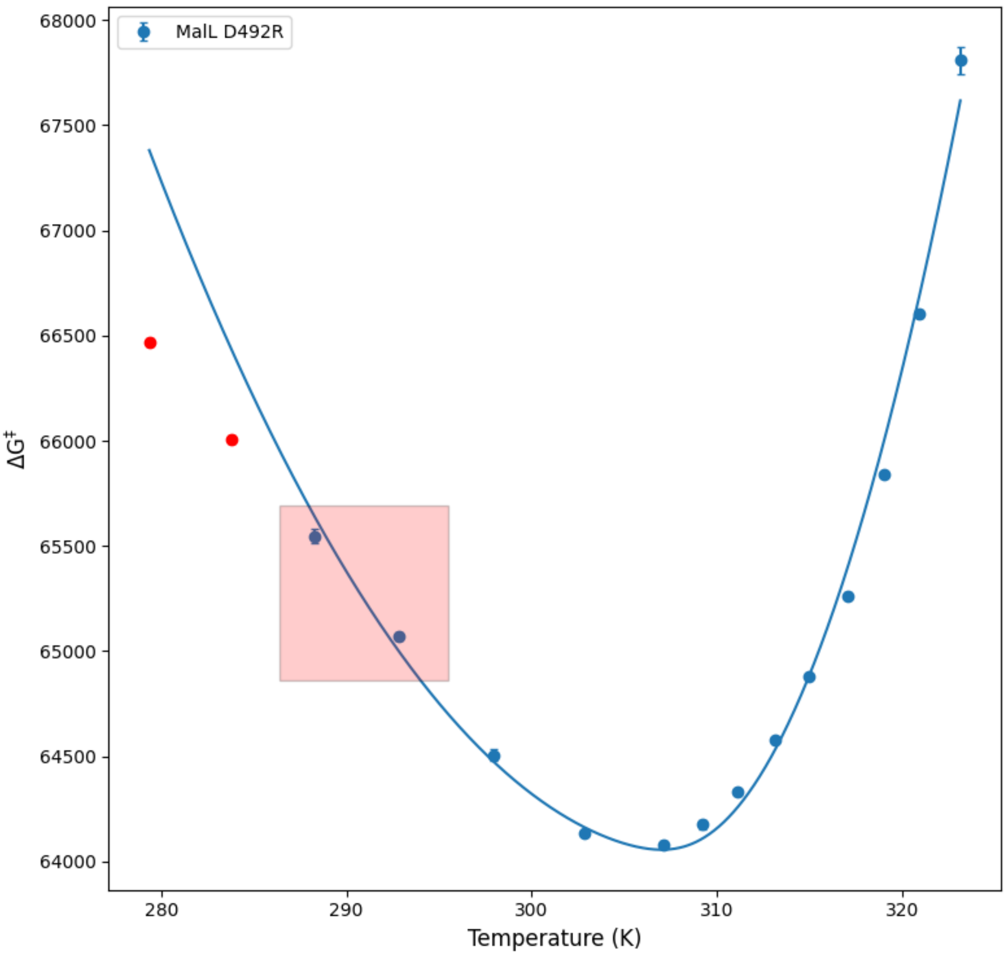


Figure 6.13. Exclusion of data points. Excluded data points are shown in red. Multiple points can be excluded by clicking and dragging (red box).

6.3.9 Viewing temperature dependence of fit parameters

The temperature dependence of the fit parameters can be seen by clicking on ‘Analyse Fit’ in the ‘Tools’ menu. This will bring up a window showing the plots for the temperature dependence of the activation Gibbs free energy, activation enthalpy, activation entropy (as TΔS‡), and activation heat capacity (Figure 6.14). The plots for any dataset selected using the dataset selector panel will be displayed, including any overlaid plots. The rate type displayed in the top left plot can be changed in the 'Tools' menu under 'Rate Type'. The rate data can be plotted as the activation Gibbs free energy, rate, or log(rate) versus temperature, as well as log(rate) versus the inverse temperature (1/T).

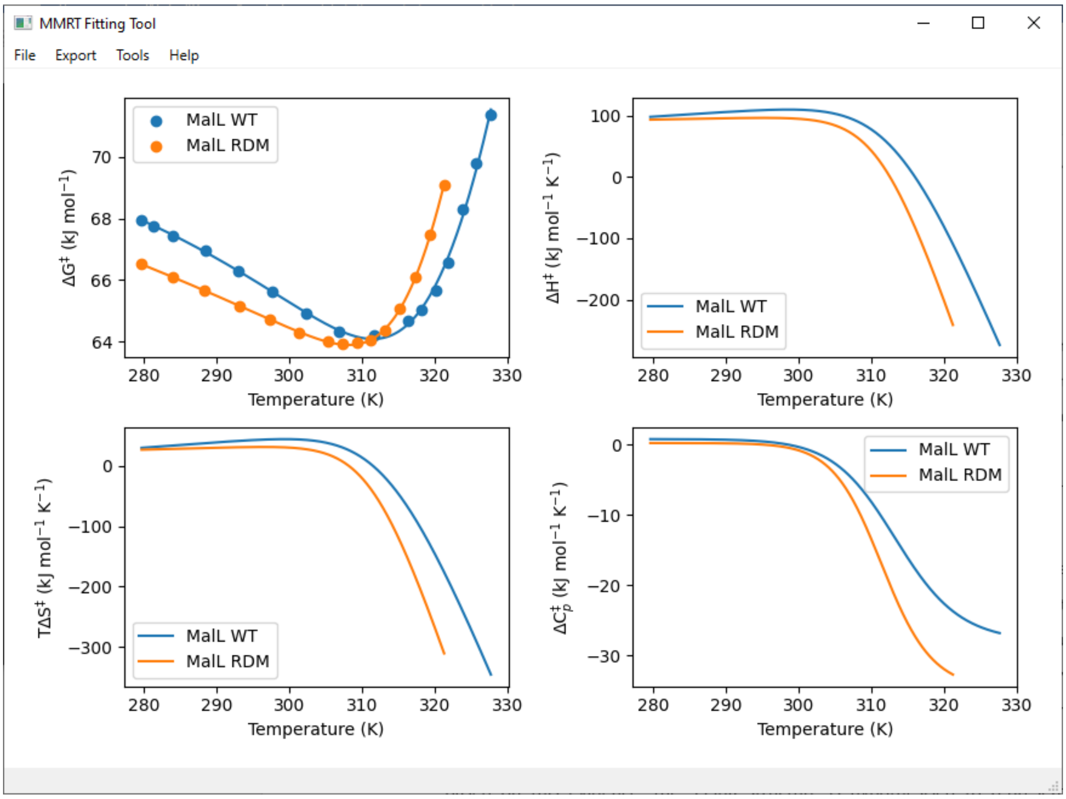


Figure 6.14. Parameter display window, showing temperature dependence of fit parameters.

6.3.10 Exporting data

MFit has four options for exporting data from the program (Figure 6.15). Export session will export an '.mmrt' file containing the information necessary to recreate the session. This includes the data points that were imported as well as the values and constraints used in the fit for each dataset. Session files generated from either MFit1 or MFit2 are cross-compatible with each other and datasets imported in this way will be fitted with the default parameters for that program.

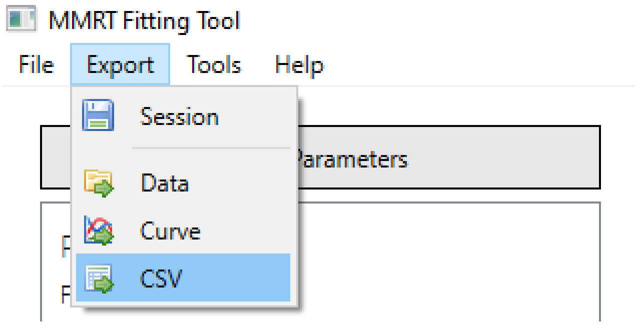


Figure 6.15. Export options

The export 'Data' option will prompt the user to select a directory to which data will be exported. The program will then create a directory and export PNG formatted images of the activation Gibbs free energy fit curve for each dataset, as well as an image containing all datasets. A '.csv' file containing all of the fit information will also be exported. This includes the fit parameter values, standard error of the parameter fit, 95% confidence interval for each parameter, and fit statistics including the R-squared, adjusted R-squared, Akaike Information Criterion (AIC), corrected AIC (AICc), chi-squred, reduced chi-squared, and the Bayesian Information Criterion (BIC). R-squared and the adjusted R-squared are calculated using Equation 6.8 and Equation 6.9. The other statistics are generated automatically during the fitting procedure (Newville *et al.*, 2014). This '.csv' file can also be exported individually using the export 'CSV' option.

Equation 6.8. R-squared

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - f(x_i))^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Equation 6.9. Adjusted R-squared

$$AR^2 = 1 - \frac{(n-1)\sum_{i=1}^{n}(y_i - f(x_i))^2}{(n-k)\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

The export 'Curves' options will export a '.csv' file containing the fit and parameter curves for each dataset, for plotting and analysis by the user in another program.

For each dataset 250 points will be exported for temperature (K), activation free energy (J mol^{-1}), activation enthalpy (J mol^{-1} K^{-1}), activation entropy [TΔS‡ (J mol^{-1} K^{-1})], and activation heat capacity (J mol^{-1} K^{-1}).

6.4 About MFit

MFit is written entirely in the Python programming language. It uses the PyQt6 GUI framework. The packages numpy, pandas, and openpyxl are used for data importing, processing and manipulation. The package lmfit (built on top of cipy) is used for data fitting and minimisation (Newville *et al.*, 2014). The package scipy is also used for integration. The package numba is used to optimise the residual function for data fitting, and improve code performance. The package matplotlib is used for data plotting. The code is open source and made available under the GNU GPL v3.0 software licence. The code can be downloaded from https://github.com/CarlinHamill.

6.5 Discussion

6.5.1 MFit

MalL wildtype was fit with MMRT 1.5, MMRT 1.0 and MMRT 2.0 using MFit as an example case. The MMRT 1.0 fit was to the low temperature section between 279 K and 307 K, as this model poorly fits the full temperature range. The results of these fits are presented alongside results from fitting in Graphpad Prism or RStudio, as a comparison to the standard MMRT fitting processes (Table 6.5, Table 6.6, and Table 6.7).

Table 6.5. Comparison of MMRT 1.0 fits with standard error

Parameter	MFit	GraphPad
$\Delta H^{\ddagger}_{T_0}$ $(kJ\ mol^{-1})$	100.7 ± 2.0	100.7 ± 1.1
$\Delta S^{\ddagger}_{T_0}$ $(J\ mol^{-1}K^{-1})$	116.9 ± 7.2	116.9 ± 4.0
$\Delta C^{\ddagger}_{P,T_0}$ $(kJ\ mol^{-1}K^{-1})$	0.4 ± 0.1	0.4 ± 0.1
T_0 (K)	278.15	278.15

Table 6.6. Comparison of MMRT 1.5 fits with standard error

Parameter	MFit	GraphPad
$\Delta H^{\ddagger}_{T_0}$ $(kJ\ mol^{-1})$	19.0 ± 16.0	19.0 ± 8.4
$\Delta S^{\ddagger}_{T_0}$ $(J\ mol^{-1}K^{-1})$	-174.0 ± 56.5	-174.0 ± 29.7
$\Delta C^{\ddagger}_{P,0}$ $(kJ\ mol^{-1}K^{-1})$	210 ± 18.0	210.0 ± 9.5
m $(J\ mol^{-1}K^{-2})$	-709.9 ± 59.3	-709.9 ± 31.2
T_0 (K)	278.15	278.15

Both methods produce best-fit values that are consistent with each other although the standard error is calculated differently. The standard error for MFit is calculated using the average of the parameters only, while GraphPad Prism is calculated using the replicate data.

Table 6.7. Comparison of MMRT 2.0 fits with standard erorr

Parameter	MFit	RStudio
$\Delta H^{\ddagger}_{T_0}$ $(kJ\ mol^{-1})$	96.6 ± 2.3	96.6 ± 2.3
$\Delta S^{\ddagger}_{T_0}$ $(J\ mol^{-1}K^{-1})$	102.3 ± 8.2	102.3 ± 8.2
$\Delta\Delta H^{\ddagger}$ $(kJ\ mol^{-1})$	186.7 ± 45.5	186 ± 46
$\Delta C^{\ddagger}_{P,lowT}$ $(kJ\ mol^{-1}K^{-1})$	0.8 ± 0.2	0.8 ± 0.2
$\Delta C^{\ddagger}_{P,highT}$ $(kJ\ mol^{-1}K^{-1})$	-28.1 ± 6.3	-28.1 ± 6.3
T_C (K)	313.5 ± 2.7	313.5 ± 2.7
$\Delta C^{\ddagger}_{P,lowT}$ **Fit Range**	1-6	1-6
T_0 (K)	278.15	278.15

The values and errors of the MMRT fit are identical to those fit using R studio (Walker *et al.*, 2023).

The MFit programs allow for the fitting of MMRT equations to biological rate data by non-technical users, and allows for easy visualisation and analysis of the results in a consistent manner. The MMRT fitting software exists as two versions, which fit either MMRT 1.0 and MMRT 1.5 (MFit1) or MMRT 2.0 (MFit2). MMRT 2.0 is complicated to fit and generally requires high quality enzyme kinetic data. For this reason, it has been split into its own program to allow the fitting of MMRT 1.0 and MMRT 1.5 to be more accessible for fitting a wider variety of data.

6.5.2 MMRT as a model of temperature dependence – an example case

MMRT models the temperature dependence of enzyme rates, and allows access to thermodynamic parameters of the catalysed reaction. The temperature dependence of the thermodynamic parameters $\Delta\mathrm{H}^{\ddagger}$, $\mathrm{T}\Delta\mathrm{S}^{\ddagger}$, $\Delta C_{\mathrm{P}}^{\ddagger}$, and $\Delta\mathrm{G}^{\ddagger}$ are shown in Figure 6.16, Figure 6.17, Figure 6.18 and Figure 6.19, respectively. $\Delta\mathrm{G}^{\ddagger}$ is the energy barrier for the reaction and is related to the measured rate (Equation 1.7). It is composed of enthalpic and entropic terms. The activation enthalpy is the difference in energy between the reactant state and the transition state, and represents the difference in the number and strength of bonding interactions. For MalL $\Delta\mathrm{H}^{\ddagger}$ is large and positive at low temperatures (where the TLC is favoured) and negative and steeply temperature dependent at high temperatures (where ES is favoured) (Figure 6.16). The activation entropy represents the loss of conformational freedom in the constrained transition state. For MalL at high temperatures the activation entropy is steeply temperature dependent, and is the main energy barrier (Figure 6.17). The activation heat capacity is also a result of the constrained transition state. For MalL at low temperatures it is near zero, and the rate approaches Arrhenius-like kinetics (Figure 6.18). At high temperatures it becomes large and negative, imparting a steep temperature dependence on the activation enthalpy and activation entropy, resulting in a non-linear reaction rate (Figure 6.19).

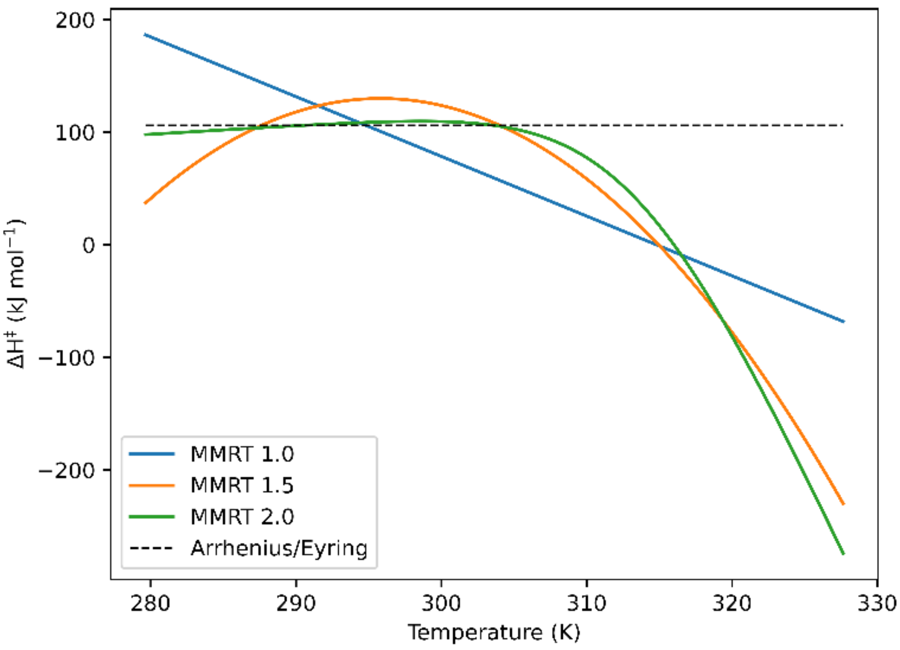


Figure 6.16. Comparison of the temperature dependence of the activation Enthalpy ($\Delta H^{\ddagger}$) between different MMRT forms. Data are calculated from the fit of each model to MalL wildtype. The black dotted line shows the Eyring-Polanyi equation ($\Delta C_P^{\ddagger} = 0$).

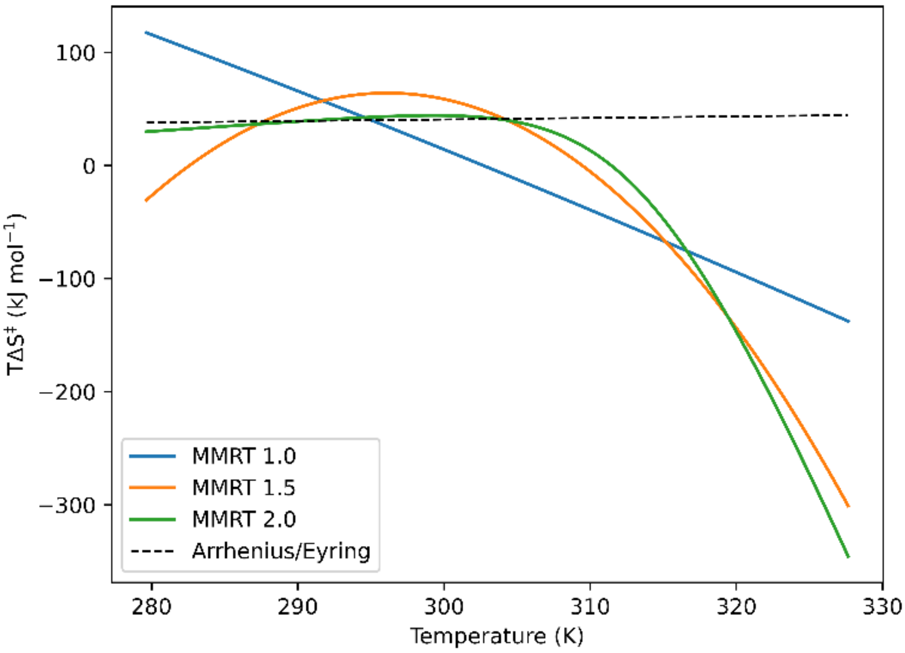


Figure 6.17. Comparison of the temperature dependence of the activation Entropy ($T\Delta S^{\ddagger}$) between different MMRT forms. Data are calculated from the fit of each model to MalL wildtype. The black dotted line shows the Eyring-Polanyi equation ($\Delta C_P^{\ddagger} = 0$).

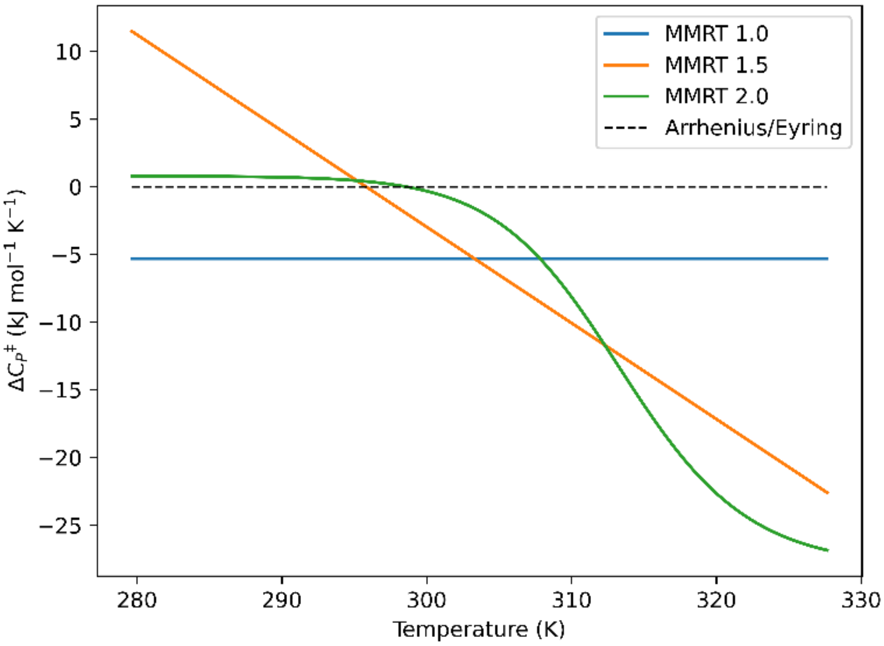


Figure 6.18. Comparison of the temperature dependence of the activation heat capacity ($\Delta C_P^\ddagger$) between different MMRT forms. Data are calculated from the fit of each model to MalL wildtype. The black dotted line shows the Eyring-Polanyi equation ($\Delta C_P^\ddagger = 0$).

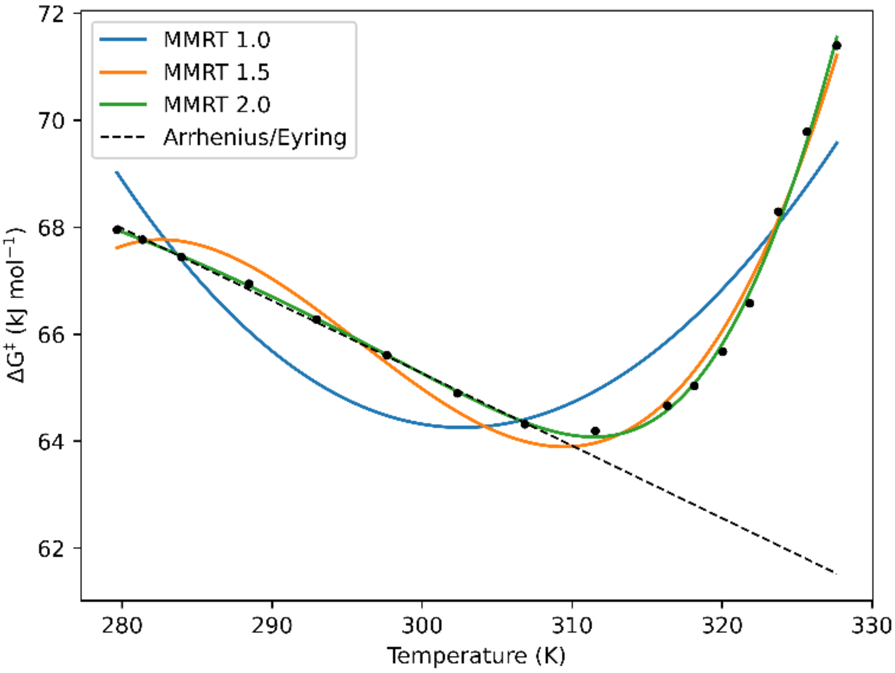


Figure 6.19. Comparison of the temperature dependence of the Gibbs free energy ($\Delta G^\ddagger$) between different MMRT forms. Data are calculated from the fit of each model to MalL wildtype. Gibbs free energy is the sum of the entropy and enthalpy curves according to the equation ($\Delta G^\ddagger = \Delta H^\ddagger - T\Delta S^\ddagger$). The black dotted line shows the Eyring-Polanyi equation ($\Delta C_P^\ddagger = 0$). Experimental data for MalL wildtype is shown as black dots.

The graphs above show the differences in thermodynamic parameters between different forms of MMRT compared to an Arrhenius/Eyring model. The initial portion of the data, before the T_{opt} (315.1 K), is fit well by MMRT 1.0, MMRT 1.5 and the Arrhenius/Eyring models (Figure 6.19). In MMRT 2.0 the $\Delta C_P^{\ddagger}$ in this region is near zero (0.8 kJ mol^{-1} K^{-1}) and approximates the Arrhenius/Eyring model. At temperatures above the T_{opt} $\Delta G^{\ddagger}$ increases, which is not captured by the Arrhenius/Eyring model. For the Arrhenius/Eyring model the increase in $\Delta G^{\ddagger}$ and corresponding decrease in the rate is attributed to denaturation. The change in $\Delta G^{\ddagger}$ is captured in all three MMRT models. A non-zero activation heat capacity defines the temperature dependence of the activation enthalpy and activation entropy, and results in curvature in the Gibbs free energy/ rate. MMRT 1.0 is equivalent to fitting a quadratic to $\Delta G^{\ddagger}$. Thus, the first derivative ($\Delta S^{\ddagger}$) is linear, and the second derivative ($\Delta C_P^{\ddagger}$) is constant. Similarly, MMRT 1.5 is equivalent to fitting a cubic to $\Delta G^{\ddagger}$, yielding a quadratic $\Delta S^{\ddagger}$, and linear $\Delta C_P^{\ddagger}$.

A constant $\Delta C_P^{\ddagger}$ (MMRT 1.0) fails to capture the temperature dependence of the heat capacity over the full temperature range resulting in a poor fit across the full temperature range of the data (Figure 6.19). MMRT 1.0 is therefore most useful in cases where the data are distributed over a narrow temperature range, where $\Delta C_P^{\ddagger}$ does not significantly vary, or where the number of experimental data points is limited, and not able to justify the fitting of the extra parameters in MMRT 1.5 or MMRT 2.0.

Although MMRT 2.0 can provide the closest fit to experimental data, it requires high quality data, and the greater number of parameters may result in difficulty in converging on a true fit, since many parameters are closely correlated. MMRT 1.5 therefore offers a suitable middle ground, balancing the number of fit parameters while still capturing the temperature dependence of the activation heat capacity.

7 Discussion

7.1 Overview

The traditional view of the temperature dependence of enzyme catalysis posits that enzyme reaction rates vary exponentially with temperature. This has been modelled using the Arrhenius and Eyring-Polanyi equations. Under these models the observed drop in enzyme activity at high temperatures is attributed solely to thermal denaturation. An increasing body of evidence, particularly for psychrophilic enzymes, which experience reduced reaction rates well below the temperature at which they become thermally denatured, has shown that denaturation alone is insufficient to fully account for these deviations (D'Amico *et al.*, 2003).

Macromolecular rate theory (MMRT) describes the temperature dependence of biological reaction rates by adding a term for the activation heat capacity (Hobbs *et al.*, 2013). The heat capacity of a protein may be defined by the number of accessible vibrational modes at a given temperature (Gómez *et al.*, 1995). As an enzyme moves from the ground state to the transition state, the conformational flexibility of the enzyme is restricted, and vibrational modes move to higher frequencies and consequently the heat capacity of the enzyme is lowered. Thus, the activation heat capacity, which describes the change in heat capacity between the enzyme substrate (ES) complex and enzyme transition state complex (TS), is generally negative for an enzyme catalysed reaction. This is analogous to protein folding where a flexible ground-state moves to a more compact transition state in the course of folding resulting in a negative activation heat capacity for this process. The introduction of the activation heat capacity adds curvature to the free energy barrier with temperature and thus, curvature to the rate equation independent of thermal denaturation.

The activation heat capacity describes the temperature dependence of the activation enthalpy and activation entropy and can itself be considered temperature dependent. The three MMRT models (MMRT 1.0, MMRT 1.5 and MMRT 2.0) vary in their temperature dependence of the activation heat capacity.

Under MMRT 1.0 the activation heat capacity is constant at all temperatures. In MMRT 1.5 the activation heat capacity is linearly dependent on temperature. Finally, under MMRT 2.0 the activation heat capacity has a sigmoidal temperature dependence. This sigmoidal activation heat capacity defines the behaviour of the activation heat capacity as being distinct at high and low temperatures, with a cooperative transition in-between. The MMRT 2.0 model posits that the enzyme must pass through a transition state like complex (TLC) prior to catalysis, and it is the temperature dependence of the equilibrium between this complex and the ES complex that drives the change in activation heat capacity, and ultimately the temperature dependence of the catalytic rate.

Each MMRT model has different characteristics that lend themselves to specific applications. MMRT 1.0 is most suitable for datasets with a limited temperature range and can be regarded as a first approximation. MMRT 1.5 is generally suitable for many applications with a balance of a small number of parameters to reduce overfitting, whilst still capturing the trend of the variation in the activation heat capacity with temperature. MMRT 2.0 is the most detailed model, suitable for high quality enzyme data, however, is prone to poor convergence and overfitting. MMRT in general has found wide ranging applications across many levels of biological rate data, from isolated enzymes (Hobbs *et al.*, 2013; Walker *et al.*, 2023), *in situ* soil enzymes (Alster *et al.*, 2016a), metabolism (Prentice *et al.*, 2020), membrane conductance (Pahlavan *et al.*, 2023), soil and leaf respiration (Liang *et al.*, 2018; Schipper *et al.*, 2014), and ecosystem level photosynthesis and respiration (Duffy *et al.*, 2021).

To aid in the application of MMRT across all these diverse disciplines, we have developed a graphical user interface to fit MMRT equations to user supplied data. We have developed two software packages, ‘MFit1’ and ‘MFit2’. MFit1 fits the MMRT 1.0 and MMRT 1.5 equations, while MFit2 fits the MMRT 2.0 equation. MMRT 2.0 requires high quality enzyme data, and we have thus separated it into its own program (MFit2) to leave MFit1 accessible to fitting a wider variety of data. These programs allow for the fitting of MMRT equations to become part of a routine workflow by non-technical users and allows for easy visualisation and analysis of fitting results.

7.2 Arginine mutations and psychrophile evolution

Psychrophilic enzymes have many adaptations that optimise their activity at low temperatures. These enzymes in general, have increased activity at low temperatures, decreased relative activity at moderate temperatures, and a decreased temperature optimum (Collins & Feller, 2023). Psychrophilic enzymes often display an increase in conformational flexibility, and an associated decrease in stability presumably as a mechanism to lower the activation barrier at low temperatures. Decreasing the activation enthalpy at the cost of the activation entropy (a so-called enthalpy-entropy trade-off) is thought to be one such mechanism to lower the activation free energy. A decrease in activation enthalpy results in a broader distribution of ES states, and therefore requires a greater conformational restriction into the TS state, which increases the entropy barrier. This will also increase the magnitude of the activation heat capacity and increase curvature in the temperature dependence of the rate. This has been dubbed the psychrophilic trap insofar as the temperature-rate curve will become increasingly narrow at low temperatures and this will be limiting for organisms adapting to varying temperatures (Arcus *et al.*, 2016). The activation enthalpy and activation entropy are thus locked together in a trade-off, where changes to either parameter will be offset by changes in the other.

This thesis built on previous results studying the dynamic effects of urea binding on enzyme activity. Urea acts as a general modifier of enzyme dynamics and activity. Urea has been shown to act as an activator to the model enzyme MalL at low temperatures and an inhibitor at high temperatures. Several crystal structures of urea bound to MalL were previously collected and the urea binding sites characterised (Prentice, unpublished). A binding site had previously been identified in the auxiliary C-terminal domain. The mutation S536R was designed to place the guanidine group of the mutated arginine into the urea binding site. The kinetics of this designed mutant were unchanged from that of MalL wildtype, with the fit parameters for both MMRT 1.0 and MMRT 1.5 being within standard error of each other.

Mall S536R significantly changed the crystal structure when compared to MalL wildtype, with a 1.2 Å improvement in resolution, from 2.3 to 1.10 Å, as well as significantly lowered temperature factors. This is consistent with an overall increase in order in the crystal state. The combination of increased resolution (and order) with the observation that the dynamics of S536R did not demonstrate a change in activation heat capacity led to the proposal that this structure may be representative of the TLC complex (Walker *et al.*, 2023). The S536R structure has a significantly greater number and strength of hydrogen bonds, consistent with the greater order of the TLC. A majority of the additional hydrogen bonds found in MalL S536R are located in regions that are highly dynamic in molecular dynamics simulations of MalL wildtype. The constriction of these regions in the transition state contribute greatly to the activation heat capacity (van der Kamp *et al.*, 2018; Walker *et al.*, 2023).

Arginine mutants based on urea binding sites were further explored, with the design of an additional four mutants (of which three were fully characterised) and one double mutant. The kinetics of these four mutants produced two distinct profiles. One group consisted of the mutations V376R and D492R, and the second group consisted of the mutant T150R and the arginine double mutant (RDM, T150R and D492R). MalL T150R and RDM show analogous behaviour to traditional psychrophilic enzymes. They show a reduced temperature optimum, from 315.1 K in MalL wildtype, to 312.6 and 311.6 K in MalL T150R and RDM, respectively. They also show an activation enthalpy-entropy trade off, with a decrease in both the activation enthalpy (favourable) and the activation entropy (unfavourable). A decrease in the activation Gibbs free energy may be achieved by either lowering the activation enthalpy or raising the activation entropy. At high temperatures an equivalent activation Gibbs free energy can be achieved with an entropy-enthalpy trade-off. When this occurs to lower the activation enthalpy, then the rate constant is automatically increased at low temperatures, as it is the enthalpic component of the rate which is temperature dependent, in the absence of activation heat capacity effects (Equation 7.1) (Stockbridge *et al.*, 2010).

Equation 7.1. Eyring-Polanyi equation

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

MalL V376R and D492R are characterised by an overall flattening of the temperature dependence profile. This can be seen in the slope of the activation heat capacity, in the MMRT 1.5 fit where the slope decreases from -710 ± 60 J mol^{-1} K^{-2} for MalL wildtype to -437 ± 44.0 and -372 ± 25 J mol^{-1} K^{-2}, for MalL V376R and D492R, respectively. This flattening of the temperature dependence is associated with a decrease in the flexibility of the ES complex, allowing for a lowered barrier into the constrained TLC complex, and thus a rate acceleration at low temperatures (Figure 7.1).

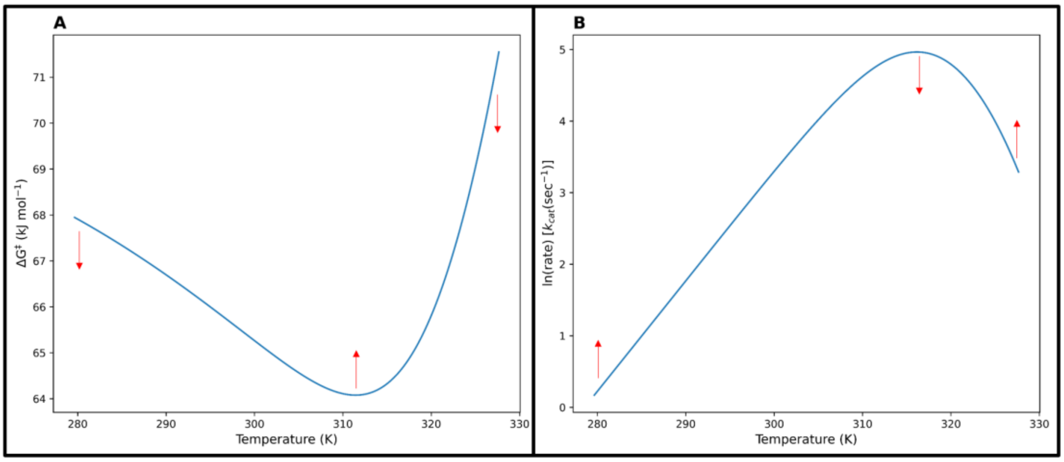


Figure 7.1. Flattening of the temperature dependence of the activation Gibbs free energy ($\Delta G^\ddagger$) decreases the $\Delta G^\ddagger$ at low temperatures, causing an increase in the rate (B).

At low to moderate temperatures, we see the same enthalpy-entropy trade off as with MalL T150R and RDM; however, at higher temperatures the flattened profile leads to an increase in both activation enthalpy and activation entropy. MalL D492R acts as a stabilising mutation in the ground state, with a lowered barrier of entry into the TLC. Thus, at low temperatures the equilibrium between the ES and TLC states is shifted towards the TLC, resulting in a rate acceleration (Figure 7.2).

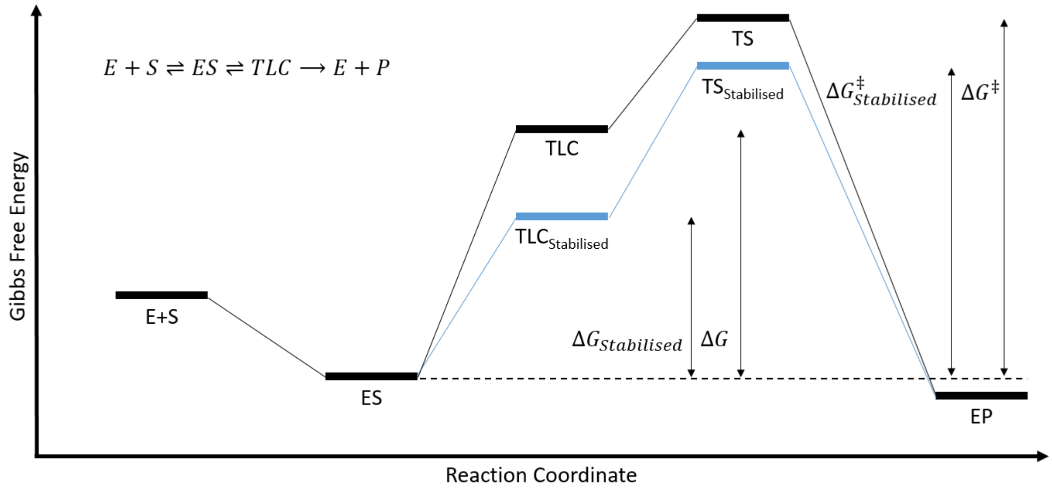


Figure 7.2. Gibbs free energy change along an enzyme-substrate reaction coordinate. MalL D492R and V376R are stabilised relative to MalL wildtype (represented here as a decrease in the Gibbs free energy of the TLC). This reduced $\Delta G_{Stabilised}$ into the TLC shifts the ES $\rightleftharpoons$ TLC equilibrium towards TLC at low temperatures, promoting the rate.

The T150R and D492R mutants were also structurally characterised. Both mutations were designed with the intention to place the guanidine group of the mutated arginine residue into a urea binding site. In both structures the mutated arginine did not occupy the urea binding site. As with the S536R mutation, both of the T150R and D492R mutants, crystallised with significantly improved resolution. The MalL T150R and D492R structures were solved at 1.46 and 1.04 Å resolution, a 0.84 and 1.26 Å improvement, respectively. In addition, the temperature factors for each mutant were significantly lowered, indicating greater order in the crystal state. MalL D492R exerts a stabilising effect by binding to the bottom loop of the C-terminal domain, stabilising the enzyme and flattening the curvature in the activation Gibbs free energy. In contrast the T150R mutation acts to add entropy to the system, destabilising the enzyme, adding flexibility and decreasing the activation enthalpy, as with traditional psychrophiles (Collins & Feller, 2023). In addition, this mutation packs up against Lys165, acting to bury some hydrophobic surface area, contributing to a stabilisation of the local area.

Each of the S536R, T150R and D492R mutations were designed to mimic urea binding. Each mutation significantly improved the resolution of the crystal structure compared to MalL wildtype. Neither of the T150R or D492R mutations bound in the urea binding site; however, these mutations still improved the crystal structure, without changing the overall structure. Thus, while the applicability of using arginine mutants that replicate urea binding to improve crystal structures has limited support, these mutations indicate that surface engineering of arginine mutations may be a viable general mechanism for protein stability engineering and resolution improvement.

7.3 Future work

Molecular dynamics of the arginine mutants would help greatly in understanding the dynamic changes of each mutation. Molecular dynamics simulations allow for an atomistic level overview of how structural changes influence enzyme dynamics and would provide insight into the mechanisms by which these arginine mutations influence the rate, given that the mutation sites are far from the active site. Molecular dynamics of each mutant enzyme in the presence of the substrate or a transition state analogue, could be used to calculate the activation heat capacity (van der Kamp *et al.*, 2018). This would also allow for the identification of regions whose dynamics contribute to catalysis and expand on how the introduced mutations exert their effects.

In this context, the behaviour of S376R and D492R mutants appears to be different from traditional psychrophiles, flattening the temperature profile to enhance rates at low temperature. This behaviour warrants further investigation as an alternative mechanism in adapting to low temperatures.

A more extensive analysis of the dynamics of the arginine mutants would be valuable. The SAXS results for MalL D492R showed some interesting trends, regarding the decrease in the volume and radius of gyration compared to MalL wildtype, especially with the further decrease at elevated temperatures.

Therefore, SAXS characterisation of all arginine mutations at multiple temperatures would provide valuable insights into the dynamics of these mutations. This work would also be completed by further REES spectroscopy characterisation, which is a sensitive reporter of changes to protein dynamics and flexibility (Kwok *et al.*, 2021).

Further design work to test the general method of urea binding and arginine mutations would be beneficial. The mutant enzymes T150R, D492R, and V376R, were designed simply, based on their *in silico* energy score, and the potential ability of the introduced mutation to overlay the urea binding site. For the generation of further mutants, the design phase could include further considerations to increase the likelihood that the introduced mutation will bind into the urea binding site, for example energy scoring of different rotameric configurations, or molecular dynamics simulations to check that the introduced mutation occupies the urea binding site. It would be useful to test the effect of introduced arginine mutations on a different enzyme other than MalL. Testing using molecular dynamics to identify the urea binding sites, without an initial structure would also be beneficial.

Directed evolution of MalL at low temperatures would be an interesting further experiment. It would be interesting to see if similar mutations to the designed arginine mutations would be generated by directed evolution. It would also be interesting to run molecular dynamics on evolved enzymes to see if we can match the mechanism of action of evolution towards psychrophily, with the action of our designed mutants.

References

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., & Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography, 66*(Pt 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.

Agarwal, P. K., Bernard, D. N., Bafna, K., & Doucet, N. (2020). Enzyme Dynamics: Looking Beyond a Single Structure. *ChemCatChem, 12*(19), 4704-4720.

Aghajari, N., Feller, G., Gerday, C., & Haser, R. (1998). Structures of the psychrophilic Alteromonas haloplanctis α-amylase give insights into cold adaptation at a molecular level. *Structure, 6*(12), 1503-1516.

Alster, C. J., Baas, P., Wallenstein, M. D., Johnson, N. G., & von Fischer, J. C. (2016a). Temperature Sensitivity as a Microbial Trait Using Parameters from Macromolecular Rate Theory. *Frontiers in Microbiology, 7*.

Alster, C. J., Koyama, A., Johnson, N. G., Wallenstein, M. D., & von Fischer, J. C. (2016b). Temperature sensitivity of soil microbial communities: An application of macromolecular rate theory to microbial respiration. *Journal of Geophysical Research: Biogeosciences, 121*(6), 1420-1433.

Alster, C. J., Robinson, J. M., Arcus, V. L., & Schipper, L. A. (2022). Assessing thermal acclimation of soil microbial respiration using macromolecular rate theory. *Biogeochemistry, 158*(1), 131-141.

Alster, C. J., von Fischer, J. C., Allison, S. D., & Treseder, K. K. (2020). Embracing a new paradigm for temperature sensitivity of soil microbes. *Global Change Biology, 26*(6), 3221-3229.

Åqvist, J., Sočan, J., & Purg, M. (2020). Hidden Conformational States and Strange Temperature Optima in Enzyme Catalysis. *Biochemistry, 59*(40), 3844-3855.

Arcus, V. L., & Mulholland, A. J. (2020). Temperature, Dynamics, and Enzyme-Catalyzed Reaction Rates. *Annual Review of Biophysics, 49*(1), 163-180.

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), 1681-8.

Arrhenius, S. (1889). Über die Reaktionsgeschwindigkeit bei der Inversion von Rohrzucker durch Säuren. *Zeitschrift für Physikalische Chemie, 4U*(1), 226-248.

Bagshaw, C. R. (2013). Stopped-Flow Techniques. In G. C. K. Roberts (Ed.), *Encyclopedia of Biophysics* (pp. 2460-2466). Berlin, Heidelberg: Springer Berlin Heidelberg.

Baldwin, R. L. (1986). Temperature dependence of the hydrophobic interaction in protein folding. *Proceedings of the National Academy of Sciences, 83*(21), 8069-8072.

Blow, D. M. (2002). *Outline of crystallography for biologists*. Oxford ; New York: Oxford University Press.

Catici, D. A. M., Amos, H. E., Yang, Y., van den Elsen, J. M. H., & Pudney, C. R. (2016). The red edge excitation shift phenomenon can be used to unmask protein structural ensembles: implications for NEMO–ubiquitin interactions. *The FEBS Journal, 283*(12), 2272-2284.

Collins, T., & Feller, G. (2023). Psychrophilic enzymes: strategies for cold-adaptation. *Essays in Biochemistry*, EBC20220193.

Cooper, D. R., Boczek, T., Grelewska, K., Pinkowska, M., Sikorska, M., Zawadzki, M., & Derewenda, Z. (2007). Protein crystallization by surface entropy reduction: optimization of the SER strategy. *Acta Crystallographica Section D, 63*(5), 636-645.

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.

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.

Dale, G. E., Oefner, C., & D’Arcy, A. (2003). The protein as a variable in protein crystallization. *Journal of Structural Biology, 142*(1), 88-97.

Dang Ngoc Chan, C. (2011). *Braggs Law*. Retrieved November, 2019, from https://en.wikipedia.org/wiki/File:Braggs_Law.svg.

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.

Davis, S. J., Brady, R. L., Barclay, A. N., Harlos, K., Dodson, G. G., & Williams, A. F. (1990). Crystallization of a soluble form of the rat T-cell surface glycoprotein CD4 complexed with Fab from the W3/25 monoclonal antibody. *Journal of molecular biology, 213*(1), 7-10.

Derewenda, Z. S. (2010). Application of protein engineering to enhance crystallizability and improve crystal properties. *Acta crystallographica. Section D, Biological crystallography, 66*(Pt 5), 604-615.

Derewenda, Z. S., Lee, L., & Derewenda, U. (1995). The Occurence of C–H · · · O Hydrogen Bonds in Proteins. *Journal of Molecular Biology, 252*(2), 248-262.

Dodson, E. J. (1991). Baverage.

Drenth, J. (1999). *Principles of protein x-ray crystallography*. Springer advanced texts in chemistry (2nd ed.). New York: Springer.

Duffy, K. A., Schwalm, C. R., Arcus, V. L., Koch, G. W., Liang, L. L., & Schipper, L. A. (2021). How close are we to the temperature tipping point of the terrestrial biosphere? *Science Advances, 7*(3), eaay1052.

Dunbar, J., Yennawar, H. P., Banerjee, S., Luo, J., & Farber, G. K. (1997). The effect of denaturants on protein structure. *Protein Science, 6*(8), 1727-1733.

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., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica Section D, 66*(4), 486-501.

Evans, P. (2011). An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallographica Section D, 67*(4), 282-292.

Evans, P. R., & Murshudov, G. N. (2013). How good are my data and what is the resolution? *Acta crystallographica. Section D, Biological crystallography, 69*(Pt 7), 1204-1214.

Eyal, E., Gerzon, S., Potapov, V., Edelman, M., & Sobolev, V. (2005). The Limit of Accuracy of Protein Modeling: Influence of Crystal Packing on Protein Structure. *Journal of Molecular Biology, 351*(2), 431-442.

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

Fersht, A. (1999). *Structure and mechanism in protein science : a guide to enzyme catalysis and protein folding*. New York: W.H. Freeman.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. e., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In J. M. Walker (Ed.), *The Proteomics Protocols Handbook* (pp. 571-607). Totowa, NJ: Humana Press.

Gómez, J., Hilser, V. J., Xie, D., & Freire, E. (1995). The heat capacity of proteins. *Proteins: Structure, Function, and Bioinformatics, 22*(4), 404-412.

Haldar, S., Chaudhuri, A., & Chattopadhyay, A. (2011). Organization and Dynamics of Membrane Probes and Proteins Utilizing the Red Edge Excitation Shift. *The Journal of Physical Chemistry B, 115*(19), 5693-5706.

Hamill, C. J. (2020). *Understanding allosteric enzyme regulation using macromolecular rate theory.* Masters thesis, The University of Waikato, Hamilton, New Zealand.

Harding, M. M., Nowicki, M. W., & Walkinshaw, M. D. (2010). Metals in protein structures: a review of their principal features. *Crystallography Reviews, 16*(4), 247-302.

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.

Huang, A., Lu, F., & Liu, F. (2023). Discrimination of psychrophilic enzymes using machine learning algorithms with amino acid composition descriptor. *Frontiers in Microbiology, 14*.

Huynh, K., & Partch, C. L. (2015). Analysis of Protein Stability and Ligand Interactions by Thermal Shift Assay. *Current Protocols in Protein Science, 79*(1), 28.9.1-28.9.14.

Isaksen, G. V., Åqvist, J., & Brandsdal, B. O. (2016). Enzyme surface rigidity tunes the temperature dependence of catalytic rates. *Proceedings of the National Academy of Sciences, 113*(28), 7822.

Johnson, M. L. (1992). Why, when, and how biochemists should use least squares. *Analytical Biochemistry, 206*(2), 215-225.

Jones, H. B. L., Crean, R. M., Matthews, C., Troya, A. B., Danson, M. J., Bull, S. D., Arcus, V. L., van der Kamp, M. W., & Pudney, C. R. (2018). Uncovering the relationship between the change in heat capacity for enzyme catalysis and vibrational frequency through isotope effect studies. *ACS Catalysis, 8*(6), 5340-5349.

Jones, H. B. L., Wells, S. A., Prentice, E. J., Kwok, A., Liang, L. L., Arcus, V. L., & Pudney, C. R. (2017). A complete thermodynamic analysis of enzyme turnover links the free energy landscape to enzyme catalysis. *The FEBS Journal, 284*(17), 2829-2842.

Joosten, R. P., Long, F., Murshudov, G. N., & Perrakis, A. (2014). The PDB_REDO server for macromolecular structure model optimization. *IUCrJ, 1*(4), 213-220.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature, 596*(7873), 583-589.

Justi, R., & Gilbert, J. K. (1999). History and Philosophy of Science through Models: The Case of Chemical Kinetics. *Science & Education, 8*(3), 287-307.

Kabsch, W. (2010). Integration, scaling, space-group assignment and post-refinement. *Acta Crystallographica Section D, 66*(2), 133-144.

Kantardjieff, K. A., & Rupp, B. (2003). Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein science : a publication of the Protein Society, 12*(9), 1865-1871.

Kikhney, A. G., & Svergun, D. I. (2015). A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Letters, 589*(19, Part A), 2570-2577.

Knight, M. J., Woolley, R. E., Kwok, A., Parsons, S., Jones, H. B. L., Gulácsy, C. E., Phaal, P., Kassaar, O., Dawkins, K., Rodriguez, E., Marques, A., Bowsher, L., Wells, S. A., Watts, A., van den Elsen, J. M. H., Turner, A., O'Hara, J., & Pudney, C. R. (2020). Monoclonal antibody stability can be usefully monitored using the excitation-energy-dependent fluorescence edge-shift. *Biochemical Journal, 477*(18), 3599-3612.

Kobayashi, M., Hondoh, H., Mori, H., Saburi, W., Okuyama, M., & Kimura, A. (2011). Calcium Ion-Dependent Increase in Thermostability of Dextran Glucosidase from Streptococcus mutans. *Bioscience, Biotechnology, and Biochemistry, 75*(8), 1557-1563.

Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., & Svergun, D. I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography, 36*(5), 1277-1282.

Kwok, A., Camacho, I., Winter, S., Knight, M., Meade, R., Van der Kamp, M., Turner, A., O’Hara, J., Mason, J., Jones, A., Arcus, V., & Pudney, C. (2021). A Thermodynamic Model for Interpreting Tryptophan Excitation-Energy-Dependent Fluorescence Spectra Provides Insight Into Protein Conformational Sampling and Stability. *Frontiers in Molecular Biosciences, 8*.

Laidler, K. J. (1965). *Chemical kinetics*. (2d ed.. ed.). New York: New York, McGraw-Hill.

Laidler, K. J., & King, M. C. (1983). Development of transition-state theory. *The Journal of Physical Chemistry, 87*(15), 2657-2664.

Langridge, T. D., Tarver, M. J., & Whitten, S. T. (2014). Temperature effects on the hydrodynamic radius of the intrinsically disordered N-terminal region of the p53 protein. *Proteins: Structure, Function, and Bioinformatics, 82*(4), 668-678.

Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography, 26*(2), 283-291.

Laurell, H., Contreras, J. A., Castan, I., Langin, D., & Holm, C. (2000). Analysis of the psychrotolerant property of hormone-sensitive lipase through site-directed mutagenesis. *Protein Engineering, Design and Selection, 13*(10), 711-717.

Levie, R. d. (2000). Curve Fitting with Least Squares. *Critical Reviews in Analytical Chemistry, 30*(1), 59-74.

Liang, L. L., Arcus, V. L., Heskel, M. A., O'Sullivan, O. S., Weerasinghe, L. K., Creek, D., Egerton, J. J. G., Tjoelker, M. G., Atkin, O. K., & Schipper, L. A. (2018). Macromolecular rate theory (MMRT) provides a thermodynamics rationale to underpin the convergent temperature response in plant leaf respiration. *Global Change Biology, 24*(4), 1538-1547.

Liepinsh, E., & Otting, G. (1994). Specificity of Urea Binding to Proteins. *Journal of the American Chemical Society, 116*(21), 9670-9674.

Longenecker, K. L., Garrard, S. M., Sheffield, P. J., & Derewenda, Z. S. (2001). Protein crystallization by rational mutagenesis of surface residues: Lys to Ala mutations promote crystallization of RhoGDI. *Acta Crystallographica Section D, 57*(5), 679-688.

Manalastas-Cantos, K., Konarev, P. V., Hajizadeh, N. R., Kikhney, A. G., Petoukhov, M. V., Molodenskiy, D. S., Panjkovich, A., Mertens, H. D. T., Gruzinov, A., Borges, C., Jeffries, C. M., Svergun, D. I., & Franke, D. (2021). ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. *Journal of Applied Crystallography, 54*(1), 343-355.

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

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*(Pt 4), 658-674.

McPherson, A., and, A. J. M., & Kuznetsov, Y. G. (2000). Atomic Force Microscopy in the Study of Macromolecular Crystal Growth. *Annual Review of Biophysics and Biomolecular Structure, 29*(1), 361-410.

McPherson, A., & Gavira, J. A. (2013). Introduction to protein crystallization. *Acta crystallographica. Section F, Structural biology communications, 70*(Pt 1), 2-20.

McPhillips, T. M., McPhillips, S. E., Chiu, H.-J., Cohen, A. E., Deacon, A. M., Ellis, P. J., Garman, E., Gonzalez, A., Sauter, N. K., Phizackerley, R. P., Soltis, S. M., & Kuhn, P. (2002). Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines. *Journal of Synchrotron Radiation, 9*(6), 401-406.

Møller Marie, S., Fredslund, F., Majumder, A., Nakai, H., Poulsen Jens-Christian, N., Lo Leggio, L., Svensson, B., & Abou Hachem, M. (2012). Enzymology and Structure of the GH13_31 Glucan 1,6-α-Glucosidase That Confers Isomaltooligosaccharide Utilization in the Probiotic Lactobacillus acidophilus NCFM. *Journal of Bacteriology, 194*(16), 4249-4259.

Mrabet, N. T., Van den Broeck, A., Van den Brande, I., Stanssens, P., Laroche, Y., Lambeir, A. M., Matthijssens, G., Jenkins, J., & Chiadmi, M. (1992). Arginine residues as stabilizing elements in proteins. *Biochemistry, 31*(8), 2239-2253.

Newville, M., Stensitzki, T., Allen, D. B., & Ingargiola, A. (2014). LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python.

Nicholls, R. A., Fischer, M., McNicholas, S., & Murshudov, G. N. (2014). Conformation-independent structural comparison of macromolecules with ProSMART. *Acta crystallographica. Section D, Biological crystallography, 70*(Pt 9), 2487-2499.

Pahlavan, B., Buitrago, N., & Santamaria, F. (2023). Macromolecular rate theory explains the temperature dependence of membrane conductance kinetics. *Biophysical Journal, 122*(3), 522-532.

Panjkovich, A., & Svergun, D. I. (2018). CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. *Bioinformatics, 34*(11), 1944-1946.

Park, H., Bradley, P., Greisen, P., Jr., Liu, Y., Mulligan, V. K., Kim, D. E., Baker, D., & DiMaio, F. (2016). Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules. *Journal of Chemical Theory and Computation, 12*(12), 6201-6212.

Pauw, B. R. (2013). Everything SAXS: small-angle scattering pattern collection and correction. *Journal of Physics: Condensed Matter, 25*(38), 383201.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry, 25*(13), 1605-1612.

Pike, A. C. W., & Acharya, K. R. (1994). A structural basis for the interaction of urea with lysozyme. *Protein Science, 3*(4), 706-710.

Portet, S. (2020). A primer on model selection using the Akaike Information Criterion. *Infectious Disease Modelling, 5*, 111-128.

Prabhu, N. V., & Sharp, K. A. (2005). HEAT CAPACITY IN PROTEINS. *Annual Review of Physical Chemistry, 56*(1), 521-548.

Prentice, E. J., Hicks, J., Ballerstedt, H., Blank, L. M., Liáng, L. n. L., Schipper, L. A., & Arcus, V. L. (2020). The Inflection Point Hypothesis: The Relationship between the Temperature Dependence of Enzyme-Catalyzed Reaction Rates and Microbial Growth Rates. *Biochemistry, 59*(38), 3562-3569.

Pu, J., Gao, J., & Truhlar, D. G. (2006). Multidimensional Tunneling, Recrossing, and the Transmission Coefficient for Enzymatic Reactions. *Chemical Reviews, 106*(8), 3140-3169.

Ratnaparkhi, G. S., & Varadarajan, R. (1999). X-ray crystallographic studies of the denaturation of ribonuclease S. *Proteins: Structure, Function, and Bioinformatics, 36*(3), 282-294.

Rossum, G. V., & Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace.

Ryan, T. M., Trewhella, J., Murphy, J. M., Keown, J. R., Casey, L., Pearce, F. G., Goldstone, D. C., Chen, K., Luo, Z., Kobe, B., McDevitt, C. A., Watkin, S. A., Hawley, A. M., Mudie, S. T., Samardzic Boban, V., & Kirby, N. (2018). An optimized SEC-SAXS system enabling high X-ray dose for rapid SAXS assessment with correlated UV measurements for biomolecular structure analysis. *Journal of Applied Crystallography, 51*(1), 97-111.

Sanfelice, D., & Temussi, P. A. (2016). Cold denaturation as a tool to measure protein stability. *Biophysical Chemistry, 208*, 4-8.

Schipper, L. A., Hobbs, J. K., Rutledge, S., & Arcus, V. L. (2014). Thermodynamic theory explains the temperature optima of soil microbial processes and high Q10 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.

Schorsch (Compiler) (2005). *Interferenz*. wikipedia.org.

Schrödinger, L. (2000). The PyMOL Molecular Graphics System.

Sočan, J., Purg, M., & Åqvist, J. (2020). Computer simulations explain the anomalous temperature optimum in a cold-adapted enzyme. *Nature Communications, 11*(1), 2644.

Sokalingam, S., Raghunathan, G., Soundrarajan, N., & Lee, S.-G. (2012). A Study on the Effect of Surface Lysine to Arginine Mutagenesis on Protein Stability and Structure Using Green Fluorescent Protein. *PLOS ONE, 7*(7), e40410.

Stockbridge, R. B., Lewis, C. A., Yuan, Y., & Wolfenden, R. (2010). Impact of temperature on the time required for the establishment of primordial biochemistry, and for the evolution of enzymes. *Proceedings of the National Academy of Sciences, 107*(51), 22102-22105.

Strub, C., Alies, C., Lougarre, A., Ladurantie, C., Czaplicki, J., & Fournier, D. (2004). Mutation of exposed hydrophobic amino acids to arginine to increase protein stability. *BMC Biochemistry, 5*(1), 9.

Struvay, C., & Feller, G. (2012). Optimization to Low Temperature Activity in Psychrophilic Enzymes. *International Journal of Molecular Sciences, 13*(9).

Sun, Z., Liu, Q., Qu, G., Feng, Y., & Reetz, M. T. (2019). Utility of B-Factors in Protein Science: Interpreting Rigidity, Flexibility, and Internal Motion and Engineering Thermostability. *Chemical Reviews, 119*(3), 1626-1665.

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

Truhlar, D. G. (2015). Transition state theory for enzyme kinetics. *Archives of Biochemistry and Biophysics, 582*, 10-17.

van der Kamp, M. W., Prentice, E. J., Kraakman, K. L., Connolly, M., Mulholland, A. J., & Arcus, V. L. (2018). Dynamical origins of heat capacity changes in enzyme-catalysed reactions. *Nature Communications, 9*(1), 1177.

Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Žídek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., Figurnov, M., Cowie, A., Hobbs, N., Kohli, P., Kleywegt, G., Birney, E., Hassabis, D., & Velankar, S. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research, 50*(D1), D439-D444.

Vekilov, P. G. (2016). Nucleation of protein crystals. *Progress in Crystal Growth and Characterization of Materials, 62*(2), 136-154.

Walker, E. J., Hamill, C. J., Crean, R., Connolly, M. S., Warrender, A. K., Kraakman, K. L., Prentice, E. J., Steyn-Ross, A., Steyn-Ross, M., Pudney, C. R., van der Kamp, M. W., Schipper, L. A., Mulholland, A. J., & Arcus, V. L. (2023). Cooperative conformational transitions and the temperature dependence of enzyme catalysis. *bioRxiv*, 2023.07.06.548038.

Warrender, A. K., Pan, J., Pudney, C. R., Arcus, V. L., & Kelton, W. (2023). Constant domain polymorphisms influence monoclonal antibody stability and dynamics. *Protein Science, 32*(3), e4589.

Weichenberger, C. X., & Rupp, B. (2014). Ten years of probabilistic estimates of biocrystal solvent content: new insights via nonparametric kernel density estimate. *Acta Crystallographica Section D, 70*(6), 1579-1588.

Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., Hartsch, T., & Ramakrishnan, V. (2000). Structure of the 30S ribosomal subunit. *Nature, 407*(6802), 327-339.

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., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., & Wilson, K. S. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D, 67*(4), 235-242.

Wong, C., Sridhara, S., Bardwell, J. C. A., & Jakob, U. (2000). Heating Greatly Speeds Coomassie Blue Staining and Destaining. *BioTechniques, 28*(3), 426-432.

Wyttenbach, T., Witt, M., & Bowers, M. T. (1999). On the question of salt bridges of cationized amino acids in the gas phase: glycine and arginine11Dedicated to the memory of Ben Freiser. *International Journal of Mass Spectrometry, 182-183*, 243-252.

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

Zheng, X., Bi, C., Li, Z., Podariu, M., & Hage, D. S. (2015). Analytical methods for kinetic studies of biological interactions: A review. *Journal of Pharmaceutical and Biomedical Analysis, 113*, 163-180.

Žoldák, G., Šut'ák, R., Antalík, M., Sprinzl, M., & Sedlák, E. (2003). Role of conformational flexibility for enzymatic activity in NADH oxidase from

Thermus thermophilus. *European Journal of Biochemistry, 270*(24), 4887-4897.

8 Appendix

8.1 Gene and protein sequences

Mutation positions for MalL variants are bolded and underlined

8.1.1 MalL Wildtype

8.1.1.1 Nucleotide sequence – 1764 bp

ATGTCGTACTACCATCACCATCACCATCACGATTACGATATCCCAACGACCGAAAACCTG
TATTTTCAGGGCGCCATGGGTGAATGGTGGAAAGAAGCCGTTGTTTATCAGATTTATCC
GCGTAGCTTTTATGATGCCAATGGTGATGGTTTTGGTGATCTGCAAGGTGTTATTCAGA
AACTGGATTACATCAAAAATCTGGGTGCCGATGTGATTTGGCTGAGTCCGGTTTTTGAT
AGTCCGCAGGATGATAATGGTTATGATATTAGCGATTACAAAAACATGTATGAAAAATT
TGGCACCAATGAAGATATGTTTCAGCTGATTGATGAAGTGCATAAACGCGGTATGAAA
ATTGTTATGGATCTGGTGGTTAATCATACCAGTGATGAACATGCATGGTTTGCAGAAAG
CCGTAAAAGCAAAGATAATCCGTATCGCGATTATTATCTGTGGAAAGATCCGAAACCG
GATGGTAGCGAACCGAATAATTGGGGTAGCATTTTTAGCGGTAGCGCATGGACCTATG
ATGAAGGCACCGGTCAGTATTATCTGCATTACTTTAGCAAAAAACAGCCGGATCTGAAT
TGGGAAAATGAAGCAGTTCGTCGTGAAGTGTATGATGTTATGCGTTTTTGGATGGATC
GTGGTGTTGATGGTTGGCGTATGGATGTTATTGGTAGCATTAGCAAATATACCGATTTT
CCGGATTATGAAACCGATCATAGCCGTAGCTATATTGTGGGTCGTTATCATAGCAATGG
TCCGCGTCTGCATGAATTTATTCAGGAAATGAATCGCGAAGTTCTGAGCCATTATGATT
GTATGACCGTTGGTGAAGCAAATGGCAGCGATATTGAAGAAGCCAAAAAATACACAGA
TGCCAGTCGCCAAGAACTGAATATGATTTTTACCTTTGAACATATGGATATTGATAAAG
AACAGAATAGCCCGAATGGCAAATGGCAGATTAAACCGTTTGATCTGATTGCCCTGAA
AAAAACCATGACCCGTTGGCAGACAGGTCTGATGAATGTTGGTTGGAATACCCTGTATT
TTGAAAATCATGATCAGCCTCGTGTTATTAGCCGTTGGGGTAATGATCGTAAACTGCGT
AAAGAATGTGCAAAAGCATTTGCAACCGTTCTGCATGGTATGAAAGGCACCCCGTTTAT
CTATCAGGGTGAAGAAATTGGTATGGTGAATAGCGATATGCCGCTGGAAATGTATGAT
GATCTGGAAATCAAAAATGCCTATCGCGAACTGGTGGTGGAAAACAAAACCATGAGCG
AAAAAGAATTTGTGAAAGCCGTGATGATTAAAGGTCGTGATCATGCACGTACCCCGAT
GCAGTGGGATGCAGGTAAACATGCAGGTTTTACCGCAGGCGATCCGTGGATTCCGGTT

AATAGCCGTTATCAGGATATTAATGTGAAAGAAAGCCTGGAAGATCAGGATAGCATTT
TCTTTTACTATCAGAAACTGATTCAGCTGCGCAAACAGTATAAAATCATGATTTATGGCG
ATTATCAGCTGCTGCAAGAAAATGATCCGCAGGTTTTTAGCTATCTGCGTGAATATCGT
GGTGAAAAACTGCTGGTTGTTGTTAATCTGAGCGAAGAAAAAGCACTGTTTGAAGCAC
CTCCGGAACTGATTCATGAACGTTGGAAAGTTCTGATTAGCAATTATCCGCAGGAACGT
GCAGATCTGAAAAGCATTAGCCTGAAACCGTATGAAGCAGTTATGGGCATTAGCATCT
AATAA

8.1.1.2 Protein sequence – 586 amino acids – 69.2 kDa

MSYYHHHHHHDYDIPTTENLYFQGAMGEWWKEAVVYQIYPRSFYDANGDGFGDLQGVI
QKLDYIKNLGADVIWLSPVFDSPQDDNGYDISDYKNMYEKFGTNEDMFQLIDEVHKRGMK
IVMDLVVNHTSDEHAWFAESRKSKDNPYRDYYLWKDPKPDGSEPNNWGSIFSGSAWTYD
EGTGQYYLHYFSKKQPDLNWENEAVRREVYDVMRFWMDRGVDGWRMDVIGSISKYTDF
PDYETDHSRSYIVGRYHSNGPRLHEFIQEMNREVLSHYDCMTVGEANGSDIEEAKKYTDAS
RQELNMIFTFEHMDIDKEQNSPNGKWQIKPFDLIALKKTMTRWQTGLMNVGWNTLYFE
NHDQPRVISRWGNDRKLRKECAKAFATVLHGMKGTPFIYQGEEIGMVNSDMPLEMYDDL
EIKNAYRELVVENKTMSEKEFVKAVMIKGRDHARTPMQWDAGKHAGFTAGDPWIPVNS
RYQDINVKESLEDQDSIFFYYQKLIQLRKQYKIMIYGDYQLLQENDPQVFSYLREYRGEKLLV
VVNLSEEKALFEAPPELIHERWKVLISNYPQERADLKSISLKPYEAVMGISI*

8.1.2 MalL S536R

8.1.2.1 Nucleotide sequence -1764 bp

ATGTCGTACTACCATCACCATCACCATCACGATTACGATATCCCAACGACCGAAAACCTG
TATTTTCAGGGCGCCATGGGTGAATGGTGGAAAGAAGCCGTTGTTTATCAGATTTATCC
GCGTAGCTTTTATGATGCCAATGGTGATGGTTTTGGTGATCTGCAAGGTGTTATTCAGA
AACTGGATTACATCAAAAATCTGGGTGCCGATGTGATTTGGCTGAGTCCGGTTTTTGAT
AGTCCGCAGGATGATAATGGTTATGATATTAGCGATTACAAAAACATGTATGAAAAATT
TGGCACCAATGAAGATATGTTTCAGCTGATTGATGAAGTGCATAAACGCGGTATGAAA
ATTGTTATGGATCTGGTGGTTAATCATACCAGTGATGAACATGCATGGTTTGCAGAAAG
CCGTAAAAGCAAAGATAATCCGTATCGCGATTATTATCTGTGGAAAGATCCGAAACCG
GATGGTAGCGAACCGAATAATTGGGGTAGCATTTTTAGCGGTAGCGCATGGACCTATG
ATGAAGGCACCGGTCAGTATTATCTGCATTACTTTAGCAAAAAACAGCCGGATCTGAAT

TGGGAAAATGAAGCAGTTCGTCGTGAAGTGTATGATGTTATGCGTTTTTGGATGGATC
GTGGTGTTGATGGTTGGCGTATGGATGTTATTGGTAGCATTAGCAAATATACCGATTTT
CCGGATTATGAAACCGATCATAGCCGTAGCTATATTGTGGGTCGTTATCATAGCAATGG
TCCGCGTCTGCATGAATTTATTCAGGAAATGAATCGCGAAGTTCTGAGCCATTATGATT
GTATGACCGTTGGTGAAGCAAATGGCAGCGATATTGAAGAAGCCAAAAAATACACAGA
TGCCAGTCGCCAAGAACTGAATATGATTTTTACCTTTGAACATATGGATATTGATAAAG
AACAGAATAGCCCGAATGGCAAATGGCAGATTAAACCGTTTGATCTGATTGCCCTGAA
AAAAACCATGACCCGTTGGCAGACAGGTCTGATGAATGTTGGTTGGAATACCCTGTATT
TTGAAAATCATGATCAGCCTCGTGTTATTAGCCGTTGGGGTAATGATCGTAAACTGCGT
AAAGAATGTGCAAAAGCATTTGCAACCGTTCTGCATGGTATGAAAGGCACCCCGTTTAT
CTATCAGGGTGAAGAAATTGGTATGGTGAATAGCGATATGCCGCTGGAAATGTATGAT
GATCTGGAAATCAAAAATGCCTATCGCGAACTGGTGGTGGAAAACAAAACCATGAGCG
AAAAAGAATTTGTGAAAGCCGTGATGATTAAAGGTCGTGATCATGCACGTACCCCGAT
GCAGTGGGATGCAGGTAAACATGCAGGTTTTACCGCAGGCGATCCGTGGATTCCGGTT
AATAGCCGTTATCAGGATATTAATGTGAAAGAAAGCCTGGAAGATCAGGATAGCATTT
TCTTTTACTATCAGAAACTGATTCAGCTGCGCAAACAGTATAAAATCATGATTTATGGCG
ATTATCAGCTGCTGCAAGAAAATGATCCGCAGGTTTTTAGCTATCTGCGTGAATATCGT
GGTGAAAAACTGCTGGTTGTTGTTAATCTGAGCGAAGAAAAAGCACTGTTTGAAGCAC
CTCCGGAACTGATTCATGAACGTTGGAAAGTTCTGATT**CGC**AATTATCCGCAGGAACGT
GCAGATCTGAAAAGCATTAGCCTGAAACCGTATGAAGCAGTTATGGGCATTAGCATCT
AATAA

8.1.2.2 Protein sequence – 586 amino acids – 69.2 kDa

MSYYHHHHHHDYDIPTTENLYFQGAMGEWWKEAVVYQIYPRSFYDANGDGFGDLQGVI
QKLDYIKNLGADVIWLSPVFDSPQDDNGYDISDYKNMYEKFGTNEDMFQLIDEVHKRGMK
IVMDLVVNHTSDEHAWFAESRKSKDNPYRDYYLWKDPKPDGSEPNNWGSIFSGSAWTYD
EGTGQYYLHYFSKKQPDLNWENEAVRREVYDVMRFWMDRGVDGWRMDVIGSISKYTDF
PDYETDHSRSYIVGRYHSNGPRLHEFIQEMNREVLSHYDCMTVGEANGSDIEEAKKYTDAS
RQELNMIFTFEHMDIDKEQNSPNGKWQIKPFDLIALKKTMTRWQTGLMNVGWNTLYFE
NHDQPRVISRWGNDRKLRKECAKAFATVLHGMKGTPFIYQGEEIGMVNSDMPLEMYDDL
EIKNAYRELVVENKTMSEKEFVKAVMIKGRDHARTPMQWDAGKHAGFTAGDPWIPVNS

RYQDINVKESLEDQDSIFFYYQKLIQLRKQYKIMIYGDYQLLQENDPQVFSYLREYRGEKLLV

VVNLSEEKALFEAPPELIHERWKVLI**R**NYPQERADLKSISLKPYEAVMGISI*

8.1.3 MalL D492R

8.1.3.1 Nucleotide sequence – 1749 bp

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCC

ATATGATGGGTGAATGGTGGAAAGAAGCTGTGGTGTACCAAATCTATCCTCGTAGTTTC

TACGACGCGAACGGCGACGGCTTCGGCGACTTACAGGGCGTCATCCAAAAGTTAGACT

ATATTAAGAACCTCGGAGCGGACGTCATCTGGTTGTCGCCTGTCTTCGACTCGCCCCAA

GACGACAACGGCTACGACATCAGTGACTATAAGAATATGTACGAGAAGTTCGGTACGA

ACGAGGACATGTTCCAACTTATCGACGAGGTTCACAAGCGTGGCATGAAGATCGTGAT

GGACCTTGTCGTCAACCACACTTCCGACGAGCACGCGTGGTTCGCGGAGTCCCGCAAG

TCTAAGGACAACCCTTACCGTGACTACTACTTATGGAAGGACCCAAAGCCAGACGGCTC

GGAGCCCAACAACTGGGGCTCAATCTTCTCAGGCAGTGCCTGGACATACGACGAGGGT

ACGGGCCAATACTACTTGCACTATTTCTCTAAGAAGCAACCCGACCTTAACTGGGAGAA

CGAGGCCGTGCGCCGCGAGGTTTACGACGTAATGCGCTTCTGGATGGACCGCGGCGTA

GACGGGTGGCGCATGGACGTGATCGGGAGTATCTCTAAGTACACGGACTTCCCAGACT

ACGAGACGGACCACTCTCGCTCCTACATCGTTGGCCGCTACCACTCAAACGGACCACGC

TTACACGAGTTCATCCAAGAGATGAACCGTGAGGTCTTAAGTCACTACGACTGCATGAC

GGTGGGAGAGGCCAACGGTAGTGACATCGAGGAAGCGAAGAAGTATACTGACGCGTC

CCGTCAGGAGCTCAACATGATCTTCACTTTCGAGCACATGGACATCGACAAGGAGCAA

AACTCACCAAACGGTAAGTGGCAAATCAAGCCCTTCGACTTGATCGCACTTAAGAAGAC

GATGACGCGCTGGCAAACCGGCTTGATGAACGTGGGCTGGAACACATTGTACTTCGAG

AACCACGACCAACCGCGCGTGATCAGTCGCTGGGGCAACGACCGCAAGTTACGCAAGG

AGTGCGCCAAGGCGTTCGCGACGGTGTTACACGGCATGAAGGGTACGCCTTTCATTTA

CCAAGGCGAGGAGATCGGAATGGTCAACTCCGACATGCCCTTAGAGATGTACGACGAC

TTAGAGATTAAGAACGCGTACCGTGAGTTAGTCGTTGAGAATAAGACAATGTCAGAGA

AGGAGTTCGTAAAGGCGGTAATGATCAAGGGCCGTGACCACGCCCGCACGCCTATGCA

ATGGGACGCGGGCAAGCACGCTGGATTCACTGCCGGTGATCCCTGGATCCCAGTGAAC

TCGCGCTATCAAGACATCAACGTAAAGGAGTCATTGGAAGACCAAGACTCTATCTTCTT

CTATTACCAAAAGCTGATCCAACTCCGTAAGCAATACAAGATTATGATCTACGGTGACT

ACCAATTATTGCAGGAGAAC**CGC**CCCCAAGTGTTCAGTTACTTACGCGAGTACCGCGGC

GAGAAGTTGCTCGTGGTCGTGAACTTATCCGAGGAGAAGGCGTTATTCGAGGCTCCGC

CAGAGCTCATCCACGAGCGCTGGAAGGTGTTGATCAGTAACTACCCACAAGAGCGCGC

GGACTTAAAGTCCATCTCATTAAAGCCATACGAGGCTGTAATGGGTATCAGTATTTAA

8.1.3.2 Protein sequence – 582 amino acids – 68.4 kDa

MGSSHHHHHHSSGLVPRGSHMMGEWWKEAVVYQIYPRSFYDANGDGFGDLQGVIQKL

DYIKNLGADVIWLSPVFDSPQDDNGYDISDYKNMYEKFGTNEDMFQLIDEVHKRGMKIV

MDLVVNHTSDEHAWFAESRKSKDNPYRDYYLWKDPKPDGSEPNNWGSIFSGSAWTYDE

GTGQYYLHYFSKKQPDLNWENEAVRREVYDVMRFWMDRGVDGWRMDVIGSISKYTDFP

DYETDHSRSYIVGRYHSNGPRLHEFIQEMNREVLSHYDCMTVGEANGSDIEEAKKYTDASR

QELNMIFTFEHMDIDKEQNSPNGKWQIKPFDLIALKKTMTRWQTGLMNVGWNTLYFEN

HDQPRVISRWGNDRKLRKECAKAFATVLHGMKGTPFIYQGEEIGMVNSDMPLEMYDDLE

IKNAYRELVVENKTMSEKEFVKAVMIKGRDHARTPMQWDAGKHAGFTAGDPWIPVNSR

YQDINVKESLEDQDSIFFYYQKLIQLRKQYKIMIYGDYQLLQEN**R**PQVFSYLREYRGEKLLVV

VNLSEEKALFEAPPELIHERWKVLISNYPQERADLKSISLKPYEAVMGISI*

8.1.4 MalL T150R

8.1.4.1 Nucleotide sequence – 1749 bp

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCC

ATATGATGGGCGAGTGGTGGAAGGAAGCGGTAGTATATCAAATTTACCCACGTTCTTTC

TACGACGCGAACGGCGACGGCTTCGGCGACCTTCAAGGGGTGATCCAAAAGTTAGACT

ATATTAAGAACCTCGGCGCAGACGTTATCTGGCTTTCGCCAGTGTTCGACAGCCCCCAA

GACGACAACGGATACGACATCTCGGACTATAAGAATATGTACGAGAAGTTCGGTACGA

ACGAGGACATGTTCCAATTAATCGACGAGGTACACAAGCGTGGCATGAAGATCGTGAT

GGACTTGGTTGTGAACCACACGTCGGACGAGCACGCTTGGTTCGCTGAGTCTCGCAAG

TCTAAGGACAACCCTTACCGTGACTACTACCTTTGGAAGGACCCCAAGCCAGACGGCTC

AGAGCCTAACAACTGGGGCTCAATCTTCTCCGGCTCAGCCTGG**CGC**TACGACGAGGGT

ACGGGCCAATACTACTTGCACTATTTCAGTAAGAAGCAACCCGACCTTAACTGGGAGAA

CGAGGCGGTGCGCCGCGAGGTATACGACGTAATGCGCTTCTGGATGGACCGCGGCGT

AGACGGCTGGCGCATGGACGTAATCGGCTCTATCTCGAAGTACACGGACTTCCCTGACT

ACGAGACGGACCACAGTCGCTCCTACATCGTCGGCCGCTACCACTCAAACGGCCCACGC

TTACACGAGTTCATCCAAGAGATGAACCGTGAGGTCTTATCGCACTACGACTGCATGAC

TGTGGGCGAGGCGAACGGGAGTGACATCGAAGAGGCGAAGAAGTATACTGACGCGA
GCCGTCAGGAGTTGAACATGATCTTCACGTTCGAGCACATGGACATCGACAAGGAGCA
AAACTCTCCTAACGGTAAGTGGCAAATCAAGCCATTCGACTTAATCGCACTTAAGAAGA
CTATGACGCGCTGGCAAACCGGACTTATGAACGTAGGATGGAACACGTTATACTTCGA
GAATCACGACCAACCCCGCGTCATCTCGCGCTGGGGCAACGACCGCAAGCTGCGCAAG
GAGTGCGCGAAGGCGTTCGCTACAGTGCTCCACGGCATGAAGGGTACACCATTCATTT
ACCAAGGCGAGGAGATCGGCATGGTCAACAGTGACATGCCATTGGAGATGTACGACG
ACCTTGAGATTAAGAACGCATACCGTGAGCTCGTCGTTGAGAATAAGACGATGTCAGA
GAAGGAGTTCGTAAAGGCAGTTATGATCAAGGGCCGCGATCACGCGCGCACTCCTATG
CAATGGGACGCCGGAAAGCACGCGGGCTTCACGGCTGGGGACCCTTGGATCCCTGTGA
ACTCACGCTACCAAGACATCAACGTGAAGGAGTCGCTTGAGGACCAAGACTCAATCTTC
TTCTATTACCAAAAGTTGATCCAACTTCGTAAGCAATACAAGATTATGATCTACGGAGA
CTACCAATTGCTCCAGGAGAACGACCCCCAAGTGTTCTCGTACTTACGCGAGTACCGCG
GCGAGAAGTTATTAGTGGTCGTAAACTTGTCCGAGGAGAAGGCCCTCTTCGAGGCGCC
ACCCGAGTTGATCCACGAGCGCTGGAAGGTGTTGATCTCAAACTACCCACAAGAGCGC
GCGGACTTAAAGTCAATCTCCTTGAAGCCTTACGAGGCGGTGATGGGTATCTCAATTTA
A

8.1.4.2 Protein sequence – 582 amino acids – 68.4 kDa

MGSSHHHHHHSSGLVPRGSHMMGEWWKEAVVYQIYPRSFYDANGDGFGDLQGVIQKL
DYIKNLGADVIWLSPVFDSPQDDNGYDISDYKNMYEKFGTNEDMFQLIDEVHKRGMKIV
MDLVVNHTSDEHAWFAESRKSKDNPYRDYYLWKDPKPDGSEPNNWGSIFSGSAW**R**YDE
GTGQYYLHYFSKKQPDLNWENEAVRREVYDVMRFWMDRGVDGWRMDVIGSISKYTDFP
DYETDHSRSYIVGRYHSNGPRLHEFIQEMNREVLSHYDCMTVGEANGSDIEEAKKYTDASR
QELNMIFTFEHMDIDKEQNSPNGKWQIKPFDLIALKKTMTRWQTGLMNVGWNTLYFEN
HDQPRVISRWGNDRKLRKECAKAFATVLHGMKGTPFIYQGEEIGMVNSDMPLEMYDDLE
IKNAYRELVVENKTMSEKEFVKAVMIKGRDHARTPMQWDAGKHAGFTAGDPWIPVNSR
YQDINVKESLEDQDSIFFYYQKLIQLRKQYKIMIYGDYQLLQENDPQVFSYLREYRGEKLLVV
VNLSEEKALFEAPPELIHERWKVLISNYPQERADLKSISLKPYEAVMGISI*

8.1.5 MalL V376R

8.1.5.1 Nucleotide sequence – 1749 bp

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCC
ATATGATGGGAGAATGGTGGAAGGAAGCGGTGGTATATCAAATCTACCCGCGCAGTTT
CTACGACGCAAACGGGGACGGATTCGGAGACTTACAGGGCGTGATCCAAAAGCTGGA
CTATATTAAGAACTTAGGCGCGGACGTAATCTGGTTAAGCCCCGTATTCGACTCACCAC
AAGACGACAACGGGTACGACATCTCAGACTATAAGAATATGTACGAGAAGTTCGGAAC
AAACGAGGACATGTTCCAATTGATCGACGAGGTTCACAAGCGTGGCATGAAGATCGTG
ATGGACCTTGTTGTGAACCACACGTCCGACGAGCACGCTTGGTTCGCTGAGTCCCGCAA
GAGTAAGGACAACCCTTACCGTGACTACTACCTTTGGAAGGACCCAAAGCCTGACGGG
TCAGAGCCCAACAACTGGGGCTCCATCTTCTCGGGCTCCGCGTGGACATACGACGAGG
GTACAGGCCAATACTACTTACACTATTTCTCAAAGAAGCAACCTGACCTTAACTGGGAG
AACGAGGCGGTGCGCCGCGAGGTTTACGACGTGATGCGCTTCTGGATGGACCGCGGA
GTAGACGGCTGGCGCATGGACGTGATCGGGAGTATCTCAAAGTACACGGACTTCCCCG
ACTACGAGACGGACCACTCTCGCAGTTACATCGTCGGACGCTACCACTCCAACGGGCCA
CGCTTGCACGAGTTCATCCAAGAGATGAACCGTGAGGTGCTTTCGCACTACGACTGCAT
GACGGTAGGCGAGGCGAACGGATCCGACATCGAAGAGGCGAAGAAGTATACCGACGC
ATCTCGTCAGGAGCTTAACATGATCTTCACTTTCGAGCACATGGACATCGACAAGGAGC
AAAACAGTCCTAACGGGAAGTGGCAAATCAAGCCCTTCGACTTAATCGCGCTCAAGAA
GACGATGACACGCTGGCAAACCGGCCTCATGAACGTAGGCTGGAACACTCTCTACTTC
GAGAACCACGACCAACCGCGCGTGATCTCACGCTGGGGCAACGACCGCAAGCTCCGCA
AGGAGTGCGCCAAGGCCTTCGCGACGGTGTTGCACGGAATGAAGGGTACGCCTTTCAT
TTACCAAGGCGAGGAGATCGGCATG**CGC**AACTCCGACATGCCACTGGAGATGTACGAC
GACTTAGAGATTAAGAACGCGTACCGTGAGCTCGTTGTAGAGAATAAGACAATGTCGG
AGAAGGAGTTTGTCAAGGCGGTCATGATCAAGGGCCGCGACCATGCGCGCACTCCCAT
GCAATGGGACGCGGGCAAGCATGCCGGATTCACAGCCGGTGACCCTTGGATCCCAGTG
AACTCACGCTACCAAGACATCAACGTAAAGGAGTCTCTTGAGGACCAAGACTCGATCTT
CTTCTATTACCAAAAGCTCATCCAATTACGTAAGCAATACAAGATTATGATCTACGGAG
ACTACCAATTACTCCAGGAGAACGACCCACAAGTGTTCTCGTACCTTCGCGAGTACCGC
GGGGAGAAGTTATTGGTCGTGGTGAACCTTTCAGAGGAGAAGGCTCTTTTCGAGGCTC
CGCCAGAGTTGATCCACGAGCGCTGGAAGGTCCTTATCTCTAACTACCCTCAAGAGCGC

GCGGACTTAAAGTCAATCTCATTGAAGCCATACGAGGCGGTCATGGGGATCTCGATTTA
A

8.1.5.2 Protein sequence – 582 amino acids – 68.4 kDa

MGSSHHHHHHSSGLVPRGSHMMGEWWKEAVVYQIYPRSFYDANGDGFGDLQGVIQKL
DYIKNLGADVIWLSPVFDSPQDDNGYDISDYKNMYEKFGTNEDMFQLIDEVHKRGMKIV
MDLVVNHTSDEHAWFAESRKSKDNPYRDYYLWKDPKPDGSEPNNWGSIFSGSAWTYDE
GTGQYYLHYFSKKQPDLNWENEAVRREVYDVMRFWMDRGVDGWRMDVIGSISKYTDFP
DYETDHSRSYIVGRYHSNGPRLHEFIQEMNREVLSHYDCMTVGEANGSDIEEAKKYTDASR
QELNMIFTFEHMDIDKEQNSPNGKWQIKPFDLIALKKTMTRWQTGLMNVGWNTLYFEN
HDQPRVISRWGNDRKLRKECAKAFATVLHGMKGTPFIYQGEEIGM**R**NSDMPLEMYDDLE
IKNAYRELVVENKTMSEKEFVKAVMIKGRDHARTPMQWDAGKHAGFTAGDPWIPVNSR
YQDINVKESLEDQDSIFFYYQKLIQLRKQYKIMIYGDYQLLQENDPQVFSYLREYRGEKLLVV
VNLSEEKALFEAPPELIHERWKVLISNYPQERADLKSISLKPYEAVMGISI*

8.2 Summary of kinetic data

8.2.1 Michaelis-Menten kinetics

Table 8.1. Michaelis-Menten parameters for arginine mutants with standard error

Parameter	MalL V376R	MalL T150R	MalL D492R	MalL RDM
k_{cat} (sec^{-1})	29.7 ± 1.1	31.2 ± 0.8	38.1 ± 1.3	51.0 ± 1.4
K_M (mM)	0.062 ± 0.005	0.053 ± 0.004	0.071 ± 0.007	0.080 ± 0.006
K_i (mM)	1.6 ± 0.2	3.4 ± 0.3	4.4 ± 0.6	3.9 ± 0.4

8.2.2 Temperature characterisation

Table 8.2. MMRT 1.5 fits of MalL variants with standard error

	MalL WT	MalL V376R	MalL T150R	MalL D492R	MalL RDM
$\Delta H^{\ddagger}_{T_0}$ $(kJ\ mol^{-1}\ K^{-1})$	19.0 ± 16.0	48.9 ± 10.2	5.7 ± 20.5	69.2 ± 5.9	18.1 ± 18.4
$\Delta S^{\ddagger}_{T_0}$ $(J\ mol^{-1}\ K^{-1})$	-174.0 ± 56.5	-65.0 ± 36.2	-216.0 ± 72.3	10.0 ± 20.7	-172.4 ± 65.2
$\Delta C^{\ddagger}_{P,0}$ $(kJ\ mol^{-1}\ K^{-1})$	210 ± 18.0	128.4 ± 13.3	215.2 ± 24.5	108.4 ± 7.4	227.8 ± 24.9
m $(kJ\ mol^{-1}\ K^{-2})$	-709.9 ± 59.3	-436.4 ± 44.0	-729.5 ± 81.0	-371.9 ± 24.6	-774.5 ± 82.9
$T_0\ (K)$	278.15	278.15	278.15	278.15	278.15
$T_{opt}\ (K)$	315.1	316.7	312.6	315.2	311.6
$T_{inf}\ (K)$	307.5	307.3	304.9	305.7	304.1

Table 8.3. MMRT 2.0 fits of MalL variants with standard error

Parameter	MalL WT	MalL V376R	MalL T150R	MalL D492R	MalL RDM
$\Delta H^{\ddagger}_{T_0}$ $(kJ\ mol^{-1}\ K^{-1})$	96.6 ± 2.3	92.0 ± 0.7	109.2 ± 0.9	95.2 ± 1.8	92.9 ± 0.7
$\Delta S^{\ddagger}_{T_0}$ $(J\ mol^{-1}\ K^{-1})$	102.3 ± 8.2	88.7 ± 2.5	150.5 ± 3.2	102.9 ± 6.2	94.3 ± 2.4
$\Delta\Delta H^{\ddagger}$ $(kJ\ mol^{-1}\ K^{-1})$	186.7 ± 45.5	170.0 ± 19.4	380.7 ± 105.4	167.3 ± 16.9	235.4 ± 21.4
$\Delta C^{\ddagger}_{P,lowT}$ $(kJ\ mol^{-1}\ K^{-1})$	0.8 ± 0.2	0.2 ± 0.2	-0.8 ± 0.2	0.1 ± 0.1	0.2 ± 0.1
$\Delta C^{\ddagger}_{P,highT}$ $(kJ\ mol^{-1}\ K^{-1})$	-28.1 ± 6.3	-28.4 ± 7.4	-21.9 ± 3.2	-11.7 ± 0.2	-34.9 ± 3.7
T_C (K)	313.5 ± 2.7	316.1 ± 2.6	309.1 ± 1.1	307	311.5 ± 0.9
Fit Range	1-6	1-6	1-6	1-5	1-6
$T_0\ (K)$	278.15	278.15	278.15	278.15	278.15

8.3 Melting temperature

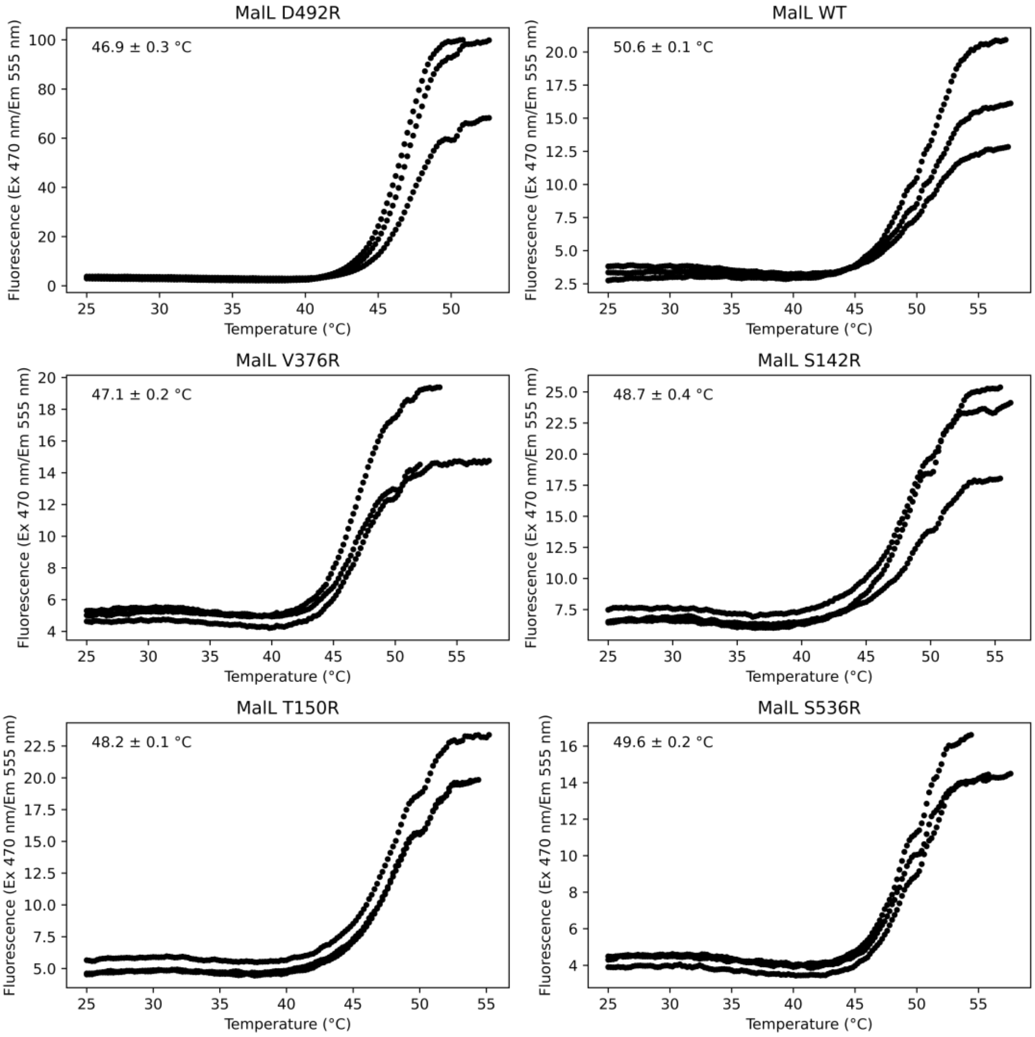


Figure 8.1. Raw scans of SYPRO temperature assay (Section 2.7).

8.4 Python scripts

Code Snippet 8.1. Analyse hydrogen bond output from USCF Chimera 1.15

```
"""
Copyright (C) 2022 Carlin Hamill, University of Waikato

Permission is hereby granted, free of charge, to any person obtaining a copy of
this software and associated documentation files (the "Software"), to deal in
the Software without restriction, including without limitation the rights to
use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies
of the Software, and to permit persons to whom the Software is furnished to do
so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all
```

```
copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR
IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY,
FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE
AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER
LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM,
OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE
SOFTWARE.
"""

from __future__ import annotations
import re
import pathlib
import typing
import argparse


class _BondMember:
    """
    Represents an atom/residue involved in forming a hydrogen bond
    """

    def __init__(self, input_string: str):
        input_string = re.split(r"\s+(?!hydrogen)|\.(?=\D)", input_string)

        if input_string == ["no hydrogen"]:
            self.res = None
            self.res_num = None
            self.chain = None
            self.atom_name = None
        else:
            self.res = input_string[0]
            self.res_num = int(input_string[1])
            self.chain = input_string[2]
            self.atom_name = input_string[3]

            if len(input_string) == 5:
                self.alt_id = input_string[4]
            else:
                self.alt_id = None


def __eq__(self, other: _BondMember) -> bool:
        """
        Checks if bond member is equivalent to another
        checks only residue and residue number match (eg TRP 4 != TYR 11)
        returns true if both are water
        """
        if self.res != other.res:
            return False
        if self.res == "HOH" and other.res == "HOH":
            return True
        return self.res_num == other.res_num

    def __repr__(self) -> str:
        return f"Residue: {self.res_num}, Atom: {self.atom_name}, Chain: {self.chain}"
```

```
    @property
    def pymol_atom_id(self) -> str:
        """pymol atom identifier string for _BondMember atom"""
        if self.alt_id is not None:
            return f"{self.chain}/{self.res_num}/{self.atom_name}`{self.alt_id}"
        return f"{self.chain}/{self.res_num}/{self.atom_name}"

    @property
    def pymol_res_id(self) -> str:
        """pymol identifier string for residue of _BondMember"""
        if self.alt_id is not None:
            return f"{self.chain}/{self.res_num}/`{self.alt_id}"
        return f"{self.chain}/{self.res_num}/"

    def __hash__(self):
        return hash((self.res, self.res_num))


class _HydrogenBond:
    """
    Represents a hydrogen bond
    """

    def __init__(self, line: str, bond_threshold: float = 0.3):
        splits = re.split(r"\s{2,}", line)
        self.donor = _BondMember(splits[0])
        self.acceptor = _BondMember(splits[1])
        self.hydrogen = _BondMember(splits[2])
        self.donor_acceptor_dist = splits[3]
        self.hydrogen_acceptor_dist = splits[4]
        self.line = line

        self.bond_threshold = bond_threshold

    def __repr__(self) -> str:
        return self.line

    def __eq__(self, other: _HydrogenBond) -> bool:
        """
        checks if hydgron bonds are equivalent. Only checks if donor residue and acceptor residue
        involved are equivalent.
        """
        if not isinstance(other, _HydrogenBond):
            return False
        if self.donor != other.donor:
            return False
        return self.acceptor == other.acceptor

    def __lt__(self, other: _HydrogenBond) -> bool:
        return (
            float(other.donor_acceptor_dist) - float(self.donor_acceptor_dist)
        ) > self.bond_threshold

    def _set_bond_threshold(self, threshold):
        self.bond_threshold = threshold
```

```
    @property
    def pymol_dist(self) -> str:
        """pymol distance command to show hydrogen bonds"""
        return f"dist {self.donor.pymol_atom_id}, {self.acceptor.pymol_atom_id}"

    @property
    def pymol_show_sticks(self) -> str:
        return f"show sticks, {self.donor.pymol_res_id} {self.acceptor.pymol_res_id}"

    def __hash__(self):
        return hash((self.donor, self.acceptor))


class HbondsFromFile(list):
    """
    Reads in hydrogen bonds from UCSF Chimera 1.15 Find_Hbond output file.
    Returns a list of HydrogenBond objects
    """

    def __init__(self, file: typing.Union[str, pathlib.Path]):
        with open(file) as f:
            for line in f.readlines()[7:]:
                self.append(_HydrogenBond(line))

    def _set_bond_threshold(self, threshold: float) -> None:
        for bond in self:
            bond._set_bond_threshold(threshold)


def compare_hbonds(
    query_list: list[_HydrogenBond], *match_lists: list[_HydrogenBond]
) -> list[_HydrogenBond]:
    """
    Returns a list of hydrogen bond objects unique to the query list
    (not present in the match list/s). Matches res name and res number (eg TRP 125) only.
    Water residue numbers are ignored.
    """
    return [
        bond for bond in query_list if not any([bond in lst for lst in match_lists])
    ]


def compare_bond_lengths(
    query_list: list[_HydrogenBond],
    *match_lists: list[_HydrogenBond],
    bond_threshold: float = None,
) -> list[_HydrogenBond]:
    """
    Retruns any bond in query list that is significantly shorter
    """
    if bond_threshold is not None:
        query_list._set_bond_threshold(bond_threshold)
        for match_list in match_lists:
            match_list._set_bond_threshold(bond_threshold)

    smaller_hbond = []
    same_bonds = _get_matching_hbonds(query_list, *match_lists)
```

```
    for bond in same_bonds:
        matched_bonds = _return_matching_hbonds(bond, *match_lists)
        if all([bond < match_bond for match_bond in matched_bonds]):
            smaller_hbond.append(bond)
    return smaller_hbond


def _get_matching_hbonds(
    query_list: list[_HydrogenBond], *match_lists: list[_HydrogenBond]
) -> list[_HydrogenBond]:
    """
    Retruns any bond in query list that is present all match lists
    """
    return [
        bond
        for bond in query_list
        if all([bond in match_list for match_list in match_lists])
    ]


def _return_matching_hbonds(
    query_bond: _HydrogenBond, *match_lists: list[_HydrogenBond]
) -> list[_HydrogenBond]:
    """
    Returns any bond in any match list that is equivalent to the query bond
    """
    matched_bonds = []
    for match_list in match_lists:
        for match_bond in match_list:
            if query_bond == match_bond:
                matched_bonds.append(match_bond)
    return matched_bonds


def write_output(
    output_func: typing.Callable,
    unique: list[_HydrogenBond],
    short: list[_HydrogenBond],
    bond_threshold: float,
    *args,
    **kwargs,
) -> None:
    output_func(
        "Unique Hydrogen Bonds\nH-bonds (donor, acceptor, hydrogen, D..A dist, D-H..A dist):\tPyMOL Distance command\n",
        *args,
        **kwargs,
    )
    [
        output_func(f"{bond.line.rstrip()}\t{bond.pymol_dist}\n", *args, **kwargs)
        for bond in unique
    ]
    output_func(
        f"\nShortened Hydrogen Bonds: Threshhold = {bond_threshold} Angstroms\nH-bonds (donor, acceptor, hydrogen, D..A dist, D-H..A dist):\tPyMOL Distance command\n",
        *args,
        **kwargs,
```

```
    )
    [
        output_func(f"{bond.line.rstrip()}\t{bond.pymol_dist}\n", *args, **kwargs)
        for bond in short
    ]


def main() -> None:
    parser = argparse.ArgumentParser()
    parser.add_argument(
        "query_bond_file", type=str, help="Chimera bond list to act as query"
    )
    parser.add_argument(
        "match_bond_file", type=str, help="Chimera bond list to act as match"
    )
    parser.add_argument(
        "--output_file", "-o", type=str, help="Filepath to export output txt file"
    )
    parser.add_argument(
        "--bond_threshold",
        "-b",
        type=float,
        default=0.3,
        help="Bond length difference in Angstrom to consider as significantly different.",
    )
    args = parser.parse_args()

    query = HbondsFromFile(args.query_bond_file)
    query._set_bond_threshold(args.bond_threshold)
    match = HbondsFromFile(args.match_bond_file)
    match._set_bond_threshold(args.bond_threshold)

    unique = compare_hbonds(query, match)
    short = compare_bond_lengths(query, match)

    if args.output_file is None:
        write_output(print, unique, short, args.bond_threshold, end="")
        return

    with open(args.output_file, "w") as output:
        write_output(output.write, unique, short, args.bond_threshold)
if __name__ == "__main__":
    main()
```

Code Snippet 8.2. Linear regression of absorbance rate data

```
"""
Copyright (C) 2023 Carlin Hamill, University of Waikato

Permission is hereby granted, free of charge, to any person obtaining a copy of
this software and associated documentation files (the "Software"), to deal in
the Software without restriction, including without limitation the rights to
use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies
of the Software, and to permit persons to whom the Software is furnished to do
```

```
so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all
copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR
IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY,
FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE
AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER
LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM,
OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE
SOFTWARE.
"""

from dataclasses import dataclass
import copy
from sklearn.linear_model import LinearRegression
import numpy as np

@dataclass
class Trace:
    x: np.ndarray # time
    y: np.ndarray # Absorbance

    def linear_regression(self, max_x:float=None, rsq:float=0.99)->None:
                            #set regression range (0 -> max_x)
        if max_x is None or max_x > max(self.x):
            max_x = max(self.x)
        store_max_x = copy.deepcopy(maxs)
        score = 0
        max_x+=0.1

        #iterate until fit score (score) is > goal R-squared (rsq)
        while score < rsq:
            #decrease regrassion range by 0.1 seconds each iteration
            max_x -= 0.1
            x = self.x[np.where(self.x<=maxs)].reshape(-1,1)
            y = self.y[np.where(self.x<=maxs)]
            model = LinearRegression().fit(x, y)
            score = model.score(x,y)
            #reset regression range and decrease rsq if range decreases by half
            if max_x < store_max_x/2:
                rsq -= 0.1
                max_x = copy.deepcopy(store_max_x)
        self.max_x = max_x
        self.slope = float(model.coef_)
```

8.5 Hydrogen bond differences

Hydrogen bonds present or significantly shortened (>0.3 Å) in one structure, that are absent or lengthened in the other structure. Hydrogen bonds are represented as donor atom, acceptor atom, hydrogen, donor-acceptor distance, hydrogen-acceptor distance.

8.5.1 MalL S536R versus MalL wildtype Chain A

8.5.1.1 MalL S536R

GLN 12.B NE2	SER 50.B OG.C	no hydrogen	3.160	N/A
GLY 75.B N	TYR 71.B O	no hydrogen	3.349	N/A
TYR 125.B N	ARG 122.B O	no hydrogen	3.524	N/A
VAL 177.B N	ASN 174.B O	no hydrogen	2.944	N/A
HIS 283.B N	HIS 283.B ND1	no hydrogen	2.806	N/A
LYS 348.B NZ	ASP 461.B OD2	no hydrogen	3.454	N/A
GLU 349.B N	GLU 349.B OE1	no hydrogen	2.840	N/A
GLN 369.B NE2	TYR 14.B O	no hydrogen	3.616	N/A
LYS 391.B N	ASP 387.B O	no hydrogen	3.220	N/A
LYS 402.B NZ	GLU 400.B O	no hydrogen	3.348	N/A
ARG 422.B NH2	ASP 332.B OD1	no hydrogen	3.190	N/A
THR 423.B N	ASP 419.B O	no hydrogen	3.307	N/A
ALA 433.B N	GLY 430.B O	no hydrogen	2.841	N/A
ASN 445.B ND2	ASN 377.B O	no hydrogen	3.023	N/A
GLU 458.B N	GLU 455.B O	no hydrogen	3.201	N/A
SER 547.B N	ASP 544.B O	no hydrogen	3.411	N/A

8.5.1.2 MalL S536R shortened

GLY 27.B N	ARG 16.B O	no hydrogen	2.789	N/A
LYS 40.B NZ.A	ARG 90.B O	no hydrogen	2.788	N/A
ARG 114.B NH2	ASP 152.B OD1	no hydrogen	2.879	N/A
LYS 265.B NZ	GLU 262.B OE2	no hydrogen	2.734	N/A
ASN 294.B N	SER 292.B OG	no hydrogen	2.911	N/A
LYS 506.B NZ	GLU 529.B OE2	no hydrogen	2.985	N/A

8.5.1.3 MalL wildtype

GLY 75.A N	GLU 72.A O	no hydrogen	3.087	N/A
ASP 79.A N	THR 76.A OG1	no hydrogen	3.306	N/A
LYS 89.A NZ	GLU 86.A OE2	no hydrogen	3.313	N/A
SER 104.A OG	ASP 105.A OD1	no hydrogen	3.509	N/A
TYR 207.A OH	ASN 229.A O	no hydrogen	3.216	N/A
TYR 221.A OH	GLU 136.A OE2	no hydrogen	2.966	N/A
MET 277.A N	THR 252.A O	no hydrogen	3.322	N/A

ARG 335.A NE HIS 420.A O no hydrogen 3.481 N/A
ASP 459.A N GLU 455.A O no hydrogen 3.098 N/A

8.5.1.4 MalL wildtype shortened

HIS 247.A N VAL 244.A O no hydrogen 2.662 N/A
ARG 313.A NH1 GLU 282.A OE1 no hydrogen 2.610 N/A
VAL 495.A N ASP 492.A OD2 no hydrogen 2.636 N/A

8.5.2 MalL S536R versus MalL wildtype Chain B

8.5.2.1 MalL S536R

GLY 75.B N TYR 71.B O no hydrogen 3.349 N/A
PHE 110.B N HIS 107.B O no hydrogen 2.976 N/A
TYR 125.B N ARG 122.B O no hydrogen 3.524 N/A
ARG 186.B NH2 GLU 243.B O no hydrogen 3.021 N/A
HIS 283.B N HIS 283.B ND1 no hydrogen 2.806 N/A
LYS 288.B NZ GLU 289.B O no hydrogen 3.170 N/A
LYS 348.B NZ ASP 461.B OD2 no hydrogen 3.454 N/A
GLU 349.B N GLU 349.B OE1 no hydrogen 2.840 N/A
LYS 391.B N ASP 387.B O no hydrogen 3.220 N/A
LYS 402.B NZ VAL 399.B O no hydrogen 2.972 N/A
LYS 402.B NZ GLU 400.B O no hydrogen 3.348 N/A
ARG 422.B NH2 ASP 332.B OD1 no hydrogen 3.190 N/A
ASN 445.B ND2 ASN 377.B O no hydrogen 3.023 N/A
GLU 458.B N GLU 455.B O no hydrogen 3.201 N/A
LYS 469.B NZ.A LEU 534.B O no hydrogen 3.449 N/A
ILE 535.B N MET 557.B O no hydrogen 3.494 N/A
SER 547.B N ASP 544.B O no hydrogen 3.411 N/A
LYS 551.B NZ GLN 540.B OE1 no hydrogen 3.427 N/A

8.5.2.2 MalL S536R shortened

GLN 30.B NE2 GLN 82.B OE1 no hydrogen 2.933 N/A
LYS 40.B NZ.A ARG 90.B O no hydrogen 2.788 N/A
LYS 73.B NZ GLU 72.B OE2 no hydrogen 2.872 N/A
LYS 265.B NZ GLU 262.B OE2 no hydrogen 2.734 N/A
ASN 294.B N SER 292.B OG no hydrogen 2.911 N/A
LYS 348.B NZ ASP 459.B OD2 no hydrogen 2.748 N/A
LYS 475.B NZ GLY 43.B O no hydrogen 2.788 N/A

8.5.2.3 MalL wildtype

GLY 75.B N GLU 72.B O no hydrogen 3.135 N/A
ASP 79.B N THR 76.B OG1 no hydrogen 3.331 N/A
SER 104.B OG ASP 105.B OD1 no hydrogen 3.511 N/A
TYR 207.B OH ASN 229.B O no hydrogen 3.334 N/A

TYR 221.B OH GLU 136.B OE2 no hydrogen 3.083 N/A
MET 277.B N THR 252.B O no hydrogen 3.321 N/A
ARG 335.B NE HIS 420.B O no hydrogen 3.469 N/A
ILE 415.B N LYS 411.B O no hydrogen 3.291 N/A
LYS 431.B NZ ALA 437.B O no hydrogen 3.169 N/A
ASP 459.B N GLU 455.B O no hydrogen 3.053 N/A
TRP 531.B N ALA 543.B O no hydrogen 2.791 N/A
GLN 540.B NE2 GLU 554.B OE2 no hydrogen 2.638 N/A
ARG 542.B NH2 SER 547.B O no hydrogen 2.769 N/A

8.5.2.4 MalL wildtype shortened

HIS 161.B ND1 PHE 163.B O no hydrogen 2.653 N/A
HIS 247.B N VAL 244.B O no hydrogen 2.661 N/A
LYS 300.B NZ GLU 289.B OE1 no hydrogen 3.057 N/A
LYS 300.B NZ GLU 289.B OE2 no hydrogen 2.862 N/A
ARG 313.B NH1 GLU 282.B OE1 no hydrogen 2.588 N/A
VAL 495.B N ASP 492.B OD2 no hydrogen 2.654 N/A

8.5.3 MalL D492R Chain A versus MalL wildtype Chain A

8.5.3.1 MalL D492R

GLU 3.A N GLU 3.A OE1 no hydrogen 2.821 N/A
GLN 12.A NE2 SER 50.A OG no hydrogen 3.143 N/A
GLY 75.A N TYR 71.A O no hydrogen 3.355 N/A
GLN 82.A NE2 GLU 86.A OE2 no hydrogen 3.031 N/A
ASP 118.A N SER 116.A OG no hydrogen 2.935 N/A
TYR 125.A N ARG 122.A O no hydrogen 3.431 N/A
ARG 186.A NH2 GLU 243.A O no hydrogen 2.985 N/A
HIS 247.A NE2 ARG 242.A O no hydrogen 2.996 N/A
LYS 309.A NZ GLU 490.A OE2 no hydrogen 2.404 N/A
LYS 391.A N ASP 387.A O no hydrogen 3.156 N/A
LYS 402.A NZ GLU 400.A O no hydrogen 3.064 N/A
ARG 422.A NH2 ASP 332.A OD1 no hydrogen 3.186 N/A
LYS 475.A NZ ASP 45.A OD1 no hydrogen 3.227 N/A
LYS 475.A NZ ASP 45.A OD2 no hydrogen 2.783 N/A
VAL 495.A N PRO 493.A O no hydrogen 2.803 N/A

8.5.3.2 MalL D492R shortened

VAL 9.A N ASP 45.A OD2 no hydrogen 2.895 N/A
LYS 40.A NZ ARG 90.A O no hydrogen 2.776 N/A
ARG 114.A NH2 ASP 152.A OD1 no hydrogen 2.858 N/A
HIS 234.A NE2 GLU 263.A OE1 no hydrogen 2.713 N/A
ASN 294.A N SER 292.A OG no hydrogen 2.907 N/A
ALA 306.A N ASP 303.A OD1 no hydrogen 3.059 N/A
LYS 391.A NZ TYR 385.A O no hydrogen 2.864 N/A

8.5.3.3 MalL wildtype

```
LYS 89.A NZ    GLU 86.A OE2   no hydrogen  3.313  N/A
SER 104.A OG   ASP 105.A OD1  no hydrogen  3.509  N/A
SER 116.A OG   ASP 118.A OD1  no hydrogen  2.811  N/A
TYR 207.A OH   ASN 229.A O    no hydrogen  3.216  N/A
LYS 266.A NZ   GLU 263.A OE1  no hydrogen  3.522  N/A
LYS 308.A NZ   GLU 490.A O    no hydrogen  2.941  N/A
GLU 383.A N    GLU 383.A OE1  no hydrogen  2.796  N/A
ARG 418.A NH2  ASP 332.A OD1  no hydrogen  2.953  N/A
ILE 451.A N    TYR 448.A O    no hydrogen  3.255  N/A
ASN 452.A ND2  GLN 449.A O    no hydrogen  3.375  N/A
LYS 475.A NZ   GLY 43.A O     no hydrogen  2.926  N/A
GLU 529.A N    LEU 526.A O    no hydrogen  3.117  N/A
```

8.5.3.4 MalL wildtype shortened

```
ASN 342.A ND2  GLU 515.A OE1  no hydrogen  2.945  N/A
```

8.5.4 MalL D492R Chain B versus MalL wildtype Chain A

8.5.4.1 MalL D492R

```
GLU 3.B N       GLU 3.B OE1     no hydrogen  2.885  N/A
GLN 12.B NE2    SER 50.B OG     no hydrogen  3.143  N/A
GLN 82.B NE2    GLU 86.B OE2    no hydrogen  3.059  N/A
ASP 118.B N     SER 116.B OG    no hydrogen  2.937  N/A
TYR 125.B N     ARG 122.B O     no hydrogen  3.415  N/A
ARG 186.B NH2   GLU 243.B O     no hydrogen  3.011  N/A
HIS 247.B NE2   ARG 242.B O     no hydrogen  2.974  N/A
LYS 288.B NZ    GLU 289.B O     no hydrogen  3.561  N/A
LYS 309.B NZ    ASN 491.B OD1   no hydrogen  2.828  N/A
LYS 391.B N     ASP 387.B O     no hydrogen  3.178  N/A
LYS 402.B NZ    GLU 400.B O     no hydrogen  3.365  N/A
LYS 454.B NZ    GLU 458.B OE1   no hydrogen  3.430  N/A
LYS 454.B NZ    GLU 458.B OE2   no hydrogen  2.693  N/A
VAL 495.B N     PRO 493.B O     no hydrogen  2.802  N/A
LYS 532.B NZ    TYR 477.B OH    no hydrogen  3.386  N/A
```

8.5.4.2 MalL D492R shortened

```
LYS 40.B NZ     ARG 90.B O      no hydrogen  2.772  N/A
ARG 114.B NH2   ASP 152.B OD1   no hydrogen  2.835  N/A
HIS 234.B NE2   GLU 263.B OE2   no hydrogen  2.715  N/A
ASN 294.B N     SER 292.B OG    no hydrogen  2.910  N/A
SER 378.B N     HIS 420.B ND1   no hydrogen  2.895  N/A
LYS 391.B NZ    TYR 385.B O     no hydrogen  2.843  N/A
LYS 506.B NZ    GLU 529.B OE2   no hydrogen  2.642  N/A
```

8.5.4.3 MalL wildtype

```
ASN 69.A ND2   ASP 66.A OD2   no hydrogen  2.771  N/A
LYS 89.A NZ    GLU 86.A OE2   no hydrogen  3.313  N/A
SER 104.A OG   ASP 105.A OD1  no hydrogen  3.509  N/A
SER 116.A OG   ASP 118.A OD1  no hydrogen  2.811  N/A
TYR 207.A OH   ASN 229.A O    no hydrogen  3.216  N/A
LYS 308.A NZ   GLU 490.A O    no hydrogen  2.941  N/A
GLU 383.A N    GLU 383.A OE1  no hydrogen  2.796  N/A
ASN 452.A ND2  GLN 449.A O    no hydrogen  3.375  N/A
GLU 529.A N    LEU 526.A O    no hydrogen  3.117  N/A
```

8.5.5 MalL T150R versus MalL wildtype Chain A

8.5.5.1 MalL T150R

```
GLN 12.A NE2     SER 50.A OG    no hydrogen  3.146  N/A
GLN 82.A NE2     GLU 86.A OE2   no hydrogen  3.179  N/A
LYS 131.A NZ     ASP 212.A OD2  no hydrogen  3.056  N/A
ARG 186.A NH2    GLU 243.A O    no hydrogen  2.991  N/A
HIS 247.A NE2    ARG 242.A O    no hydrogen  2.987  N/A
SER 292.A OG     GLN 298.A OE1  no hydrogen  3.340  N/A
LYS 348.A NZ     ASP 461.A O    no hydrogen  3.541  N/A
LYS 348.A NZ     ASP 461.A OD2  no hydrogen  2.989  N/A
LYS 391.A N      ASP 387.A O    no hydrogen  3.201  N/A
LYS 391.A NZ     ASP 386.A O    no hydrogen  3.191  N/A
LYS 402.A NZ     GLU 400.A O    no hydrogen  3.399  N/A
ARG 422.A NH2    ASP 332.A OD1  no hydrogen  3.127  N/A
LYS 454.A NZ     GLU 458.A OE1  no hydrogen  3.320  N/A
LYS 454.A NZ     GLU 458.A OE2  no hydrogen  2.872  N/A
LYS 475.A NZ     ASP 45.A OD2   no hydrogen  2.589  N/A
LYS 532.A NZ     TYR 477.A OH   no hydrogen  3.265  N/A
```

8.5.5.2 MalL T150R shortened

```
GLY 27.A N       ARG 16.A O     no hydrogen  2.770  N/A
SER 116.A OG     ASP 118.A OD1  no hydrogen  2.442  N/A
HIS 234.A NE2    GLU 263.A OE2  no hydrogen  2.719  N/A
LYS 265.A NZ     GLU 262.A OE2  no hydrogen  2.838  N/A
ASN 294.A N      SER 292.A OG   no hydrogen  3.035  N/A
ALA 306.A N      ASP 303.A OD1  no hydrogen  3.125  N/A
LYS 546.A NZ     ASP 544.A OD1  no hydrogen  2.692  N/A
```

8.5.5.3 MalL wildtype

```
LYS 89.A NZ    GLU 86.A OE2   no hydrogen  3.313  N/A
ARG 114.A NH2  ASP 152.A OD1  no hydrogen  3.204  N/A
TYR 124.A OH   GLU 180.A OE1  no hydrogen  2.783  N/A
```

```
TYR 158.A OH   GLY 134.A O    no hydrogen  2.890  N/A
LYS 266.A NZ   GLU 263.A OE1  no hydrogen  3.522  N/A
LYS 391.A NZ   TYR 385.A O    no hydrogen  3.552  N/A
ARG 418.A NH2  ASP 332.A OD1  no hydrogen  2.953  N/A
LYS 475.A NZ   GLY 43.A O     no hydrogen  2.926  N/A
```

8.5.5.4 MalL wildtype shortened

```
ARG 313.A NH1  GLU 282.A OE1  no hydrogen  2.610  N/A
VAL 495.A N    ASP 492.A OD2  no hydrogen  2.636  N/A
```

8.5.6 MalL D492R Chain A versus MalL wildtype Chain B

8.5.6.1 MalL D492R

```
GLU 3.A N       GLU 3.A OE1     no hydrogen  2.821  N/A
GLN 82.A NE2    GLU 86.A OE2    no hydrogen  3.031  N/A
ASP 118.A N     SER 116.A OG    no hydrogen  2.935  N/A
TYR 125.A N     ARG 122.A O     no hydrogen  3.431  N/A
ARG 186.A NH2   GLU 243.A O     no hydrogen  2.985  N/A
ASP 209.A N     ASP 209.A OD1   no hydrogen  2.681  N/A
HIS 247.A NE2   ARG 242.A O     no hydrogen  2.996  N/A
LYS 309.A NZ    GLU 490.A OE2   no hydrogen  2.404  N/A
LYS 391.A N     ASP 387.A O     no hydrogen  3.156  N/A
SER 405.A N     GLU 408.A OE1   no hydrogen  2.949  N/A
ILE 451.A N     ARG 447.A O     no hydrogen  3.007  N/A
LYS 475.A NZ    ASP 45.A OD1    no hydrogen  3.227  N/A
LYS 475.A NZ    ASP 45.A OD2    no hydrogen  2.783  N/A
LYS 478.A NZ    LYS 475.A O     no hydrogen  2.715  N/A
VAL 495.A N     PRO 493.A O     no hydrogen  2.803  N/A
GLN 540.A NE2   TYR 538.A O     no hydrogen  3.579  N/A
```

8.5.6.2 MalL D492R shortened

```
VAL 9.A N       ASP 45.A OD2    no hydrogen  2.895  N/A
LYS 40.A NZ     ARG 90.A O      no hydrogen  2.776  N/A
LYS 73.A NZ     GLU 72.A OE2    no hydrogen  2.977  N/A
ARG 114.A NH2   ASP 152.A OD1   no hydrogen  2.858  N/A
ALA 306.A N     ASP 303.A OD1   no hydrogen  3.059  N/A
GLU 525.A N     GLU 525.A OE1   no hydrogen  2.683  N/A
GLN 540.A NE2   SER 536.A OG    no hydrogen  2.820  N/A
```

8.5.6.3 MalL wildtype

```
SER 104.B OG   ASP 105.B OD1  no hydrogen  3.511  N/A
TYR 207.B OH   ASN 229.B O    no hydrogen  3.334  N/A
LYS 308.B NZ   GLU 490.B O    no hydrogen  3.137  N/A
GLU 383.B N    GLU 383.B OE1  no hydrogen  2.769  N/A
```

```
ARG 418.B NH2  ASP 332.B OD1  no hydrogen  2.982  N/A
ASN 452.B ND2  GLN 449.B O    no hydrogen  3.380  N/A
LYS 506.B NZ   GLU 529.B OE1  no hydrogen  3.250  N/A
GLU 529.B N    LEU 526.B O    no hydrogen  3.113  N/A
TRP 531.B N    ALA 543.B O    no hydrogen  2.791  N/A
```

8.5.6.4 MalL wildtype shortened

```
HIS 161.B ND1  PHE 163.B O    no hydrogen  2.653  N/A
ASN 342.B ND2  GLU 515.B OE1  no hydrogen  2.943  N/A
```

8.5.7 MalL D492R Chain B versus MalL wildtype Chain B

8.5.7.1 MalL D492R

```
GLU 3.B N        GLU 3.B OE1      no hydrogen  2.885  N/A
GLN 82.B NE2     GLU 86.B OE2     no hydrogen  3.059  N/A
ASP 118.B N      SER 116.B OG     no hydrogen  2.937  N/A
TYR 125.B N      ARG 122.B O      no hydrogen  3.415  N/A
ARG 186.B NH2    GLU 243.B O      no hydrogen  3.011  N/A
HIS 247.B NE2    ARG 242.B O      no hydrogen  2.974  N/A
LYS 309.B NZ     ASN 491.B OD1    no hydrogen  2.828  N/A
LYS 391.B N      ASP 387.B O      no hydrogen  3.178  N/A
VAL 495.B N      PRO 493.B O      no hydrogen  2.802  N/A
LYS 546.B NZ     ILE 527.B O      no hydrogen  3.273  N/A
```

8.5.7.2 MalL D492R shortened

```
GLN 30.B NE2     GLN 82.B OE1     no hydrogen  2.915  N/A
LYS 40.B NZ      ARG 90.B O       no hydrogen  2.772  N/A
ASN 60.B ND2     TYR 14.B OH      no hydrogen  3.182  N/A
TRP 109.B N      HIS 107.B ND1    no hydrogen  2.915  N/A
ARG 114.B NH2    ASP 152.B OD1    no hydrogen  2.835  N/A
LYS 454.B NZ     GLU 458.B OE2    no hydrogen  2.693  N/A
LYS 506.B NZ     GLU 529.B OE2    no hydrogen  2.642  N/A
GLN 540.B NE2    SER 536.B OG     no hydrogen  2.898  N/A
```

8.5.7.3 MalL wildtype

```
SER 104.B OG   ASP 105.B OD1  no hydrogen  3.511  N/A
LYS 308.B NZ   GLU 490.B O    no hydrogen  3.137  N/A
GLU 383.B N    GLU 383.B OE1  no hydrogen  2.769  N/A
GLU 529.B N    LEU 526.B O    no hydrogen  3.113  N/A
TRP 531.B N    ALA 543.B O    no hydrogen  2.791  N/A
```

8.5.7.4 MalL wildtype shortened

```
HIS 161.B ND1  PHE 163.B O    no hydrogen  2.653  N/A
ARG 313.B NH1  GLU 282.B OE1  no hydrogen  2.588  N/A
```

8.5.8 MalL T150R versus MalL wildtype Chain B

8.5.8.1 MalL T150R

```
GLN 82.A NE2     GLU 86.A OE2   no hydrogen  3.179  N/A
ARG 186.A NH2    GLU 243.A O    no hydrogen  2.991  N/A
ASP 209.A N      ASP 209.A OD1  no hydrogen  2.514  N/A
HIS 247.A NE2    ARG 242.A O    no hydrogen  2.987  N/A
SER 292.A OG     GLN 298.A OE1  no hydrogen  3.340  N/A
LYS 348.A NZ     ASP 461.A O    no hydrogen  3.541  N/A
LYS 348.A NZ     ASP 461.A OD2  no hydrogen  2.989  N/A
LYS 391.A N      ASP 387.A O    no hydrogen  3.201  N/A
ARG 422.A NH2    ASP 332.A OD1  no hydrogen  3.127  N/A
LYS 475.A NZ     GLU 7.A O      no hydrogen  2.997  N/A
LYS 475.A NZ     ASP 45.A OD2   no hydrogen  2.589  N/A
LYS 532.A NZ     TYR 477.A OH   no hydrogen  3.265  N/A
```

8.5.8.2 MalL T150R shortened

```
GLY 27.A N       ARG 16.A O     no hydrogen  2.770  N/A
LYS 73.A NZ      GLU 72.A OE2   no hydrogen  3.194  N/A
LYS 265.A NZ     GLU 262.A OE2  no hydrogen  2.838  N/A
LYS 348.A NZ     ASP 459.A OD2  no hydrogen  2.765  N/A
GLN 540.A NE2    SER 536.A OG   no hydrogen  2.959  N/A
```

8.5.8.3 MalL wildtype

```
ARG 114.B NE   ASP 152.B OD1  no hydrogen  3.527  N/A
ARG 114.B NH2  ASP 152.B OD1  no hydrogen  3.169  N/A
TYR 124.B OH   GLU 180.B OE1  no hydrogen  2.697  N/A
TYR 158.B OH   GLY 134.B O    no hydrogen  2.806  N/A
ARG 418.B NH2  ASP 332.B OD1  no hydrogen  2.982  N/A
ASN 452.B ND2  GLN 449.B O    no hydrogen  3.380  N/A
TRP 531.B N    ALA 543.B O    no hydrogen  2.791  N/A
```

8.5.8.4 MalL wildtype shortened

```
ARG 313.B NH1  GLU 282.B OE1  no hydrogen  2.588  N/A
VAL 495.B N    ASP 492.B OD2  no hydrogen  2.654  N/A
```

8.6 Stabilising hydrogen bonds

Hydrogen bonds that are found in stabilised structures [MalL D492R (chain A and chain B), MalL T150R and MalL S536R], that are absent in MalL wildtype (chain A and chain B). Hydrogen bonds are represented as donor atom, acceptor atom, hydrogen, donor-acceptor distance, hydrogen-acceptor distance.

8.6.1 All four stabilised structures

```
LYS 402 NZ     GLU 400 O      no hydrogen  3.064  N/A
GLY 75 N       TYR 71 O       no hydrogen  3.355  N/A
LYS 391 N      ASP 387 O      no hydrogen  3.156  N/A
ARG 186 NH2    GLU 243 O      no hydrogen  2.985  N/A
TYR 125 N      ARG 122 O      no hydrogen  3.431  N/A
```

8.6.2 Three stabilised structures

```
HIS 247 NE2    ARG 242 O      no hydrogen  2.996  N/A
LYS 475 NZ     ASP 45 OD1     no hydrogen  3.227  N/A
ARG 422 NH2    ASP 332 OD1    no hydrogen  3.186  N/A
ASP 118 N      SER 116 OG     no hydrogen  2.935  N/A
GLN 82 NE2     GLU 86 OE2     no hydrogen  3.031  N/A
```

8.6.3 Shortened in all stabilised structures

```
ASN 294 N      SER 292 OG     no hydrogen  2.907  N/A
```

8.7 Hydrogen bonds from the asymmetric unit

Hydrogen bonds to adjacent monomers from the asymmetric unit. Hydrogen bonds are represented as donor atom, acceptor atom, hydrogen, donor-acceptor distance, hydrogen-acceptor distance.

8.7.1 MalL wildtype

LYS 35.A NZ	GLU 541.B OE2	no hydrogen	2.899	N/A
LYS 68.A NZ	GLU 515.A OE1	no hydrogen	3.235	N/A
LYS 68.A NZ	GLU 515.A OE2	no hydrogen	3.058	N/A
LYS 68.B NZ	LYS 402.B O	no hydrogen	3.244	N/A
TYR 121.B OH	GLU 408.B OE2	no hydrogen	2.795	N/A
ARG 191.B NH2	GLU 400.B O	no hydrogen	3.121	N/A
LYS 402.B NZ	ASP 183.B O	no hydrogen	3.508	N/A
ARG 542.B NH1	GLU 458.A OE2	no hydrogen	2.840	N/A
SER 549.B N	GLU 458.A OE2	no hydrogen	3.399	N/A
LYS 551.B NZ	LEU 457.A O	no hydrogen	2.803	N/A

8.7.2 MalL D492R

LYS 40.A NZ	PRO 132.B O	no hydrogen	2.891	N/A
LYS 68.A NZ.A	ASP 379.B OD2	no hydrogen	2.683	N/A
LYS 117.A NZ	GLU 521.A OE1	no hydrogen	2.791	N/A
TYR 121.A OH	ASP 450.B OD1	no hydrogen	2.429	N/A
ARG 179.A NE	GLN 449.B OE1.B	no hydrogen	2.763	N/A
ARG 179.A NH2	GLN 449.B OE1.B	no hydrogen	3.176	N/A
ARG 191.A NH2	ASP 379.B O	no hydrogen	2.686	N/A
LYS 300.A NZ	GLU 175.A OE1	no hydrogen	3.469	N/A
LYS 300.A NZ	GLU 175.A OE2	no hydrogen	2.829	N/A
LYS 345.A NZ	ASP 54.B OD2	no hydrogen	3.180	N/A
TYR 394.A OH.B	PRO 524.A O	no hydrogen	2.541	N/A
TYR 448.A OH	ASP 183.B OD2.A	no hydrogen	3.268	N/A
TYR 448.A OH	ASP 183.B OD2.B	no hydrogen	2.687	N/A
GLN 449.A NE2	ASP 183.B OD2.A	no hydrogen	3.415	N/A
GLN 449.A NE2	ASP 183.B OD2.B	no hydrogen	3.025	N/A
GLN 476.A NE2	HIS 217.B NE2	no hydrogen	3.001	N/A
TYR 477.A OH	HIS 217.B NE2	no hydrogen	3.240	N/A
ASN 491.A ND2	ASP 118.A O	no hydrogen	2.892	N/A
ARG 492.A NE	ASP 118.A OD1	no hydrogen	2.843	N/A
ARG 492.A NH2	ASP 118.A OD2	no hydrogen	2.944	N/A
HIS 528.A NE2	GLU 406.A OE2	no hydrogen	2.855	N/A
LYS 532.A NZ	THR 215.B O	no hydrogen	3.058	N/A
ARG 542.A NH1	ASP 249.B OD2	no hydrogen	2.805	N/A
LYS 546.A NZ	GLU 383.A OE1	no hydrogen	2.665	N/A
LYS 40.B NZ	PRO 132.A O	no hydrogen	2.985	N/A

LYS 117.B NZ	GLU 521.B OE2	no hydrogen	3.009	N/A
ARG 191.B NE	ASP 379.A O	no hydrogen	3.050	N/A
ARG 191.B NH2	ASP 379.A OD1	no hydrogen	2.920	N/A
HIS 217.B NE2	TYR 477.A OH	no hydrogen	3.240	N/A
ARG 242.B NH2	GLU 289.B OE2	no hydrogen	3.178	N/A
LYS 300.B NZ	GLU 175.B OE1.A	no hydrogen	3.044	N/A
LYS 300.B NZ	GLU 175.B OE1.B	no hydrogen	2.835	N/A
LYS 300.B NZ	GLU 175.B OE2.A	no hydrogen	3.053	N/A
LYS 300.B NZ	GLU 175.B OE2.B	no hydrogen	3.493	N/A
TYR 394.B OH.B	PRO 524.B O	no hydrogen	2.588	N/A
GLN 449.B NE2.A	ALA 176.A O	no hydrogen	3.194	N/A
GLN 449.B NE2.B	ALA 176.A O	no hydrogen	3.190	N/A
GLN 449.B NE2.B	GLU 180.A OE2	no hydrogen	2.830	N/A
ASN 491.B ND2	ASP 118.B O	no hydrogen	2.894	N/A
ARG 492.B NE	ASP 118.B OD1	no hydrogen	2.815	N/A
ARG 492.B NH2	ASP 118.B OD2	no hydrogen	2.914	N/A
HIS 528.B NE2	GLU 406.B OE2	no hydrogen	2.850	N/A
LYS 532.B NZ	THR 215.A O	no hydrogen	2.473	N/A
LYS 546.B NZ	GLU 383.B OE1	no hydrogen	2.829	N/A

8.7.3 MalL T150R

LYS 68.A NZ	ASP 379.A OD2	no hydrogen	2.860	N/A
TRP 109.A NE1	ASP 450.A OD2	no hydrogen	3.194	N/A
TYR 121.A OH	ASP 450.A OD1	no hydrogen	2.555	N/A
ARG 179.A NE	GLN 449.A OE1	no hydrogen	2.832	N/A
ARG 179.A NH2	GLN 449.A OE1	no hydrogen	3.200	N/A
ARG 191.A NH1	ASP 379.A O	no hydrogen	3.525	N/A
ARG 191.A NH2	ASP 379.A O	no hydrogen	2.732	N/A
LYS 300.A NZ	GLU 175.A OE1	no hydrogen	3.276	N/A
LYS 300.A NZ	GLU 175.A OE2	no hydrogen	2.698	N/A
TYR 394.A OH	PRO 524.A O	no hydrogen	2.894	N/A
LYS 407.A NZ	GLU 78.A O	no hydrogen	3.041	N/A
LYS 411.A NZ	GLU 78.A OE2	no hydrogen	3.142	N/A
GLN 449.A NE2	ALA 176.A O	no hydrogen	3.132	N/A
GLN 449.A NE2	GLU 180.A OE2	no hydrogen	2.878	N/A
LYS 454.A NZ	GLU 490.A OE2	no hydrogen	3.274	N/A
ARG 503.A NH2	GLU 106.A OE2	no hydrogen	3.041	N/A
HIS 528.A NE2	GLU 406.A OE1	no hydrogen	2.565	N/A
ARG 542.A NE	SER 246.A O	no hydrogen	3.150	N/A

8.8 Publications

Cooperative conformational transitions and the temperature dependence of enzyme catalysis

Emma J. Walker[1], Carlin J. Hamill*[1], Rory Crean[2], Michael S. Connolly[2], Annmaree K. Warrender[1], Kirsty L. Kraakman[1], Erica J. Prentice[1], Alistair Steyn-Ross[3], Moira Steyn-Ross[3], Christopher R. Pudney[5], Marc W. van der Kamp[4], Louis A. Schipper[1], Adrian J. Mulholland[2‡], Vickery L. Arcus[1‡].*

[1]Te Aka Mātuatua School of Science, University of Waikato, Hamilton, New Zealand.
[2]Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol BS8 1TS, United Kingdom
[3]School of Engineering, University of Waikato, Hamilton New Zealand.
[4]School of Biochemistry, University of Bristol, University Walk, Bristol BS8 1TD, United Kingdom.
[5]Department of Biology and Biochemistry, Centre for Biosensors, Bioelectronics and Biodevices, University of Bath, Bath, UK.
‡Corresponding authors: vic.arcus@waikato.ac.nz, Adrian.Mulholland@bristol.ac.uk
*These authors contributed equally to the work.

Key Words: *Enzyme catalysis, Activation heat capacity, Enzyme kinetics, Crystallography, Molecular dynamics, Macromolecular rate theory.*

Abstract

Many enzymes display non-Arrhenius behaviour with curved Arrhenius plots in the absence of denaturation. There has been significant debate about the origin of this behaviour and recently the role of the activation heat capacity ($\Delta C_P^\ddagger$) has been widely discussed. If enzyme-catalysed reactions occur with appreciable negative values of $\Delta C_P^\ddagger$ (arising from narrowing of the conformational space along the reaction coordinate), then curved Arrhenius plots are a consequence. To investigate these phenomena in detail, we have collected high precision temperature-rate data over a wide temperature interval for a model glycosidase enzyme MalL, and a series of mutants that change the temperature-dependence of the enzyme-catalysed rate. We use these data to test a range of models including macromolecular rate theory (MMRT) and an equilibrium model. In addition, we have performed extensive molecular dynamics (MD) simulations to characterise the conformational landscape traversed by MalL in the enzyme-substrate complex and an enzyme-transition state complex. We have crystallised the enzyme in a transition state-like conformation in the absence of a ligand and determined an X-ray crystal structure at very high resolution (1.10 Å). We show (using simulation) that this enzyme-transition state conformation has a more restricted conformational landscape than the wildtype enzyme. We coin the term "transition state-like conformation (TLC)" to apply to this state of the enzyme. Together, these results imply a cooperative conformational transition between an enzyme-substrate conformation (ES) and a transition-state-like conformation (TLC) that precedes the chemical step. We present a two-state model as an extension of MMRT (MMRT-2S) that describes the data along with a convenient approximation with linear temperature dependence of the activation heat capacity (MMRT-1L) that can be used where fewer data points are available. Our model rationalises disparate behaviour seen for MalL and a thermophilic alcohol dehydrogenase and is consistent with a raft of data for other enzymes. Our model can be used to characterise the conformational changes required for enzyme catalysis and provides insights into the role of cooperative conformational changes in transition state stabilisation that are accompanied by changes in heat capacity for the system along the reaction coordinate. TLCs are likely to be of wide importance in understanding the temperature dependence of enzyme activity, and other aspects of enzyme catalysis.

Introduction

Scientific discourse on the temperature dependence of enzyme catalysis has a long history. In the last decade, debate has been a reignited regarding the origins of non-Arrhenius behaviour seen for many enzymes.[1–5] We have argued previously that the changes in the conformational landscape along the reaction coordinate lead to changes in heat capacity that explain curved Arrhenius plots and we have given this scheme the name "macromolecular rate theory (MMRT)".[6,7] Others have used similar arguments of conformational complexity to explain deviations from Arrhenius behaviour and we have suggested that these proposals are complementary.[5,6,8]

Conformational changes can be envisaged as 'many-state' or 'two-state', where the latter often implies cooperativity for macromolecules. Cooperative phenomena are ubiquitous in biology and are a feature of protein folding, ligand binding, enzyme catalysis and allosteric regulation.[9] Two-state protein folding is an archetype for cooperativity, with an equilibrium between a unfolded state and a folded state involving large numbers of intramolecular interactions. The observed folding kinetics for two-state protein folding are the sum of the forward and reverse rate constants, and the steady state equilibrium is the quotient of these two rate constants at a given temperature.[9] The cooperativity of the transition leads to a change in heat capacity between the two states and a curved temperature-dependence of the equilibrium free energy (ΔG) giving rise to denaturation at both low and high temperatures.[10,11] The position of the transition state for folding with respect to the folded and unfolded states also dictates the temperature-dependence of the activation free energy ($\Delta G^{\ddagger}$) and hence the temperature-dependence of the folding and unfolding kinetics.[12] For example, for barnase, the temperature dependence of the folding kinetics deviates significantly from Arrhenius kinetics due to the transition state for folding closely resembling the folded state, giving rise to a significant value of the activation heat capacity ($\Delta C_P^{\ddagger}$) for folding ($\Delta C_P^{\ddagger}$= -1.25 kJ.mol^{-1}.K^{-1}).[12] In contrast, the unfolding kinetics are close to Arrhenius in behaviour for the same reason – the transition state closely resembles the folded state and hence a negligible value of $\Delta C_P^{\ddagger}$ upon unfolding. Importantly, the temperature dependence of the folding kinetics for barnase gives rise to positive activation enthalpies at low temperatures and negative activation enthalpies at high temperatures due to the steep temperature dependence of $\Delta H^{\ddagger}$ (a consequence of $\Delta C_P^{\ddagger} \ll 0$). This is a hallmark of the kinetics of cooperative conformational changes involving changes in heat capacity along the reaction coordinate.

Large negative values for ΔC_P are seen for the binding of transition-state analogues to the enzyme 5'-methylthioadenosine phosphorylase (MTAP).[13] This results in positive enthalpies for binding at low temperatures and negative enthalpies for binding at high temperatures due to the steep temperature dependence of ΔH for binding. The impressively small equilibrium binding constants for this system ($K_d < 1$ nM) highlight the very essence of enzyme catalysis, namely, stabilisation of the transition state species by the enzyme. The cooperative nature of this tight binding is manifested in the large, negative value for ΔC_P for this interaction (~ $-2.3\ kJ.mol^{-1}.K^{-1}$). This equilibrium ΔC_P is reflected in the temperature dependence of the rate with a commensurate value of the activation heat capacity for MTAP with 2-amino-5'-methylthioadenosine as substrate ($\Delta C_P^{\ddagger} = -2.3\ kJ.mol^{-1}.K^{-1}$).[13]

The importance of activation heat capacity for enzyme kinetics has been the subject of debate recently.[3,5,6,14] The origin of this phenomenon lies in differences in fluctuations and populations of conformations between the enzyme-substrate complex (ES) and the enzyme-transition state complex (E-TS). We have argued that a consensus is emerging whereby the enzyme-substrate complex fluctuates between at least two conformations, one that favours the substrate and one that favours the transition state and that there is a difference in heat capacity between these two conformations.[6] Arguments for an equilibrium between two states in the context of enzyme kinetics date back at least 70 years[1] and also underpins the MWC and KNF models for allostery.[15] Cooper and Dryden proposed (37 years ago) increased fluctuations about a mean conformation as an equivalent mechanism underlying allostery[16] and indeed, an ensemble view of allostery has recently gained attention[17] and acceptance.[18] More recently, it has been shown that apparently puzzling temperature dependence of kinetic isotope effects in enzyme-catalysed reaction can be accounted for by a transition state theory model including two states with different reactivity.[19] It can be argued that all these mechanisms involving multiple conformational states, coupled with changes to the number of states along the pathway, can give rise to a change in heat capacity if they involve changes in conformational fluctuations along the pathway.

Many investigators have invoked multiple conformations in the ES complex to explain deviations from Arrhenius kinetics for enzyme catalysed reactions. Truhlar and Kohen postulated an equilibrium between reactive and non-reactive states as an underlying mechanism giving rise to these deviations and noted similar behaviour for non-biochemical systems.[4] This had been previously suggested by Massey and colleagues in 1966 for an amino acid oxidase.[20] Daniel and Danson applied an "equilibrium" model (involving active and inactive states, ES and ES') to explain deviations from Arrhenius kinetics for a wide range of enzymes.[21] Mulholland and colleagues have also invoked two reactive states in enzyme catalysis and use transition state theory to rationalise deviations from Arrhenius kinetics.[19] Åqvist and colleagues have made an effort to differentiate between an "equilibrium" model (with active and inactive states) and an activation heat capacity model (MMRT) proposed by us (also invoking multiple conformations).[5] Although Åqvist presents these two models in opposition, we have argued that they are actually consistent, and effectively two sides of the same coin, whereby a constriction of the conformational space along the reaction coordinate (i.e. a two-conformation to one-conformation transition from ES to E-TS) can give rise to an appreciable activation heat capacity.[6] More recently, Åqvist has used molecular simulations to argue that there is no evidence for $\Delta C_P^{\ddagger}$ in enzyme-catalysed reactions.[22] In contrast, Warshel and colleagues have used extensive MD/EVB simulations to show temperature-dependent activation enthalpies and entropies for alcohol dehydrogenase (ADH) and thus, by definition, an activation heat capacity for this enzyme.[8] They argue that pre-organisation is generally important for enzyme catalysis.[8]

Here, we analyse these questions in detail through a combination of structural and biochemical experiments, molecular simulations and kinetic modelling on a well characterized enzyme, which has been the focus of debates and proposals in this area. We present high-resolution temperature-rate data for the model glycosidase enzyme MalL (over a wide temperature range) and for a range of mutants that alter the temperature-dependence of the enzyme catalysed reaction rate. We combine these kinetic data with detailed molecular dynamics simulations, and high-resolution X-ray crystal structures where different conformations are trapped. Our MD simulations show that the ES complex accesses a range of conformations whereas the E-TS complex is much more constrained. We call this conformation the transition-state-like conformation (TLC). This is a conformation that favours the chemical transition state and is visited by the enzyme when the substrate is bound and is the conformational bottleneck in phase space before the chemical step. For MalL, this is a cooperative transition that involves the shortening of more than 28 hydrogen bonds, an increase in correlated motions, and thus a significant increase in order immediately prior to the chemical step. This order-disorder transition (the ES-TLC transition) is temperature dependent, and we argue that this is the origin of the activation heat capacity.

We present and test alternative models to describe the temperature dependence of reaction for MalL including: a two-state model that is an extension of MMRT (MMRT-2S); and a convenient function with linear temperature-dependence for $\Delta C_P^{\ddagger}$ that serves as a good approximation to describe the kinetic data (MMRT-1L). We also apply this model to the intriguing case of alcohol dehydrogenase, where we show that the temperature dependent transition between ES and TLC is reversed (compared to MalL) due to the ES complex being more ordered than the TLC complex leading to the opposite slope for $\Delta C_P^{\ddagger}$. We show how our analysis for the temperature dependence of $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ for alcohol dehydrogenase is in agreement with the results of Warshel and colleagues, emphasising the emerging consensus.[8]

Results

Temperature-dependent enzyme kinetics and analysis

We collected high-quality enzyme kinetics data over a wide temperature range (6-56 °C), giving a dataset that allows a thorough analysis of the temperature dependence of MalL and its mutants. We extensively optimised our stopped-flow experimental protocols to ensure highly reproducible data and precise temperature control. A feature of the optimised protocol is the use of several dummy shots before each data collection run to ensure a very stable temperature throughout the stopped flow lines. The temperature-dependence data for WT MalL (in the absence of denaturation), collected at pH 7.0, are shown in Figure 1A&C. We have plotted all three replicates: note how tightly constrained the errors are for each point. We use transition state theory (equation 1) with a transmission coefficient (γ) of 1 to convert temperature-rate data (Figure 1C) into the activation free energy at a given temperature (Figure 1A). The temperature-dependence of $\Delta G^{\ddagger}$ shows two regimes, with a clear transition at ~313 K. There are sufficient data to initially treat these regimes separately. Each shows curvature (non-Arrhenius behaviour) and therefore MMRT fitting is appropriate (using equation 2).[7] This gives $\Delta C_P^{\ddagger}$ values of ~0.8 (i.e. near zero) and ~$^-$20.8 kJ.mol^{-1}.K^{-1} for the low and high temperature regimes respectively (blue and red curves Figure 1A). This immediately suggests that $\Delta C_P^{\ddagger}$ is temperature dependent and that there is a two-state cooperative transition at ~313 K (Figure 1B).

$$k = \frac{\gamma k_B T}{h} e^{\frac{-\Delta G^{\ddagger}}{RT}} \qquad (1)$$

$$\Delta G^{\ddagger} = \Delta H_{T_0}^{\ddagger} - T\Delta S_{T_0}^{\ddagger} + \Delta C_P^{\ddagger}\big(T - T_0 - Tln(T/T_0)\big) \qquad (2)$$

To establish a model to describe these data, we begin with two conformations: the enzyme-substrate conformational ensemble (ES) and a transition-state-like conformation (TLC) that favours the chemical transition state and the chemical reaction. In this context we differentiate between TLC and E-TS to highlight the fact that the TLC state is a conformational state visited from the ES complex and these conformations (ES and TLC) are on pathway. For the E-TS state in MalL in crystallography and in MD simulations, we use the complex of the enzyme with a transition-state analogue.

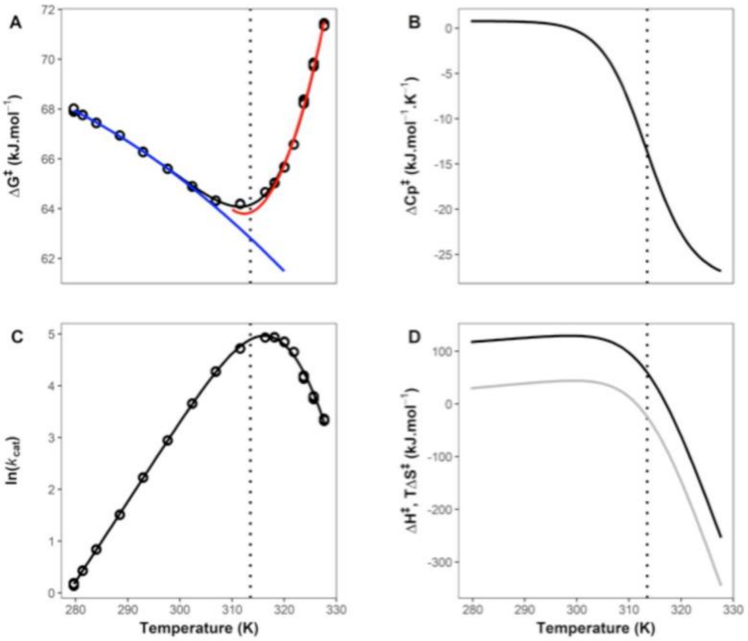


Figure 1. High-resolution temperature dependence of activity for MalL. **A.** $\Delta G^{\ddagger}$ versus temperature. All data are plotted as circles (triplicates at each temperature). The blue curve is a plot using MMRT[7] (equation 2) for the first six temperature points; the red curve is an MMRT fit for the last six temperature points. The black curve in **A** uses the two-state model proposed here (MMRT-2S), which shows a smooth transition between the blue and red curves (equations 4 & 5). **B.** $\Delta C_P^{\ddagger}$ versus T illustrating the two-state model with constant $\Delta C_P^{\ddagger}$ values at low and high temperatures and a cooperative transition between the two. **C.** ln k_{cat} versus T. Experimental data points in triplicate are plotted as circles. The curve is fitted to the two-state model (equations 4 & 5). **D.** Derived values for $\Delta H^{\ddagger}$(black) and $T\Delta S^{\ddagger}$ (grey) versus T. The transition between low and high temperature regimes is clear. In all panels, the vertical dotted line shows the transition temperature, T_C (313.5 K, determined from fitting equations 4 & 5).

A minimal model is:

$$E + S \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} ES \underset{k_{con\_r}}{\overset{k_{con\_f}}{\rightleftharpoons}} TLC \overset{k_{chem}}{\longrightarrow} E + P \qquad \text{(Scheme 1)}$$

The substrate binding equilibrium is followed by a conformational equilibrium between ES and TLC and the chemical step proceeds from the TLC. The rate constants in Scheme 1 are for forward and reverse binding of the substrate (k_{on}, k_{off}), the forward and reverse conformational transitions between ES and TLC (k_{con_f}, k_{con_r}) and the rate constant for the chemical step (k_{chem}). Formally, the rate equation for this system is:

$$k_{obs} = \frac{k_{con\_f} k_{chem}}{k_{con\_f} + k_{con\_r} + k_{chem}} \qquad (3)$$

However, this is unwieldy as each rate constant potentially has three variables ($\Delta H^{\ddagger}_{T_0}$, $\Delta S^{\ddagger}_{T_0}$ and $\Delta C^{\ddagger}_{P}$, equations 1 & 2) giving an expression with potentially nine variables. We therefore take a simplified approach.

The two arms of the rate and activation free energy data suggest a two-state transition between low temperatures and high temperatures with a transition at ~313 K. The value of $\Delta C^{\ddagger}_{P}$ at low temperatures is near zero, while the value at high temperatures is large and negative. We can construct a suitable model for such a two-state transition by treating the $\Delta C^{\ddagger}_{P}$ as a function of temperature thus:

$$\Delta C^{\ddagger}_{P}(T) = \frac{\left(\Delta C^{\ddagger}_{P,lowT}\right) + \left(\Delta C^{\ddagger}_{P,highT}\right) e^{\frac{-\Delta\Delta H^{\ddagger}(1 - T/T_C)}{RT}}}{1 + e^{\frac{-\Delta\Delta H^{\ddagger}(1 - T/T_C)}{RT}}} \qquad (4)$$

…where T_C is the temperature at the midpoint of the transition, $\Delta C^{\ddagger}_{P,lowT}$ is the value of $\Delta C^{\ddagger}_{P}$ for the low temperature arm, $\Delta C^{\ddagger}_{P,highT}$ is the value of $\Delta C^{\ddagger}_{P}$ for the high temperature arm and $\Delta\Delta H^{\ddagger}$ is the difference in $\Delta H^{\ddagger}$ between the two arms (at T_C). Having defined $\Delta C^{\ddagger}_{P}$ as a function of temperature, the activation enthalpy, entropy and free energy terms are given by standard expressions:

$$\Delta H^{\ddagger} = \Delta H^{\ddagger}_{T_0} + \int \Delta C^{\ddagger}_P dT, \qquad \Delta S^{\ddagger} = \Delta S^{\ddagger}_{T_0} + \int \frac{\Delta C^{\ddagger}_P}{T} dT,$$

$$\Delta G^{\ddagger} = \Delta H^{\ddagger}_{T_0} - T\Delta S^{\ddagger}_{T_0} + \int \Delta C^{\ddagger}_P dT \; - T \int \frac{\Delta C^{\ddagger}_P}{T} dT \tag{5}$$

The rate coefficient at a particular temperature is given by equation 1. This provides a rate equation with five terms, $\Delta H^{\ddagger}_{T_0}$, $\Delta S^{\ddagger}_{T_0}$, T_C, $\Delta\Delta H^{\ddagger}$ and $\Delta C^{\ddagger}_{P,highT}$ with $\Delta C^{\ddagger}_{P,lowT}$ having been found first from an independent fit of the low temperature arm of the data (blue curve in figure 1A).

We have collected very accurate kinetic data for WT MalL and for the mutant series V200T, V200S and V200A and fitted these data using our proposed two-state model, MMRT-2S (Figure 2 and Table 1). The data are very well described by this model, and in all cases show a transition between $\Delta C^{\ddagger}_P$ near zero for low temperatures and large negative values for $\Delta C^{\ddagger}_P$ at high temperatures (Figures 1B&2B). The values for $\Delta C^{\ddagger}_P$ at high temperatures for the mutant series get less negative along the series (WT-V200A-V200T-V200S) which is consistent with these mutations increasing the rigidity of the ES complex as we have described previously[3,7] (Figure 2B). The value found for $\Delta C^{\ddagger}_P$ at T_C is also consistent with previous results measured over a narrow temperature range around 310 K.[3] The significant increase in rate for V200A comes at the cost of K_M for this mutation: the K_M value is more than 7-times that of the WT (K_M = 1.56 mM, c.f. 0.21 mM for WT).[3] Mutation from valine at this position in all cases significantly changes the temperature dependence of the rate and shifts T_{opt} up by ~6 degrees. It is interesting to note that the WT enzyme, which has a more flexibile ES state, has a slightly higher rate at intermediate temperatures (c.f. V200T and V200S): this is consistent with previous hypotheses regarding the evolution to psychrophily (cold adapted activity) whereby higher flexibility is observed for psychrophilic enzymes than their mesophilic counterparts.[7,23,24] This increased flexibility of the ES comes at the cost of a lower T_C and a larger, more negative $\Delta C^{\ddagger}_P$ consistent with the 'psychrophilic trap' that we have previously described.[7] The steepness of the transition from low temperature to high temperature is determined by the difference in $\Delta H^{\ddagger}$ between the two processes ($\Delta\Delta H^{\ddagger}$) (although this parameter is not particularly well determined by the data: relatively large standard errors of fitting, see Table 1). As expected $\Delta\Delta H^{\ddagger}$ is significantly lower for V200S than for the WT: the V200S ES complex is more rigid.

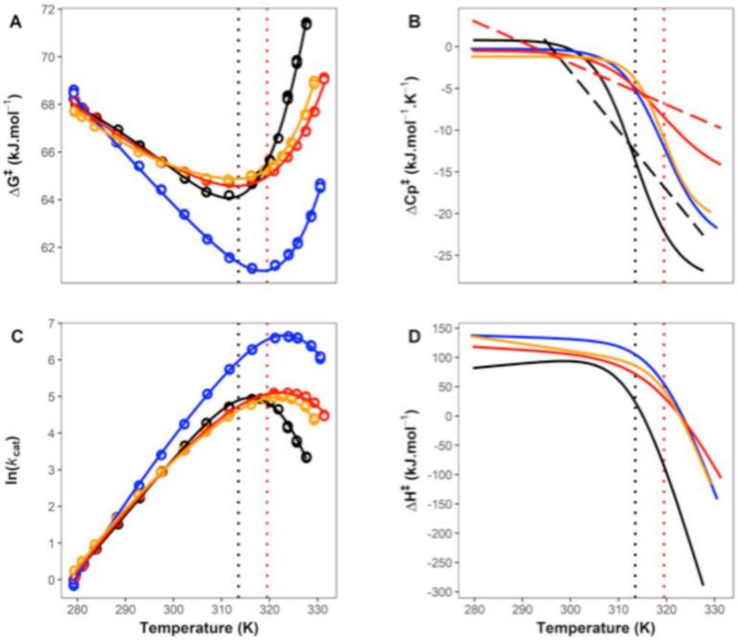


Figure 2. Temperature-dependent enzyme kinetic measurements for MalL WT (black), V200T (orange), V200S (red) and V200A (blue) mutations. **A.** $\Delta G^{\ddagger}$ versus T. **B.** $\Delta C_P^{\ddagger}$ versus *T*. **C.** ln(k_{cat}) versus T. **D.** $\Delta H^{\ddagger}$ versus T. All smooth curves are fits of MMRT-2S (equations 4 & 5) to the data and in the case of $\Delta C_P^{\ddagger}$ and $\Delta H^{\ddagger}$, are derived from these fits (see Table 1). Vertical dotted lines are values for T_C for WT (black) and the V200T, V200S and V200A mutants (red; all three have the same T_C values). Linear dashed lines (panel B) are linear $\Delta C_P^{\ddagger}$ values from fitting the data using MMRT-1L (equations 6 & 1) for WT and V200S.

Table 1. Activation parameters from the fitting to the two-state model (MMRT-2S).

	$\Delta C^{\ddagger}_{P,lowT}$ kJ.mol^{-1}.K^{-1}	$\Delta C^{\ddagger}_{P,highT}$ kJ.mol^{-1}.K^{-1}	$\Delta H^{\ddagger}_{T_0}$ ^{3} kJ.mol^{-1}	$\Delta S^{\ddagger}_{T_0}$ ^{3} J.mol^{-1}.K^{-1}	$\Delta\Delta H^{\ddagger}$ kJ.mol^{-1}	T_C K
WT	0.8 (0.2)	–28.1 (6.3)	96.6 (2.3)	102.3 (8.2)	186 (46)	313.5 (2.7)
V200A	–0.3 (0.9)^1	–23.8 (8.4)	132.7 (1.3)	230.3 (4.5)	185 (48)	319.5 (4.0)
V200T	–1.2 (0.1)	–20.4 (0.8)	113.6 (1.5)	164.2 (2.5)	258 (24)	319.5^2
V200S	–0.8 (0.3)^1	–16.3 (0.9)	112.9 (1.5)	160.5 (5.2)	154 (16)	319.5^2
WT^4	0.09 (1)^1	–9.1 (10)	105.4 (8)	134 (28)	282^5 (27) 888^5 (80)	

[1] Not significantly different from zero. [2] Fixed during fitting. [3] T_0 = 278.15 K. [4] T_{eq} model.
[5] ΔH_{eq}, ΔS_{eq}

Figure 4D clearly illustrates the pitfalls of discussions based on activation enthalpies and entropies for enzyme-catalysed reactions: activation enthalpy curves for WT MalL and the V200 variants cross, with the order at low temperatures being reversed at moderate temperatures and changing significantly again at high temperatures. Discussions of enzyme activity, and e.g. of the evolution of enzyme catalysis, must consider this temperature dependence; arguments based on enthalpy or entropy are not independent of temperature. Notwithstanding this, large positive values of $\Delta H^{\ddagger}$ at low temperatures are consistent with a bond-breaking chemical step and large negative values of $\Delta H^{\ddagger}$ at high temperatures suggest that a cooperative conformational step contributes to the observed rate with formation of hydrogen bonds along the reaction coordinate. In chemical systems, negative activation enthalpies have been attributed to the formation of multiple hydrogen bonds *en route* to the transition state.[25]

A convenient approximation, with linear temperature dependence of $\Delta C_P^{\ddagger}$ *(MMRT-1L).*

The two-state model presented above requires fitting of five or six parameters; at least two of the parameters are tightly correlated giving rise to practical issues with the fitting procedure (e.g. parameters not converging). For example, in the case of V200T and V200S, the value of T_C must be fixed to prevent this from happening. Indeed, most enzyme rate-temperature datasets contain fewer data points and thus cannot be reasonably fitted using a five-parameter model. Hence, a simpler model that captures the trends in the data is warranted. We propose a linear temperature-dependence of $\Delta C_P^\ddagger$ as a useful approximation to MMRT-2S, because the slope of $\Delta C_P^\ddagger$ captures the dynamics of the transition. For example, the slope of $\Delta C_P^\ddagger$ for such a linear model would encompass both the size of $\Delta\Delta C_P^\ddagger$ across the temperature range and the steepness of the transition as reflected in the size of $\Delta\Delta H^\ddagger$. For a such a linear model, we take $\Delta C_P^\ddagger = \Delta C_{P,0}^\ddagger + mT$, where $\Delta C_{P,0}^\ddagger$ is the y-axis intercept at 0 K and *m* is the slope of $\Delta C_P^\ddagger$. We derive a function for $\Delta G^\ddagger$ by integration in the usual way (equation 5). This results in an additional T^2 term in the activation free energy (equation 6):

$$\Delta G^\ddagger = \Delta H_{T_0}^\ddagger - T\Delta S_{T_0}^\ddagger + \Delta C_P^\ddagger\big(T - T_0 - Tln(T/T_0)\big) - \frac{m}{2}(T - T_0)^2 \tag{6}$$

The enthalpy and entropy are given by:

$$\Delta H^\ddagger(T) = \Delta H_{T_0}^\ddagger + \Delta C_P^\ddagger(T - T_0) + m/2(T^2 - T_0^2) \tag{7}$$

$$\Delta S^\ddagger(T) = \Delta S_{T_0}^\ddagger + \Delta C_P^\ddagger ln(T/T_0) + m(T - T_0) \tag{8}$$

In the context of the rate equation, the maximum temperature for reaction (T_{opt}) and the inflection points (T_{inf})[26] are:

$T_{opt}\ \ when\ \Delta H^\ddagger = -RT$

$T_{inf}\ \ when\ (\Delta H^\ddagger)^2 = -\Delta C_P^\ddagger RT^2$ i.e. the criteria for the variance where $\Delta C_P^\ddagger < 0$.

Here, we have used MMRT-1L to fit the rate data for WT and V200S MalL. The linear temperature dependence of $\Delta C_P^\ddagger$ is shown in Figure 2B (black and red dashed lines and shows slopes of –710 and –252 J.mol^{-1}.K^{-2} respectively). These lines capture the trend for the two-state transition and their slopes capture both the magnitude of the change in $\Delta C_P^\ddagger$ and the steepness of the transition ($\Delta\Delta H^\ddagger$).

As an example of the utility of MMRT-1L, we take data from the literature for alcohol dehydrogenase (ADH) which has been extensively analysed by Klinmann *et al.* and Warshel *et al.* but has relatively few temperature points (7-8 points).[8,19,27,28] Fitting equation 6 to these data for the protiated and deuterated substrates makes it immediately obvious that the temperature dependence of $\Delta C_P^\ddagger$ is opposite to that seen for MalL: for ADH, at low temperatures

$\Delta C_P^\ddagger$ is negative and at higher temperatures it approaches zero (Figure 3). This is consistent with Warshel *et al*.'s analysis and hypothesis[8] that the ES state is more ordered than (what we denote here as) the TLC state for ADH. Thus, the ES state is favoured at low temperatures, accompanied by negative values for $\Delta C_P^\ddagger$ (in contrast to MalL where the TLC state is favoured at low temperatures; $\Delta C_P^\ddagger \sim 0$). The cooperative transition between ES and TLC that we invoke is also consistent with the details of the differences in the temperature dependence of the rate between the protiated and deuterated substrates (over and above zero-point energy differences) where small changes can manifest as large differences in cooperative systems. We have previously observed differences of this magnitude for thermophilic glucose dehydrogenase with protiated and deuterated substrates.[29] Warshel and colleagues highlight this point when they suggest a "phase transition" for the ADH system.[8] We illustrate this 'phase transition' for ADH data by estimating the two-state transition for $\Delta C_P^\ddagger$ in Figure 3B (dashed lines). Here, at low temperatures, our model indicates that a cooperative conformational transition contributes to the free energy barrier and can be seen in large negative values for $\Delta C_P^\ddagger$. At high temperatures, the TLC state is favoured and the free energy barrier is dominated by the chemical step, with $\Delta C_P^\ddagger$ close to zero. Our analysis closely reproduces the analysis of Warshel and colleagues albeit via a completely different route and the two-state transition shows steep temperature-dependence of $\Delta H^\ddagger$ and $\Delta S^\ddagger$ below the T_C value of ~305 K and constant values of $\Delta H^\ddagger$ and $\Delta S^\ddagger$ above the T_C with $\Delta S^\ddagger$ close to zero at these higher temperatures (Figure 3D).

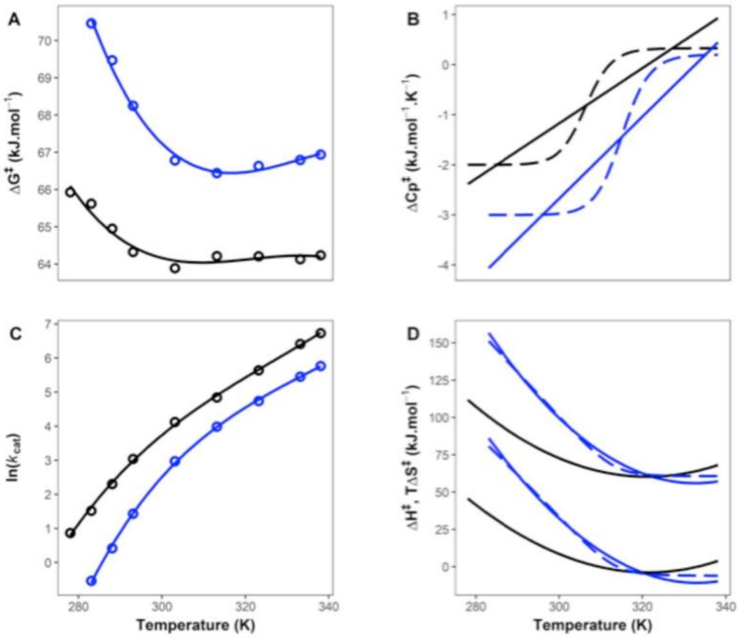


Figure 3. Temperature dependence of ADH for protiated and deuterated substrates. **A.** $\Delta G^{\ddagger}$ versus T for protiated (black) and deuterated (blue) substrates. **B.** Derived values of $\Delta C_P^{\ddagger}$ from fitting the linear model (MMRT-1L) to the data (equation 6).[28] Two-state curves are shown as dashed lines in panel B and are for illustration as there are insufficient data to fit the two-state model (MMRT-2S). **C.** ln(k_{cat}) versus T for protiated and deuterated substrates. **D.** Derived values of $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ for protiated and deuterated substrates (equations 7&8). The dashed lines are for the two-state model for the deuterated substrate by way of illustration.

Two conformations on the pathway for the enzyme-catalysed reaction

Previously, we have used extensive molecular dynamics simulations to characterise the conformational landscape in two states (ES and E-TS) for the model enzymes ketosteroid isomerase (KSI) and MalL and designer Kemp eliminase enzymes including an evolved variant.[14,30] This involved repeated independent simulations (x10) of 500 ns each to characterise the enzyme bound to either the substrate (ES complex) or a transition state analogue (E-TS complex). The difference in heat capacity between these states was calculated from the difference in fluctuations in the enthalpy for the two conformational ensembles. While calculation of heat capacities from this fluctuation approach has recently been criticised, we note that it has been applied successfully e.g. in the context of protein folding simulations[31] and, for KSI and MalL, it gives results in very good agreement with experimental observations from MMRT fitting.

Here, we have applied these simulation and analysis approaches to characterise WT MalL in detail and to compare it with mutant MalL (see below). This involved 20 independent simulations of 500 ns each for both the ES complex and the E-TS complex. The latter is simulated by binding a (physical) transition-state analogue.[14] Simulations were also performed for the mutant MalL S536R (see below). Principal component analysis was used to characterise the conformational landscape in each of the two states (Figure 4). These simulations show that the ES complex has a much broader conformational landscape than the E-TS complex and that the ES complex visits the E-TS conformation in traversing the landscape (Figure 4C&D). The E-TS complex is characterised by a constrained conformation with a significant number of additional H-bonds clustered in the centre of the structure when compared to the ES complex. The first principal component (PC1) for the ES complex is characterised by translational movement of the lid domains above the active site of the enzyme (Figure 4A&B). These movements are large and can be broadly captured by the Cα-Cα distance between His217 in domain 2 and Glu400 in domain 3 on opposite sides of the active site. This distance fluctuates between 35 Å and 15 Å and clusters around 25 Å and 17.5 Å corresponding to the two peaks in Figure 4C at -18 Å and 16 Å respectively. Thus, broadly speaking, the ES state fluctuates between "open" and "closed" conformations with domain movements of up to 20 Å. The E-TS complex corresponds to the "closed" conformation and our hypothesis is that the transition between open and closed is cooperative involving the establishment of a large number of intramolecular interactions (e.g. ~28 shortened H-bonds, see below). This emphasises the extraordinary preorganisation necessary to achieve a rate enhancement of ~10^{15} (i.e. a stabilisation of the transition state of ~110 kJ.mol^{-1}). The second principal component (PC2) is characterised by an opening and closing of the loops of domains 1 and 2 with the lid domain of domain 3 rotating as it closes (Figure 4B). The PC2 histogram shows that both states fluctuate around the same mean position and again, that the E-TS complex traverses a narrower conformational space when compared to the ES complex (see supplementary Figure 1). The two-dimensional plot of PC1 versus PC2 clearly shows the additional conformational space traversed by the ES complex when compared to the E-TS complex (Figure 4D). The origin of the heat capacity difference between ES and E-TS lies in the broad conformational fluctuations for the ES complex (primarily in lid and loop domains above the active site) in contrast to the narrowed fluctuations for the E-TS complex. Importantly, the E-TS conformation is visited by the ES conformations and formation of the E-TS complex is accompanied by the strengthening of a large number of hydrogen bonds. Indeed, these observations are commensurate with the very tight binding of the transition state chemical species by the enzyme which is important for catalysis.

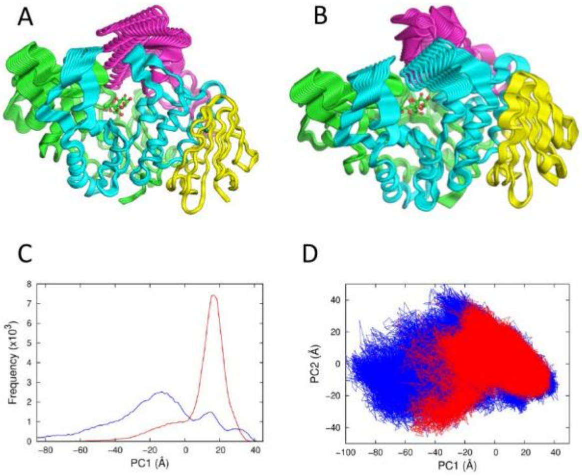


Figure 4. Conformational landscapes of WT MalL ES and E-TS complexes from MD simulations. **A.** Structure of MalL WT showing the first principal component (PC1) projections illustrating the movement of domains for the ES complex. The structure is coloured to indicate 4 regions (1-193, green; 194-321, blue; 322-459 magenta; 460-561, yellow). The substrate (isomaltose) is shown as ball-and-stick in green and red. **B.** Principal component projections illustrating the movement of domains for the ES complex for PC2. **C**. Principal component analysis showing PC1 histogram for the ES complex (blue) and E-TS complex (red). **D.** Two-dimensional plot of PC1 versus PC2 for the ES complex (blue) and the E-TS complex (red).

Very high resolution structure of apo-S536R that favours the TLC

We designed several mutants based on WT MalL X-ray structures determined in the presence of urea. We designed arginine mutations to substitute the guanidinium group of the arginine side chain for urea thereby introducing new hydrogen bond networks at the surface of the protein. Our aim was to assess the allosteric effects of these new interactions on the enzymatic rate and conformational dynamics. In the case of S536R we determined a structure at very high resolution in the absence of ligand (1.10 Å resolution, Supplementary Table 1, PDB code 7LV6). Notably, there are only very small differences between the temperature-dependent enzymatic activity of the mutant enzyme (S536R) and WT MalL (Supplementary Figure 2).

Although the overall RMSD between the WT and S536R crystal structures is small (0.185 Å over 496 Cα positions), this masks very significant differences in the H-bonding patterns for the two conformations. The structures show significant shortening (>0.3 Å) of 28 H-bonds in the S536R structure compared to the WT structure, which is consistent with increased order of the mutant structure (Figure 5, for H-bond criteria and bond lengths see Methods section). The majority of the shortened H-bonds (20 H-bonds) are found in two regions of the enzyme: region 1, 1-193 (green, Figure 5A) and region 3, 322-459 (magenta, Figure 5A). Further, an additional H-bond between the mainchain of Tyr14 and the sidechain of Gln369 connects these two regions. These two regions are contiguous in the structure and are the most mobile in the MD simulations for the WT enzyme discussed above. Their reduced mobility in the E-TS complex makes the greatest contribution to the calculated $\Delta C_P^\ddagger$ from the WT simulations.[14]

To check whether these shortened H-bonds are simply a function of the higher resolution of the S566R structure, we refined the S536R structure after truncating the diffraction data to a resolution equivalent to the WT structure (2.30 Å). We repeated the comparison and this also showed a significant shortening of 26 H-bonds at this resolution, confirming that the analysis was not biased based on the very high resolution of the S536R structure.

We carried out extensive MD simulations (20 trajectories of 500 ns each) for the S536R mutant using this high-resolution structure with substrate (ES) or transition state analogue (E-TS) bound to calculate $\Delta C_P^\ddagger$ for this mutant. The calculated values for $\Delta C_P^\ddagger$ for this mutant are close to zero for all window sizes. This implies that S536R MalL does not fully escape the energetic basin of the transition-state like conformation, either with the substrate or the transition state analogue. PC analysis indicates that conformational sampling of the substrate-complex is indeed more restricted than for WT MalL, with the main conformation being similar to that found with the transition state analogue (Figure 5B&C).

The very high resolution of the crystal structure combined with the reduced conformational dynamics over the time course of the simulations suggests that this conformation is similar to that of the TLC for MalL.

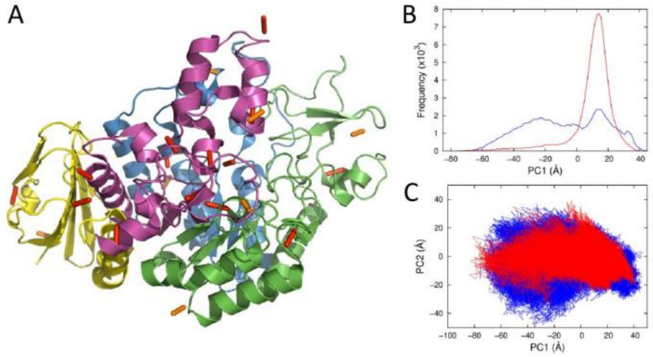


Figure 5. MalL S536R crystal structure and principal component analyses. **A.** The structure of MalL S536R determined at 1.10 Å resolution. The view is from the back (c.f. Figure 4) and regions are coloured the same as Figure 4. H-bonds that are significantly shorter (>0.3 Å) in the S536R structure compared to WT are shown as orange and red bars. The majority of these are in domains 1 (green) and 3 (magenta). **B.** Principal component analysis showing PC1 histogram for the ES complex (blue) and E-TS complex (red). **C.** Two-dimensional plot of PC1 versus PC2 for the ES complex (blue) and the E-TS complex (red). Note the similarity between the two areas when compared to Figure 4D giving a calculated value for $\Delta C_P^{\ddagger} \sim 0$.

Discussion

Interpretation of kinetics data and analysis

There are several possible interpretations to account for the two state model that we present here (MMRT-2S). There is clearly a transition between low temperature behaviour and high temperature behaviour (independent of denaturation). We propose that this transition is cooperative and two-state. The simplest explanation for the low temperature behaviour is that this constitutes the chemical step from TLC to products and that the TLC is favoured at low temperatures (see Scheme 1), as we would expect for the more ordered state (when TLC is compared to ES). At high temperatures, the activation barrier $\Delta G^{\ddagger}$ involves a large, negative $\Delta C_P^{\ddagger}$ value and steeply temperature-dependent values for $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$. The fact that $\Delta H^{\ddagger}$ proceeds from positive to negative values suggests that this step involves a cooperative conformational process consistent with the ES-to-TLC transition. Whether there is a change in rate determining step or whether the conformational transition is combined with the chemical step (c.f. the equilibrium model) is not clear. Nonetheless, the emergence of negative activation enthalpies under either the equilibrium model or a change in rate determining step implies that the conformational transition plays a significant role in the observed kinetics at these intermediate and high temperatures.

Relationship to other models

The equilibrium model postulates two conformations, one of which is inactive (E_{act} and E_{inact}).[32] This is consistent with the ES and TLC conformations postulated here. If we assume that the ES state is saturated, and that k_{chem} is rate limiting at all temperatures, then the rate equation simplifies to:

$$k_{obs} = \frac{k_{con\_f}k_{chem}}{k_{con\_f}+k_{conf\_r}} = \frac{k_{chem}}{1+K_{conf}} \qquad (10)$$

Under conditions where the conformational equilibrium, K_{conf}, is very small (in the case of MalL, at low temperatures) the rate is primarily a function of the chemical step. When the temperature increases, the rate is modulated by the equilibrium between the ES (E_{inact}) and TLC (E_{act}) states. Thus, at high temperatures the activation barrier is a combination of conformational changes and the chemical step. As we have determined that there is a significant $\Delta C_P^{\ddagger}$ for the equilibrium between ES and TLC, then the expression for $\Delta G^{\ddagger}$ for the equilibrium model becomes:

$$\Delta G^{\ddagger} = \left[\Delta H^{\ddagger}_{T_0} - T\Delta S^{\ddagger}_{T_0} + \Delta C^{\ddagger}_{P}\left(T - T_0 - T\ln\left(\frac{T}{T_0}\right)\right)\right]_{chem} + RT\ln\left(1 + e^{\frac{-\left[\Delta H^{\ddagger}_{T_0} - T\Delta S^{\ddagger}_{T_0} + \Delta C^{\ddagger}_{P}\left(T - T_0 - T\ln\left(\frac{T}{T_0}\right)\right)\right]_{conf}}{RT}}\right) \qquad (11)$$

…where the first term in square brackets refers to the chemical step (with $\Delta C^{\ddagger}_{P}{\sim}0$) and the exponential term refers to the conformational step (with $\Delta C^{\ddagger}_{P} \ll 0$). This model fits the data very well (final line in Table 1, Figure 6) although the fitting errors are large as we would expect from fitting 6 parameters. Further problems with fitting also arise from the exponential term making this model very difficult to unambiguously determine. Nonetheless, there are significant differences in the predicted temperature-dependence of the activation heat capacity when comparing our two-state model (MMRT-2S) and the equilibrium model. It is not possible to discriminate between the equilibrium model and the two-state model given the data (Figure 6A&B, the sum of squares for the residuals are similar for the two fits). If the observed changes were genuinely due to a change in the rate determining step from the chemical step to the conformational step then a suitable test would be to find mutants that showed an increase in rate at high temperatures where k_{conf_r} becomes significant and thus, show that the conformational step is rate limiting at high temperatures.

In their analyses, Åqvist and colleagues have suggested that the "correct" model in the case of α-amylase AHA and MalL is an equilibrium model with an off-pathway dead-end conformation and that the chemical step is rate limiting at all temperatures.[5] However, they present $\Delta H^{\ddagger}$ as being large and positive at low temperatures and then large and negative at high temperatures. It is difficult to imagine a chemical step where this is the case without contributions to the activation barrier coming from a conformational change. It is also hard to rationalise the presence of an off pathway conformation from an evolutionary perspective. These issues are resolved if, instead, the equilibrium is on-pathway between ES and TLC conformations and involves a cooperative conformational transition contributing to the activation barrier.

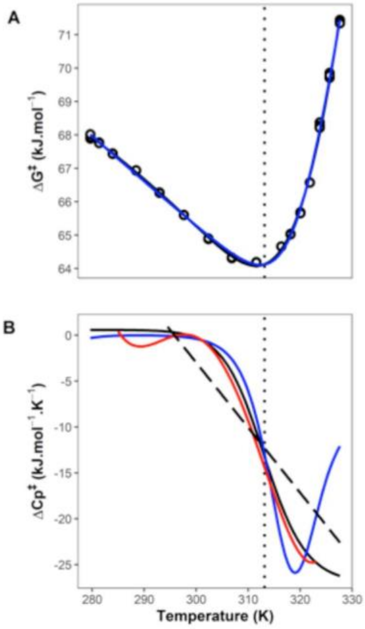


Figure 6. Comparison between two-state (MMRT-2S) and equilibrium models. **A.** $\Delta G^{\ddagger}$ versus T fitted using MMRT-2S (black) and equilibrium model (blue). **B.** $\Delta C_P^{\ddagger}$ versus T derived from the fits in A. Black is the MMRT-2S, dashed black line is MMRT-1L (i.e. linear $\Delta C_P^{\ddagger}$, equation 6), blue is the equilibrium model (equation 11) and red is derived from a polynomial fit of order 6 to $\Delta G^{\ddagger}$ and calculating $\Delta C_P^{\ddagger}$ from the second derivative.

The combination of high-resolution temperature-dependent enzyme kinetics, molecular dynamics simulations and X-ray crystal structures provide evidence for an on-pathway equilibrium between ES and TLC conformations for MalL. This makes intuitive sense insofar as the enzyme-substrate complex must visit a conformation that favours the chemical transition state species in order to facilitate catalysis. If the on-pathway equilibrium involves a cooperative transition, then we would expect a change in heat capacity for this equilibrium. A proxy for this ES-TLC equilibrium can be found in the binding of a transition state analogue to the enzyme MTAP and this binding is accompanied by a large negative value of ΔC_P as measured by isothermal titration calorimetry ($^{-}$2.4 kJ.mol^{-1}.K^{-1}).[13] In this case, the enthalpy of binding is large and positive at low temperatures (~20 kJ.mol^{-1} at 293 K) and large and negative at high temperatures (~$^{-}$40 kJ.mol^{-1} at 319 K). Depending on the position of the transition state for the equilibrium, we would also expect a change in activation heat capacity for the kinetics of this transition. In the case of MTAP, curvature in the observed temperature-dependence of the kinetics for this enzyme give very similar values for the activation heat capacity, $\Delta C_P^\ddagger$ ($^{-}$2.3 kJ.mol^{-1}.K^{-1}). The equilibrium constant for the interaction between MTAP and various transition state analogues is in the range ~10^{-9} – 10^{-10} M. By comparison, in the case of glycosidases, we are expecting an equilibrium constant for transition state binding of ~10^{-19} M based on the rate enhancement over that in water at pH 7.0.[33,34] Thus, we might expect larger absolute values of ΔC_P for the ES-TLC equilibrium and by implication, the *kinetics* of the ES-TLC transition will also occur with $\Delta C_P^\ddagger$ depending on the position of the transition state with respect to the ES and TLC conformations. This will entail a curved temperature-dependence of $\Delta G^\ddagger$. Notably, the kinetics of cooperative processes are often slow (akin to folding kinetics) and thus, may be on a similar timescale to the chemical step for catalysis.

Conceptually, the ES-to-TLC transition can be considered under three potential regimes. The first is if the barrier between ES and TLC is very low at all temperatures (compared to the chemical step). In this case the ES/TLC can be considered as one species that is simply conformationally dynamic. Here, the equilibrium model is valid (equation 10), the TLC is always accessible and the rate is simply the chemical step with expected Arrhenius-like behaviour. The second scenario is that there is a moderate barrier between ES and TLC and that this barrier is temperature-dependent. In this case the observed rate will be a combination of the chemical step and the conformational step and that this will be more marked at low or high temperatures. This is the case in our analysis of MalL, and the analysis of Warshel for ADH, giving significant $\Delta C_P^\ddagger$ values at high and low temperatures respectively. When the ES state is favoured, the free energy landscape is broad and must proceed through a TLC bottleneck to reach the chemical step giving rise to large negative values for $\Delta C_P^\ddagger$ and this could occur at low temperatures (e.g. ADH) or high temperatures (e.g. MalL) (see Figure 7). The third scenario is that the ES-TLC barrier is extremely temperature-dependent and thus is negligible at one temperature and dominates at another. This constitutes a genuine change in the rate determining step. It is difficult to distinguish between the latter two scenarios and this requires further investigation. An obvious route here would be to find mutations that change the conformational barrier which would alter the temperature-dependence of the rate in the case where there has been change in rate determining step.

Our two-state model (MMRT-2S) describes this process well. Its interpretation is consistent with the conformational change as the main contributor to $\Delta C_P^{\ddagger}$ irrespective of the microscopic details of the mechanism. We present scheme 1 as a general scheme for enzymes: the role of an enzyme is to bind the substrate relatively weakly and the transition state much more strongly.[35,36] The scheme is a minimal model. The reality of enzyme conformational behaviour is of course probably much more complicated, with many possible conformations distinguishable in various ways. Nonetheless, the ES conformation must visit a conformation that facilitates catalysis (i.e. the TLC) and this is discussed further below. This argument is consistent with previous arguments that highlight the role of a specific conformation of the enzyme which lowers the chemical reaction barrier.[37,38] Our argument is that forming this conformation involves a cooperative transition to instigate the very large apparent K_M values for the transition state (i.e. precise preorganisation). This can be seen in the extensive molecular dynamics simulations of the ES complex and the E-TS complex, for which principal component analysis shows that the ES complex visits the TLC and the TLC is much more constrained in comparison to ES (Figure 4).

MMRT-2S is rather complex to fit without a large number of data points. Much of the temperature-rate data in the literature contain relatively few points and for such typical cases, we recommend using a linear $\Delta C_P^{\ddagger}$ model as a convenient approximation (MMRT-1L). This linear model clearly reveals details of the ES-TLC equilibrium and temperatures at which the conformational component contributes significantly to the activation free energy. In the case of ADH, this occurs at low temperatures (with a positive slope for $\Delta C_P^{\ddagger}$, Figure 3) and in the case of MalL this occurs at higher temperatures (with a negative slope for $\Delta C_P^{\ddagger}$, Figure 2). Indeed, this approach may be used to glean evidence for the nature of the TLC conformation as the steeper the slope for $\Delta C_P^{\ddagger}$, the greater the cooperativity of the transition.

Our two-state model (MMRT-2S) is consistent with the numerous historical arguments that postulate two states for enzymes with and without allosteric regulation. Here, we extend these arguments to identify and characterise the two states, ES and TLC. Our arguments are consistent with those of Åqvist and colleagues (i.e. the equilibrium model), except that the equilibrium is on path for ES and TLC. In the case of the directed evolution of the activity of a designer Kemp eliminase enzyme, which gives rise to curved Arrhenius plots, we argue that this is most obviously the result of the optimisation of the TLC to improve preorganisation and catalysis with increases in ES-TLC cooperativity and attendant correlated motions at the TLC.[30,39] The increases in cooperativity to reach the TLC also rationalise remote mutations that improve catalysis as seen in many directed evolution studies.[40,41]

Evidence for TLCs in other enzymes

Molecular simulations show that specific reactive conformations of enzyme-substrate complexes are involved in many enzymes, and identify structural features of these complexes, providing details of TLCs. One example is fatty acid amide hydrolase, in which hydrolysis of oleamide occurs by a distinct, high energy conformation: the barrier to reaction is significantly lower in this conformation, and so reaction will proceed via this TLC, even though it is much less populated than other conformations of the ES complex of FAAH at 300K.[42,43] In lactate dehydrogenase, while reaction in the direction of lactate formation can proceed with the active site loop open or closed, oxidation of lactate to pyruvate requires loop closure.[44] In thymidylate synthase, different experimentally observed conformations of the enzyme show different reaction barriers, associated with different product stabilization.[45] From simulations of HIV-1 protease, Ribeiro *et al.* noted that the reaction will be dominated "by a very few transient enzyme conformations that provide very low barriers".[46]

QM/MM and MD simulations of ketosteroid isomerase (KSI) show that changes in active site structure cause changes in solvation of the catalytic base (Asp38): these changes significantly lower the barrier to reaction; the reaction proceeds via the TLC in which the base is less solvated.[38] In triosephosphate isomerase (TIM), a crucial active site loop adopts various different conformations in the ES complex, but a low barrier to reaction is only found when the loop is fully closed: in the TLC, the catalytic base (Glu165) is desolvated and so is more basic.[37] Reactive conformations (TLCs), involving distinct conformations of the substrate and active site, and desolvation of the catalytic base, are also important in the antibiotic breakdown activity of β-lactamases.[47,48] Mhashal *et al.* showed that reaction in glycerol-3-phosphate dehydrogenase proceeds via a conformation in which the active site is also less solvated.[49] Changes in basicity associated with changes in solvation, and active site loop behaviour, have also been found to be important in evolution and activity of dihydrofolate reductase[50], and in differences in reactivity between thermophilic and mesophilic DHFR, which are also modulated by dynamical/entropic changes caused by dimerization of the thermophilic enzyme.[51] Changes in solvation (e.g. desolvation of catalytic carboxylate groups) are likely to be a common feature of TLCs. In the TLC, specific changes in solvation increase reactivity of important groups, achieving desolvation and ground state destabilization within the overall context of a polar active site.

It is important to note that a TLC is a distinct conformation of the enzyme-substrate complex, involving structural changes throughout the protein, and is not limited to trivial substrate conformational changes or simple loop opening and closing motions. The conformation of the protein as a whole is different. The dynamics of different parts of the protein change in different ways in the TLC: some regions becoming less ordered, and others becoming more flexible. This is shown by MD simulations of KSI: in this homodimeric enzyme, the dynamics of the monomer in which reaction is not occurring are strongly affected by changes in the reactive monomer.[14] As noted above, the dynamics of the small domain of MalL, far from the active site, are different in the TLC. The relative stabilities of TLCs may be modulated by evolution, allosteric ligands and solvent effects as well as by temperature[52]

Conclusions

Warshel and colleagues have suggested that the abrupt changes seen for the temperature-dependence of $\Delta H^{\ddagger}$and $\Delta S^{\ddagger}$ for ADH are indicative of a phase transition.[8] This description is useful because the thermodynamic tools used to describe phase transitions are very well studied.[53] Indeed, the two-state model that we present here is the analogue of a second order, finite-size phase transition.[54] In this case, the kinetics of the system are fundamentally altered from when the temperature favours the TLC state (e.g. MalL at low T) to when ES state is favoured (e.g. MalL at high T). The kinetics of phase transitions have also been very well studied and are generally described in terms of a sphere of nucleation which must reach a critical size in order to affect the phase transition. Similar hypotheses have been put forward for protein folding (e.g. the nucleation-condensation model[55]) and may be a feature of the ES-to-TLC transition for enzyme-catalysed reactions.

Here, high resolution temperature-rate data provide sufficient detail to characterise a two-state cooperative conformational transition prior to the chemical step for enzyme catalysis (Figure 7). From high resolution structures and molecular dynamics simulations, we have characterised these two states (ES and TLC). At temperatures that favour the TLC (Figure 7A), enzyme kinetics are a function of the chemical step (e.g. low temperatures for MalL and high temperatures for ADH). In contrast, at temperatures that favour the ES (Figure 7B), the enzyme kinetics are a combination of both a conformational step and a chemical step leading to significant negative values of the activation heat capacity ($\Delta C_P^\ddagger$) at these temperatures (high temperatures for MalL and low temperatures for ADH). Finally, a model that uses a linear change in $\Delta C_P^\ddagger$ (MMRT-1S) as an approximation can discriminate between MalL-like behaviour and ADH-like behaviour and is sufficiently simple to use with most published data. This approach should find wide application in analysing and understanding the temperature dependence of enzyme-catalysed reactions.

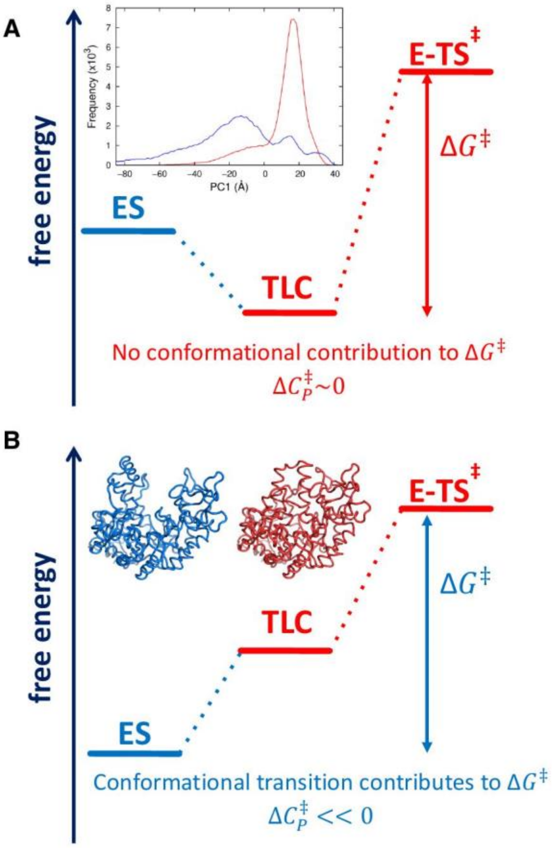


Figure 7. Free energy schematic at two different temperatures. **A.** If the temperature favours the TLC, then there is no contribution from conformational changes to $\Delta G^\ddagger$ and $\Delta C_P^\ddagger$ will be near 0 (i.e. the heat capacity for E-TS‡ is similar to that for TLC). This is the case for MaL at low temperatures and ADH at high temperatures. **B.** If the temperature favours the ES state, then conformational fluctuations will be a component of $\Delta G^\ddagger$ and $\Delta C_P^\ddagger$ will be significantly less than 0. This is the case for MaL at high temperatures and ADH at low temperatures.

Methods

Protein Expression

Bacillus subtilis MalL, and single amino acid variants, were expressed with N-terminal hexa-histidine tags in *E. coli* BL21 DE3 cells. Luria Broth cultures in exponential phase were induced at 18 °C with 0.75 mM Isopropyl β-D-1-thiogalactopyranoside and grown overnight.

Protein Purification

Protein purification was carried out in two steps via immobilised metal affinity chromatography and size exclusion chromatography (IMAC and SEC respectively) at pH 7.0. Initially, cell pellets were lysed via sonication on ice. A 25 mM to 0.5 M imidazole gradient over 50 mL was used to elute MalL during IMAC. SEC was carried out in 20 mM HEPES buffer. Enzymes were dialysed into 40 mM $NaPO_4$ buffer with 150 mM NaCl.

X-ray Structure Determination

Crystallisation of MalL S536R was performed using hanging-drop vapour diffusion at 18 °C. Crystals were obtained in 0.1 M Tris pH 8.0, 0.2 M ammonium acetate and 18% w/v PEG 10,000. Data collection was performed on flash-cooled crystals on the MX2 beamline at the Australian synchrotron. A solution of 0.1M Tris pH 8.0, 0.2 M ammonium acetate and 17% w/v PEG 10,000 with 20% v/v glycerol was used as cryoprotectant. Data was indexed, integrated and scaled in XDS[56] and further scaled and merged in Aimless.[57] The structure was solved by molecular replacement in Molrep[58] with wildtype MalL (PDB code: 4M56) as the search model. This was followed by iterative cycles of manual building in COOT,[59] structure correction using PDB-REDO,[60] and further refinement using Phenix.Refine[61] and Refmac5.[62]

MalL Temperature Assay

The KinetAsyst™ Stopped-Flow System (TgK Scientific, UK) with a connected circulating water bath for temperature control was used to characterise the temperature profiles of MalL via cleavage of saturating concentrations of *p*-nitrophenyl-α-D-glucopyranoside at 405 nm. Reactions were completed in triplicate with five 0.2 s dummy shots in between. Each reaction was carried out for 45 s. Temperature values reported are those from the thermostat control monitoring the reaction chamber. Enzyme stability over the experimental time-period (~5 hours) was confirmed by a mid-range temperature assay at the end of the experimental time-period.

Rate Calculation

Linear regression of (at most) the first 10 s of the reaction was carried out using Kinetic Studio (TgK Scientific, UK). Catalytic rates (k_{cat}; s^{-1}) were determined using an extinction coefficient (L mol^{-1} cm^{-1}) of 7413. Rate data were converted to change in Gibbs free energy ($\Delta G^{\ddagger}$) for model fitting using the Eyring equation with transmission coefficient set to 1.

Model Fitting

Levenberg-Marquardt non-linear regression was carried out in RStudio.

Molecular Dynamics Simulations and Principal Component Analysis

We previously performed 10, 500 ns long molecular dynamics (MD) simulations of WT-MalL with both isomaltose bound (reactant state, RS) and with a transition state analogue bound (TSA).[14] Here, we performed an additional 10 MD simulations of WT-MalL alongside 20 new MD simulations of the point variant S536R (with the starting coordinates based on the new X-ray structure), in both their ES and E-TS states, meaning each state was sampled with twenty 500 ns replicas each. The same protocols, force-field parameters (ff99SB-ILDN for protein, TIP4P-Ew for water, GLYCAM 06j-1 for isomaltose and combination of GLYCAM 06j-1 and GAFF for the TSA) and protonation states were used as in our previous work.[14]

Trajectory analysis was performed using CPPTRAJ (part of the AmberTools suite of programmes https://ambermd.org/AmberTools.php) using snapshots taken every 10 ps from all trajectories unless otherwise stated. The Cα RMSFs were determined by RMSD fitting (to the Cα of residues 7-561) to a running average coordinates using a time window of 10 ns. A hydrogen bond (HB) was defined to exist using typical criteria if the donor-acceptor distance was within 3.5 Å, and if the donor-hydrogen-acceptor angle was within 180±45°, . Hydrogen bonds between all residues were separated into main chain (MC) and side chain (SC) contributions from each residue, giving rise to either MC-MC, MC-SC and/or SC-SC HBs between residues. The average difference (from the 20 replicas) between the RS and TSA was determined from the 20 runs, and the significance of the differences was evaluated using a *t*-test. Principal component analysis (PCA) was performed on the Cα of every residue for all states simulated (WT, and S536R in both RS and TSA forms) combined. RMS fitting was first performed to a crystal structure of WT-MalL (PDB: 5WCZ) using the Cα of residues 7-561, to create an average structure. Following this, all snapshots were then re-fitted to this average structure for the subsequent calculation (again to the Cα of residues 7-561).

Hydrogen bond analysis

Hydrogen bonds were analysed for WT MalL and S536R structures. All hydrogen bonds were found using FindHBond in Chimera 1.15.[63] Hydrogen bond criteria are described in Mills & Dean (1996)[64] with criteria relaxed by 0.4 Å and 20°. Existing explicit hydrogens were removed from the structure prior to analysis. Unique H-bonds and those that were at least 0.3 Å shorter than their equivalent were extracted. To discount the effects of the resolution the structures were solved at (WT: 2.3 Å, S536R: 1.1 Å) the process was repeated with the S536R structure where the data were truncated to 2.3 Å resolution and the structure refined. In addition, bonds involving multiple rotamers, or in regions not modelled in the other structure, were discounted.

Acknowledgements

This research was undertaken in part using the MX2 beamline at the Australian Synchrotron, part of ANSTO, and made use of the Australian Cancer Research Foundation (ACRF) detector. VLA, LS & AJM are investigators funded by a Marsden Fund Council grant from the Marsden Fund of New Zealand. EJW, CJH and AW acknowledge doctoral funding from the University of Waikato. M.C. and A.J.M. thank the EPSRC Centre for Doctoral Training in Theory and Modelling in Chemical Sciences (EP/L015722/1). We further acknowledge EPSRC funding for CCP-BioSim (EP/M022609/1). This work is part of a project that has received funding from the European Research Council under the European Horizon 2020 research and innovation programme (PREDACTED Advanced Grant Agreement no. 101021207) to A.J.M. The simulation work was conducted using the computational facilities of the Advanced Computing Research Centre, University of Bristol. MWvdK thanks BBSRC for funding (BB/M026280/1).

References

(1) Kavanau, J. L. Enzyme Kinetics and the Rate of Biological Processes. *J. Gen. Phys.* 1950, *34* (2), 193–209.

(2) Peterson, M. E.; Eisenthal, R.; Danson, M. J.; Spence, A.; Daniel, R. M. A New Intrinsic Thermal Parameter for Enzymes Reveals True Temperature Optima. *J. Biol. Chem.* 2004, *279* (20), 20717–20722.

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

(4) Truhlar, D. G.; Kohen, A. Convex Arrhenius Plots and Their Interpretation. *Proc. Natl Acad. Sci. USA* 2001, *98* (3), 848–851.

(5) Åqvist, J.; Sočan, J.; Purg, M. Hidden Conformational States and Strange Temperature Optima in Enzyme Catalysis. *Biochemistry* 2020, 59 (40), 3844–3855.

(6) Arcus, V. L.; Mulholland, A. J. Temperature, Dynamics, and Enzyme-Catalyzed Reaction Rates. *Ann. Rev. Biophys.* 2020, 49, 163-180.

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

(8) Roy, S.; Schopf, P.; Warshel, A. Origin of the Non-Arrhenius Behavior of the Rates of Enzymatic Reactions. *J. Phys. Chem. B* 2017, *121* (27), 6520–6526.

(9) Fersht, A. R. Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding, 4th ed.; World Scientific Publishing Co., 2017.

(10) Dijk, E. van; Varilly, P.; Knowles, T. P. J.; Frenkel, D.; Abeln, S. Consistent Treatment of Hydrophobicity in Protein Lattice Models Accounts for Cold Denaturation. *Phys. Rev. Lett.* 2016, *116* (7), 078101.

(11) Szyperski, T.; Mills, J. L.; Perl, D.; Balbach, J. Combined NMR-Observation of Cold Denaturation in Supercooled Water and Heat Denaturation Enables Accurate Measurement of ΔCp of Protein Unfolding. *Eur. Biophys. J.* 2006, *35* (4), 363–366.

(12) Oliveberg, M.; Tan, Y. J.; FERSHT, A. Negative Activation Enthalpies in the Kinetics of Protein Folding. *Proc. Natl Acad. Sci. USA* 1995, *92* (19), 8926–8929.

(13) Firestone, R. S.; Cameron, S. A.; Karp, J. M.; Arcus, V. L.; Schramm, V. L. Heat Capacity Changes for Transition-State Analogue Binding and Catalysis with Human 5'-Methylthioadenosine Phosphorylase. *ACS Chem. Biol.* 2017, *12* (2), 464–473.

(14) Kamp, M. W. van der; Prentice, E. J.; Kraakman, K. L.; Connolly, M.; Mulholland, A. J.; Arcus, V. L. Dynamical Origins of Heat Capacity Changes in Enzyme-Catalysed Reactions. *Nature Comms* 2018, *9* (1), 1177.

(15) Changeux, J.-P. Allostery and the Monod-Wyman-Changeux Model after 50 Years. *Ann. Rev. Biophys.* 2012, *41*, 103–133.

(16) Cooper, A.; Dryden, D. Allostery Without Conformational Change - a Plausible Model. *Eur. Biophys. J. Biophys. Lett.* 1984, *11* (2), 103–109.

(17) Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J. The Ensemble Nature of Allostery. *Nature* 2014, *508* (7496), 331–339.

(18) Nussinov, R. Introduction to Protein Ensembles and Allostery. *Chem. Rev.* 2016, *116* (11), 6263–6266.

(19) Glowacki, D. R.; Harvey, J. N.; Mulholland, A. J. Taking Ockham's Razor to Enzyme Dynamics and Catalysis. *Nature Chem.* 2012, *4* (3), 169–176.

(20) Massey, V.; Curti, B.; Ganther, H. A Temperature-Dependent Conformational Change in D-Amino Acid Oxidase and Its Effect on Catalysis. *J. Biol. Chem.* 1966, *241* (10), 2347–2357.

(21) Daniel, R. M.; Danson, M. J.; Eisenthal, R. The Temperature Optima of Enzymes: A New Perspective on an Old Phenomenon. *Trends Biochem.Sci.* 2001, *26* (4), 223–225.

(22) Åqvist, J.; Ent, F. van der. Calculation of Heat Capacity Changes in Enzyme Catalysis and Ligand Binding. *J. Chem. Theory Comput.* 2022. 18 (10), 6345–6353.

(23) Feller, G. Psychrophilic Enzymes: From Folding to Function and Biotechnology. *Scientifica* 2013, *2013* (2), 1–28.

(24) Feller, A. C. F. D. J.-L. D. L. G.; Delbrassine, F.; Lage, J.-L. D.; Feller, G. Temperature Adaptations in Psychrophilic, Mesophilic and Thermophilic Chloride-Dependent Alpha-Amylases. *Biochimie* 2012, *94* (9), 1943–1950.

(25) Han, X.; Lee, R.; Chen, T.; Luo, J.; Lu, Y.; Huang, K.-W. Kinetic Evidence of an Apparent Negative Activation Enthalpy in an Organocatalytic Process. *Sci. Rep. UK* 2013, *3* (1), 2557.

(26) Prentice, E. J.; Hicks, J.; Ballerstedt, H.; Blank, L. M.; Liáng, L. N. L.; Schipper, L. A.; Arcus, V. L. The Inflection Point Hypothesis: The Relationship between the Temperature Dependence of Enzyme-Catalyzed Reaction Rates and Microbial Growth Rates. *Biochemistry* 2020, *59* (38), 3562–3569.

(27) Nagel, Z. D.; Dong, M.; Bahnson, B. J.; Klinman, J. P. Impaired Protein Conformational Landscapes as Revealed in Anomalous Arrhenius Prefactors. *Proc. Natl Acad. Sci. USA* 2011, *108* (26), 10520–10525.

(28) Kohen, A.; Cannio, R.; Bartolucci, S.; Klinman, J. P. Enzyme Dynamics and Hydrogen Tunnelling in a Thermophilic Alcohol Dehydrogenase. *Nature* 1999, *399* (6735), 496–499.

(29) Winter, S. D.; Jones, H. B. L.; Răsădean, D. M.; Crean, R. M.; Danson, M. J.; Pantoş, G. D.; Katona, G.; Prentice, E.; Arcus, V. L.; Kamp, M. W. van der; Pudney, C. R. Chemical Mapping Exposes the Importance of Active Site Interactions in Governing the Temperature Dependence of Enzyme Turnover. *ACS Catalysis* 2021, *11* (24), 14854–14863.

(30) Bunzel, H. A.; Anderson, J. L. R.; Hilvert, D.; Arcus, V. L.; Kamp, M. W. van der; Mulholland, A. J. Evolution of Dynamical Networks Enhances Catalysis in a Designer Enzyme. *Nat. Chem.* 2021, *13* (10), 1017–1022.

(31) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. Protein Folding Kinetics and Thermodynamics from Atomistic Simulation. *Proc. Natl Acad. Sci. USA* 2012, *109* (44), 17845–17850.

(32) Daniel, R. M.; Danson, M. J. A New Understanding of How Temperature Affects the Catalytic Activity of Enzymes. *Trends Biochem. Sci.* 2010, *35* (10), 584–591.

(33) Wolfenden, R.; Lu, X.; Young, G. Spontaneous Hydrolysis of Glycosides. *J. Am. Chem. Soc.* 1998, *120* (27), 6814–6815.

(34) Stockbridge, R. B.; Lewis, C. A.; Yuan, Y.; Wolfenden, R. Impact of Temperature on the Time Required for the Establishment of Primordial Biochemistry, and for the Evolution of Enzymes. *Proc. Natl Acad. Sci. USA* 2010, *107* (51), 22102–22105.

(35) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M. Electrostatic Basis for Enzyme Catalysis. *Chem. Revs* 2006, *106* (8), 3210–3235.

(36) Edwards, D. R.; Lohman, D. C.; Wolfenden, R. Catalytic Proficiency: The Extreme Case of S-O Cleaving Sulfatases. *J. Am. Chem. Soc.* 2012, *134* (1), 525–531.

(37) Liao, Q.; Kulkarni, Y.; Sengupta, U.; Petrović, D.; Mulholland, A. J.; Kamp, M. W. van der; Strodel, B.; Kamerlin, S. C. L. Loop Motion in Triosephosphate Isomerase Is Not a Simple Open and Shut Case. *J. Am. Chem. Soc.* 2018, *140* (46), 15889–15903.

(38) Kamp, M. W. van der; Chaudret, R.; Mulholland, A. J. QM/MM Modelling of Ketosteroid Isomerase Reactivity Indicates That Active Site Closure Is Integral to Catalysis. *FEBS J.* 2013, *280* (13), 3120–3131.

(39) Weitzner, B. D.; Kipnis, Y.; Daniel, A. G.; Hilvert, D.; Baker, D. A Computational Method for Design of Connected Catalytic Networks in Proteins. *Prot. Sci.* 2019, *28* (12), 2036–2041.

(40) Bunzel, H. A.; Kries, H.; Marchetti, L.; Zeymer, C.; Mittl, P. R. E.; Mulholland, A. J.; Hilvert, D. Emergence of a Negative Activation Heat Capacity during Evolution of a Designed Enzyme. *J. Am. Chem. Soc.* 2019, *141* (30), 11745–11748.

(41) Dalby, P. A. Strategy and Success for the Directed Evolution of Enzymes. *Curr. Opin. Struc. Biol.* 2011, *21* (4), 473–480.

(42) Lodola, A.; Mor, M.; Zurek, J.; Tarzia, G.; Piomelli, D.; Harvey, J. N.; Mulholland, A. J. Conformational Effects in Enzyme Catalysis: Reaction via a High Energy Conformation in Fatty Acid Amide Hydrolase. *Biophys. J.* 2007, *92* (2), L20-2.

(43) Lodola, A.; Sirirak, J.; Fey, N.; Rivara, S.; Mor, M.; Mulholland, A. J. Structural Fluctuations in Enzyme-Catalyzed Reactions: Determinants of Reactivity in Fatty Acid Amide Hydrolase from Multivariate Statistical Analysis of Quantum Mechanics/Molecular Mechanics Paths. *J. Chem. Theory Comp.* 2010, *6* (9), 2948–2960.

(44) Świderek, K.; Tuñón, I.; Martí, S.; Moliner, V. Protein Conformational Landscapes and Catalysis. Influence of Active Site Conformations in the Reaction Catalyzed by L-Lactate Dehydrogenase. *ACS Catal.* 2015, *5* (2), 1172–1185.

(45) Świderek, K.; Kohen, A.; Moliner, V. The Influence of Active Site Conformations on the Hydride Transfer Step of the Thymidylate Synthase Reaction Mechanism. *Phys. Chem. Chem. Phys.* 2015, *17* (46), 30793–30804.

(46) Ribeiro, A. J. M.; Santos-Martins, D.; Russo, N.; Ramos, M. J.; Fernandes, P. A. Enzymatic Flexibility and Reaction Rate: A QM/MM Study of HIV-1 Protease. *ACS Catal.* 2015, *5* (9), 5617–5626.

(47) Hirvonen, V. H. A.; Weizmann, T. M.; Mulholland, A. J.; Spencer, J.; Kamp, M. W. van der. Multiscale Simulations Identify Origins of Differential Carbapenem Hydrolysis by the OXA-48 Β-Lactamase. *ACS Catal.* 2022, *12* (8), 4534–4544.

(48) Chudyk, E.; Beer, M.; Limb, M. A. L.; Jones, C. A.; Spencer, J.; Kamp, M. W. van der; Mulholland, A. QM/MM Simulations Reveal the Determinants of Carbapenemase Activity in Class A β-Lactamases. *ACS Infect. Dis.* 2022, 8, 8, 1521–1532.

(49) Mhashal, A. R.; Romero-Rivera, A.; Mydy, L. S.; Cristobal, J. R.; Gulick, A. M.; Richard, J. P.; Kamerlin, S. C. L. Modeling the Role of a Flexible Loop and Active Site Side Chains in Hydride Transfer Catalyzed by Glycerol-3-Phosphate Dehydrogenase. *ACS Catal.* 2020, *10* (19), 11253–11267.

(50) Mhashal, A. R.; Pshetitsky, Y.; Cheatum, C. M.; Kohen, A.; Major, D. T. Evolutionary Effects on Bound Substrate PK a in Dihydrofolate Reductase. *J. Am. Chem. Soc.* 2018, *140* (48), 16650–16660.

(51) Ruiz-Pernía, J. J.; Tuñón, I.; Moliner, V.; Allemann, R. K. Why Are Some Enzymes Dimers? Flexibility and Catalysis in Thermotoga Maritima Dihydrofolate Reductase. *ACS Catal.* 2019, *9* (7), 5902–5911.

(52) Hindson, S. A.; Bunzel, H. A.; Frank, B.; Svistunenko, D. A.; Williams, C.; Kamp, M. W. van der; Mulholland, A. J.; Pudney, C. R.; Anderson, J. L. R. Rigidifying a De Novo Enzyme Increases Activity and Induces a Negative Activation Heat Capacity. *ACS Catal.* 2021, *11* (18), 11532–11541.

(53) Goldenfeld. *Lectures On Phase Transitions And The Renormalization Group*; CRC Press, 1992.

(54) Blundell, S. J.; Blundell, K. M. *Concepts in Thermal Physics*; Oxford University Press; Oxford University Press, 2006.

(55) Fersht, A. R. Optimization of Rates of Protein Folding: The Nucleation-Condensation Mechanism and Its Implications. *Proc. Natl Acad. Sci. USA* 1995, *92* (24), 10869–10873.

(56) Kabsch, W. Integration, Scaling, Space-Group Assignment and Post-Refinement. *Acta Cryst. Sect D Biol. Cryst.* 2010, *66* (2), 133–144.

(57) Evans, P. R.; Murshudov, G. N. How Good Are My Data and What Is the Resolution? *Acta Cryst. Sect D* 2013, *69* (7), 1204–1214.

(58) Vagin, A.; Teplyakov, A. Molecular Replacement with MOLREP. *Acta Cryst. Sect D* 2010, *66* (1), 22–25.

(59) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of Coot. *Acta Cryst. Sect D* 2010, *66* (Pt 4), 486–501.

(60) Joosten, R. P.; Long, F.; Murshudov, G. N.; Perrakis, A. The PDB_REDO Server for Macromolecular Structure Model Optimization. *IUCR J.* 2014, *1* (4), 213–220.

(61) 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. Towards Automated Crystallographic Structure Refinement with Phenix.Refine. *Acta Cryst. Sect D* 2012, *68* (4), 352–367.

(62) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Cryst. Sect. D* 1997, *53* (3), 240–255.

(63) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* 2004, *25* (13), 1605–1612.

(64) Mills, J. E. J.; Dean, P. M. Three-Dimensional Hydrogen-Bond Geometry and Probability Information from a Crystal Survey. *J. Comput. Aid Mol. Des.* 1996, *10* (6), 607–622.

Supplementary Material

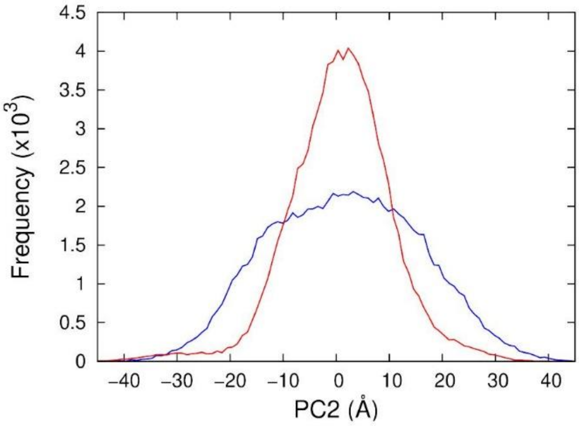


Figure S1. Principal component analysis for WT MalL in the ES and E-TS states showing a projection of the second principal component, PC2.

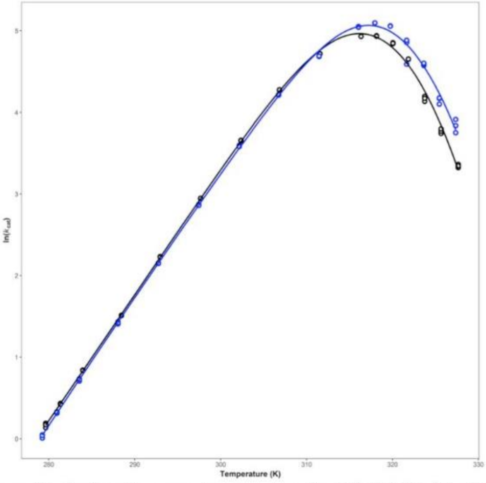


Figure S2. Ln(rate) versus temperature for WT MalL (black) and S536R (blue).

Table S1. Data collection and refinement statistics for MalL S536R.

Statistic	**MalL S536R**
Wavelength (Å)	0.953735
Space group	$P\ 1\ 2_1\ 1$
Unit cell lengths (Å)	a = 48.75 b = 101.00 c = 61.75
Unit cell angles (°)	α = 90.00 β = 113.06 γ = 90.00
Resolution (Å)	1.10 - 44.85 (1.10 - 1.12)
R_{merge}	0.107 (0.548)
Completeness (%)	94.2 (87.1)
Redundancy	10.9 (6.9)
No. of observations	2270037 (65265)
No. of unique reflections	208774 (9509)
Mean I/σI	12.7 (2.9)
R factor	0.126
R_{free}	0.145
Protein atoms	9633
Solvent atoms	785
Average temperature factor (Å^2)	16.13
RMSD bond lengths (°)	0.01
RMSD bond angles (Å)	1.182