Exploration and exploitation of non-canonical amino acid incorporation to detect or improve transketolase activity and stability

Thesis submitted to University College London for the degree of Doctor of Philosophy

By

Henry C. Wilkinson

Department of Biochemical Engineering
University College London
2019
I, Henry Wilkinson, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed:

Henry Wilkinson
2019
Firstly, I’d like to thank my supervisor Paul Dalby for his support during my PhD. He has given me great freedom to explore my project and fulfil my own interests, whilst offering invaluable guidance and in-depth discussions throughout. I have thoroughly enjoyed having him as my supervisor and his approach in this role has been instrumental in the research output of this project. I would also like to thank Dhushy Stanislaus, Facilities Manager, whose tireless efforts to maintain and improve all aspects of the laboratory, particularly the support she offered me in my role as HPLC lab manager, were greatly appreciated.

My introduction to the Dalby Group and laboratory was seamless thanks to Pierre Affaticati, who took time out of his schedule to train me on all lab facilities. Samir Aoudjane, John Hales and Pierre-Yvres Colin also freely offered invaluable technical advice whenever asked.

I also thank Dr Kersti Karu for mass spectrometric measurements in the UCL Department of Chemistry Mass Spectrometry Facility, Dr Jayesh Gor for AUC measurements in the Molecular Interaction Facility at UCL, Dr Clive Metcalfe for the peptide digest, fragment MS and analysis at NIBSC, and Peter Schultz, George Church and Dr John Ward for donation of the *E. coli* strain C321AA.exp, the plasmid *pUltra* and the transketolase plasmid pQR791, respectively. Finally, I thank the Engineering and Physical Sciences Research Council (EPSRC) and the Centre for Doctoral Training in Emergent Macromolecular Therapies (EP/L015218/1).
Site-specific, non-canonical amino acid (ncAA) incorporation into proteins is a ground-breaking methodology that facilitates the exploration of protein sequence space beyond what is possible naturally, and enables the site-specific introduction of bio-orthogonal functional groups. NcAAs can be exploited to improve, modify or create novel enzyme activities, or to probe protein structure-function relationships through genetically encoded or post-translationally modified ncAAs. A transketolase-based ncAA-incorporation platform based on an existing ncAA-incorporation plasmid, pUltra, was developed and optimised to form the foundation of this study. The thesis outlines novel approaches to quantitatively analyse incorporation fidelity, and to adjust activity parameters to account for misincorporation.

The platform was first utilised to incorporate para-substituted phenylalanine derivatives into the active site of a transketolase variant, S385Y/D469T/R520Q, previously evolved to accept aromatic substrates, to probe the effect of side-chain structure and electronics on catalytic activity, acceptor substrate inhibition, enzyme stability and binding pocket preference. A novel function, the Modified Michaelis-Menten function, was derived to describe and quantify the extensive substrate inhibition observed at high [substrate]. To this author’s knowledge, variant pAMF is the first example in which catalytic activity (2.4-fold and 2.6-fold improvements in \( K_m \) and catalytic efficiency) and stability (5 °C increase in \( T_m \)) have been simultaneously evolved via site-specific incorporation of ncAAs into an active site.

After briefly exploring the use of para-cyanophenylalanine (pCNF) as a genetically-encoded fluorescent probe, it became apparent that transketolase intrinsic fluorescence could be used to determine the binding parameters of the transketolase cofactors and substrates, independently of activity, in a novel TPP-binding assay. A second, previously uncharacterised, low-affinity TPP binding-site was observed that belonged to an inactive, low-affinity TK species, TK\(_{\text{low}}\). The two forms of the monomeric unit with high- (TK\(_{\text{high}}\)) and low- (TK\(_{\text{low}}\)) affinity could dimerise to form three compositions of dimer. This work led to the proposal of a novel Two-Species Model of transketolase activation, regulation and inhibition that describes the interconversions between apo-/holo- TK\(_{\text{high}}\) and TK\(_{\text{low}}\) in response to heat-shock and oxidative stress, and its physiological relevance. This is a significant discovery in a well-studied protein and could have major implications in both transketolase research and in a wider context, such as in cancer research.
Finally, a novel, FRET-based stability assay was developed that detected local protein unfolding and aggregation of transketolase, when both purified and in cell lysate, via incorporation and bio-orthogonal labelling of the non-canonical amino acid para-azidophenylalanine (pAzF). The stability assay was applied to transketolase in a lysate background to investigate the effect of cell-based macromolecular crowding on local stability. While the ncAA pAzF and the two Alexafluor dyes have been used previously to interrogate protein structure and stability under different conditions, to this author’s knowledge this is the first time doubly-incorporated and labelled pAzF has been used to study unfolding in real time via FRET, both in purified protein and in a lysate background.
First and foremost, the Two-Species Model of transketolase activation, regulation and inhibition developed in this thesis will significantly impact future research into transketolase in several ways. I expect future academic work will attempt to maximise the highly-active form of transketolase, TK$_{\text{high}}$, and further characterise the physiological relevance of redox- and heat-regulation. This research can potentially be leveraged to develop more-active industrial biocatalysts and novel cancer therapies, given the role of transketolase in cancer proliferation, or even exploit its redox-sensitivity as a biosensor for oxidative stress. It will also provide the backbone for decoding the structure-function relationships of what is a complex and quirky protein, such as the identification of the proton wire between active sites. The cofactor- and donor substrate-binding assays, specific to transketolase, are potentially compatible with a microplate set-up, which could be utilised to evolve cofactor and donor-substrate specificity.

The major focus of this thesis, ncAA-incorporation, will have far wider implications than just transketolase activity and stability. The development of a transketolase ncAA-incorporation platform, albeit consisting of components that were developed previously, is already being adapted for use in a number of different protein studies in the Dalby lab, as is the protein stability assay developed in this thesis. The latter has great potential as a high-throughput stability assay using lysate instead of purified protein, which could, in theory, be exploited to evolve the stability of any biologic, including therapeutic proteins, biocatalysts and vaccines, given its bio-orthogonal nature.

More generally, the data presented in this thesis demonstrates the huge potential ncAAs have in a) significantly improving enzyme activity and stability; and b) introducing bio-orthogonal chemistries into proteins as the basis of biophysical or biochemical assays. This thesis outlines a robust methodology to analyse the fidelity of incorporation and % labelling in detail, which can vary hugely from variant to variant and ncAA to ncAA, and its impact on apparent vs true activity and stability. Future ncAA-incorporation studies should endeavour to follow a similar methodology for consistency between academic studies.
ABBREVIATIONS

\%B_{\text{max(high)}} \quad \text{The percentage of all TK that is TK}_{\text{high}}

\%I_{\text{max}} \quad \text{Maximum \% inhibition activity}

\%TK_{\text{high}} \quad \text{The percentage of all TK that is TK}_{\text{high}}

3-FBA \quad 3\text{-formylbenzoic acid}

3-HBA \quad 3\text{-hydroxybenzaldehyde}

4-FBA \quad 4\text{-formylbenzoic acid}

aaRS \quad \text{amino-acyl tRNA synthetase}

AF488 \quad \text{AlexaFluor-488}

AF594 \quad \text{AlexaFluor-594}

AUC \quad \text{Analytical Ultracentrifugation}

CHP \quad \text{Cumene Hydroperoxide}

CPEC \quad \text{Circular Polymerase Extension Cloning}

DHE-TPP \quad \text{Dihydroxyethyl-TPP}

DCP-Bio1 \quad \text{Biotinylated dimedone}

EPL \quad \text{Expression Protein Ligation}

Ery \quad \text{Erythrulose}

FRET \quad \text{Förster Resonance Energy Transfer}

GA \quad \text{Glycolaldehyde}

HPA \quad \beta\text{-hydroxypyruvate}

IAM \quad \text{iodoacetamide}

IFE \quad \text{Inner Filter Effect}

k_{\text{cat}} \quad \text{Catalytic turnover}

K_{d(high)} \quad \text{Dissociation constant of TK}_{\text{high}}

K_{d(low)} \quad \text{Dissociation constant of TK}_{\text{low}}

K_i \quad \text{Inhibition binding constant}

K_m \quad \text{Michaelis-Menten constant}

LC-ESI-MS \quad \text{Liquid Chromatography Electrospray Ionisation Mass Spectrometry}

n \quad \text{Hill coefficient}

ncAA \quad \text{non-canonical amino acid}

n_{\text{high}} \quad \text{Hill coefficient of TK}_{\text{high}}
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_i$</td>
<td>Hill coefficient of inhibition</td>
</tr>
<tr>
<td>$n_{low}$</td>
<td>Hill coefficient of TK$_{low}$</td>
</tr>
<tr>
<td>$p$CNF</td>
<td>para-cyanophenylalanine</td>
</tr>
<tr>
<td>PDHc-E1</td>
<td>E1 subunit of the pyruvate dehydrogenase complex</td>
</tr>
<tr>
<td>PP-domain</td>
<td>Pyrophosphate-Binding Domain</td>
</tr>
<tr>
<td>PPP</td>
<td>Pentose Phosphsate Pathway</td>
</tr>
<tr>
<td>PTM</td>
<td>Post-Translational Modification</td>
</tr>
<tr>
<td>Pyr-domain</td>
<td>Pyrimidine-Binding Domain</td>
</tr>
<tr>
<td>SLS</td>
<td>Static Light Scattering</td>
</tr>
<tr>
<td>$T_{agg}$</td>
<td>Thermal aggregation transition mid-point</td>
</tr>
<tr>
<td>TFA</td>
<td>Trifluoroacetic Acid</td>
</tr>
<tr>
<td>TK</td>
<td>Transketolase</td>
</tr>
<tr>
<td>$T_m$</td>
<td>Thermal transition mid-point</td>
</tr>
<tr>
<td>$T_{\text{on}}$</td>
<td>The temperature at which 10% of the sample was aggregated</td>
</tr>
<tr>
<td>TPP</td>
<td>Thiamine Pyrophosphate</td>
</tr>
<tr>
<td>TyrRS</td>
<td>Tyrosyl-tRNA Synthetase</td>
</tr>
<tr>
<td>X5P</td>
<td>D-xylulose-5-phosphate</td>
</tr>
<tr>
<td>$\Delta H_{vh}$</td>
<td>van’t Hoff enthalpy</td>
</tr>
</tbody>
</table>
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS .................................................................................................................... 3

ABSTRACT ........................................................................................................................................ 4

IMPACT STATEMENT ....................................................................................................................... 6

ABBREVIATIONS ............................................................................................................................. 7

TABLE OF CONTENTS ...................................................................................................................... 9

TABLE OF FIGURES ........................................................................................................................ 15

TABLE OF TABLES .......................................................................................................................... 19

CHAPTER ONE: Introduction ............................................................................................................. 21

   Enzymes as biocatalysts ................................................................................................................ 21

   Transketolase as a biocatalyst ...................................................................................................... 21

      Industrial applications of transketolase: Asymmetric carbon-carbon bond elongation .......... 22

      Structure and function of wild-type transketolase ................................................................. 22

      Structural changes to apo-transketolase and TPP upon cofactor binding ......................... 23

      The fluorescence intensity characteristics of E. coli transketolase .................................... 25

      The catalytic mechanism of transketolase ............................................................................. 25

      Substrate specificity, stereospecificity & stereoselectivity of transketolase ....................... 26

      Directed evolution of transketolase to alter substrate specificity ...................................... 28

      Improved activity towards aromatic aldehyde substrates .................................................... 29

      Thermal stability of transketolase ......................................................................................... 33

   Improving enzyme kinetics and protein stability ....................................................................... 34

RESEARCH AIMS ............................................................................................................................. 35

CHAPTER TWO: General experimental procedures ........................................................................... 36

   Media & solutions ...................................................................................................................... 36

      Chemicals & reagents ............................................................................................................ 36

      Ampicillin (Amp) stock solution ......................................................................................... 36

      Spectinomycin (Spectin) stock solution ............................................................................ 36

      Lysogenic broth (LB) medium ......................................................................................... 36

      Lysogenic broth (LB) agar ............................................................................................... 37
TABLE OF CONTENTS

Tris-HCl buffer ............................................................................................................ 37
100x ncAA stock solution .......................................................................................... 37
1000x IPTG stock solution ......................................................................................... 37
Standard transketolase 10x cofactor stock solution .................................................. 37

Standard molecular biology techniques ................................................................... 37
Glycerol stock preparation and maintenance ............................................................. 37
Plasmid extraction, quantification & storage .............................................................. 37
Transformation of commercial ultra-competent cells by heat-shock ......................... 38
Quikchange Mutagenesis ......................................................................................... 38
Transformation ......................................................................................................... 39

Protein expression & purification .............................................................................. 40
Generation of single bacterial colonies on LB-agar ..................................................... 40
Overnight starter culture ............................................................................................ 40
50 mL shake flask culture (in 250 mL flask) .............................................................. 40
Cell pellet extraction & storage .................................................................................. 40
Preparation of clarified lysate .................................................................................... 40
Transketolase His-Tag purification ............................................................................ 41
Protein dialysis .......................................................................................................... 41
Protein ultra-filtration & concentration ..................................................................... 41

Protein characterisation ............................................................................................ 42
Quantification of cell lysate protein concentration by Bradford assay ....................... 42
Quantification of purified protein concentration by A280 using a Nanodrop-2000 ......... 42
Mass spectrometry ..................................................................................................... 42
SDS-PAGE analysis .................................................................................................... 43

CHAPTER THREE: Development & optimisation of a genetically encoded E. coli ncAA-
incorporation platform for para-substituted phenylalanine derivatives .................... 44
INTRODUCTION ......................................................................................................... 44
Expanding the genetic code ....................................................................................... 44
Methods of UAA incorporation ................................................................................. 45
In vivo UAA incorporation ......................................................................................... 45
Advances in site-specific UAA incorporation ............................................................. 49

EXPERIMENTAL PROCEDURES ............................................................................. 50
Mutagenesis at residue 385Y, 105Y and 603K ............................................................ 50
Preparation of and co-transformation into competent C321.DA.exp “amberless” cells .... 51
TABLE OF CONTENTS

Mass spectrometry ........................................................................................................... 51
SDS-PAGE analysis ........................................................................................................... 51
Circular polymerase extension cloning (CPEC) ................................................................. 51

RESULTS ........................................................................................................................... 53
The efficiency and fidelity of ncAA-incorporation ............................................................ 53
Comparison of incorporation of pAMF, pCNF and pNTF relative to Y ......................... 54
Comparison of incorporation of pAzF at residue 603K and 105Y relative to S385Y/D469T/R520Q ................................................................. 55
Optimisation of the ncAA-incorporation platform using tacI ........................................... 55

DISCUSSION ....................................................................................................................... 56

SUMMARY .......................................................................................................................... 56

CHAPTER FOUR: Fine-tuning the activity and stability of an evolved enzyme active site through ncAAs ......................................................................................................................... 58

INTRODUCTION .................................................................................................................. 58

EXPERIMENTAL PROCEDURES ...................................................................................... 60

Mutagenesis at residue 385 ............................................................................................... 60
Preparation of and co-transformation into competent C321ΔA.exp “amberless” cells........ 60
Enzyme preparation & enzyme kinetics ........................................................................... 60
Derivation of the kinetic model for activity and substrate inhibition .............................. 61
Mass spectrometry ........................................................................................................... 63
SDS-PAGE analysis ......................................................................................................... 63
Thermal denaturation measurements .............................................................................. 63
Determination of kinetic parameters of pAMF, pCNF and pNTF variants ...................... 63
Computational docking .................................................................................................. 64

RESULTS ........................................................................................................................... 65
Electronic properties of the single-ringed aromatic amino acids utilised in this study ........ 65
Efficiency and fidelity of ncAA incorporation ................................................................... 66
Determination of the apparent, ncAA-incorporated and true (TK_{high}) kinetic parameters ...... 67
Thermal stability of holo-transketolase variants ................................................................. 73
Computational docking of 3-HBA into triple-mutant variants ......................................... 75

DISCUSSION ....................................................................................................................... 80

SUMMARY .......................................................................................................................... 81
CHAPTER FIVE: Novel insights into activation by cofactor binding suggests a Two-Species Model of transketolase interconversion

INTRODUCTION .................................................................................................................. 82

EXPERIMENTAL PROCEDURES ....................................................................................... 84

Enzyme preparation ........................................................................................................ 84
Fluorescence assay to detect TPP binding .................................................................. 85
Analytical Ultracentrifugation (AUC) Measurements .................................................. 85
Transketolase activity assay ......................................................................................... 86
Oxidation of transketolase using cumene hydroperoxide (CHP) .................................. 86
Mass spectrometry ........................................................................................................ 86
Dimedone labelling ........................................................................................................ 86
Peptide digest-Mass spectrometry (performed by Clive Metcalfe at NIBSC) .............. 86

RESULTS ............................................................................................................................ 88

A novel fluorescence quenching-based cofactor-binding assay .................................. 88
Characterisation of TPP binding to wild–type E. coli transketolase ............................ 89
Detection and characterisation of a novel low-affinity species, TK\textsubscript{low} .................. 94
TK\textsubscript{high} and TK\textsubscript{low} are two distinct forms of transketolase ......................... 97
TK\textsubscript{low} is a low-activity form of transketolase .................................................... 103
Ruling out alternative possible origins of \(\% B_{\text{max(\text{high})}}\) ........................................ 104

DISCUSSION .................................................................................................................... 107

SUMMARY ......................................................................................................................... 110

CHAPTER SIX: Transketolase donor substrate-binding, inhibition and heat-activation to refine the Two-Species Model of transketolase interconversion

INTRODUCTION .................................................................................................................. 111

EXPERIMENTAL PROCEDURES ....................................................................................... 113

Enzyme preparation ........................................................................................................ 113
Fluorescence assay to detect TPP binding .................................................................. 114
Transketolase activity assay ......................................................................................... 114
Fluorescence assay to detect HPA binding .................................................................. 114
Mass spectrometry (LC-ESI-MS) ................................................................................ 114

RESULTS ............................................................................................................................ 115

The two-species transketolase model is persistent across variants ............................ 115
Adaption of the fluorescence quenching-based TPP-binding assay for donor substrate binding ......................................................................................................................... 118
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Characterisation of E. coli transketolase binding to HPA</td>
<td>119</td>
</tr>
<tr>
<td>Heat-induced activation and conversion of TK&lt;sub&gt;low&lt;/sub&gt; to TK&lt;sub&gt;high&lt;/sub&gt;</td>
<td>122</td>
</tr>
<tr>
<td>The existence of a TK&lt;sub&gt;high&lt;/sub&gt;-TK&lt;sub&gt;low&lt;/sub&gt; mixed dimer species that mediates HPA substrate inhibition</td>
<td>124</td>
</tr>
<tr>
<td>The unified Two-Species Model of transketolase activation, regulation and inhibition</td>
<td>126</td>
</tr>
<tr>
<td><strong>DISCUSSION</strong></td>
<td>127</td>
</tr>
<tr>
<td><strong>SUMMARY</strong></td>
<td>128</td>
</tr>
<tr>
<td><strong>CHAPTER SEVEN: A novel cell-lysate-compatible protein stability assay using bio-orthogonal FRET-labelling of a non-canonical amino acid (ncAA) reveals the impact of macromolecular crowding on local structural stability of E. coli transketolase</strong></td>
<td>130</td>
</tr>
<tr>
<td><strong>INTRODUCTION</strong></td>
<td>130</td>
</tr>
<tr>
<td><strong>EXPERIMENTAL PROCEDURES</strong></td>
<td>134</td>
</tr>
<tr>
<td>Mutagenesis at residue 603 and 105</td>
<td>134</td>
</tr>
<tr>
<td>Preparation of and co-transformation into competent C321.ΔA.exp “amberless” cells</td>
<td>135</td>
</tr>
<tr>
<td>Enzyme preparation and labelling</td>
<td>135</td>
</tr>
<tr>
<td>Mass spectrometry</td>
<td>135</td>
</tr>
<tr>
<td>SDS-PAGE analysis</td>
<td>135</td>
</tr>
<tr>
<td>Thermal denaturation and Static light scattering (SLS) measurements using the UNit</td>
<td>136</td>
</tr>
<tr>
<td>Global thermal denaturation measurements using the Fluoromax-4 via intrinsic fluorescence (350/330)</td>
<td>136</td>
</tr>
<tr>
<td>Local thermal denaturation measurements using the Fluoromax-4 via the AF488-594 FRET system</td>
<td>136</td>
</tr>
<tr>
<td>Local thermal denaturation measurements of TK in cell lysate using a Fluoromax-4 via the AF488-594 FRET system</td>
<td>137</td>
</tr>
<tr>
<td><strong>RESULTS</strong></td>
<td>137</td>
</tr>
<tr>
<td>Site-specific incorporation and labelling of para-azidophenylalanine (pAzF)</td>
<td>137</td>
</tr>
<tr>
<td>Incorporation and labelling of pAzF minimally perturbs global stability</td>
<td>139</td>
</tr>
<tr>
<td>Effect of [TK] on global stability for S385Y/D469T/R520Q, 603-AF488/594 and 105-AF488/594</td>
<td>142</td>
</tr>
<tr>
<td>Utilisation of the AF488/AF594 FRET system to probe the effects of macromolecular crowding on the local thermal stability of purified transketolase</td>
<td>144</td>
</tr>
<tr>
<td>Local denaturation profiles provide information on the drivers of global unfolding and aggregation</td>
<td>149</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>Utilisation of the AF488/AF594 FRET system as a reporter of local</td>
<td>150</td>
</tr>
<tr>
<td>thermal stability in cell-like conditions (cell lysate)</td>
<td></td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>152</td>
</tr>
<tr>
<td>SUMMARY</td>
<td>153</td>
</tr>
<tr>
<td>FINAL SUMMARY, DISCUSSION &amp; FUTURE WORK</td>
<td>154</td>
</tr>
<tr>
<td>FINAL SUMMARY</td>
<td>154</td>
</tr>
<tr>
<td>FINAL DISCUSSION &amp; FUTURE WORK</td>
<td>156</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>158</td>
</tr>
<tr>
<td>APPENDICES</td>
<td>170</td>
</tr>
<tr>
<td>APPENDIX ONE: CPEC schematic diagram</td>
<td>170</td>
</tr>
<tr>
<td>APPENDIX TWO: pCNF as a genetically-encoded fluorescent probe</td>
<td>171</td>
</tr>
<tr>
<td>Introduction</td>
<td>171</td>
</tr>
<tr>
<td>Results</td>
<td>174</td>
</tr>
<tr>
<td>APPENDIX THREE: Trypsin digest and mass spectrometry supplementary</td>
<td>179</td>
</tr>
<tr>
<td>material</td>
<td></td>
</tr>
<tr>
<td>Labelling with dimedone</td>
<td>179</td>
</tr>
<tr>
<td>Labelling with DCP-Bio1</td>
<td>182</td>
</tr>
<tr>
<td>APPENDIX FOUR: Evaluation of the pyruvate-binding parameters of wild-</td>
<td>187</td>
</tr>
<tr>
<td>type and pyruvate-accepting variants</td>
<td></td>
</tr>
<tr>
<td>Introduction</td>
<td>187</td>
</tr>
<tr>
<td>Results</td>
<td>187</td>
</tr>
</tbody>
</table>
## TABLE OF FIGURES

**Figure 1:** A general transketolase-catalysed reaction. .......................................................... 22

**Figure 2:** The crystal structure of *E. coli* holo-transketolase in complex with D-xylulose-5-phosphate (X5P) (2R80) (23). ......................................................................................................................... 23

**Figure 3:** a) Free TPP b) Enzyme-bound TPP (V-conformation). .............................................. 24

**Figure 4:** The transketolase reaction mechanism. ........................................................................ 26

**Figure 5:** The natural substrates and products of transketolase-catalysed reactions. .............. 27

**Figure 6:** Examples of the possible substrates transketolase has been engineered to accept using a substrate-walking approach. ........................................................................................................ 29

**Figure 7:** Schematic representation of a potential reaction catalysed by a transketolase able to accept an aromatic aldehyde in combination with a transaminase. ....................................................... 30

**Figure 8:** Aromatic aldehyde substrates 3-formylbenzaldehyde (3-FBA), 4-formylbenzaldehyde (4-FBA) and 3-hydroxybenzaldehyde (3-HBA), and their dihydroxyketone products. .......... 31

**Figure 9:** Summary of site-specific incorporation of unnatural amino acids. .............................. 47

**Figure 10:** Directed evolution of orthogonal amber suppressor tRNA<sub>CUA</sub>^{Tyr} in *E. coli* using the ‘double sieve’ negative-positive selection process (64). ............................................... 48

**Figure 11:** A general double-sieve positive-negative selection scheme for evolving aaRS variants specific to an ncAA (64). .............................................................................................................. 48

**Figure 12:** Vector maps for pEVOL and pUltra (74). ................................................................. 50

**Figure 13:** Summary of the efficiency (grey) and fidelity (red) of ncAA-incorporation into the active site residue 385, and the protein surface residues 603 and 105, of the transketolase variant S385Y/D469T/R520Q. ................................................................................................................. 54

**Figure 14:** Chemical structures of the five single-ringed aromatic amino acids utilised in this study. ........................................................................................................................................ 65

**Figure 15:** Summary of the efficiency and fidelity of incorporation at residue 385 (A), and the mass spectra of pNTF, pCNF, and pAMF (B), (C) and (D), respectively). ........................................... 66

**Figure 16:** The mass spectra of mutants Y and F. Each mass spectrum (black) is fitted to the sum of multiple Gaussian functions (smaller individual peaks) corresponding to unmodified, single- and double-oxidised forms, as well as over-oxidised forms. ......................................................... 67

**Figure 17:** The 3-HBA apparent specific activity and thermal stability of residue 385-substituted variants pAMF, Y, F, pCNF and pNTF. ................................................................. 71
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 18</td>
<td>Summary of the true catalytic performance (A &amp; B) and acceptor substrate inhibition (C) of the TK$_{high}$ species, and the stability (D) of the ncAA-incorporated mutants $p$AMF, Y, F, $p$CNF and $p$NTF.</td>
</tr>
<tr>
<td>Figure 19</td>
<td>The energy-minimised active site structures of $p$AMF, Y (A; green and cyan), F, $p$CNF and $p$NTF (B; magenta, yellow and wheat).</td>
</tr>
<tr>
<td>Figure 20</td>
<td>Computational docking of 3-HBA into the active site of mutants $p$AMF, Y (A; green and cyan), F (C; magenta), $p$CNF (B; yellow) and $p$NTF (C; wheat).</td>
</tr>
<tr>
<td>Figure 21</td>
<td>The predicted enzyme-substrate interactions of Pocket A (A and C) and Pocket B (B and D) of variants $p$AMF and Y, respectively.</td>
</tr>
<tr>
<td>Figure 22</td>
<td>The predicted enzyme-substrate interactions of Pocket B (A) and the extended region of pocket B (B) of variant $p$CNF.</td>
</tr>
<tr>
<td>Figure 23</td>
<td>The predicted Pocket B enzyme-substrate interactions of variants F and $p$NTF, respectively.</td>
</tr>
<tr>
<td>Figure 24</td>
<td>A schematic diagram of the oligomeric states of <em>E. coli</em> apo-transketolase (apoTK) and holo-transketolase (holoTK).</td>
</tr>
<tr>
<td>Figure 25</td>
<td>Evidence of fluorescence quenching upon cofactor-binding.</td>
</tr>
<tr>
<td>Figure 26</td>
<td>The fluorescence intensity of free TPP in 50 mM Tris-HCl buffer ($\lambda_{ex} = 240$ nm; $\lambda_{em} = 330$ nm), fitted to the correction function described by MacDonald et al. (118).</td>
</tr>
<tr>
<td>Figure 27</td>
<td>Experimental data of 0.05 mg/mL wild-type TK binding to TPP at 0 mM (black), 1 mM (red), 4.5 (blue), 9 mM (magenta) and 18 mM (green) Mg$^{2+}$.</td>
</tr>
<tr>
<td>Figure 28</td>
<td>Experimental data of 0.05 mg/mL wild-type transketolase binding to TPP at A) and B) 0 mM; C) and D) 1 mM; E) and F) 4.5 mM; G) and H) 9 mM; and I) and J) 18 mM Mg$^{2+}$.</td>
</tr>
<tr>
<td>Figure 29</td>
<td>Dependence of TPP-dissociation constants and Hill coefficients, of the high and low affinity binding sites for 0.05 mg/mL TK, at 0 mM, 1 mM, 4.5 mM, 9 mM and 18 mM Mg$^{2+}$.</td>
</tr>
<tr>
<td>Figure 30</td>
<td>AUC interference data and the derived c(s) distribution and sedimentation coefficients.</td>
</tr>
<tr>
<td>Figure 31</td>
<td>Comparison of A) apo-dimer formation as a function of [apoTK] and B) TPP-binding at 0.2 mg/mL.</td>
</tr>
<tr>
<td>Figure 32</td>
<td>The mass spectra of wild-type transketolase in the absence and presence of 0.005% CHP added half-way through an eight hour fermentation, and its impact on %TK$_{high}$, TPP-binding and activity.</td>
</tr>
<tr>
<td>Figure 33</td>
<td>The $2F_oF_c$ (green) and $F_oF_c$ (orange) electron density difference maps around residue Cys157 in subunit A and B of wild-type <em>E. coli</em> TK (1QGD) at 3σ omit level.</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>---------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Figure 34</td>
<td>Evidence for existence and location of sulfenic acid modifications.</td>
</tr>
<tr>
<td>Figure 35</td>
<td>Experimental data of 0.05 mg/mL wild-type transketolase binding to TPP at 9 mM Mg²⁺, purified from a fermentation supplemented with 0.5 mM thiamine.</td>
</tr>
<tr>
<td>Figure 36</td>
<td>A schematic diagram of the Two-Species Model of TK cofactor binding.</td>
</tr>
<tr>
<td>Figure 37</td>
<td>Experimental data of 0.05 mg/mL TK S385Y/D469T/R520Q binding to TPP at 9 mM Mg²⁺, purified from a fermentation supplemented with 0.5 mM thiamine.</td>
</tr>
<tr>
<td>Figure 38</td>
<td>Dependence of TPP-dissociation constants and Hill coefficients, of the high and low affinity binding sites for 0.05 mg/mL TK (S385Y/D469T/R520Q), at 0 mM, 4.5 mM, 9 mM and 18 mM Mg²⁺.</td>
</tr>
<tr>
<td>Figure 39</td>
<td>Evidence of fluorescence quenching upon HPA-binding.</td>
</tr>
<tr>
<td>Figure 40</td>
<td>The fluorescence intensity of free HPA in 50 mM Tris-HCl and 0.3 mM TPP pH 7.0 (λₑₓ = 240 nm; λₑₘ = 330 nm), fitted to the correction function described by MacDonald et al.</td>
</tr>
<tr>
<td>Figure 41</td>
<td>Experimental data of 0.05 mg/mL A) wild-type and B) S385Y/D469T/R520Q transketolase binding to HPA in the presence of 9 mM Mg²⁺ and 0.3 mM TPP in 50 mM Tris-HCl buffer.</td>
</tr>
<tr>
<td>Figure 42</td>
<td>The impact of heat-activation on %TK↑high, TPP-binding and activity.</td>
</tr>
<tr>
<td>Figure 43</td>
<td>Experimental data of 0.05 mg/mL transketolase variant binding to TPP at 9 mM Mg²⁺, after two consecutive rounds of heat activation at 42 °C followed by re-equilibration at 22°C for 30 minutes.</td>
</tr>
<tr>
<td>Figure 44</td>
<td>The Two-Species Model of transketolase regulation, activation and inhibition.</td>
</tr>
<tr>
<td>Figure 45</td>
<td>A schematic diagram summarising site-specific incorporation &amp; labelling (azide-alkyne click chemistry) of the UAA pAzF into transketolase &amp; the detection of unfolding and/or inter-subunit aggregation via an increase in FRET distance.</td>
</tr>
<tr>
<td>Figure 46</td>
<td>The site-specific incorporation and labelling of pAzF at residue 603 and 105.</td>
</tr>
<tr>
<td>Figure 47</td>
<td>Experimental data of A) the normalised intrinsic fluorescence ratio (350/330) and B) SLS data (266 nm) of TK variants S385Y/D469T/R520Q (black), 603unlabelled, (red), 603-AF488/594 (blue), 105unlabelled, (magenta), and 105-AF488/594 (green), measured using the UNit.</td>
</tr>
<tr>
<td>Figure 48</td>
<td>The global (350/330) and local (FRET) denaturation curves of purified TK variants S385Y/D469T/R520Q, 603-AF488/594 and 105-AF488/594, measured using the Fluoromax-4.</td>
</tr>
<tr>
<td>Figure 49</td>
<td>The change in global and local thermal stability of S385Y/D469T/R520Q, 603-AF488/594 and 105-AF488/594 as a function of [holoTK].</td>
</tr>
<tr>
<td>Figure 50</td>
<td>Schematic diagram of the aggregation of two transketolase molecules.</td>
</tr>
</tbody>
</table>
**Figure 51:** The local denaturation (FRET) profiles and change in the normalised FRET magnitude or fractional aggregation of 603-AF488/594 and 105-AF488/594 as a function of [holoTK]. ................................................................. 147

**Figure 52:** The local denaturation (FRET) profiles and change in the normalised FRET magnitude or fractional aggregation of 603-AF488/594 and 105-AF488/594 as a function of [lysate]. 150

**Figure 53:** The change in local thermal stability (FRET) of residue 603 and 105 as a function of [lysate]. ........................................................................................................... 151

**Figure A1:** Absorption spectra of pCNF in different solvents. ......................................................... 172

**Figure A2:** Fluorescence emission spectra of pCNF in different solvents........................................ 173

**Figure A3:** Fluorescence spectra of the apo-form of mutant Y and mutant pCNF at 0.08 mg/mL TK, 9 mM Mg²⁺ and 2.4 mM TPP in 50 mM Tris-HCl and pH 7.0. .................................................. 175

**Figure A4:** The energy-minimised structure (see Chapter Four) of mutant pCNF. ....................... 176

**Figure A5:** Absorption spectra of 0.8 mM TPP in solution at pH 6.5, 6.8, 7.0, 7.2, 7.5 and 7.8... 177

**Figure A6:** Fluorescence spectra of 2.4 mM TPP (green), and the holo-forms of mutant Y (black) and mutant pCNF (red) at 0.08 mg/mL TK, 9 mM Mg²⁺ and 2.4 mM TPP in 50 mM Tris-HCl and pH 7.0. ........................................................................................................... 178

**Figure A7:** Experimental data of 0.05 mg/mL wild-type TK (black), mutant P-P (red) and mutant P-F (blue), 9 mM Mg²⁺ and 0.3 mM TPP binding to 0-80 mM HPA. .............................................. 188

**Figure A8:** Experimental data of 0.05 mg/mL wild-type TK (black), mutant P-P (red) and mutant P-F (blue), 9 mM Mg²⁺ and 0.3 mM TPP binding to 0-100 mM pyruvate. Experimental data-points were fitted to the Hill function. ...................................................................................... 189
### TABLE OF TABLES

**Table 1**: Kinetic parameters of D469T/R520Q and S385Y/D469T/R520Q towards 3-FBA, 4-FBA and 3-HBA and the former’s improvement against the latter. Data from (48,49).......................... 32

**Table 2**: The true ($TK_{\text{high}}$) kinetic parameters of variants pAMF, Y, F, pCNF and pAMF towards 3-HBA, having accounted for both misincorporation and the ratio of $TK_{\text{high}}$: $TK_{\text{low}}$. .................. 70

**Table 3**: Summary of the thermal stability of the correctly incorporated species of wild-type, pAMF, Y, F, pCNF and pNTF, having accounted for misincorporation. .................................................. 74

**Table 4**: Summary of the TPP-binding parameters when fitted to a single Hill function. ............... 92

**Table 5**: Summary of the binding parameters of the high affinity binding sites of $TK_{\text{high}}$. .................. 92

**Table 6**: Summary of the binding parameters of the low affinity binding sites of $TK_{\text{low}}$. ................ 95

**Table 7**: Summary of the %$B_{\text{max(high)}}$, %$TK_{\text{modified}}$ and specific activity of wild-type transketolase in the presence and absence of 0.005% CHP.......................................................... 101

**Table 8**: Summary of the TPP-binding parameters of the high affinity binding sites of the $TK_{\text{high}}$ species of S385Y/D469T/R520Q. .......................................................... 117

**Table 9**: Summary of the TPP-binding parameters of the low affinity binding sites of the $TK_{\text{low}}$ species of S385Y/D469T/R520Q. .......................................................... 117

**Table 10**: Summary of the binding parameters of the high and low affinity binding sites of $TK_{\text{high}}$ and $TK_{\text{low}}$ of wild-type and S385Y/D469T/R520Q transketolase............................................. 120

**Table 11**: The %$B_{\text{max(high)}}^{\text{TPP}}$ and %$B_{\text{max(high)}}^{\text{HPA}}$ and %$TK_{\text{modified}}$ of wild-type transketolase (From Chapter Five) and variant S385Y/D469T/R520Q, determined by TPP-binding, HPA-binding and mass spectrometry data. .......................................................... 122

**Table 12**: Summary of the TPP-dissociation constants of wild-type $TK_{\text{low}}$ and $TK_{\text{high}}$ pre- and post-incubation at 42 °C for 1 h.......................................................... 123

**Table 13**: Summary of the predicted % dimer of $TK_{\text{high}}$-$TK_{\text{high}}$, $TK_{\text{high}}$-$TK_{\text{low}}$ and $TK_{\text{low}}$-$TK_{\text{low}}$, calculated from TPP-binding data after heat activation and HPA-inhibition data. .......... 125

**Table 14**: Summary of the efficiency and fidelity of incorporation, the %$TK_{\text{high}}$ and the labelling efficiency when incorporating at residue 603 or 105.......................................................... 139

**Table 15**: Summary of the thermal stability of mutants S385Y/D469T/R520Q, 603_unlabelled, 105_unlabelled, 603-AF488/594 and 105-AF488/594. .......................................................... 141

**Table 16**: The absorbance of 0.8 mM TPP and 0.2 mg/mL apoTK (S385Y/D469T/R520Q) at 240 nm and 285 nm........................................................................ 177
Table 17: Summary of the kinetic and binding parameters of wild-type, mutant P-P and mutant P-F using HPA as the donor substrate using 0.05 mg/mL TK, 9 mM Mg2+, 0.3 mM TPP and 0-80 mM HPA.......................................................... 188

Table 18: Summary of the kinetic and binding parameters of wild-type, mutant P-P and mutant P-F using HPA as the donor substrate using 0.05 mg/mL TK, 9 mM Mg2+, 0.3 mM TPP and 0-100 mM pyruvate......................................................... 190
Enzymes as biocatalysts

Biocatalysis using enzymes has become a powerful weapon in the synthetic organic chemistry toolbox, most notably because they can catalyse reactions at low temperatures with up to $10^{17}$-fold improvements in reaction rates, and with high stereo-, regio- and chemo-selectivity (1). Biocatalysis also has the advantage of limited to no use of protecting groups, minimised side-reactions, easier separation, eco-friendly processes and mild operating conditions relative to harsh chemical synthesis (2,3). This is especially beneficial in the production of complex, biologically active molecules with multiple stereocentres and functional groups, such as antibiotics.

Although remarkable enzyme diversity exists in nature, with over 2000 classes of enzyme-catalysed chemical reactions currently recognized (4), naturally occurring variants often lack the characteristics required for a specific, human-constructed function. Protein engineering is frequently utilised to develop enzymes with improved or altered properties, for example improved enzyme stability or increased activity towards non-natural substrates. Initially held back by low protein yield, narrow substrate scope, poor stereo-/regio-selectivity and/or poor stability, the field has been revolutionised by recombinant DNA technologies and directed evolution (5). These advances, in combination with whole genome sequencing capabilities and an extraordinary surge in the structural knowledge of many enzyme classes, have facilitated the development of enzymes with capabilities far beyond their natural limits.

Transketolase as a biocatalyst

Transketolase (TK) is ubiquitous in all known organisms and provides a unique link between glycolysis and the non-oxidative phase of the pentose phosphate pathway and Calvin cycle (6). It catalyses the reversible transfer of a two-carbon ketol group from a donor substrate to an aldehyde acceptor substrate, forming a new asymmetric carbon-carbon bond with high regio- and stereo-specificity (7) (Figure 1). Wild-type transketolase has been shown to have a relatively low specificity for the aldehyde substrate, but favours the transfer of the dihydroethyl group from 2R hydroxylated aldehydes to a 2R hydroxylated aldehyde, producing a dihydroxyketone with a 3S,4R or D-threo configuration (8–10). While inherently reversible, the reaction can be rendered irreversible through use of β-hydroxypyruvate (HPA) as the donor substrate, thus increasing biocatalytic product yields.
Although the higher yields achieved make HPA an attractive donor for industrial use, chemical synthesis of HPA is expensive.

![Figure 1: A general transketolase-catalysed reaction.](image)

**Industrial applications of transketolase: Asymmetric carbon-carbon bond elongation**

Asymmetric carbon-carbon bond synthesis is a key tool in organic chemistry because it enables the production of a carbon backbone to which desired functional groups can be added. The synthesis of α,α'-dihydroxyketone products provide a versatile backbone as a precursor to ketosugars, chiral aminodiols and other high-value molecules such as fragrances and flavours (11–14), while the stereo-selectivity of transketolase is highly appealing when attempting C-C bond formation in order to form more complex biologics with multiple stereo-centres, such as antibiotics. Chemical synthesis of asymmetric carbon-carbon bonds via dihydroxyketones has several limitations; they are expensive multistep reactions, have poor atom efficiency (12), utilise poor catalysts that fail to stabilise the hydroxyl carbanion intermediate sufficiently (12,15) and usually require the use of protective groups. Chemo-synthesis is especially inadequate when attempting to synthesise complex, multifunctional target structures where chirality is paramount, such as antibiotics. As a consequence, enzymes capable of catalysing asymmetric carbon-carbon bond elongation reactions with high stereoselectivity such as transketolase have received plenty of attention from industry and academia alike.

A major research and industrial output has been the development of transketolase variants, through rational design, semi-rational design and directed evolution, that can synthesise industrially-relevant biologics or their precursors. A key goal of industry is to create a “biocatalytic asymmetric C-C bond toolbox” consisting of many transketolase variants that have, collectively, broad substrate specificity and, as such, can be utilised in a wide range of industrial processes.

**Structure and function of wild-type transketolase**

Originally sourced from *Saccharomyces cerevisiae* (7) and spinach (16) at low yields, endogenous transketolase was subsequently expressed in *Escherichia coli* (*E. coli*) (E.C. 2.2.1.1) in
much greater yields (9). For this reason, recent work has mostly been carried out on the *E. coli* and *S. cerevisiae* enzymes. In most organisms and species, including *E. coli* and *S. cerevisiae*, active holo-transketolase exists as homodimers, while spinach and *Candida boidinii* holo-transketolases exist as a monomer and tetramer, respectively (17).

Transketolase (E.C. 2.2.1.1) is a member of the thiamine pyrophosphate (TPP)-dependent enzyme family. All TPP-dependent enzymes have two highly conserved domains - the pyrophosphate (PP)-binding domain (2-322 aa in *E. coli* TK) and the pyrimidine (Pyr)-binding domain (323-539 aa in *E. coli* TK), whose interface forms the TPP-binding sites (Figure 2). The third, C-terminal domain is less conserved and has an unknown biological function, but has been proposed as a regulatory binding site (18). Multiple sequence alignment of the conserved PP and Pyr domains highlighted a number of highly conserved residues critical in metal binding, TPP binding and TPP activation (19). More specifically, a common TPP-binding motif was also identified (GDGX26NN or GDGX26NCN) (20) that forms an αβα fold to facilitate the binding of divalent cations (21,22) which in turn bind to the pyrophosphate group of the TPP molecule.

**Figure 2**: The crystal structure of *E. coli* holo-transketolase in complex with D-xylose-5-phosphate (X5P) (2R80) (23). A) The quaternary structure of the dimeric holo-transketolase. The domains of subunit A and B are shown as the following, respectively: PP-binding domain (bright and pale red); Pyr-binding domain (bright and pale green); C-terminal domain (bright and pale blue). Calcium ions and the TPP-X5P intermediate complex are shown as magenta spheres and purple sticks, respectively. B) The TPP-X5P intermediate complex within a dimeric holoTK active site. The two, more ordered cofactors loops (N185-W196 from subunit A and L382-G392 from subunit B) are shown as sticks.

**Structural changes to apo-transketolase and TPP upon cofactor binding**

TPP is a derivative of vitamin B1 and consists of a pyrimidine ring, a thiazolium ring and a diphosphate group. In all TPP-dependent enzyme reactions, TPP is directly involved in catalysis
without being released from the active site (24). Although the free coenzyme can aid TPP-dependent reactions, the enzyme-bound cofactor can increase the overall reaction rate considerably - by as much as $10^{12}$ in the case of pyruvate decarboxylase (25). The protein environment that surrounds the cofactor plays a crucial role in enhancing the catalytic activity of the cofactor.

Crystal structures of both *E. coli* and *S. cerevisiae* transketolase showed two TPP binding pockets located at the subunit interface with TPP almost completely buried inside the protein (23,26). The V-conformation of enzyme-bound TPP (Figure 3B) positioned the catalytically important C2 atom of the thiazolium ring closer to the 4'-NH$_2$ group of the pyrimidine ring when compared to the crystal structure of free TPP (Figure 3A) (27). At the active site entrance, there is a narrow, funnel-shaped channel just wide enough for sugar substrate entry, which leads to the C2 atom of the thiazolium ring – the only part of the cofactor accessible to solvent (23,26). Until recently, it was thought that the positioning of the 4'-NH$_2$ of the pyrimidine ring was critical in the mechanism for cofactor activation, but this is now believed to be incorrect (28).

Although no crystal structure of *E. coli* apo-transketolase has been resolved, the crystal structure of the holo-form has been determined to 1.9Å and identified two cofactor-binding loops (N185-W196 and L382-G392) that stabilise the cofactor (29) (Figure 2B). Comparative structural studies between the apo- (30) and holo- (26) forms of yeast transketolase indicated a shift from disordered to ordered in the cofactor loops upon cofactor-binding. The high degree of sequence and structure conservation between *S. cerevisiae* and *E. coli* transketolase strongly implies the same structural change occurs in both orthologues.

Transketolase cofactor-binding has been studied extensively over the years, yet certain mysteries remain, such as a lack of consensus on the cooperativity of thiamine pyrophosphate (TPP) binding into the two active sites, in the presence and absence of the divalent cation, Mg$^{2+}$, the unusual phenomenon of increased transketolase activity after heat shock, and the conflicting structural and thermodynamic/kinetic evidence for non-equivalence between active sites. The current literature on
transketolase cofactor-binding is expanded upon in Chapter Five and Six, where novel data is presented that explains, at least in part, these unsolved phenomena.

**The fluorescence intensity characteristics of E. coli transketolase**

Little direct literature is available on the fluorescence spectrum of *E. coli* transketolase. The reconstitution of holoTK from apoTK quenched the intrinsic fluorescence of transketolase when excited at 280 nm (31). A number of potentially fluorescence quenching interactions were suggested, including Tyr182 and Mg$^{2+}$ (6.4 Å); Tyr182 and the TPP phosphate (6.8 Å); Tyr440 and the methyl group of the TPP pyridine ring (3.4 Å); and between Trp196 and the TPP phosphate (7.6 Å), TPP thiazolium ring (8.5 Å), and the Mg$^{2+}$ ion (8.4 Å). It is also possible that TPP-binding quenched fluorescence of Trp196 and Trp390 on the cofactor-binding loops as loops became structured around each TPP molecule. The structuring of these loops positioned the two residues in a more polar environment near the side-chains of Asn403 and Glu366, which likely quenched their fluorescence (31).

**The catalytic mechanism of transketolase**

The transketolase reaction occurs via a ping-pong reaction mechanism that can be divided into two stages (32,33) (Figure 4). First, enzyme-bound TPP is activated via deprotonation of the reactive C2 atom, forming a carbanion, which then acts as the nucleophile and attacks the carbonyl group of the donor substrate. A recent study, using a quantum mechanical/molecular mechanical method based on crystallographic structures of yeast and *E. coli* transketolase, together with experimental kinetic data reported in the literature with wild-type and mutant transketolase, contradicted the more-widely-accepted method of proton extraction (28). Previously it was thought that the highly conserved glutamate residue extracts a proton from the N1’ atom of the pyrimidine ring, which in turn extracts a proton from the 4’NH$_2$ group of the pyrimidine ring. The activated 4’-NH$_2$ is potentially an efficient proton-acceptor of the C2 proton because of their close proximity when the TPP is bound to transketolase in the V-conformation (34,35). However, recent evidence suggested His481B deprotonates the C2 atom, while the 4’-NH$_2$ group changes from an amino- to an imino-group that subsequently stabilises the carbanion (28). In addition to the conserved residues involved in TPP-activation and acid/base catalysis, a number of highly conserved residues located in and around the active site are important in substrate recognition, stereoselectivity, transition-state stabilisation, TPP binding, divalent cation binding, phosphate binding or have multiple functions through complex H-bond networks.
Figure 4: The transketolase reaction mechanism. The acid/base catalysts are denoted as B1, B2 and B3; four highly conserved histidine residues (H26, H69 and H261 on the PP domain; H473 on the Pyr domain) have been identified as playing a role in the proton transfer events (23,33).

**Substrate specificity, stereospecificity & stereoselectivity of transketolase**

Physiologically, transketolase facilitates the interconversion between the 3C-, 5C- and 7C-phosphorylated sugar intermediates of the pentose phosphate pathway (Figure 5). However, cross-species analysis of transketolase substrate specificity has shown that, in general, transketolases can also accept a wide variety of substrates, including non-phosphorylated sugars (11), non-hydroxylated aldehydes, cyclic aldehydes (9,36) and heteroaromatic substrates (37). Generally, of the non-phosphorylated substrates it can accept, transketolase has the greatest activity towards those with (R)-α-hydroxyaldehyde configuration (10). Hydroxylated substrates are more readily accepted because highly conserved histidine and aspartate residues directly interact with the hydroxyl group(s) of acceptor substrates. Transketolase variants have already been utilised for the biosynthetic production of desirable products, for example spinach transketolase has been used to biosynthesise 6-deoxy-l-sorbose, a precursor to the important food flavour furaneol (14). Industry has ambitions to widen the
The scope of transketolase substrate specificity further to create a “biocatalytic asymmetric C-C bond toolbox” of variants that can be utilised in a myriad of industrial processes. Over the years, a detailed – but still incomplete – map of transketolase structure-function relationships has been built up, to facilitate the development of transketolase variants with wide-ranging substrate specificity and stereospecificity.

**Figure 5:** The natural substrates and products of transketolase-catalysed reactions. The reversible transfer of a 2-carbon ketol group from a donor substrate to an acceptor aldehyde substrate in A) the non-oxidative pentose phosphate pathway (PPP); and B) the Calvin cycle and non-oxidative PPP.

Transketolase interacts with various non-phosphorylated substrates through interactions between the conserved, positively charged residues (H26, H66, H100 and H261) and the substrate hydroxyl moieties. Additional contacts between phosphorylated substrates and phosphate-binding residues (R358, H461, R520 and S385) generally increase transketolase affinity, and hence decrease the Michaelis-Menten constant, $K_m$, for phosphorylated substrates. The crystal structure of an *E. coli* TK-ribose-5-phosphate complex (2R5N) suggested H26 bound the C3 hydroxyl group of the donor substrate via H-bonding (23) and is also important in stereoselectivity (38). Crystal structures of
transketolase (2R8P, 2R8O) suggested H66 is located close to the phosphate group of TPP; the two imidazole-ring nitrogen atoms formed a H-bonding network, and interacted directly with TPP and indirectly with the C1- hydroxyl group of the donor-TPP complex via a water molecule. H66 was also bound to the DHE-TPP intermediate in yeast TK (1GPU) but not acceptor substrates in both E. coli (2R5N) and yeast (1NGS). The structural evidence was supported by kinetic data, which showed a significant decrease in affinity, or increase in $K_m$, towards the donor substrate and a >95% decrease in specific activity of H66 variants relative to wild-type. The kinetic importance of H66 suggested it may be important in maintaining the appropriate orientation of the donor-TPP and DHE-TPP intermediates. The donor-TPP intermediate is also stabilised by H100 via H-bonding to the C1-hydroxyl group (2R8O & 2R8P).

The additional interactions between phosphorylated substrates and phosphate-binding residues R358, S385, H461 and R520 decreased the $K_m$ values for phosphorylated substrates (0.1 – 7 mM) 1000-fold compared to non-phosphorylated substrates (39). The substantial reduction in substrate affinity also impacted substrate conversion rates. Targeted mutagenesis at these residues shifted the substrate specificity, and hence improved activity, further towards non-phosphorylated substrates, such as glycolaldehyde. S385 is situated within one of the two co-factor binding loops located in each active site (N185-W196 and L382-G392 in E. coli), and forms a H-bond with the substrate phosphate group via its side chain hydroxyl group.

The high stereospecificity and stereoselectivity of E. coli transketolase is dictated by a H-bond network between H26, H261 and the substrate, as well as important interactions between D469 and the C3 and C4 hydroxyl group of the donor and C2 hydroxyl group of the acceptor substrate (23,32,42). Mutations at these critical residues significantly altered substrate specificity; for example, replacement of histidine by tyrosine at residue 26 strongly reversed the stereoselectivity from 3S to 3R using propionaldehyde as an acceptor substrate (38).

**Directed evolution of transketolase to alter substrate specificity**

The Dalby lab has spent a number of years evolving E. coli transketolase to accept a number of different substrates through a substrate-walking directed evolution approach away from its natural substrates, phosphorylated ketose sugars. Initially saturation mutagenesis libraries of different active site residues were designed using structural and phylogenetic information tailored to the genre of the target substrate. Through various mutagenesis libraries, single-mutant transketolases were identified with improved activity towards polar non-phosphorylated substrates such as glycolaldehyde (40),
aliphatic non-phosphorylated substrates such as propionaldehyde (43,44) and heteroaromatic substrates (45) (Figure 6). However, all attempts to combine single mutations into double-mutants were unsuccessful and resulted in the formation of insoluble aggregates. Statistical coupling analysis of a transketolase multiple protein sequence alignment subsequently identified a coevolved network of 9 residues within the active site that affected all but one of the double-mutants created, and R520Q was found to stabilise the previous D469T mutation to give the most active mutant screened for to date (46). This opened up the potential to engineer triple mutants with further enhanced activity.

![Chemical structures](image)

**Figure 6:** Examples of the possible substrates transketolase has been engineered to accept using a substrate-walking approach.

**Improved activity towards aromatic aldehyde substrates**

Identification of the stabilising R520Q mutation, in combination with D469T, created a double-mutant that was used as the base variant for directed evolution towards accepting aromatic aldehydes. An aromatic aldehyde-accepting transketolase variant, in combination with an aromatic dihydroxyketone-accepting transaminase variant, has the potential to synthesise complex molecules, such as precursors to the antibiotic chloramphenicol (Figure 7), and is therefore of interest to industry (47,48). Wild-type *E. coli* transketolase has poor specificity, low product yields and often produces
by-products when accepting aromatic aldehydes (36). Therefore, the substrate specificity of the double-mutant was evolved towards accepting aromatic aldehydes. Nevertheless, the extensive through saturation mutagenesis at functionally important residues (48,49).

The stabilisation of the D469T mutation by R520Q facilitated a further round of saturation mutagenesis with libraries created for the active site residues R358 and S385. Three benzaldehyde derivatives, 3-formylbenzoic acid (3-FBA), 4-formylbenzoic acid (4-FBA) and 3-hydroxybenzaldehyde (3-HBA), were chosen as model aromatic substrates to test the new active site features (Figure 8). The rationale behind the use of these substrates in particular was to introduce moieties into the substrates that would mimic the electrostatic and H-bonding interactions that were removed when moving away from natural phosphorylated substrates (48,49).

Figure 7: Schematic representation of a potential reaction catalysed by a transketolase able to accept an aromatic aldehyde in combination with a transaminase. Chloramphenicol is shown to illustrate the potential implications of engineering such a pathway.
Figure 8: Aromatic substrates 3-formylbenzaldehyde (3-FBA), 4-formylbenzaldehyde (4-FBA) and 3-hydroxybenzaldehyde (3-HBA), and their dihydroxyketone products.

A library of R358X/D469T/R520Q mutants were screened for activity towards the three substrates, but none had improved activity relative to the double-mutant because of the disruption of stabilising interactions between R358 and the substrate-carboxylate moiety (49). Conversely, analysis of the S385X/D469T/R520Q library identified mutants with improved activity towards the substrates, in particular S385Y (Table 1), S385T and S385E. The mechanisms of improvement in the activity of S385Y/D469T/R520Q were divergent for 3-FBA and 4-FBA; 3-HBA was incomparable as previous mutants were essentially inactive towards it. Kinetic improvements towards 3-FBA were driven by a 10-fold improvement in $K_m$ with only a modest 16% increase in catalytic turnover, $k_{cat}$, while 4-FBA was mainly from an 8.5-fold improvement in $k_{cat}$ rather than a 40% improvement in $K_m$. It was hypothesised that the divergent mechanisms arose from altered binding modes of each substrate, with each substrate orientated differently relative to the C2 atom of the thiazolium ring of TPP (48).
Table 1: Kinetic parameters of D469T/R520Q and S385Y/D469T/R520Q towards 3-FBA, 4-FBA and 3-HBA and the former’s improvement against the latter. Data from (48,49).

Recently, the X-ray crystal structure of the S385Y/D469T/R520Q triple-mutant was solved to a resolution of 1.5 Å and substrate binding of the three aromatic aldehydes was predicted using in silico computational docking (50). Introduction of the three mutations had created an evolutionary intermediate with two distinct binding pockets that were sterically separated by T469; 3-FBA bound one, 4-FBA the other, and the least well accepted substrate, 3-HBA, had equal preference for each binding pocket. Mutation to D469T replaced the salt bridge to R91 with an H-bond, which liberated the side-chain of R91 to form H-bonds with substrates bound in the 3-FBA binding pocket. The predicted change in enzyme-substrate interactions supported the large kinetic improvements in affinity towards 3-FBA. Q520 widened the opening of the active site, compensating for the steric hindrance introduced by Y385, while Y385 itself, together with F434, formed two enclosed hydrophobic binding pockets through π–π stacking interactions with each other and the substrates in close proximity to the C2 atom of TPP. Y385 was held in position by a H-bond between the side-chain hydroxyl group and the carbonyl backbone of G262 via a network of water molecules.
**Thermal stability of transketolase**

*E. coli* is a mesophile that has evolved to function optimally at moderate temperatures and pH. Consequently, wild-type *E. coli* transketolase has low stability at elevated temperatures and at high/low pH (51). Unusually, it has a broad optimum activity between 20-55 °C, largely due to an as-yet unexplained phenomenon of increased activity after heat exposure for 1 hr and subsequent re-equilibration at 25 °C. It then rapidly loses activity above 55 °C due to irreversible aggregation.

Thermostable protein variants offer significant advantages to industrial processes, such as enhanced reaction rates (reaction rates roughly double for every 10 °C increase in temperature), increased reactant solubility and reduced risk of microbial contamination. Thermostable transketolase variants would be particularly beneficial in the bioconversion of aromatic substrates, which have low solubility in water-based buffers. Furthermore, directed evolution of enzyme activity often comes at a price; typically, mutations that improve activity can also negatively impact enzyme stability, therefore activity and stability are, ideally, co-evolved. For example, it was necessary to stabilise mutant D469T by mutation at R520Q, to form the double-mutant D469T/R520Q (46), before activity towards aromatic substrates could be significantly improved by mutation at S385Y, to form the triple-mutant S385Y/D469T/R520Q (48,49). Despite stabilisation by R520Q, S385Y/D469T/R520Q still has reduced stability relative to wild-type transketolase (52).

Recent efforts have identified other stabilising mutations in transketolase. For example, a protein sequence comparison between the cofactor-binding loops in *E. coli* and the thermophile *Thermus thermophiles* identified H192P as another stabilising mutation that both increased the optimal temperature for activity from 55 °C to 60 °C and increased the thermal aggregation transition mid-point, $T_{agg}$, measured by dynamic light scattering (DLS), from 60 °C to 61.5 °C (53). Secondly, additional thermostable variants were identified by rigidifying flexible regions of transketolase (54). Flexible sites were first located using two approaches; a “back to consensus mutations” strategy, and an *in silico* methodology based on $\Delta \Delta G$ calculations in Rosetta. Mutagenesis within these flexible regions yielded three single-variants with improved stability, I189H, A282P and D143K. Finally, A282P was combined with H192P to give the best all-round variant with a 5 °C increase in thermal transition mid-point, $T_m$, relative to wild-type.

While these targeted mutagenesis approaches were highly successful in generating thermostable mutants, each required homogeneous (i.e. purified) samples which limited the experimental throughput and hence sequence space that could be explored. This thesis will explore...
and develop novel techniques that are potentially compatible with high- and ultra-high-throughput directed evolution strategies consisting of iterative cycles of random mutagenesis and high-throughput screening to identify the most-thermostable variants.

**Improving enzyme kinetics and protein stability**

Improving the catalytic proficiency of enzymes can add significant value to industrial processes. However, how does one determine quantitatively which enzyme of a library of variants is the ‘best’ biocatalyst? Many different factors contribute to the catalytic performance of an enzyme, the most obvious being enzyme kinetics. Protein stability is also as important; a 10 °C increase in reaction temperature doubles reaction rates, and hence decreases the amount of enzyme required. In addition, high temperatures can reduce reactant viscosity, increase reactant solubility, increase the enzyme half-life and inhibit microbial contamination (if above ~60 °C). Furthermore, stability must often be co-evolved to stabilise the destabilising mutations that improve activity. Michaelis-Menten kinetics is widely accepted as the best model for enzyme kinetics. Protein engineers aim to improve several kinetic parameters, with the main emphasis on enhancing $k_{cat}$ (catalytic turnover), the Michaelis constant, $K_m$ (the substrate concentration at which the reaction rate is at half-maximum; an inverse measure of the substrate's affinity for the enzyme) and product inhibition constants. These in turn allow quantification and comparison of enzyme characteristics, such as specific activity, catalytic efficiency ($k_{cat}/K_m$), the degree of substrate and/or product inhibition, and protein stability (55).
RESEARCH AIMS

The overall aim of this research is to explore non-canonical amino acid (ncAA)-incorporation into transketolase to exploit their site-specific, bio-orthogonal properties to either improve or measure transketolase activity, stability and cofactor-/donor substrate-binding. This work will primarily utilise or develop intrinsic/extrinsic fluorescence-based assays under four broad research themes, over five chapters.

The first chapter investigates the site-specific incorporation of ring-substituted phenylalanine derivatives at the *para* position at three different transketolase positions. This research will establish a novel methodology to quantify the fidelity of incorporation by fitting mass spectra to the sum of multiple Gaussian functions. The trends in the efficiency and fidelity of incorporation with respect to the substrate specificity of the tRNA/synthetase pair will be determined and discussed.

The second chapter will utilise site-specific ncAA-incorporation into the active site of S385Y/D469T/R520Q to probe the effect of side-chain structure and electronics on catalytic activity, acceptor substrate inhibition, enzyme stability and binding pocket preference.

The third and fourth chapters will develop a novel, fluorescent ncAA (*p*CNF) or intrinsic fluorescence-quenching-based cofactor- and donor substrate-binding assay to study: the affinity and cooperativity of TPP- and HPA binding; heat-activation of transketolase; and donor-substrate inhibition.

The final chapter will aim to develop a novel, extrinsic Förster Resonance Energy Transfer (FRET)-based stability assay for use in biophysical studies into local unfolding and structural dynamics of transketolase in cellular environments. The technique involves incorporating ncAAs into the protein that can then be labelled site-specifically with the fluorescent donor and acceptor Alexafluor dyes for FRET. The assay will be utilised to investigate the effect of macromolecular crowding on the local stability in two different domains of transketolase.
CHAPTER TWO: General experimental procedures

Chapter Two outlines all of the experimental procedures common to each results chapter to avoid repetition. The experimental procedures outlined in each results chapter will contain methods specific to that chapter.

Media & solutions

Chemicals & reagents

Tris-HCl and PBS tablets were purchased from VWR International (Lutterworth, UK) and Guanidine-HCl was purchased from Life Technologies Ltd. HPA was synthesised by reacting bromopyruvic acid with LiOH, as described previously (36). Para-cyanophenylalanine (pCNF) was purchased from Bachem (California, USA). AlexaFluor-488 (AF488) and AlexaFluor-594 (AF594) dyes were purchased from ThermoFisher Scientific (Loughborough, UK). All other chemical reagents were purchased from Sigma–Aldrich (Poole, UK).

Ampicillin (Amp) stock solution

Ampicillin powder was dissolved in RO water to a final concentration of 150 mg/ml and sterilised by filtration through a 0.2 µm filter. The stock solution was aliquoted into 1.5 mL sterile Eppendorf tubes and stored at -20 °C. The working concentration of ampicillin was 150 µg/mL and added to LB and LB agar to select for bacteria carrying the plasmid pQR791.

Spectinomycin (Spectin) stock solution

Spectinomycin stock solution was prepared and stored as described for ampicillin stocks but with a stock concentration of 100 mg/mL and a working concentration of 100 µg/mL. Spectinomycin was added to LB and LB agar to select for bacteria carrying the plasmid pUltra.

Lysogenic broth (LB) medium

LB broth was prepared by dissolving 5 g yeast extract, 10 g tryptone, and 10 g NaCl in RO water to a final concentration of 1 L, pH-adjusted to 7.0, and then autoclaved for sterilisation.
**Lysogenic broth (LB) agar**

LB agar was prepared as described for LB medium with the addition of select agar to a final concentration of 15 g/L. Following autoclaving, the solution was cooled to approximately 50 °C. 20 µl of ampicillin and/or spectinomycin 1000x stock solutions were added to 20 mL LB-agar and poured into a petri dish to set. LB-Amp-Spectin plates can be stored at 4 °C for a month.

**Tris-HCl buffer**

50 mM Tris-HCl buffer was prepared by dissolving 7.88 g of Tris-HCl powder in 1 L MilliQ water, pH-adjusted to 7.0 and sterilised by filtration through a 0.2 µm filter.

**100x ncAA stock solution**

ncAA stock solution was prepared by dissolving ncAA powder in 10 mL water to a final concentration of 100 mM. The pH was then adjusted to 7.0 and the stock solution aliquoted and stored at -20 °C.

**1000x IPTG stock solution**

IPTG stock solution was prepared by dissolving 2.383 g IPTG powder in 10 mL to a final concentration of 1 M, pH-adjusted to 7.0, aliquoted and finally stored at -20 °C.

**Standard transketolase 10x cofactor stock solution**

Standard transketolase cofactor stock solution was prepared by dissolving MgCl₂ and Thiamine pyrophosphate (TPP) in 50 mM Tris-HCl buffer to final concentrations of 90 mM and 24 mM, respectively. The pH was adjusted to 7.0 and the solution aliquoted and stored at -20 °C.

**Standard molecular biology techniques**

**Glycerol stock preparation and maintenance**

25% (v/v) *E. coli* glycerol stocks were prepared by adding filter-sterilised 50% (v/v) glycerol to 0.5 mL of a 10 mL overnight culture in a 1:1 volume ratio and stored at -80 °C.

**Plasmid extraction, quantification & storage**

5 mL of overnight culture was transferred into a 50 mL falcon tube and centrifuged for 15 minutes at 5000 rpm at 4 °C. The supernatant was discarded, leaving a cell pellet. Plasmid DNA was
then extracted using a commercial plasmid mini-preparation kit (Qiagen Ltd, UK) according to the manufacturer’s protocol. The final product was eluted into 50 mL elution buffer. Plasmid DNA concentration was quantified using a Nanodrop ND-1000 spectrophotometer (Fisher Scientific, UK) and stored at -20 °C.

**Transformation of commercial ultra-competent cells by heat-shock**

Commercial super-competent *E. coli* strains were transformed as described by the provided company protocols. All tubes and solutions were thawed and kept on ice except S.O.C media (Invitrogen, UK), which was pre-heated to 42 °C.

**Quikchange Mutagenesis**

Primer design

Site-directed mutagenesis was performed on the transketolase triple mutant (S385Y/D469T/R520Q) to generate two mutants with an amber stop codon (TAG) at either the Y385, Y105 or K603 positions. Primers were designed to have the following properties: a) 24-33 base pairs long b) 40-60 % GC content c) *T*<sub>m</sub> 55-80 °C (Target of 60 °C) d) 2-3 GCs in the last 5 bases at the 3’ end d) the desired mutation was located in the middle of the primer with 10-15 base pairs either side e) No secondary structure. Primers were ordered Eurofins MWG Operon (Ebersberg, Germany) in lyophilised form, salt free grade. Primers were centrifuged and subsequently re-suspended in Milli-Q® water to a final concentration of 100 μM and stored at -20 °C.

Polymerase Chain Reaction (PCR) conditions

125 ng of each primer was required for each reaction tube and was calculated using the following formula:

\[
x \text{ pmol oligo} = \frac{\text{ng oligo}}{330 \times \# \text{ oligo bases}} \times 1000
\]

The sample reaction was prepared as indicated below:

5 μL of 10x reaction buffer

X μL (5–50 ng) of dsDNA template

X μL (125 ng) of oligonucleotide primer #1

X μL (125 ng) of oligonucleotide primer #2

1 μL of dNTP mix ddH2O to a final volume of 50 μL
Then add:

1 μL of *PfuUltra* HF DNA polymerase (2.5 U/μL)

Each PCR tube was transferred to the thermal cycler (Techgene, version 13, block 20 x 0.5 ml) using the following cycling parameters:

<table>
<thead>
<tr>
<th>Segment</th>
<th># cycle</th>
<th>Step</th>
<th>Temperature</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>Initial denaturation</td>
<td>95 °C</td>
<td>30 s</td>
</tr>
<tr>
<td>2</td>
<td>12 (single point mutation)</td>
<td>Denaturation</td>
<td>95 °C</td>
<td>30 s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Annealing</td>
<td>55 °C</td>
<td>1 min</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Elongation</td>
<td>68 °C</td>
<td>5.4 mins</td>
</tr>
<tr>
<td>3</td>
<td>Hold</td>
<td></td>
<td>4 °C</td>
<td>...</td>
</tr>
</tbody>
</table>

After successful PCR amplification, 1 μL of Dpn1 restriction enzyme was added directly to each reaction and incubated at 37 °C for 1 hour to digest the template DNA.

**Transformation**

XL10-gold super-competent cells were thawed on ice. For each reaction 50 μL was transferred to a pre-chilled 14 mL BD Falcon polypropylene round-bottom tube. 1 μL Dpn1-treated DNA was then added to separate aliquots of super-competent cells, gently swirled to mix and incubated on ice for 30 minutes.

The water bath was pre-heated to 42 °C and the transformation reactions heat-shocked for 45 seconds then placed on ice for 2 minutes. 0.5 ml of S.O.C media preheated to 42 °C was added and incubated at 37 °C and agitated at 250 rpm for 1 hour. 100-250 μL of each transformation reaction was spread on LB agar (Amp+) plates in order to generate single colonies. The culture was then incubated at 37 °C for 18 hours. At least 6 transformed colonies were picked to prepare glycerol stocks. Mutated DNA plasmids were extracted and the DNA sequenced using the service of Source
Biosciences to confirm successful mutagenesis. Samples were prepared according to the service guidelines. DNA sequences were compared to original and expected sequences and stored electronically.

**Protein expression & purification**

*Generation of single bacterial colonies on LB-agar*

Glycerol stocks of *E. coli* strains were streaked out onto Petri dishes containing LB agar (Amp+ or Amp+/Spectin+) using a sterile inoculation loop in order to generate single colonies. The plates were incubated at 37 °C overnight and subsequently stored at 4 °C for up to a week.

*Overnight starter culture*

A single *E. coli* colony was picked from the agar plate using a sterile 200 µL pipette tip and transferred into a 50 mL falcon tube containing 10 mL LB medium (Amp+ Spectin+). The tube was incubated for 16-18 hours at 37 °C and agitated at 250 rpm.

*50 mL shake flask culture (in 250 mL flask)*

5 mL of starter culture was added to 45 mL LB medium (Amp+ Spectin+) in a 250 mL baffled shake flask, with ncAA stock solution added to a final concentration of 1 mM, if appropriate. The shake flask was incubated at 37 °C and agitated at 250 rpm until the culture reached an OD600 of 0.8-0.95. At this point IPTG stock solution was added to a final concentration of 1 mM, if appropriate. The culture was then incubated at 37 °C and agitated at 250 rpm for 7 hours (or until OD600 ≈ 3-4).

*Cell pellet extraction & storage*

The 50 mL culture was transferred to a 50 mL Falcon tube and centrifuged at 5,000 rpm for 15 minutes at 4 °C. The supernatant was discarded, leaving a cell pellet which was stored at -20 °C for at least 30 minutes.

*Preparation of clarified lysate*

The frozen cell pellet was thawed on ice and resuspended in 5 mL 50 mM Tris-HCl buffer, pH 7.0. The cells were lysed using sonication (MSE Soniprep 150 probe, Sanyo) on ice using 10 cycles of 10 seconds on 10 seconds off. Clarified lysate containing transketolase was separated from the insoluble fraction/cell debris by centrifugation at 18,000 g for 10 minutes at 4 °C. The clarified lysate (supernatant) was used fresh or stored at -20 °C in 750 mL aliquots for up to 1 month.
**Transketolase His-Tag purification**

Transketolase samples were purified using Ni-NTA spin columns (Qiagen Ltd., UK) according to the company protocol, leaving 600 µL of purified transketolase in elution buffer. The lysis, wash and elution buffers consist of the following:

<table>
<thead>
<tr>
<th>Buffers</th>
<th>Lysis (500 mL) pH 8.0</th>
<th>Wash (500 mL) pH 8.0</th>
<th>Elution (100 ml) pH 7.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>500 mM NaCl (14.61 g)</td>
<td>500 mM NaCl (2.92 g)</td>
<td>20 mM Tris-HCl (0.315 g)</td>
<td></td>
</tr>
<tr>
<td>20 mM Tris-HCl (1.576 g)</td>
<td>20 mM Imidazole (0.68 g)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 mM Imidazole (0.17 g)</td>
<td>1 M Imidazole (6.808 g)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Protein dialysis**

Protein dialysis was performed after purification to remove the imidazole and to generate the fully apo-form of transketolase variants for fluorescence experiments. The Slide-A-Lyzer dialysis cassette was immersed in 50 mM Tris-HCl for 2 minutes and then removed before carefully blotting the outer edges to remove excess liquid with care taken not to blot the membrane. A hypodermic needle was attached to a 2 mL syringe and air was sucked up equivalent to the volume of the sample. The syringe was then filled with the sample and the gasket was penetrated through a syringe port and the sample injected. Once injected, excess air was removed to compress the membrane windows to maximise surface area. The cassette was slipped into the groove of a buoy and dialysed in Tris-HCl in a volume 400 times the sample volume and dialysed for 2 hours at 4 °C. The dialysis buffer was then replaced and the sample dialysed for another 2 hours. The buffer was replaced again and left overnight. The next morning the buffer was replaced twice again with 2 hour periods of incubation. The sample was then extracted by piercing the gasket through a different syringe port and transferred to an eppendorf tube.

**Protein ultra-filtration & concentration**

To simultaneously concentrate transketolase for use in experiments and remove salts and imidazole, 600 µL of purified transketolase was transferred to a 4 mL 10,000 NMWL Amicon® Ultra-4 centrifugal filter unit and made up to 4 mL with 50 mM Tris-HCl. This was then centrifuged at 7,500 g until the volume was reduced to ~50-100 µL. The filtrate containing imidazole and salts was removed and the retentate made up to 4 mL, thus diluting the imidazole and salts in the sample.
The sample was centrifuged and made up to 4 mL a further two times, with the fourth and final centrifugation step generating concentrated transketolase in 50 mM Tris-HCl with negligible concentrations of imidazole and NaCl present in the sample. Alternatively, a 0.5 mL 10,000 NMWL Amicon® Ultra-4 centrifugal filter unit was used.

**Protein characterisation**

*Quantification of cell lysate protein concentration by Bradford assay*

Total protein concentration was measured using the Bradford assay using BSA as a standard protein. 50 µL of 0.2, 0.5, 0.8 and 1.0 mg/mL BSA and 50 µL of sample (diluted to a concentration between 0.2 – 1 mg/ml if required) were added to separate cuvettes. 1.5 mL of Bradford reagent warmed to room temperature was added to each cuvette. Absorbance was measured at A595 after 15 minutes when the absorbance was stable. Any A595 values above 1.0 were diluted to be within the appropriate absorbance range. The A595 at the four concentrations of BSA was used to plot a standard graph and used to calculate the total protein concentration of the sample. This method was used to determine the concentration of both clarified lysate and concentrated protein.

*Quantification of purified protein concentration by A280 using a Nanodrop-2000*

Protein concentration was determined by absorbance at 280 nm in 6 M Guanidine-HCl and 20 mM Sodium Phosphate, pH 6.5. Absorbance was measured using a Nanodrop-2000 spectrophotometer; the molecular weight of each variant was based on the wild-type monomeric molecular weight of 73035.5 g mol⁻¹ and an extinction coefficient (ε) of 92630 L mol⁻¹ cm⁻¹, modified for each variant.

**Mass spectrometry**

LC-MS was performed using an Agilent 1100/1200 LC system connected to a 6510A QTOF mass spectrometer (Agilent, UK). Samples of 10 µL TK at 0.2 µg/µL were injected onto an Agilent PLRP-S (150 mm x 2.1 mm, 1000 Å, 8 µm) column, maintained at 30 °C. Two mobile phases A (5% MeCN in aqueous 0.1% formic acid) and B (95% MeCN, 5% water, 0.1% formic acid) were used at 0.3 mL/min. The column was pre- equilibrated at 25% B for 1.9 min, before injection, held for 1 min further at 25% B, and then a gradient elution increased B to 99% over 16 min. After 2 min, B was decreased to 25% over 0.1 min. The QTOF mass spectrometer scanned m/z from 100 to 3100 Da. Positive electrospray ionisation (ESI) was used with 4000 V capillary voltage, fragmentor at 175 V, skimmer at 65 V and octopole RF peak at 750 V. Nitrogen was used as the nebuliser and desolvation
gas at a flow of 5 L/min. Spectra were acquired every second with an acquisition time of 1000 msec/spectrum. Lockspray was used during analysis to maintain mass accuracy. Data were processed in MassHunter software (version B.07.00) and deconvolved using the maximum entropy deconvolution algorithm.

**SDS-PAGE analysis**

The total concentration of prepared lysate was determined by a Bradford assay, and diluted to 1 mg/mL. 25 µL 2x Laemmli buffer (Bio-Rad Laboratories, UK) was added to 25 µL sample and heated for 10 minutes at 95 °C to denature the protein. 20 µL sample was then loaded into each well and analysed by an Amersham Imager 600.
CHAPTER THREE: Development & optimisation of a genetically encoded *E. coli* ncAA-incorporation platform for para-substituted phenylalanine derivatives

As stated in the research aims, this project is focused on the exploration of ncAA-incorporation into transketolase to exploit their site-specific, bio-orthogonal properties. To this purpose, the initial goal was to develop and optimise a genetically-encoded *E. coli* ncAA-incorporation platform for para-substituted phenylalanine derivatives.

INTRODUCTION

*Expanding the genetic code*

Traditionally, three main protein-engineering strategies have been utilised; rational design, directed evolution and the two combined (semi-rational) (2,56,57). These exploit the staggering natural sequence - and therefore structural and functional - diversity of proteins, generated by the use of only 20 common amino acids, and have generated countless protein-engineering success stories. However, the extent to which proteins can be engineered is ultimately limited by the sequence and spatial arrangement of those 20 amino acids. Therefore, the genetic code, as found in nature, hinders our progress towards the holy grail of protein engineering; the ability to make precise changes at specific sites in a protein through altered steric and electronic properties of amino acids, beyond those possible with side chain functional groups of the 20 naturally-encoded amino acids.

Furthermore, protein-engineering can improve biocatalysis indirectly, by using either a genetically encoded or post-translationally incorporated molecular probe that can detect a desired property, such as activity or stability. These genetic systems can, for example, provide a new functional screen for an otherwise-undetectable enzyme trait, probe structure-function relationships to provide information for future (semi-)rational engineering, and/or increase the throughput of functional screens for the rapid directed evolution of proteins. Many analytical techniques used to study the properties of macromolecules require functional groups or atoms that are not found in canonical amino acids, such as the use of heavy atoms to improve the resolution of X-ray crystal structures (58) or fluorescent probes to study protein dynamics. Therefore, the genetic code limits
both the scope of potential biocatalytic reactions and the biophysical assays available to engineer biocatalysts.

Incorporation of non-canonical amino acids (ncAAs) (also called unnatural amino acids, UAAs) into proteins has the potential to remove many of the discussed limitations of protein-engineering. NcAAs can have a wide range of structures and functions that are distinct from the 20 common amino acids, and thus provide the means to improve, modify or create novel enzyme activities and also to generate probes of protein structure and function. Transketolase is a great example of an enzyme that can potentially benefit from exploitation of the additional diversity offered by ncAA-incorporation.

**Methods of UAA incorporation**

There are three main methods of ncAA incorporation; chemical synthesis using native chemical ligation (59) and expression protein ligation (EPL) (60,61); *in vitro* biosynthetic approaches using cell-free translation systems (62,63); and *in vivo* ncAA incorporation. Chemical synthesis is non-trivial and cannot be utilised universally because of the need for protecting group chemistry and additional complications such as restrictions on the sites of ligation and issues with protein folding, and is ultimately unfeasible at large-scale. Cell-free systems are generally expensive, protein yields are low, the process of charging tRNAs for use *in vitro* is complex, protein misfolding may occur and the input of energy is expensive (64). This report will focus on *in vivo* ncAA incorporation.

Initially, *in vivo* ncAA-incorporation was limited by low efficiency of incorporation (i.e. the yield of full-length (misincorporated and ncAA-incorporated) protein relative to that of wild-type) and fidelity of incorporation (proportion of ncAA-incorporated protein relative to misincorporated protein), and a narrow range of ncAAs. Over the last decade, however, the field has taken giant strides forward and has largely overcome these limitations.

**In vivo UAA incorporation**

*In vivo* ncAA incorporation has gained significant traction in recent years as a means to introduce novel structural and functional properties to proteins geared towards a multitude of applications. It offers substantial advantages over *in vitro* biosynthetic and chemical strategies; namely higher yields, greater fidelity, native folding and the compatibility with *in vivo* studies into protein structure, function and localisation.
a) **General principles of protein synthesis**

For each amino acid that is incorporated at a defined site in a protein, the host requires:

a) A unique tRNA-codon pair  
b) A corresponding amino-acyl tRNA synthetase (aaRS)  
c) Significant intracellular levels of a stable amino acid.

To ensure high fidelity and efficiency of amino acid incorporation:

a) The tRNA must be orthogonal; it must not be recognised by other endogenous aaRSs but must interact efficiently with the ribosome during translation.  
b) The mRNA sense codon with which the tRNA anticodon interacts must be unique to that tRNA and must not encode any of the other common amino acids.  
c) The cognate aaRS must be orthogonal and only aminoacylate the orthogonal tRNA with only the desired amino acid.  
d) No other host aaRSs can have the desired amino acid as a substrate.  
e) Efficient import mechanisms or host biosynthesis of the amino acid is required.

*In vivo* incorporation of ncAAs follows the same principles as those outlined above. Generally, there are two different incorporation methods based on these principles; residue- and site-specific incorporation. Residue-specific incorporation exploits an existing codon-anticodon interaction by incorporating an isostructural analogue rather than the cognate amino acid, thus replacing the latter in a residue-specific manner. To minimise cognate amino acid misincorporation, a defined media lacking said amino acid must be used. While this method has its uses, such as incorporation of selenomethionine for protein X-ray crystallography, it has a number of drawbacks, such as global incorporation and amino acid limitations.

b) **Site-specific incorporation of UAA*s**

Site-specific incorporation methods also exploit tRNA, ribosome and mRNA interactions but instead of utilising existing interactions, it introduces novel ones (Figure 9). All possible codons encode a specific amino acid and have a corresponding tRNA and aaRS except the three stop (nonsense) codons; (ochre (TAA), opal (TGA) and amber (TAG)); and frameshift quadruplet codons. These can be exploited by evolving an orthogonal tRNA/synthetase pair that incorporates the ncAA in response to a particular nonsense or quadruplet codon, thus creating a genetic code with 21 uniquely encoded amino acids and two fully functioning stop codons.
c) **Generation of orthogonal suppressor tRNA/synthetase pairs**

The first orthogonal tRNA/synthetase pair was derived from a tyrosyl pair from *Methanococcus jannaschii*. The anticodon sequence of the *Methanococcus jannaschii* tyrosyl-tRNA synthetase (TyrRS) has a minimalist anticodon loop-binding domain, therefore the anticodon sequence of its cognate tRNA was mutated to CUA (tRNA$_{\text{Tyr CUA}}$) without significantly impacting the tRNA-binding affinity of the TyrRS (65,66). Previous work had shown that the major identity elements of *Methanococcus jannaschii* tRNA$_{\text{Tyr}}$, including the first base pair of the acceptor stem (C1-G72) and A73, differ from those in *E. coli* tRNA$_{\text{Tyr}}$, which uses a different first base pair (G1-C72), A73, a long variable arm, and the anticodon as major recognition elements (67). Moreover, the *Methanococcus jannaschii* TyrRS lacks an editing mechanism (68) which simplifies the directed
evolution of substrate specificity. The orthogonality of both the \textit{Methanococcus jannaschii} suppressor tRNA and TyrRS were subsequently improved using two general double sieve selection processes (Figure 10 and 11) (65).

\textbf{Figure 10:} Directed evolution of orthogonal amber suppressor tRNA\textsubscript{CUA}\textsuperscript{Tyr} in \textit{E. coli} using the ‘double sieve’ negative-positive selection process (64).

\textbf{Figure 11:} A general double-sieve positive-negative selection scheme for evolving aaRS variants specific to an ncAA (64).
**Advances in site-specific UAA incorporation**

The initial genetic constructs, encoding the orthogonal translation machinery, resulted in a low efficiency of amber suppression and hence produced little full-length protein relative to wild-type. Fundamentally, this was because; a) the decoding of nonsense codons is not as efficient as the decoding of sense codons; b) the vectors controlling the expression of tRNA/synthetase pairs were not optimised; and c) sub-optimal orthogonal ncAA-tRNA-Ef-Tu interactions during translation. While the former two have been addressed to differing degrees, sub-optimal Ef-Tu binding has remained an issue because alteration of Ef-Tu to improve UAA incorporation is likely to compromise the fidelity of incorporation of the common amino acids.

**a) Removing or reducing RF1 inhibition**

RF1 facilitates termination at amber (UAG, essential) and ochre (UAA; non-essential) stop codons by directly binding to the mRNA stop codon in the A site of the ribosome in a conformation mimicking a tRNA (69). Once bound to the ribosome, the polypeptide is released and the ribosome and release factors disassemble, terminating translation (70). Hence the orthogonal suppressor tRNA_{ncAA}^{CUA} and RF1 compete for the amber stop codon.

An approach used to negate competitive inhibition of ncAA incorporation by RF1 was to systematically replace all 321 amber codons (UAG) with ochre codons (UAA) within the *E. coli* genome, permitting the deletion of the non-redundant *prfA*, which encodes RF1 (71). During the construction of the ‘amber-less’ strain, dubbed C321. Δa, 355 additional mutations were acquired and were the most likely cause for a 60% increase in doubling time and a 13% decrease in max OD600 compared to *E. coli* MG1655. A new and improved strain, C321. Δa.exp, had mutS reverted and hence had greater genomic stability.

In addition, orthogonal ribosomes (ribo-X) with an evolved 16S rRNA have been shown to selectively translate orthogonal mRNAs containing an orthogonal Shine-Dalgarno sequence, bind RF1 with a lower affinity than native 16S rRNA and hence facilitate efficient ncAA incorporation by an orthogonal tRNA/synthetase pair at amber codons (72).

**b) Improved incorporation through plasmid optimisation**

Schultz and co-workers iteratively optimised each element of the tRNA/aaRS expression plasmid, and first designed the plasmid *pEVOL* (73), and later *pUltra* (74) (Figure 12). The *pUltra*
plasmid improved the efficiency and fidelity of incorporation by using \textit{tacI}, an inducible promoter that allows low levels of leaky expression in the absence of inducer and high expression levels upon induction. The basal level of aaRS expressed during pre-and early-induction incorporation minimised unfavourable interactions with the endogenous translational machinery during cell growth (e.g. Ef-Tu, ribosomal components). Additionally, the origin of replication was replaced with \textit{CloDF13 (cdf)} (20-40 copies per cell) to increase plasmid copy number, and a strong ribosome binding-site (RBS) was inserted before the aaRS start codon. The efficiency of \textit{para}-acetylphenylalanine incorporation into GFP(Tyr161TAG) in BL21 (DE3) was 50%. No band corresponding to full-length protein was observed by SDS-PAGE, which indicated an absolute fidelity of incorporation (i.e. the level of full-length (i.e. misincorporated) protein in the absence of ncAA relative to that in the presence of ncAA) of 100%. However, the absolute fidelity of incorporation was reduced to >80% when detected by GFP fluorescence due to the higher sensitivity of the assay. The functional fidelity of incorporation (i.e. the level of misincorporation in the presence of ncAA) was not determined.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure12.png}
\caption{Vector maps for \textit{pEVOL} and \textit{pUltra} (74).}
\end{figure}

The initial aim was to develop and optimise a genetically-encoded \textit{E. coli} ncAA-incorporation platform for para-substituted phenylalanine derivatives, to facilitate the subsequent research into transketolase activation, inhibition, stability and catalytic activity.

\section*{EXPERIMENTAL PROCEDURES}

\textit{Mutagenesis at residue 385Y, 105Y and 603K}

Mutagenesis was carried out using Quikchange site directed mutagenesis using the manufacturers protocol (Stratagene, Cambridge, UK) and the following primer and its reverse complement, designed for specific mutations at residue 385:

\begin{align*}
5\prime-\text{GCTGACCTGCGCCGTAGACCTGACCTGTGG-3\prime} & \quad \text{Mutants 385ncAA} \\
5\prime-\text{GCTGACCTGCGCGCCGTTTACCTGACCTGTGG-3\prime} & \quad \text{Mutant 385F} \\
5\prime-\text{GCTGACCTGCGCCGCGACAGACCTGACCTGTGG-3\prime} & \quad \text{Mutant 385Q}
\end{align*}
CHAPTER THREE

5'-GGAAGTGGGTAGACCGCTGGTGAGAAACC-3'    Mutant 105pAzF
5'-GTACTGCCGTTAGGCAGTTACTGCAC-3'    Mutant 603pAzF

The dpnI-digested PCR product was transformed into XL10-gold competent cells and the plasmid subsequently isolated using a Qiagen miniprep kit (Stratagene, Cambridge, UK).

**Preparation of and co-transformation into competent C321.AA.exp “amberless” cells**

The “amberless” *E. coli* strain, C321.AA.exp, a gift from George Church (Addgene plasmid #49018), was used as the expression strain for all variants. A 50 mL culture of C321.AA.exp in LB was grown to an OD$_{600}$ ≈ 0.5 in a 250 mL shake flask and subsequently transferred to two pre-chilled 50 mL falcon tube and cooled for 10 minutes on ice. All consumables required were cooled on ice for the duration of the procedure. The cells were centrifuged at 2700g for 10 minutes at 4 °C, the supernatant discarded and the pellet re-suspended in 1.6 mL pre-cooled 100 mM CaCl$_2$ for 30 minutes on ice. Centrifugation followed by re-suspension and incubation was repeated for each falcon tube. Finally, the cells were combined into a single tube and 0.5 mL pre-chilled 80% glycerol was added. The resulting chemically competent cells were frozen in liquid nitrogen and stored at -80 °C until required. The chemically competent C321.AA.exp cells were co-transformed with pUltra, a gift from Peter Schultz (Addgene plasmid #48215), encoding the ncAA-incorporation system (74), and pQR791, encoding transketolase variants.

**Mass spectrometry**

LC-ESI-MS was performed as described in General Experimental Procedures.

**SDS-PAGE analysis**

SDS-PAGE was performed as described in General Experimental Procedures.

**Circular polymerase extension cloning (CPEC)**

CPEC was performed as described previously (75). The insert (1531 bp; the tacI promoter region plus overlapping regions) was amplified from the pUltra plasmid using primers 1 and 2. The linear vector (4958 bp; pQR791-TK(S385Y/D469T/R520Q) minus the tktA1 constitutive promoter region plus overlapping regions) was amplified from the pQR791-TK(S385Y/D469T/R520Q) plasmid using primers 3 and 4. The overlapping regions are shown in lower case. An annealing temperature
gradient was used between 57-70 ºC to optimise the PCR reaction. Amplified PCR products were visualised using gel electrophoresis on either a 2% (> 1 kb) or 1% (< 1 kb) agarose gel, and subsequently purified and concentrated.

Primer 1: 5’- gacttgagetagtccatgacacGACTGGGTTGAAAGCTCTCAAGG-3’
Primer 2: 5’-gtaactgcctcagctgctcagtcgACCTTGGTTCAAAAAACCTATCAGAGCAGA-3’
Primer 3: 5’-cgaatgcgaactgcgagttcagtcgAGCTGTCGTCAGTAACTCAAGG-3’
Primer 4: 5’-gtgctcgactgcagctcaagtcGGAAAGAACATGTGAGCAAAAGGC-3’

The circular plasmid, pQR791-tacI_TK(S385Y/D469T/R520Q) (6439 bp), which consisted of the combined sequences of the insert, the linear vector and their overlapping regions, was generated using a CPEC cloning reaction, as described previously (75), with insert:vector molar ratios of 7:1, 14:1 and 21:1. The shift in molecular weight from 4958 bp to 6439 bp was visualised via gel electrophoresis and the CPEC product purified and concentrated.

XL1-Blue competent cells were transformed with the CPEC product and grown on agar plates. Plasmid DNA was isolated from six single-colony cultures using a miniprep kit, and glycerol stocks were stored at -80 ºC. The circularised plasmid DNA was visualised using gel electrophoresis to confirm which colonies had been successfully transformed with pQR791-tacI_TK(S385Y/D469T/R520Q) rather than pQR791-TK(S385Y/D469T/R520Q). Despite being circularised DNA, only one of the six colonies had a clear shift in molecular weight of +300 bp and hence corresponded to pQR791-tacI_TK(S385Y/D469T/R520Q). The successful pQR791-tacI_TK(S385Y/D469T/R520Q) plasmid DNA was sequenced fully by Source Biosciences in the forward and backwards direction (and the sequences stored electronically) to confirm a) the presence of the tacI insert and b) no additional mutations have been introduced, using the following primers:

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Primer Sequence</th>
<th>Position relative to start codon</th>
</tr>
</thead>
<tbody>
<tr>
<td>TK_Seq_F1</td>
<td>CTAGCCCGATCCAGAGATTTCTGAAG</td>
<td>-160 bp</td>
</tr>
<tr>
<td>TK_Seq_F2</td>
<td>CTTCCGTCGACGTGCACTCTAAAACCTC</td>
<td>+284 bp</td>
</tr>
<tr>
<td>TK_Seq_F3</td>
<td>CATGCTCACTGGGGATGGAAAGAAG</td>
<td>+887 bp</td>
</tr>
<tr>
<td>TK_Seq_F4</td>
<td>GGTAACATCATCCACAGCGGTCTGAG</td>
<td>+1221 bp</td>
</tr>
<tr>
<td>TK_Seq_F5</td>
<td>GAACGGTGTGCTATTGCTCGGTATGAC</td>
<td>+1892 bp</td>
</tr>
<tr>
<td>TK_Seq_R1</td>
<td>CGTGCGCGTTATGAAAGCAATTGAG</td>
<td>+2243 bp</td>
</tr>
<tr>
<td>TK_Seq_R2</td>
<td>CACGGACGCGCAGTAGACATGTTC</td>
<td>+1475 bp</td>
</tr>
</tbody>
</table>
RESULTS

The efficiency and fidelity of ncAA-incorporation

The optimised ncAA-incorporation plasmid, pUltra, which encoded an evolved, orthogonal Methanococcus jannaschii tRNA/synthetase pair (65) that incorporates ring-substituted phenylalanine derivatives at amber stop codons with high fidelities (i.e. low levels of misincorporation) and efficiencies (i.e. high protein yield relative to ‘wild-type’) of incorporation (74), but considerable ncAA promiscuity (73), was co-expressed with the TK expression plasmid, pQR791-TK-S385X/D469T/R520Q, in C321 AA.exp. The “amberless” E. coli strain, which had all amber codons replaced and RF1 deleted and thus cannot terminate translation at amber stop codons, had previously improved the efficiency of ncAA incorporation significantly (71).

The ncAA-incorporation machinery was utilised to successfully incorporate para-aminophenylalanine (pAMF), para-cyanophenylalanine (pCNF), para-nitrophenylalanine (pNTF), and para-azidophenylalanine (pAzF). In all ncAA variants, a low but significant level of misincorporation of glutamine, phenylalanine, and tyrosine was observed (Figure 13). Glutamine misincorporation was attributed to amber suppression by GlnV of the amberless strain (71), while phenylalanine and tyrosine were misincorporated into transketolase by the promiscuous aminoacyl-tRNA synthetase, even though this enzyme reportedly gave no misincorporation in another protein (74). While this could be due to my use of higher-resolution mass spectrometry, or perhaps a site-specific dependence of ncAA incorporation, it is also possible that the “amberless” strain increased the level of misincorporation in this study, since the deletion of RF1 could result in ribosomal stalling at amber stop codons and increase the likelihood of misincorporation.
CHAPTER THREE

Figure 13: Summary of the ncAA-incorporation performance into the active site residue 385, and the protein surface residues 603 and 105, of the transketolase variant S385Y/D469T/R520Q. The protein yield relative to 385Y (similar to the efficiency of incorporation), fidelity of incorporation, and %TK\textsubscript{high} are displayed in grey, red and blue, respectively. The %TK\textsubscript{high} (blue), which equals the % of transketolase that existed in the active, oxidised species, TK\textsubscript{high}, relative to the inactive, unmodified species, TK\textsubscript{low}, is examined in more detail in Chapter Five and Six.

Comparison of incorporation of pAMF, pCNF and pNTF relative to Y

The specific rationale for ncAA-incorporation at residue 385, and the choice of ncAAs, are explained in Chapter Four. The efficiency (the yield of full-length ncAA-incorporated and misincorporated TK relative to the protein yield of S385Y/D469T/R520Q) and fidelity (the proportion of ncAA-incorporated relative to misincorporated) of incorporation, both decreased as the electron density of the ncAA aromatic ring decreased, from 71.3% and 81.3% in pNTF, to 23.3% and 48.1% in pAMF, respectively. The oxidised forms of ncAA-incorporated transketolase, characterised as the most active form of transketolase, TK\textsubscript{high}, in Chapter Five/Six, were included in this calculation.
Comparison of incorporation of pAzF at residue 603K and 105Y relative to S385Y/D469T/R520Q

The specific rationale for ncAA-incorporation at residue 603 and 105, and the choice of ncAA, are explained in Chapter Seven. The efficiency and fidelity of incorporation of pAzF was very similar at residues 603 and 105, despite the replacement of two very different amino acids. The location and surrounding amino acid sequence therefore dictated the efficiency and fidelity of incorporation more than the replaced natural amino acid. The efficiency and fidelity of incorporation was among the best of the ncAAs tested, presumably because the tRNA/synthetase pair utilised was originally evolved to accept and incorporate pAzF (65,66).

The proportion of transketolase that is in its oxidised, active form, TK\textsubscript{high}, calculated from the sum of the peak areas of singly- (+16 Da), and doubly- (+32 Da) hydroxylated molecular weight peaks relative to the unmodified peak area, were roughly the same for each pAzF-incorporated variant, but was less than that observed for other transketolase variants. The reason for this is unclear, but would suggest that both variants were less sensitive to oxidative stress and/or addition of pAzF to the fermentation broth reduces oxidative stress during fermentation.

Optimisation of the ncAA-incorporation platform using tacI

High expression of transketolase via the native, constitutive promoter, tkt\textsubscript{A}p\textsubscript{1}, was initially hypothesised to facilitate misincorporation before and shortly after induction of the tRNA/synthetase system, and was therefore switched to the weaker tacI promoter in an attempt to improve the fidelity of incorporation via CPEC (Circular Polymerase Extension Cloning) (75).

CPEC has three stages. First, linearised destination vector and linearised insert are generated by PCR amplification. The PCR primers are designed to introduce overlapping DNA sequences at the 5’ and 3’ ends of each fragment. The overlapping regions should have annealing temperatures of approximately 60 °C (or approximately 25 bp). In stage two, the linearised vector and insert (with overlapping regions) are added together and the two fragments prime each other via their overlapping DNA sequences. As such, no oligo-primers are required in the second step. Phusion polymerase is used to amplify the entire vector-insert circularised product using a single polymerase without the need for restriction digestion, ligation, or single-stranded homologous recombination, but consequently the circularised vector-insert product has two nicks. The third and final step involves transformation back into \textit{E. coli} to exploit the cellular DNA repair machinery to repair the DNA
nicks. A schematic diagram is provided in Appendix One. The protein yield of mutant Y and pCNF increased by 91% and 113%, respectively, but had little impact on the fidelity of incorporation (Figure 13).

**DISCUSSION**

The *Methanococcus jannaschii* aminoacyl-tRNA synthetase/tRNA incorporation system was found to prefer ring-substituted phenylalanine derivatives with strong electron-withdrawing groups at the para position over those with electron-donating groups. Furthermore, the incorporation machinery appeared to favour large, polar or charged ring-substitutions, with the best incorporation efficiencies and fidelities observed with pNTF and pAzF. However, a direct comparison between pAzF and the other ncAAs was difficult as the site of incorporation was different.

Despite improved protein yields, the constitutive expression system was utilised in all future protein expression to allow a fair comparison with previous transketolase data from the Dalby lab, and because the inducible expression system failed to improve the fidelity of incorporation, deemed the most important metric of incorporation in this study. The lack of improvement suggested misincorporation was likely caused by the structural similarity between the ring-substituted phenylalanine derivatives and the native amino acids, and the stalling of mRNA and nascent amino acid chain resulting from the deletion of RF1, rather than the constitutive TK promoter itself.

Overall, a genetically encoded *E. coli* ncAA-incorporation platform for para-substituted phenylalanine derivatives was successfully developed and optimised through replacing the constitutive TK promoter with an inducible tacI promoter. Each ncAA was successfully incorporated with mostly excellent, but generally acceptable, levels of misincorporation. In each case, analyses of mass spectra were used to determine the misincorporation of Gln, Phe and Tyr, which were required for future analysis and deconvolution of the catalytic and stability properties of ncAA mutants.

**SUMMARY**

This chapter established the methodology of ncAA incorporation utilised in the remainder of this thesis. The methodology itself – a combination of the pUltra plasmid and the amberless strain – is not new, but it’s application in transketolase research is. It is unclear how many previous ncAA-incorporation studies have quantified incorporation fidelity; in some cases, only the m/z rather than deconvoluted (i.e. relative abundance) spectra were provided, making analysis and quantification of
misincorporated species very challenging. In other studies, no supporting mass spectra data was provided at all. Therefore, to my knowledge the in-depth quantitative analysis of the fidelity of incorporation by fitting mass spectra to the sum of multiple Gaussian functions is a novel approach.

The output from the chapter includes the analysis of the efficiency and fidelity of incorporation of four para-substituted phenylalanine derivatives and the creation of a new, inducible (\textit{tacI}) transketolase expression plasmid with improved protein yield but, unfortunately, similar fidelities of incorporation.
CHAPTER FOUR: Fine-tuning the activity and stability of an evolved enzyme active site through ncAAs

Having developed the *E. coli* transketolase ncAA-incorporation platform, the concept of site-specific saturation mutagenesis in the active-site of an enzyme was extended with a set of aromatic ncAAs. Previously, an *E. coli* transketolase variant (S385Y/D469T/R520Q) was evolved to accept aromatic aldehydes, unlike the wild type. The aromatic residue Y385 was critical to novel acceptor substrate binding, and was therefore explored beyond the natural aromatic residues, to probe the effect of side-chain structure and electronics on enzyme function and stability.

INTRODUCTION

As our understanding and control of biocatalytic reactions improves, organic chemists are increasingly able to employ naturally-occurring or engineered enzymes to catalyse otherwise difficult synthetic chemical reactions. Transketolase has considerable potential for asymmetric C-C bond formation by catalysing the transfer of a two-carbon ketol group from a donor substrate to an aldehyde acceptor substrate (76).

The natural substrates of transketolase are generally phosphorylated sugars such as ribose-5-phosphate and xylulose-5-phosphate, but transketolase has been engineered to accept a plethora of novel substrates by employing various directed evolution strategies. For example, the substrate specificity of transketolase was successfully shifted from phosphorylated sugars, first towards non-phosphorylated, polar substrates (40), followed by aliphatic non-phosphorylated and heteroaromatic substrates (44,45,77), and finally to three aromatic benzaldehyde derivatives; 3-formylbenzoic acid (3-FBA), 4-formylbenzoic acid (4-FBA) and 3-hydroxybenzaldehyde (3-HBA) (48,49).

A transketolase variant with high activity towards aromatic aldehydes, S385Y/D469T/R520Q, was the culmination of three successive rounds of directed evolution which used different smart-library design strategies, including active-site saturation mutagenesis, statistical coupling analysis, and molecular docking (48,49). Structural analysis of the triple mutant crystal structure, coupled with *in silico* molecular docking of the three benzaldehyde derivatives, revealed the creation of two distinct binding pockets that were sterically separated by the D469T mutation (50). The S385Y mutation was predicted to play a crucial role in aromatic substrate binding to both
pockets through π-π stacking interactions with F434 and the aromatic ring of the substrate. While 3-FBA bound into one pocket and 4-FBA the other, both with relatively high affinity and catalytic turnover, 3-HBA was found to bind to both pocket 1 and 2 with low affinity and poor catalytic productivity.

In previous rounds of directed evolution, activity gains were realised by semi-rational engineering of important active-site residues through relatively large saturation mutagenesis libraries. The aim of this study was to explore the critical aromatic ring of Y385, through altered aromatic-ring electron density, to probe, and potentially improve, the catalytic activity, substrate inhibition, enzyme stability and binding pocket preference of the S385Y/D469T/R520Q variant, and then to rationalise any observed trends in the context of active-site electronics, hydrophobic packing, and the size, polarity and hydrogen binding potential of the ring-substituted functional group.

The genetic code limits us to only 20 amino acids. Only phenylalanine has a less electron-dense single aromatic ring than tyrosine, and none are more electron-dense. The genetic code was therefore expanded beyond its natural limits via incorporation of ncAAs to create a series of five variants with highly electron-donating to highly electron-withdrawing aromatic ring substitutions at the para- position, \( p \)-aminophenylalanine \( (pAMF) \) > tyrosine \( (Y) \) > phenylalanine \( (F) \) > \( p \)-cyanophenylalanine \( (pCNF) \), and \( p \)-nitrophenylalanine \( (pNTF) \), in order of decreasing aromatic ring electron density).

Global incorporation of ncAAs has enhanced the activity and thermal stability of a number of enzymes, mainly through incorporation of fluorinated natural amino-acid analogues \((78–83)\). Site-specific, active-site ncAA incorporation is a nascent and promising field of research that was initially held back by low incorporation efficiencies and fidelities of incorporation. Over the last decade, the genetic code has been expanded to incorporate a myriad of ncAAs with much improved efficiencies and fidelities. To date, only a handful of studies have successfully improved or introduced novel enzyme activity via site-specific, active-site ncAA incorporation \((84–87)\). This is one of the first examples in which both catalytic activity and stability are simultaneously improved via site-specific ncAA incorporation into an enzyme active site. This study report ncAA-incorporated variants with both enhanced stability and activity, and provides an example of the benefits of including ncAAs in site-specific, directed evolution libraries.
EXPERIMENTAL PROCEDURES

Mutagenesis at residue 385

Mutagenesis was carried out as described in Experimental Procedures, Chapter Three.

Preparation of and co-transformation into competent C321.ΔA.exp “amberless” cells

Competent C321.ΔA.exp “amberless” cells were prepared as described in Experimental Procedures, Chapter Three.

Enzyme preparation & enzyme kinetics

Variants of the transketolase mutant S385Y/D469T/R520Q were co-expressed with the ncAA-incorporation machinery from the pUltra plasmid for eight hours in C321.ΔA.exp cells in the presence of 1 mM ncAA and 1 mM IPTG. The resulting cell pellet was lysed and purified as described previously (41). Purified transketolase was ultrafiltrated four times using Amicon Ultra-4 10k MWCO centrifugal filter to remove excess imidazole and cofactors. Protein concentration was determined by absorbance at 280 nm in 6 M Guanidine-HCl and 20 mM Sodium Phosphate, pH 6.5. Absorbance was measured using a Nanodrop spectrophotometer; the molecular weight of each variant was based on the wild-type monomeric molecular weight of 73035.5 g mol⁻¹ and an extinction coefficient (ε) of 92630 L mol⁻¹ cm⁻¹, modified for each variant.

Kinetic parameters were obtained at saturating 50 mM HPA and 3-50 mM 3-HBA. 80 µL 0.6-1.0 mg/mL TK was incubated with 20 µL of 10× cofactor solution (24 mM TPP, 90 mM MgCl₂) for 30 minutes, and the reaction initiated with 100 µL 2x 3-HBA in 50 mM Tris-HCl, pH 7.0. All reactions were carried out in triplicate in glass vials at 22 °C. Samples of each reaction were quenched every 30 minutes for 180 minutes by addition of 380 µL 0.1% TFA to 20 µL sample and centrifuged at 13,000 rpm for 3 min, and the supernatant analysed by HPLC, as described previously (48). The TK concentration in each reaction was between 0.24 and 0.4 mg/ml. Higher TK concentrations were used for the 3-HBA reaction due to the slower conversion. All data were fitted by non-linear regression to the modified Michaelis–Menten equation to determine the kinetic parameters of all the variants with each substrate.
Derivation of the kinetic model for activity and substrate inhibition

The Modified Michaelis-Menten function was derived in a similar way to the standard Michaelis-Menten function, except using the following chemical equilibria, left, that describe a single catalytically productive enzyme-substrate binding event, \( n_1 = 1 \), that is completely inhibited at high [S] by multiple \( (n_2) \) inhibitory enzyme-substrate interactions. The function makes the assumption that the catalytically productive binding event is kinetically much faster than the inhibitory binding event, and hence the reaction order is kinetically sequential. The chemical equilibria marked with a cross are those which become obsolete once the reaction is defined as kinetically sequential. The respective dissociation constants, right, are derived using the Law of Mass Action.

\[
\begin{align*}
[E] + n_1[S] & \rightleftharpoons [E.Sn_1] + n_2[S] \rightleftharpoons [E.Sn_2] \\
& \downarrow k_{cat} \\
& [E] + [P]
\end{align*}
\]

\[
\begin{align*}
\checkmark [E] + n_1[S] & \rightleftharpoons [E.Sn_1] ; \ n_1 = 1 & \Rightarrow K_{d1} = K_m = \frac{[E][S]}{[E.S]} \\
\checkmark [E.Sn_1] & \rightarrow [E] + [P] ; \ n_1 = 1 & \Rightarrow v_p = k_{cat} \cdot [E.S] \\
\times [E] + n_2[S] & \rightleftharpoons [E.Sn_2] \\
\checkmark [E.Sn_1] + n_2[S] & \rightleftharpoons [E.Sn_1.Sn_2] ; \ n_2 = n - n_1 = n - 1 & \Rightarrow K_{d2}^{app} = K_i^{n-1} = \frac{[E.S][S]^{n-1}}{[E.S.Sn_2]} \\
\times [E.Sn_2] + [S_1] & \rightarrow [E.Sn_2.Sn_1]
\end{align*}
\]

where [E] is the enzyme concentration, [S] is the substrate concentration, \( n_1 \) and \( n_2 \) are the theoretical number of molecules, or number of orientations, of substrate binding in the first (catalytically productive) and second (inhibitory) binding event, \( n \) is the theoretical total number, or total orientations, of productive and inhibitory substrate molecules that can bind the enzyme and \( K_{d1} \) and \( K_{d2}^{app} \) are the apparent dissociation constants of the first catalytically productive binding event and all subsequent inhibitory binding events, respectively.

The inhibitory binding events are defined as either multiple substrate molecules that simultaneously bind to the enzyme active site, or as a single substrate molecule that interacts with the enzyme active site in multiple inhibitory binding poses. The \( n_2 \), or \( n_i \), number of inhibitory poses are assumed to have slightly different binding inhibition constants, \( K_i \), which are represented by a single inhibitory dissociation constant, \( K_{d2}^{app} \). In other words, the individual inhibitory dissociation
constants for each binding pose are convoluted or “averaged” by multiplying and dividing the binding constants together to give a single representative inhibitory binding constant, such that:

\[ K_{d1} = K_{d1}^{n_1} = K_{d1}^{-1} = K_m; \text{ and} \]
\[ K_{d2}^{app} = K_{i_1}, K_{i_2}, K_{i_3}, ..., K_{i_{n_2}} = K_i^{n-n_1} = K_i^{-1} \]

In theory, the substrate can “slide” between inhibitory binding poses. “Sliding” is significantly more likely to be observed in catalytic reactions that are dominated by hydrophobic enzyme-substrate interactions, which result in much less-well defined enzyme-substrate orientations compared to electrostatic interactions.

The total enzyme concentration, \([E_T]\), can be expressed as the following:

\[ [E_T] = [E] + [E.S] + [E.S.S_{n2}] \]

Substitution of \([E.S]\) and \([E.S.S_{n2}]\) for their dissociation constant expressions (above right) gives:

\[ [E_T] = [E] + \frac{[E][S]}{K_m} + \frac{[ES][S]^{n-1}}{K_i^{n-1}} = [E] + \frac{[E][S]}{K_m} + \frac{[E][S]^{n-1}}{K_m.K_i^{n-1}} \]

Which rearranges and simplifies to:

\[ [E] = \frac{[E_T]}{1 + \frac{[S]}{K_m} + \frac{[S][S]^{n-1}}{K_i^{n-1}}} = \frac{[E_T]}{1 + \frac{[S]}{K_m} + \frac{[S]^{n}}{K_m.K_i^{n-1}}} \]

Initial reaction velocity can be derived by substitution of \(K_m = \frac{[E][S]}{[E.S]}\) into \(v_p = k_{cat} \cdot [E.S]\):

\[ v_p = k_{cat} \cdot [E.S] = \frac{k_{cat} \cdot [E][S]}{K_m} = \frac{k_{cat} \cdot [E_T][S]}{K_m(1 + \frac{[S]}{K_m} + \frac{[S]^{n}}{K_m.K_i^{n-1}})} = \frac{k_{cat} \cdot [E_T]}{K_m[\frac{S]}{K_m} + 1 + \frac{[S]^{n-1}}{K_i^{n-1}}} \]

Finally, substitution of \(v_{max} = k_{cat} \cdot [E_T]\) into the above equation gives the Modified Michaelis-Mention function:
\[ v = \frac{V_{\text{max}}}{K_m[S]} + 1 + \frac{[S]^{n-1}}{K_i^{n-1}} \]

Separately, a Modified Hill function can be derived in a similar way that describes two sequential multi-ligand binding events that do not contribute equally to the fractional saturation (i.e. the binding events are weighted):

\[ \theta = \frac{B_{\text{max}} + \alpha B_{\text{max}} [L]^{n-n_1}}{K_d_1 [L]^{n_1} + 1 + \frac{[L]^{n-n_1}}{K_d_2^{n-n_1}}} \]

This equation is not used in this thesis, however, but is an example of the applicability of the Modified function to various scenarios.

**Mass spectrometry**

LC-ESI-MS was performed as described in General Experimental Procedures.

**SDS-PAGE analysis**

SDS-PAGE was performed as described in General Experimental Procedures.

**Thermal denaturation measurements**

The \( T_m \) values of TK variants were measured using the UNit (Unchained Laboratories, Wetherby, UK) via their intrinsic fluorescence emission ratio (350/330 nm). The microcuvette arrays were loaded with 9 µL of 0.5 mg/mL sample and excited with a 266 nm laser. The fluorescence was measured as a function of temperature in the range of 30–90 °C with steps of 1 °C, equilibration time of 30 s at each temperature. Thermal denaturation curves were analysed by fitting the baseline and single transition to a two-state model (88–90).

**Determination of kinetic parameters of pAMF, pCNF and pNTF variants**

For each variant, the apparent kinetic parameters derived from the experimental data were a convolution of the contribution from ncAA-incorporated species and the misincorporated species Q, F, and in some cases, Y. By contrast, the kinetic data for incorporation of F and Y at residue 385 were not convoluted and so were obtained directly. For the misincorporated mutant Q (D469T/S385Q/R520Q), the specific activity at 50 mM 3-HBA was only 10% that for mutant F.
Additionally, the level of misincorporation of Q was 22.4%, 5.8% and 2.8% in variants pAMF, pCNF and pNTF, respectively, and so its contribution towards the catalytic activity of each variant was negligible. The true kinetic parameters of each ncAA-incorporated variant were determined (Table 2) by globally-fitting the experimental data to the following equation in OriginLabs:

\[
v_{\text{app}} = \left( \frac{\% F_{\text{misincorporation}} \times V_{\max}(F)}{K_m(F) \left[ S \right] + 1 + \frac{\left[ S \right]^{n-1}}{K_i(F)^n}} \right) + \left( \frac{\% Y_{\text{misincorporation}} \times V_{\max}(Y)}{K_m(Y) \left[ S \right] + 1 + \frac{\left[ S \right]^{n-1}}{K_i(Y)^n}} \right) + \left( \frac{\% \text{ncAA}_{\text{incorporation}} \times V_{\max}(\text{ncAA})}{K_m(\text{ncAA}) \left[ S \right] + 1 + \frac{\left[ S \right]^{n-1}}{K_i(\text{ncAA})^n}} \right)
\]

where \(v_{\text{app}}\) is the apparent specific activity of the ncAA-incorporated variant, \(\% F_{\text{misincorporation}}, \% Y_{\text{misincorporation}}\) and \(\% \text{ncAA}_{\text{incorporation}}\) are the known proportions of F and Y misincorporation and ncAA incorporation in that ncAA variant, \(V_{\max}(F), V_{\max}(Y), K_m(F), K_m(Y), K_i(F)\) and \(K_i(Y)\) are the experimentally determined, true kinetic parameters of Y and F, and \(V_{\max}(\text{ncAA}), K_m(\text{ncAA}), K_i(\text{ncAA})\) are the unknown, true kinetic parameters of that ncAA variant.

**Computational docking**

The crystal structure of S385Y/D469T/R520Q was mutated at position 385 to pAMF, Y, F, pCNF and pNTF using the Swiss-Sidechain-plugin for PyMol Molecular Graphics System (Schrödinger, USA), and energy-minimised using Charm forcefield Adopted Basic NR, Implicit Generalised Born solvent model, True SHAKE constant, and 1000 steps in Discovery Studio 2.0 (Accelrys, Inc. San Diego, California, USA). Each variant, with the enamine-TPP intermediate present, was subsequently stripped of all crystallographic waters and docked in Autodock 4.2.6. Ligands were obtained as .mol2 files and assigned three-dimensional coordinates. The explorable space for docking was defined as previously (50). For each search, a Lamarckian genetic algorithm was run 200 times with a maximum number of 25 million energy evaluations. Resulting poses were analysed and checked for hydrogen bonding in PyMol and Discovery Studios 2017. Poses were first split into clusters that differed in energy by 0.04 (kcal/mol), and designated as energy clusters (e.g. EC1). These energy clusters were sub-divided into sub-clusters, defined as catalytically productive when the aldehyde moiety was oriented 0° or 180° given the almost-symmetrical structure of 3-HBA, and catalytically inhibitive when oriented 90° or 270°, relative to the DHE-TPP intermediate. All sub-clusters of EC2 or lower were defined as energetically unfavourable and hence inhibitory. For molecular analysis, interactions facing 180° from the DHE-TPP intermediate were rotated 180° about the centre and plane of the substrate aromatic ring.
RESULTS

Electronic properties of the single-ringed aromatic amino acids utilised in this study

The strength of π-π stacking interactions is influenced by aromatic ring electron-density. Electron-withdrawing groups (e.g. -NO₂) strengthen hydrophobic interactions between aromatic rings (91), while the opposite is true for electron-donating groups (e.g. -NH₂). The amino acids used in this study can be ranked in order of ring electron-density from most to least dense: p-aminophenylalanine (pAMF) > tyrosine (Y) > phenylalanine (F) > p-cyanophenylalanine (pCNF) > p-nitrophenylalanine (pNTF) (Figure 14). The aromatic ring electron density at position 385 may therefore influence: a) active site hydrophobicity; b) the overall hydrophobic packing of active-site residues; and c) the strength of substrate binding due to π-π stacking interactions, in either catalytically productive or inhibitory orientations.

Figure 14: Chemical structures of the five single-ringed aromatic amino acids utilised in this study. In order of decreasing aromatic ring electron density: p-aminophenylalanine (pAMF) (A), tyrosine (Y) (B), phenylalanine (F) (C), p-cyanophenylalanine (pCNF) (D), and p-nitrophenylalanine (pNTF) (E).

Nevertheless, substrate binding and catalytic turnover are dictated by a multiplicity of additional factors, including hydrogen bonding (H-bonds) with other active-site residues, steric hindrance, and orientation of the acceptor substrate relative to the DHE-TPP intermediate in three-dimensional space. Furthermore, the characteristics of the ring-substituted functional groups may also impact both catalysis and stability, as they differ in size, polarity, and H-bonding potential from donating (pAMF and Y), to non-bonding (F), and finally accepting (pCNF and pNTF). While it is relatively hard to predict the outcome of such changes without far more structural information and computational power, one can predict with some confidence that one or more properties of an enzyme will at least change when altering the electronics of an active-site residue that is critical to the acceptance of aromatic substrates.


Efficiency and fidelity of ncAA incorporation

In Chapter Three, pAMF, pCNF and pNTF were successfully incorporated into the active site of S385X/D469T/R520Q. Figure 15 summarises the efficiency and fidelity of incorporation observed by SDS-PAGE and LC-ESI-MS, respectively, for each ncAA. In all ncAA variants a low but significant level of misincorporation of glutamine, phenylalanine, and sometimes tyrosine was observed, and accounted for, in addition to higher molecular-weight, oxidised forms of both UAA-incorporated and misincorporated species, in the enzyme kinetic measurements. Overall, the Methanococcus jannaschii aminoacyl-tRNA synthetase/tRNA incorporation system preferred ring-substituted phenylalanine derivatives with strong electron-withdrawing groups at the para position over those with electron-donating groups.

Figure 15: Summary of the efficiency and fidelity of incorporation at residue 385 (A), and the mass spectra of pNTF, pCNF, and pAMF (B, C) and D), respectively). Each mass spectrum (black) was fitted to the sum of multiple Gaussian functions (smaller individual peaks) corresponding to unmodified, single- and double-oxidised forms of variants Q (red), F (magenta), Y (cyan), and ncAA (blue). The ratio of the peak area of unmodified relative to single- and double-oxidised peaks of F and Y were determined from their mass spectra (Figure 16), and applied to the mass spectra of ncAA
variants. The cumulative fit of all peaks is shown in wine-red. Higher-molecular weight peaks to the right of the ncAA-incorporated transketolase represent oxidised forms of mutants $p$NTF, $p$CNF, and $p$AMF (Chapter Five), also observed in the wild-type transketolase. Only the first two oxidised peaks were included in activity ($TK_{\text{high}}$) and fidelity of incorporation calculations.

**Figure 16:** The mass spectra of mutants Y and F when co-expressed with $p$Ultra in *E. coli* C321. $\Delta A.exp$. Each mass spectrum (black) is fitted to the sum of multiple Gaussian functions (smaller individual peaks) corresponding to unmodified, single- and double-oxidised forms, as well as over-oxidised forms. The cumulative fit of all peaks is shown in wine-red. Only the first two oxidised peaks were included in activity ($TK_{\text{high}}$) and fidelity of incorporation calculations.

**Determination of the apparent, ncAA-incorporated and true ($TK_{\text{high}}$) kinetic parameters**

Previously, the $K_m$ and $k_{cat}$ of the triple mutant towards 3-HBA were reported as $390 \pm 10$ mM and $2.1 \pm 0.2$ s$^{-1}$, respectively, using a substrate concentration up to 25 mM 3-HBA owing to insolubility at higher concentrations (48). As a result, the reported $K_m$ and $k_{cat}$ were overestimated because saturating concentrations of substrate were never reached. By sonication and careful pH
adjustment, I have now improved the solubility limit of 3-HBA to at least 50 mM, allowing an accurate calculation of the true kinetic parameters.

Significant substrate inhibition was observed at above 25 mM for all variants that appeared to indicate the binding of multiple molecules or inhibitory poses, given the “cooperative” shape of the data points at high [S]. The experimental data was first fitted to the standard Michaelis-Menten-derived function describing the randomly-ordered non-competitive inhibition model, where binding to a single inhibitor molecule (in this case also the substrate) decreases enzyme activity and the inhibitor binds equally well to the enzyme regardless of whether the enzyme has already bound to the substrate. As expected, the function fitted poorly to the experimental data, with the exception of variant pNTF, indicating that the other variants could not be explained by the standard non-competitive inhibition model alone.

Subsequently, the Modified Michaelis-Menten function was derived ([Experimental Procedures] that described one catalytically productive enzyme-substrate binding event, $n_1$, that is completely inhibited at high [S] by multiple inhibitory enzyme-substrate interactions, $n_2$. The function makes the assumption that the catalytically productive binding event is kinetically much faster, and hence the reaction order is kinetically sequential. The inhibitory interactions can be defined as multiple substrate molecules simultaneously binding to the enzyme active site, which is highly unlikely in the case of transketolase and most other globular enzymes, or as a single substrate molecule that interacts with the enzyme active site in multiple inhibitory binding poses. These inhibitory poses are assumed to have slightly different binding constants ($K_i$'s) and, in theory, the substrate can “slide” between inhibitory binding poses. “Sliding” is significantly more likely to be observed in catalytic reactions that are dominated by hydrophobic enzyme-substrate interactions, such as the interactions between the hydrophobic S385Y/D469T/R520Q active site and 3-HBA, which result in much less-well defined enzyme-substrate orientations compared to electrostatic interactions. The $n_2$, or $n_i$, number of inhibitory poses are assumed to have slightly different binding inhibition constants that are represented by a single inhibitory dissociation constant, $K_i$.

The experimental data of each variant fitted extremely well to the Modified Michaelis-Menten function, providing confirmation that it was the correct kinetic model for the reaction. While the kinetic data of variant pNTF fitted well to the Modified Michaelis-Menten function, the data fitted better to the standard inhibition model (i.e. $n_i=1$ in the Modified Michaelis-Menten function):
\[
v = \frac{V_{\text{max}} \cdot [S]}{(K_m + [S]).(1 + \frac{[S]}{K_i})}
\]

The standard inhibition model was hence used only to derive the kinetic parameters of ρNTF, while the Modified Michaelis-Menten function was used for all other variants. The experimental data of each variant fitted equally well to the randomly-ordered non-competitive inhibition model modified to describe multiple inhibition modes, below, which indicated that the additional random-order term in red is negligible, thus validating the prior assumption that the reaction is kinetically sequential in order:

\[
v = \frac{V_{\text{max}} \cdot [S]}{(K_m + [S]).(1 + \frac{[S]}{K_i})} = \frac{V_{\text{max}} \cdot [S]}{K_m + [S] + \frac{K_m \cdot [S]^{n-1}}{K_i^{n-1}} + \frac{[S] \cdot [S]^{n-1}}{K_i^{n-1}}}
\]

For each variant, the apparent kinetic parameters derived from the experimental data were a convolution of the contribution from ncAA-incorporated species and misincorporated species Q, F, and in some cases, Y. By contrast, the kinetic data for incorporation of F and Y at residue 385 were obtained directly. The deconvolution methodology and apparent kinetic parameters for each ncAA-incorporated variant are provided in the Experimental Procedures.

Furthermore, it is later established in Chapter Five and Six that transketolase exists as a two-species system in which inactive TK\textsubscript{low} is oxidised to the active form of transketolase, TK\textsubscript{high}, in response to oxidative stress during cell growth. Therefore, the kinetic parameters of active TK\textsubscript{high} were calculated from the mass spectrometric ratio of TK\textsubscript{modified}:TK\textsubscript{unmodified}, which was equivalent to the ratio of TK\textsubscript{high}:TK\textsubscript{low} (Table 2). While the absolute parameter values were significantly altered between the ncAA-incorporated species and the TK\textsubscript{high}-ncAA-incorporated species, the relative parameters and parameter trends were relatively constant as the %TK\textsubscript{high} was, on the whole, constant between variants. All kinetic parameters referred to henceforth thus correspond to those of only the TK\textsubscript{high} species of each variant (Table 2), unless stated otherwise.
Table 2: The true (TK\textsubscript{high}) kinetic parameters of variants \(p\)AMF, Y, F, \(p\)CNF and \(p\)AMF towards 3-HBA, having accounted for both misincorporation and the ratio of TK\textsubscript{high}:TK\textsubscript{low} when expressed in \textit{E. coli} C321.\(\Delta\)A.\textsubscript{exp}. Associated errors are the fitting error for the modified Michaelis-Menten function, or the standard inhibition function \((n_2=n_1=1)\) (\(p\)NTF). The \%TK\textsubscript{high} was determined from the peak areas of TK\textsubscript{modified} relative to TK\textsubscript{unmodified} from LC-ESI-MS data.

Initially, the specific activity of each variant was determined at 50 mM 3-HBA – above which the substrate was insoluble at 22 °C. The specific activity of \(p\)CNF was an impressive 50-fold greater than that of variant Y, while those of \(p\)AMF and \(p\)NTF increased 10-fold and 5-fold, respectively. Conversely, variant F gave a 10% decrease in relative specific activity. The catalytic performance of each variant was investigated further over a range of [3-HBA] and their kinetic parameters determined using the modified Michaelis-Menten derived in the \textit{Experimental Procedures}, to explain these observations (\textit{Figure 17}).
**Figure 17:** The 3-HBA apparent specific activity and thermal stability of residue 385-substituted variants pAMF, Y, F, pCNF and pNTF. The initial rate of each variant was determined at between 0-50 mM 3-HBA, 9 mM Mg$^{2+}$ and 2.4 mM TPP and fitted to the modified Michaelis-Menten function ([Experimental Procedures](#)) (black) and compared to the theoretical Michaelis-Menten function in the absence of acceptor substrate inhibition (red). The apparent thermal stability of pAMF (black), Y (red), F (blue), pCNF (green) and pNTF (magenta) was determined by the change in the fluorescence emission ratio (350/330) as a function of temperature, and fitted to a two-state model of thermal denaturation (88–90).

While a change in the catalytic performance was expected upon tweaking the electronics of an important residue in the active site, it was difficult to predict what trend – if any – may emerge given the complexity of substrate binding and catalytic turnover. Indeed, no clear trend was observed in $K_m$ nor $k_{cat}$ as a function of the aromatic ring electron density of residue 385 ([Figure 17](#) and 18; [Table 2](#)). However, incorporation of pAMF had the greatest positive impact on catalytic efficiency as measured by $k_{cat}/K_m$, with a 2.6-fold improvement relative to mutant Y, largely due to a 2.4-fold improvement in $K_m$. Conversely, the 2.5-fold increase in the $k_{cat}/K_m$ of variant pCNF was instead
CHAPTER FOUR

driven by a 2.3-fold improvement in $k_{\text{cat}}$. Variants F and pNTF resulted in large improvements in $K_m$, but even larger decreases in $k_{\text{cat}}$, and consequently a decrease in catalytic efficiency.

Figure 18: Summary of the true catalytic performance (A & B) and acceptor substrate inhibition (C) of the TK$_{\text{high}}$ species, and the stability (D) of the ncAA-incorporated mutants pAMF, Y, F, pCNF and pNTF. The mutants are ranked in order of decreasing aromatic ring electron density from left to right. A) Catalytic performance relative to mutant Y, in terms of improvement of $K_m$ (grey), $k_{\text{cat}}$ (red) and $k_{\text{cat}}/K_m$ (blue); B) specific activities at 50 mM 3-HBA relative to mutant Y; C) acceptor substrate inhibition profiles, in terms of $K_i$ (grey, left y-axis) and $n_i$ (red, right y-axis); and D) thermal stability, in terms of $T_m$ and van’t Hoff enthalpy, $\Delta H_{\text{ch}}$. Associated errors are the fitting error for the modified Michaelis-Menten function or the thermal denaturation function.

The theoretical total number of inhibitory substrate orientations and/or binding events, $n_i$, was strongly correlated with the aromatic ring electron density of residue 385 (Figure 17 and 18; Table 2). 3-HBA could bind pAMF in as many as ten inhibitory orientations, while that was reduced to one in pNTF. The presence of potentially multiple orientations may reflect the hydrophobicity of both the substrate and binding pocket, which provide a less well-defined interaction location.
compared to, for example, electrostatic interactions. Consequently, a 3-HBA molecule may ‘slide’ between orientations within the binding pocket. Again, there appeared to be no obvious trend in $K_i$ as a function of aromatic ring electron density; variants Y, F and pNTF had similar $K_i$-values, while the introduction of pAMF and pCNF improved the $K_i$ by 42.4% and 75.8%, respectively.

Overall, the substantially improved specific activity of pCNF at 50 mM 3-HBA relative to variant Y appears to be explained by a combination of a 2.3-fold improvement in $k_{cat}$, with a substantially reduced substrate inhibition (through both $n_i$ and $K_i$).

**Thermal stability of holo-transketolase variants**

Previously, the introduction of mutations that conferred rigidity into the cofactor binding loop (e.g. H192P/H282P), led to improved thermal stability of wild-type transketolase by 5 °C (54). It was postulated that a decrease in aromatic ring electron-density via electron-withdrawing substitutions may also strengthen $\pi$-$\pi$ stacking interactions between residue 385 and active-site non-polar aromatic residues, with the potential to increase active-site rigidity and hence thermal stability. The modified Michaelis-Menten function within the expression, described in the ‘Determination of the apparent, ncAA-incorporated and true (TK$_{high}$) kinetic parameters’ in the Experimental Procedures section, was substituted for a function describing a two-state model of thermal degradation ((88–90)), to determine the stability parameters of each ncAA-incorporated species (Figure 17; Table 3) from the overall apparent stability of each variant sample using the known stability parameters and relative proportions of variants Q, F and Y within the ncAA sample. It was not possible to deconvolve the thermal stability of TK$_{high}$ and TK$_{low}$ because the relative stability of each is unknown and it is not currently possible to isolate each sub-species from the other. The $T_m$ did not correlate linearly to other properties of the variants or mutation types, with the best observed with the $n_i$ of the variants ($R^2$=0.65). However, the thermal transition mid-point, $T_m$, increased non-linearly with aromatic ring electron density (Figure 17; Table 3), in the opposite direction to that predicted based on the strength of $\pi$-$\pi$ stacking. The most stable mutant, pAMF, was 7 °C more stable than the least stable mutant, pNTF, 5.4 °C higher than mutant Y, and, remarkably, 2.9 °C higher than wild-type transketolase. This is the first time that a triple-mutant-based transketolase variant has been engineered to be more stable than the wild-type.
Table 3: Summary of the thermal stability of the correctly incorporated species of wild-type, pAMF, Y, F, pCNF and pNTF, having accounted for misincorporation. Parameters were determined fitting the change in the fluorescence emission ratio (350/330) as a function of temperature to a two-state model of thermal denaturation, as described previously (88–90). Associated errors are the fitting error for the thermal denaturation function.

<table>
<thead>
<tr>
<th>Variant</th>
<th>$T_m$ (°C)</th>
<th>±</th>
<th>$\Delta H_{vh}$ (kcal/mol)</th>
<th>±</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type</td>
<td>65.6</td>
<td>0.6</td>
<td>77.2</td>
<td>3.8</td>
</tr>
<tr>
<td>$p\text{AMF}_{\text{incorporated}}$</td>
<td>68.5</td>
<td>0.8</td>
<td>253.3</td>
<td>118.5</td>
</tr>
<tr>
<td>$Y_{\text{incorporated}}$</td>
<td>63.1</td>
<td>0.5</td>
<td>86.6</td>
<td>7.8</td>
</tr>
<tr>
<td>$F_{\text{incorporated}}$</td>
<td>62.5</td>
<td>1.0</td>
<td>70.3</td>
<td>8.6</td>
</tr>
<tr>
<td>$p\text{CNF}_{\text{incorporated}}$</td>
<td>61.6</td>
<td>0.7</td>
<td>90.1</td>
<td>10.9</td>
</tr>
<tr>
<td>$p\text{NTF}_{\text{incorporated}}$</td>
<td>61.5</td>
<td>0.7</td>
<td>93.1</td>
<td>8.6</td>
</tr>
</tbody>
</table>

The thermal denaturation of $p\text{AMF}$ had increased cooperativity, as indicated by the high value of $\Delta H_{vh}$, which is itself indicative of tight overall packing and increased rigidity. The additional rigidity in $p\text{AMF}$ was not expected from the increased aromatic ring electron density, which would in theory form weaker hydrophobic interactions than the other four aromatic amino acids. Molecular modelling instead suggested that the shift in the $para$-substituted functional group, from an H-bond acceptor ($p\text{CNF}$ and $p\text{NTF}$), to non-polar (F), to an H-bond donor ($p\text{AMF}$ and Y), increased the active site and cofactor loop rigidity. Analysis of the energy-minimised structures of $p\text{AMF}$ and Y showed the carbonyl backbone of G262 was within H-bonding distance of the $–\text{NH}_2$ and $–\text{OH}$ side-chain groups, at 3.6 Å and 4.2 Å, respectively (Figure 19A, C & D). The latter distance is large for an H-bond, and, if formed, such an interaction would be extremely weak. In both energy-minimised structures, a helix-turn was formed between residues D259-H261, as a result of stabilisation by the interaction between residue 385 and G262. Neither an H-bond nor a helix-turn were observed in the other three less-stable variants (Figure 19B, E, F & G). G262 is located in the opposite subunit of the transketolase dimer, and therefore provides both an anchor to increase the rigidity of the cofactor-binding loop, increasing $\Delta H_{vh}$, and an additional inter-subunit interaction, to increase $T_m$. The loss of secondary structure in at least three variants may explain why the $\Delta H_{vh}$ was much lower and roughly equivalent across these variants. Furthermore, the weakened or absent inter-subunit H-bond, in
addition to the introduction of bulkier, and polar, ring-substituted functional groups, may explain the gradual decrease in $T_m$ observed from $p$AMF to $p$NTF.

**Figure 19:** The energy-minimised active site structures of $p$AMF, Y (A; green and cyan), F, $p$CNF and $p$NTF (B; magenta, yellow and wheat). Important active site residues are labelled in each. Note that the only significant structural change is the formation of a helix-turn between residues D259-H261 of subunit B. The possible H-bonds between residue 385 of subunit A and G262 of subunit B is shown for each variant in C) D) E) and F), respectively.

**Computational docking of 3-HBA into triple-mutant variants**

In silico molecular docking of 3-HBA into the active site of mutant Y and the computationally mutated, energy-minimised active sites of $p$AMF, F, $p$CNF and $p$NTF provided insights into the molecular basis of the experimentally-determined kinetic and stability parameters. From these, plausible structural explanations for the experimentally-determined kinetic parameters are proposed for each variant.

For all analyses, the highest-energy substrate sub-cluster, energy cluster one (EC1) (Figure 20) were considered catalytically productive when the aldehyde moiety was oriented 0° or 180°, and catalytically inhibitive when oriented 90° or 270°, relative to the DHE-TPP intermediate. All sub-clusters of EC2 or lower were defined as inhibitory. Active site interactions with the catalytically productive substrate poses were analysed to rationalise the experimentally-determined data, in terms
of: the number of productive, unproductive, and inhibitory poses; binding pocket preference (% Pocket A and B); and active-site interactions.

**Figure 20:** Computational docking of 3-HBA into the active site of mutants pAMF, Y (A: green and cyan), F (C; magenta), pCNF (B; yellow) and pNTF (C; wheat). Only sub-cluster poses in catalytically productive orientations are shown. 3-HBA was bound in both Pocket A (red and salmon red) and Pocket B (blue and light blue) in mutants pAMF and Y, respectively (A); in two orientations in Pocket B of pCNF (B; blue and orange); and a single orientation in Pocket B of F and pNTF (C; purple blue and blue).

**Comparison of variant Y and pAMF**

*In silico* analysis predicted 3-HBA to bind to the active site of pAMF and Y in near-identical orientations (**Figure 20A**), including two that were catalytically productive and either three or two, respectively, that were inhibitory. The high number of inhibitory sub-clusters in both pAMF and Y was consistent with the high theoretical number of inhibitory binding orientations \( n_i \) observed experimentally. The two catalytically productive sub-clusters of each variant were split between the two binding pockets that had been previously identified (50) (**Figure 20A**). The decreased aromatic ring electron density in variant Y relative to pAMF shifted the binding pocket preference from Pocket A to Pocket B, with 52% of productive poses in Pocket B of pAMF compared to 82% in Y. This shift may have been due to the loss of two interactions between 3-HBA and Pocket A of mutant Y compared to that of mutant pAMF (**Figure 21**) and was consistent with the decreased \( K_m \) of pAMF, observed experimentally.
Figure 21: The predicted enzyme-substrate interactions of Pocket A (A and C) and Pocket B (B and D) of variants pAMF and Y, respectively. Strong (<3.4 Å) and weak (<4.0 Å) H-bonds are shown as dark and light green circles with dotted lines; Van der Waal’s forces are shown as light green circles with no dotted lines; π-π stacking and π-charge interactions are shown as pink and orange circles with dotted lines; and interactions between 0.6 and 0.7 Å are shown as red circles with dotted lines. Interactions were predicted in Discovery Studio 2017 using the minimised mutant structures and the catalytically active sub-cluster poses of 3-HBA in the mutant active sites.

Analysis of variant pCNF

3-HBA was predicted to bind pCNF in two catalytically productive and one inhibitory sub-cluster of poses that were unique to pCNF (Figure 20B) and hence could not be directly compared to the sets of sub-clusters that were common between variants pAMF/Y and F/pNTF (Figure 20A). The low number of in silico inhibitory poses agreed with the low $n_i$ determined experimentally. The presence of two catalytically productive sub-clusters likely contributed to the higher $n_i$ value relative
to variant F and pNTF, and may also have contributed to the low substrate affinity as a result of a less-favourable enthalpic contribution.

The further decrease in aromatic ring electron density completely shifted the binding pocket preference from Pocket A to Pocket B, with 54% of productive poses occurring in a newly-created extended region of Pocket B. The sub-cluster was predicted to bind weakly with few active-site interactions with 3-HBA, and an unfavourable interaction between the polar aldehyde moiety of 3-HBA and the non-polar ring of F434 (Figure 22). Nevertheless, the aldehyde moiety was also positioned very close to the DHE-TPP intermediate in the extended pocket. The prediction of a weakly-binding but highly catalytic binding pocket within the active site of pCNF is in agreement with the high $K_{\text{m}}$ and $k_{\text{cat}}$, determined experimentally.

**Figure 22:** The predicted enzyme-substrate interactions of Pocket B (A) and the extended region of pocket B (B) of variant pCNF. Strong (<3.4 Å) and weak (<4.0 Å) H-bonds are shown as dark and light green circles with dotted lines; Van der Waal’s forces are shown as light green circles with no dotted lines; $\pi - \pi$ stacking and $\pi$-charge interactions are shown as pink and orange circles with dotted lines; and interactions between 0.6 and 0.7 Å are shown as red circles with dotted lines. Interactions were predicted in Discovery Studio 2017 using the minimised mutant structures and the catalytically active sub-cluster poses of 3-HBA in the mutant active sites.
Comparison of variant F and pNTF

The single catalytically active sub-cluster of variants F and pNTF was predicted to bind exclusively in Pocket B in identical orientations (Figure 20C), while each variant had two and one inhibitory sub-clusters, respectively. The inhibitory sub-clusters were low in energy and number, again agreeing with the experimental data that suggests a low number of inhibitory substrate orientations. Furthermore, the higher number of inhibitory sub-clusters and a higher overall proportion of inhibitory poses (20%) also explains the increased substrate inhibition observed in F compared to pNTF (0.5% of poses were inhibitory).

Variant F and pNTF were predicted to have near-identical interactions with 3-HBA in their catalytically productive orientations, the only difference being a stronger π-π stacking interaction with residue F385 compared to only weak Van der Waal’s forces with residue pNTF-385 (Figure 23). This difference was not observed in the experimentally-determined $k_{cat}$ and $K_m$, however, highlighting the caution that must be taken in the interpretation of computational docking, which at best provides only guidance for possible mechanistic explanations, and to inform future experimental work. Finally, the aldehyde moiety of 3-HBA was in close proximity to H473, potentially explaining the low $K_m$ values for variants F and pNTF and their low $k_{cat}$'s, with only weak interactions with the TPP_DHE moiety.

Figure 23: The predicted Pocket B enzyme-substrate interactions of variants F and pNTF, respectively. Strong (<3.4 Å) and weak (<4.0 Å) are shown as dark and light green circles with dotted
Van der Waal’s forces are shown as light green circles with no dotted lines; \( \pi \) - \( \pi \) stacking and \( \pi \) -charge interactions are shown as pink and orange circles with dotted lines; and interactions between 0.6 and 0.7 Å are shown as red circles with dotted lines. Interactions were predicted in Discovery Studio 2017 using the minimised mutant structures and the catalytically active sub-cluster poses of 3-HBA in the mutant active sites.

**DISCUSSION**

A comparison of the kinetic and stability parameters of each variant before and after accounting for misincorporation and \( \% TK_{\text{high}} \), demonstrated the impact both can have on the apparent parameters derived directly from experimental data. Previously, activity and stability studies of ncAA-incorporated variants have confirmed, but not quantified, incorporation and misincorporation by mass spectrometry. Determination of the fidelity of incorporation should now be a standard procedure in studies similar to this to allow a comparison of the true performance of ncAA-incorporated variants. The drastic change in the fidelity of incorporation from \( p \)NTF (82.9\%) to \( p \)AMF (50.5\%) has revealed a possible mechanism by which the evolved *Methanococcus jannaschii* tRNA/synthetase pair differentiated between ncAA substrates. Further investigation may pave the way for evolution towards either reduced misincorporation of particular ncAAs, or improved incorporation of new, phenylalanine-derived ncAAs.

This study began with a highly-evolved variant that had already been optimised at residue 385 through saturation mutagenesis with natural amino acids. The observed improvements in both catalytic activity and stability demonstrate the benefits of expanding directed evolution or designer libraries to include ncAAs. By targeting not only a single residue, but a particular property of an amino acid via unnatural amino incorporation, a small library of mutants was created which nonetheless had a broad range of catalytic properties. Variant \( p \)AMF was most thermostable, had a relatively high maximal activity, and strong substrate binding, but high susceptibility to substrate inhibition. Variant \( p \)CNF gave high catalytic turnover, and lower susceptibility to substrate inhibition, but weaker substrate affinity. Variants F and \( p \)NTF had lower maximum activities, but strong substrate binding and low susceptibility to substrate inhibition.

To this author’s knowledge, variant \( p \)AMF is the first example in which catalytic activity and stability have been simultaneously evolved via site-specific incorporation of ncAAs into an enzyme.
active site, and demonstrates the benefits of both fine-tuning pre-evolved residues using ncAAs, but also expanding directed evolution or designer libraries to include ncAAs in general.

**SUMMARY**

Chapter Four utilised the ncAA incorporation platform established in Chapter Three to evolve the active site of a previously-evolved transketolase variant, S385Y/D469T/R520Q, using a smart, small library of 5 para-substituted phenylalanine derivatives. A novel function - the Modified Michaelis-Menten function - was derived to describe and quantify the extensive substrate inhibition observed at high [3-HBA]. The use of simultaneous equations within OriginLabs to deconvolve the true catalytic/stability parameters of the ncAA-incorporated species from the misincorporated species was also a novel approach to the problem of misincorporation that is often present but not always reported.

The output from this chapter was a series of variants with improved catalytic parameters. The best all-round performer was pAMF, which had a 2.6-fold, 2.4-fold and 1.1-fold improvement in catalytic efficiency, $K_m$ and $k_{cat}$, respectively, and a 5.4 °C increase in $T_m$. The latter was a result of a strong inter-subunit H-bond between 385pAMF and the carbonyl backbone of G262. To date, only a handful of studies have successfully improved (84–86) or introduced novel catalytic function (87,92), or enhanced thermostability (93) via site-specific, active-site ncAA incorporation. While the latter study did also observe a modest 15% improvement in catalytic activity at 40 °C relative to wild-type, the thermostable mutant was less-active than the wild-type at 23 °C, suggesting the improved activity reflected only the improved thermostability at 40 °C rather than a genuine improvement in catalysis. To my knowledge, pAMF is the first example of an ncAA-incorporated variant with both enhanced activity and stability at 22 °C, and demonstrates the benefits of including ncAAs in site-specific, directed evolution libraries.

This research provides a great example of the power of protein engineering with ncAAs in order to augment catalytic performance and/or stability. It may serve as a useful template outlining one approach to protein engineering via rational active site mutagenesis and subsequent analysis of structure-function relationships.
CHAPTER FIVE: Novel insights into activation by cofactor binding suggests a Two-Species Model of transketolase interconversion

As the \( pCNF \) mutant had already been made, and this was reported previously to have altered fluorescence compared to tryptophan/tyrosine (94,95), it was briefly explored whether it would provide a fluorescence probe into the topology, dynamics, substrate or cofactor binding within the active site. Details are in Appendix 2, but briefly, it was found that TPP quenched the \( pCNF \) fluorescence signal, and at most the \( pCNF \) could be used to measure TPP binding. Furthermore, the signal was convoluted with fluorescence of the TPP and tryptophan/tyrosine from the protein itself. This led us to investigate whether intrinsic fluorescence at any excitation or emission wavelength combination could be used to determine the binding parameters of the cofactors, and potentially also for the substrates, independently of enzyme activity.

INTRODUCTION

Transketolase (TK) (EC 2.2.1.1) is a key thiamine pyrophosphate (TPP)-dependent enzyme in the non-oxidative phase of the pentose phosphate pathway (PPP), which branches the glycolytic pathway and diverts metabolic flux through biosynthetic pathways such as pentose sugars and ribose-5-phosphate biosynthesis, a precursor for nucleotide biosynthesis. Another important product of the PPP is the redox cofactor NADPH, which is also an antioxidant that protects against oxidative damage often caused by reactive oxygen species (ROS). The PPP is therefore upregulated during oxidative stress and has several regulatory control points (96–98), although no such regulation has been characterised previously for transketolase. There is also evidence of upregulation of the PPP in response to other cell stresses, such as osmotic, heat, and heavy metal stress (99,100).

Yeast and \( E. \) \( coli \) apo-transketolases exist as monomers that form homodimers at higher protein concentrations (101). Upon cofactor binding, both the apo-monomer and apo-dimer form a catalytically active homodimer of apparently structurally-identical subunits, with two active sites per homodimer located at the subunit interface (Figure 24) (18,29). In its fully active form, each active site is occupied by one molecule of TPP and one divalent cation (\( M^{2+} \)) such as \( Ca^{2+}, Mg^{2+} \) or \( Mn^{2+} \), meaning each catalytically-active protein homodimer can bind two TPP molecules and two \( M^{2+} \) ions.
At low concentrations, the inactive apo-transketolase monomer from yeast is activated slowly upon addition of the two cofactors, to form the active homodimer (102,103). At higher concentrations, cofactor binding to the inactive apo-transketolase dimer leads to the structural organisation of two disordered cofactor-binding loops in the active site, to form the active holoTK homodimer (103).

![Diagram of oligomeric states of E. coli apo-transketolase (apoTK) and holo-transketolase (holoTK).](image)

**Figure 24:** A schematic diagram of the oligomeric states of *E. coli* apo-transketolase (apoTK) and holo-transketolase (holoTK). For simplicity, divalent cations are not shown.

The two active sites of several TPP-dependent enzymes have been reported to be non-equivalent in terms of their cofactor affinities (104–106), substrate binding (half-of-the-sites reactivity) (107,108) and, in the case of transketolase, their inactivation profiles (101,109,110). crystallographic evidence for non-identical active sites has so far been only reported for the E1 component of the pyruvate dehydrogenase complex (PDHc-E1) (111), which highlighted differences in the flexibility of one region near the active site, in each of the two subunits, and described a proton wire between the two TPP-binding sites that potentially mediated their cooperativity in the reaction cycle. However, crystal structures obtained to date for TK have not revealed any large structural differences between the two active sites (18,23,26,29). Nevertheless, the same proton wire could be found, plus a slight difference in the temperature factor (B-factor) distribution between each subunit in the yeast TK crystal structure (26), which manifests itself as ‘noisier’ or less well-defined coordinates in one subunit. Subsequent re-analysis of the B-factor distribution, along with noted differences in orientation of Trp391, Tyr370 - which play a role in stabilising the cofactor binding loop - and a few surrounding residues in the two subunits, was later interpreted as a slight
asymmetrical strain between subunits that may be caused by the one-by-one destabilisation of the holo-sites that leads to a permanent oscillation between two asymmetric states (112,113).

Transketolase is inactive when bound to only divalent cations, while addition of TPP at relatively high concentrations in the absence of Mg\(^{2+}\) can achieve up to 70% activity relative to that in the presence of saturating Mg\(^{2+}\) concentrations and non-saturating [TPP] (114,115). No dissociation constants for TPP have been measured in the absence of Mg\(^{2+}\). Addition of Mg\(^{2+}\) results in the formation of two active sites that can both bind TPP with a significantly higher affinity relative to that in the absence of Mg\(^{2+}\) (115,116). The cooperativity of binding when Mg\(^{2+}\) is the divalent metal ion is a contentious point with no consensus view. The literature ranges from positive (117) to negative cooperativity (101), and then also from active sites with a single dissociation constant (33,115,116,118,119), to active sites with two dissociation constants (109). It is clear, however, that a degree of non-equivalence exists, whether it is in the form of positive or negative TPP-binding cooperativity. While the binding of cofactors to yeast transketolase has been studied extensively, the binding of TPP to \textit{E. coli} transketolase is less well characterised.

The present study was undertaken to directly measure TPP-binding to \textit{E. coli} transketolase in the presence and absence of magnesium ions using a novel fluorescence-quenching assay. The initial aim was to detect TPP binding directly to provide a needed consensus on both the change in dissociation constant, and the binding cooperativity, over a large range of Mg\(^{2+}\) and TPP concentrations. Following the unexpected detection of two distinct transketolase species with stark differences in affinity and activity, the origin of the ratio of the two species was probed further, and linked to a specific chemical oxidation.

**EXPERIMENTAL PROCEDURES**

\textit{Enzyme preparation}

Wild-type transketolase with an N-terminal His6-tag was expressed in \textit{E. coli} XL10-gold cells (Agilent Technologies Ltd) from the plasmid \textit{pQR791}. The cell pellet was lysed and purified as described previously (41). Purified transketolase was ultrafiltrated four times using Amicon Ultra-4 10k MWCO centrifugal filter to remove excess imidazole/cofactors and subsequently dialysed overnight at 4 °C in 50 mM Tris-HCl, pH 7.0 to obtain apoTK. Protein concentration was determined by absorbance at 280 nm in 6 M Guanidine-HCl, 20 mM Sodium Phosphate, pH 6.5. Absorbance was measured using a Nanodrop spectrophotometer, assuming a monomeric molecular weight of
73035.5 g mol\(^{-1}\) and an extinction coefficient of 92630 L mol\(^{-1}\) cm\(^{-1}\). Series of 2x concentrated cofactor solutions were prepared and purified TK was added to a final concentration of either 0.05 mg/mL or 0.2 mg/mL. The samples were incubated at 22 °C for 45 minutes to allow TK-TPP binding to reach equilibrium.

**Fluorescence assay to detect TPP binding**

After incubation, the fluorescence intensity of the TK-cofactor samples was determined in a 1.5mm x 1.5mm quartz cuvette (Hellma UK Ltd) using a Fluoromax-4 spectrofluorometer (\(\lambda_{	ext{ex}} = 240\) nm; \(\lambda_{	ext{em}} = 330\) nm; integration time = 0.1 s; slit width = 8 nm). One measurement per sample was taken, or five measurements per cofactor concentration. The data were corrected for the wavelength dependence of the Xe-lamp intensity and source intensity fluctuations. Inner filter effect (IFE) correction factors (CF) were generated as follows: A stock TPP solution was diluted with 50 mM Tris-HCl buffer to generate 5 series of TPP samples between 0 and 1.2 mM TPP. A correction factor was subsequently generated from these data-points by fitting to the function described by MacDonald et al. (120), which accounts for non-linearity in the fluorescence emission. A new correction factor was generated each day the Fluoromax-4 spectrofluorimeter was used.

**Analytical Ultracentrifugation (AUC) Measurements**

Analytical ultracentrifugation (AUC) measurements were performed in the Molecular Interaction Facility at UCL by Dr Jayesh Gor. AUC data were obtained for wild-type transketolase in the absence of cofactors Mg\(^{2+}\) and TPP at 20 °C on a Beckman XL-1 instrument equipped with AnTi50 rotors. Data were collected at rotor speeds of 40,000 rpm in two-sector cells with column heights of 12 mm. The software SEDFIT was used to analyse the sedimentation data by fitting the experimental interference data using direct boundary Lamm fits of up to 500 scans (121,122). The resulting size distributions c(s) of oligomers within samples assumed that all species have the same frictional ratio \(f/f_0\). The c(s) fit was optimised by floating \(f/f_0\) and the baseline until a sufficiently low root mean square deviation was reached and the visual appearance of the fits were satisfactory. The ratio or percent monomer/dimer within each sample was derived by integrating each peak in the c(s) integration function and the sedimentation coefficient of dimeric TK, \(S_{w,20}\), was determined as a function of [TK].
Transketolase activity assay

Purified, dialysed apo-transketolase (0.2 mg/mL) was incubated with 50 µM or 2.4 mM TPP and 9 mM Mg²⁺ for 45 minutes at 22 °C. 50 µL was added to 100 µL 150 mM GA, 150 mM HPA, giving final substrate concentrations of 50 mM. The reaction was performed in triplicate at 22 °C in a 96 well plate with shaking at 300 rpm using a Thermomixer Comfort shaker. 10 µL of the reaction was quenched with 190 µL 0.1% trifluoroacetic acid (TFA) after 3, 5, 10, 15, 20, 30, and 40 minutes. Samples were subsequently analysed by a Dionex HPLC system (Camberley, UK) with a Bio-Rad Aminex HPX-87H reverse phase column (300 x 7.8 mm²) (Bio-Rad Labs., Richmond, CA, USA), via Chromeleon client 6.60 software, to separate and analyse the change in the concentration of substrate (GA) and product (Ery) over the course of the reaction using the method described previously (123).

Oxidation of transketolase using cumene hydroperoxide (CHP)

Oxidised samples of transketolase were prepared by the addition of 0.005% CHP to E. coli cells half-way through an eight-hour fermentation. E. coli cells were subsequently harvested, and transketolase purified and dialysed, as outlined above.

Mass spectrometry

LC-ESI-MS was performed as described in General Experimental Procedures.

Dimedone labelling

Apo-transketolase was prepared as described above. 2 mg/mL apo-transketolase was incubated with 2 mM dimedone in the presence/absence of 3.8 M urea and 2 mM iodoacetamide. The samples were subsequently dialysed overnight at 4 °C in 50 mM Tris-HCl, pH 7.0 to obtain dimedone-labelled apoTK. Samples were centrifuged at 15,000 x g for 5 minutes to remove aggregates and prepared for intact protein LC-ESI-MS as described above.

Peptide digest-Mass spectrometry (performed by Clive Metcalfe at NIBSC)

Trypsin digestion

50 µg of each sample was added to a 10 kDa 500 µl centrifugal concentrator (Vivacon 500, Sartorious) and denatured with 8M urea for 1 hour at room temperature. The samples were washed x5 with 25mM ammonium bicarbonate and trypsin digested as reported previously (124). Briefly, the
sample was resuspended in 200 µl 25 mM ammonium bicarbonate and 1 µg trypsin added and left overnight at 37°C with shaking. The tryptic peptides were eluted through the membrane into a fresh Eppendorf by sequential washes with 0.1% formic acid, 50% acetonitrile containing 0.1% formic acid and 0.1% formic acid in acetonitrile. The samples were lyophilised in vacuo.

**Enrichment of dimedone peptides**

The pooled eluents containing tryptic peptides were passed over a 50 µl monomeric avidin micro-column. The flow-through that contained all biotin labelled peptides were collected and evaporated to dryness. Biotin labelled peptides were eluted with acidified acetonitrile (500 µl, 0.4% TFA in 30% acetonitrile) lyophilised in vacuo.

**Mass spectrometry**

The tryptic peptide samples were resuspended in 0.1% formic acid aided by sonication and approximately 1 µg of material injected onto an Ultimate 3000 nano HPLC system coupled to an Orbitrap XL Discovery mass spectrometer (Thermo Scientific). Samples were online desalted on a µ-Precolumn (C18 PepMap100, 300 µm id × 5 mm; 5 µm, 100 Å) at a flow rate of 25 µL/min, which was followed by separation on a nano analytical column (Acclaim PepMap100 C18, 75 µm id × 50 cm, 3 µm, 100 Å) (Thermo Scientific) using a 90-minute linear gradient from 5 to 40% solvent B (98% CH3CN/2% H2O/0.1% formic acid, v/v/v) versus solvent A (98% H2O/2% CH3CN/0.1% formic acid, v/v/v) at a flow rate of 300 nl/min. The mass spectrometer was operated in a data-dependent acquisition mode. The full survey scan (m/z 400-2000) was acquired in the Orbitrap with a resolution of 30,000 at m/z 400, which was followed by five MS/MS scans in which the most abundant peptide precursor ions detected in the preceding survey scan were dynamically selected and subjected for collision-induced dissociation (CID) in the LTQ (linear ion trap) to generate MS/MS spectra (‘top-5 method’).

**Data analysis**

The data files from the mass spectrometry runs were combined and searched against the Swiss-Prot database using Peaks 8 proteomics studio (Bioinformatics Solutions Inc. On, Canada). Precursor mass tolerance was 10 ppm and a fragment ion tolerance was 0.6 Da with up to two missed trypsin cleavage sites per peptide allowed. Variable modifications were defined as deamidation on asparagine and glutamine, oxidation on methionine, carbimidomethylation on cysteines, oxidation on cysteines and dimedone on cysteines (either dimedone or maleimide dimedone). de-novo, peaks-db,
peaks PTM and SPIDER algorithms were sequentially used to search against a concatenated target/decoy database, providing an empirical false discovery rate (FDR) and results are reported at a 1% target/decoy FDR for both peptides and proteins.

RESULTS

A novel fluorescence quenching-based cofactor-binding assay

The intrinsic fluorescence of transketolase is quenched upon TPP binding (115) when excited at 280 nm. Further analysis indicated that TPP has two strong absorption bands at 233 nm and 266 nm (Figure A5, Appendix Two) and is itself weakly fluorescent when excited at either of these wavelengths. It was therefore decided to excite samples at the lower absorption band (\(\lambda_{ex} = 240\) nm; \(\lambda_{em} = 330\) nm). As reported at 280 nm, transketolase fluorescence was quenched upon TPP binding after excitation at 240 nm (Figure 25). The signal generated was corrected for the inner filter effect (IFE), that arises from strong absorption of a proportion of the incident light by free ligand (TPP) before it can excite the sample (‘primary’ IFE), and which therefore decreases the observed fluorescence. The correction factor was determined empirically from the fluorescence intensity of free TPP in 50 mM Tris-HCl buffer as described by MacDonald et al. (120) (Figure 26).
Figure 25: Evidence of fluorescence quenching upon cofactor-binding. A) The observed fluorescence intensity of 0.05 mg/mL TK and 0 (black), 60 (red), 300 (green), and 800 (blue) μM TPP when excited at 240 nm. B) The observed (black) and corrected (red) fluorescence intensity of each sample (λ_{ex} = 240 nm; λ_{em} = 330 nm). Fluorescence intensity was corrected for the inner filter effect (IFE) according to MacDonald et al., (120).

![Graph of fluorescence intensity vs. [TPP] (µM)](image)

Figure 26: The fluorescence intensity of free TPP in 50 mM Tris-HCl buffer (λ_{ex} = 240 nm; λ_{em} = 330 nm), fitted to the correction function described by MacDonald et al. (120).

Characterisation of TPP binding to wild–type E. coli transketolase

Since the original detection of TPP binding in the absence of divalent cations (114,115) there have been no reports of the associated binding parameters. These were therefore determined, in addition to carrying out the first comprehensive study of TPP binding over large cofactor concentration ranges, for purified E. coli transketolase.

The form of the binding isotherm of TPP with TK was found to be dependent on [Mg^{2+}], and indicated a shift in binding affinity and cooperativity at higher [Mg^{2+}] (Figure 27). The weighted sum
of two Hill functions (the double-Hill function), describing two independent binding events, both cooperative in nature, was found to give a significantly better fit to the TPP-binding isotherms compared to a standard Hill function (Figure 27A and 28) and was thus used to determine the TPP binding parameters at various [Mg$^{2+}$] from 0-18 mM. This identified two dissociation constants $K_{d(high)}$ and $K_{d(low)}$, which were independent from each other, reflecting two distinct populations, TK$_{high}$ and TK$_{low}$, as described below.

The double-Hill function: 

$$\theta = \frac{B_{max(high)}[L]^{n_{high}}}{K_{d(high)}^{n_{high}} + [L]^{n_{high}}} + \frac{\alpha B_{max(high)}[L]^{n_{low}}}{K_{d(low)}^{n_{low}} + [L]^{n_{low}}}.$$

where $\theta$ is the fractional saturation, the fraction of [protein] bound to ligand; $B_{max(high)}$ is the proportion of TPP that binds TK$_{high}$; $\alpha B_{max(high)}$ is the proportion of TPP that binds TK$_{low}$; [L] is the ligand concentration; $n_{high}$ and $n_{low}$ are the Hill coefficients of TK$_{high}$ and TK$_{low}$, respectively; and $K_{d(high)}$ and $K_{d(low)}$ are the dissociation constants of TK$_{high}$ and TK$_{low}$, respectively.

**Figure 27:** Experimental data of 0.05 mg/mL wild-type TK binding to TPP at 0 mM (black), 1 mM (red), 4.5 (blue), 9 mM (magenta) and 18 mM (green) Mg$^{2+}$. Experimental data-points A) at all [Mg$^{2+}$] fitted to the double-Hill function; Ai) at 9 mM Mg$^{2+}$ plotted on a logarithmic x axis and fitted to a single Hill function (thin black) and the double-Hill function (thick black). The contributions of the
high (red) and low affinity (blue) binding sites of the double-Hill function are also shown; B) at 9 mM Mg$^{2+}$ (double-Hill function) with the contributions of the high and low affinity binding sites shown as dashed lines; normalised contributions to double-Hill functions, of the C) high affinity and D) low affinity binding sites at each [Mg$^{2+}$].

Figure 28: Experimental data of 0.05 mg/mL wild-type transketolase binding to TPP at A) and B) 0 mM; C) and D) 1 mM; E) and F) 4.5 mM; G) and H) 9 mM; and I) and J) 18 mM Mg$^{2+}$. Experimental data-points were fitted to either a single- (black) or double- (black) Hill function. A logarithmic x-axis was used to demonstrate the superior fit to the double-Hill function.

The TPP-binding parameters determined using the single-Hill function (Table 4) had no relatability to previously determined parameters; at higher [Mg$^{2+}$], the fits either had very large associated errors or the data couldn’t be fitted to the single-Hill function. However, those of TK$_{\text{high}}$, determined by fitting the data to the double-Hill function (Table 5), correlated well with previously reported values. It was therefore concluded that two independent binding events had been detected, both cooperative, which fitted best to a double-Hill function.
Table 4: Summary of the TPP-binding parameters when fitted to a single Hill function. A TK concentration of 0.05 mg/mL was used in each binding assay. Associated errors are the fitting error for the single-Hill function. The data couldn’t be fitted to the single-Hill function at 18 mM Mg$^{2+}$.

<table>
<thead>
<tr>
<th>[Mg$^{2+}$] (mM)</th>
<th>$K_d$ (µM)</th>
<th>±</th>
<th>n</th>
<th>±</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>120</td>
<td>5</td>
<td>1.20</td>
<td>0.04</td>
</tr>
<tr>
<td>1</td>
<td>139</td>
<td>9</td>
<td>0.74</td>
<td>0.02</td>
</tr>
<tr>
<td>4.5</td>
<td>739</td>
<td>1030</td>
<td>0.42</td>
<td>0.06</td>
</tr>
<tr>
<td>9</td>
<td>4390</td>
<td>10100</td>
<td>0.34</td>
<td>0.04</td>
</tr>
<tr>
<td>18</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 5: Summary of the binding parameters of the high affinity binding sites of TK$_{high}$. A TK concentration of 0.05 mg/mL was used in each binding assay. Associated errors are the fitting error for the double-Hill function.

<table>
<thead>
<tr>
<th>[Mg$^{2+}$] (mM)</th>
<th>$K_d$$_{high}$ (µM)</th>
<th>±</th>
<th>$n$$_{high}$</th>
<th>±</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>113</td>
<td>40</td>
<td>0.67</td>
<td>0.30</td>
</tr>
<tr>
<td>1</td>
<td>20.6</td>
<td>4.2</td>
<td>1.01</td>
<td>0.14</td>
</tr>
<tr>
<td>4.5</td>
<td>3.79</td>
<td>0.38</td>
<td>1.98</td>
<td>0.38</td>
</tr>
<tr>
<td>9</td>
<td>2.29</td>
<td>0.23</td>
<td>1.36</td>
<td>0.16</td>
</tr>
<tr>
<td>18</td>
<td>1.31</td>
<td>0.27</td>
<td>0.76</td>
<td>0.06</td>
</tr>
</tbody>
</table>

a) Binding affinity of TK$_{high}$

The fluorescence quenching data revealed a high-affinity TPP-binding event with a dissociation constant, $K_d$$_{high}$, of 113 ± 40 µM ((Figure 27); Table 5) at 0 mM Mg$^{2+}$, consistent with the only comparable values available in the literature for E. coli transketolase $K_d$ = 29 µM and 8 µM at 0.01 mM and 0.1 mM Mg$^{2+}$, respectively), which were obtained indirectly from kinetic enzyme activity data (39).
The affinity of TK$_{\text{high}}$ for TPP improved markedly in the presence of Mg$^{2+}$, with an 86-fold decrease in the dissociation constant at 18 mM Mg$^{2+}$ relative to that in the absence of Mg$^{2+}$ (Figure 29; Table 5). By fitting the obtained dissociation constants to a Hill function (Figure 29A), the dissociation constant at fully saturating [Mg$^{2+}$] was estimated to be 0.93 ± 0.27 μM, a 121-fold improvement compared to that in the absence of Mg$^{2+}$. The largest decrease in $K_{d(\text{high})}$ was observed at low Mg$^{2+}$ concentrations < 1 mM, as noted by Kochetov et al. (114). The dependence of $K_{d(\text{high})}$ on [Mg$^{2+}$] fitted to a Hill function with $n = 1.33 ± 0.14$, implying that TK$_{\text{high}}$ binds the two Mg$^{2+}$ ions with slightly positive cooperativity between the two Mg$^{2+}$ binding sites.

Figure 29: Dependence of TPP-dissociation constants and Hill coefficients, of the high and low affinity binding sites for 0.05 mg/mL TK, at 0 mM, 1 mM, 4.5 mM, 9 mM and 18 mM Mg$^{2+}$. A) and B) the dissociation constants of the high and low affinity binding sites, respectively. C) and D) the Hill coefficients of the high and low affinity binding sites, respectively. Solid lines represent quantified, fitted graphs; dashed lines represent clear, unquantified trends.
b) Binding cooperativity of \( \text{TK}_{\text{high}} \)

In the absence of \( \text{Mg}^{2+} \), the Hill coefficient, an indicator of the degree of cooperativity, indicated negative cooperativity between TPP binding sites \( n_{\text{high}} = 0.67 \pm 0.30 \), implying that in the absence of \( \text{Mg}^{2+} \), binding of the first TPP molecule reduces affinity for the second TPP molecule (Figure 29C). The presence of the high-affinity sites corresponding to those of catalytically active holo-transketolase in the absence of \( \text{Mg}^{2+} \) confirmed that the divalent cations are not necessary for transketolase activity at high [TPP].

The presence of \( \text{Mg}^{2+} \) ions at low concentrations appears to increase the cooperativity of TPP binding. However, at increasingly high \([\text{Mg}^{2+}] > 4 \text{ mM}\), the Hill coefficient of TPP binding shifts from exhibiting strong positive-cooperativity, with an inferred total of two TPP molecules binding, to strong negative cooperativity (Table 5; Figure 29C).

The dependence of cooperativity on \([\text{Mg}^{2+}]\) explained why there was no consensus view, as the previous studies each used different \( \text{Mg}^{2+} \) concentrations. However, a collective review of the studies on yeast transketolase that published a Hill coefficient suggested a similar trend, albeit at lower \([\text{Mg}^{2+}]\) given the lower intracellular free \([\text{Mg}^{2+}]\) of yeast relative to \( \text{E. coli} \) (125): strong positive cooperativity \( (n \approx 2) \) at 1 mM \( \text{Mg}^{2+} \) (117), slight negative cooperativity at 2 mM \( \text{Mg}^{2+} \) (109) and strong negative cooperativity \( (n = 0.61) \) at 3 mM \( \text{Mg}^{2+} \) (101).

The molecular basis of the trend reversal at \([\text{Mg}^{2+}] > 4 \text{ mM}\) is unclear but may reflect the opposing effects of binding the first and second \( \text{Mg}^{2+} \) ions on TPP binding. While the binding of the first \( \text{Mg}^{2+} \) ion may result in cooperative binding of two TPP molecules and significantly improve affinity, binding of a second \( \text{Mg}^{2+} \) may remove the positive interaction between the first and second TPP-binding sites, and hence introduce negative cooperativity between active sites. It is likely that both TPP binding sites are occupied, even at high \([\text{Mg}^{2+}]\), because a similar maximum change in raw fluorescence signal was observed across all \([\text{Mg}^{2+}]\). It therefore follows that at higher \([\text{Mg}^{2+}]\) the apparent dissociation constants may in fact be a convolution of the binding parameters of two non-equivalent sites. Given the sensitivity of fluorescence measurements, it may be possible to concentrate data points over a much smaller range at low TPP concentrations and deconvolve the two separate dissociation constants.

**Detection and characterisation of a novel low-affinity species, \( \text{TK}_{\text{low}} \)**

The fluorescence quenching data fitted best to the weighted sum of two Hill functions (the double-Hill function) that described two independent, multi-ligand binding events; one with a high
affinity dissociation constant (described above) and one with a low affinity dissociation constant (Figure 29; Table 6). This low affinity TPP-binding event has never been previously characterised in *E. coli* nor *S. cerevisiae* transketolase. As transketolase exists as both monomer and dimer, depending on [TPP] and [TK], three dimeric species are, in theory, possible: TK\textsubscript{high}-TK\textsubscript{high}, TK\textsubscript{low}-TK\textsubscript{low} and the mixed dimer species, TK\textsubscript{high}-TK\textsubscript{low}. The arrangement of TK\textsubscript{high} and TK\textsubscript{low} as monomers, homo-, and hetero-dimers before TPP-binding may potentially impact the apparent dissociation constants and Hill coefficients of TK\textsubscript{high} and TK\textsubscript{low} subunits. The double-Hill function used to fit the data made no assumptions regarding the conformational arrangements of TK\textsubscript{high} and TK\textsubscript{low}; the derived binding parameters are, therefore, a convolution of the true binding parameters of TK\textsubscript{high} and TK\textsubscript{low} in each conformational arrangement. I attempted to fit the data to a triple-Hill function, but a third binding event could not be resolved. There was insufficient evidence to rule out the mixed dimer species, however; each sub-unit may simply have identical or near-identical binding parameters as the TK\textsubscript{high} and TK\textsubscript{low} subunits of the homodimers.

<table>
<thead>
<tr>
<th>[Mg\textsuperscript{2+}] (mM)</th>
<th>$K_d$\textsubscript{low} (µM) ±</th>
<th>n\textsubscript{low} ±</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>129 ± 20</td>
<td>1.51 ± 0.26</td>
</tr>
<tr>
<td>1</td>
<td>230 ± 17</td>
<td>1.03 ± 0.13</td>
</tr>
<tr>
<td>4.5</td>
<td>247 ± 24</td>
<td>1.16 ± 0.11</td>
</tr>
<tr>
<td>9</td>
<td>276 ± 25</td>
<td>1.15 ± 0.10</td>
</tr>
<tr>
<td>18</td>
<td>310 ± 24</td>
<td>1.27 ± 0.10</td>
</tr>
</tbody>
</table>

*Table 6*: Summary of the binding parameters of the low affinity binding sites of TK\textsubscript{low}. A TK concentration of 0.05 mg/mL was used in each binding assay. Associated errors are the fitting error for the double-Hill function.

**a) Binding affinity of TK\textsubscript{low}**

Interestingly, the addition of Mg\textsuperscript{2+} to TK\textsubscript{low} had the opposite effect on TPP binding compared to TK\textsubscript{high}. The presence of Mg\textsuperscript{2+} decreased the affinity of TK\textsubscript{low} for TPP, although to a much lesser degree, with a 2.4-fold increase in $K_d$\textsubscript{low} as [Mg\textsuperscript{2+}] increased from 0 mM to 18 mM (Figure 29B; Table 6). Although the dissociation constant clearly increased with [Mg\textsuperscript{2+}], it was difficult to define anything other than the general trend when fitted to a Hill function.
b) Binding cooperativity of TK\textsubscript{low}

The trend in the binding cooperativity of TK\textsubscript{low} was also the inverse of TK\textsubscript{high}. At low [Mg\textsuperscript{2+}], $n_{\text{low}}$ initially decreased to 1 and hence removed cooperativity between active sites; above 2 mM Mg\textsuperscript{2+} $n_{\text{low}}$ increased slightly but not significantly (Figure 29D; Table 6). Generally, the cooperativity of binding to TK\textsubscript{low} was less cooperative than TK\textsubscript{high}, which may indicate a disruption in the cross-talk and connectivity between active sites in the TK\textsubscript{low} dimer, potentially through disturbance of the proton wire.

c) $\%B_{\text{max(high)}}$ was invariant to [Mg\textsuperscript{2+}]

By fitting the data to the double-Hill function, the ratio of TPP binding to TK\textsubscript{high} and TK\textsubscript{low} ($B_{\text{max}}$ ratio), which reflected the ratio of [TK\textsubscript{high}]:[TK\textsubscript{low}], was determined. For clarity, henceforth $\%B_{\text{max(high)}}$ will represent the percentage of all TK that is TK\textsubscript{high}. The proportion of fluorescence quenching attributed to the high-affinity event was independent of [Mg\textsuperscript{2+}] as the global $\%B_{\text{max(high)}}$, derived from a global fit of all TPP-binding data-sets to the double-Hill function with the $B_{\text{max}}$ parameter shared, was 33.6 ± 2.9% over all [Mg\textsuperscript{2+}].

d) $\%B_{\text{max(high)}}$ was invariant to [TK] and the apoTK monomer-dimer equilibrium

Analytical ultracentrifugation (AUC) data indicated an increase in [TK] shifted the equilibrium between the monomeric and dimeric forms of apoTK towards dimeric apoTK (Figure 30 and 31A). An increase in [TK] from 0.05 mg/mL, the TK concentration in the previous binding assays, to 0.2 mg/mL, increased the fraction of apoTK dimer from 62.4% to 83.4%. In contrast, when the TPP-binding assay was repeated at 0.2 mg/mL, the $\%B_{\text{max(high)}}$ was 32.9 ± 1.5% (Figure 31B) compared to 33.6 ± 2.9% at 0.05 mg/mL TK, and was hence invariant to both [TK] and the resulting shift in the monomer-dimer equilibrium prior to TPP-binding.
Figure 30: AUC interference data and the derived c(s) distribution and sedimentation coefficients. Plots of c(s) distribution vs sedimentation coefficient at A) 0.01-0.8 mg/mL TK. The right-hand plot shows the AUC interference fringe data (blue) and best fit (white) from 0.8 mg/mL TK.

Figure 31: Comparison of A) apo-dimer formation as a function of [apoTK] and B) TPP-binding at 0.2 mg/mL. A) The change in sedimentation coefficient, $S_{w,20}$, which is equivalent to the change in dimer fraction (right axis), at pH 7.0 as a function [apoTK]. Experimental data was fitted to the weighted sum of two dimerization functions (126) describing a two-dimer system comprised of TK$_{high}$TK$_{high}$ and TK$_{low}$TK$_{low}$, as described. B) Experimental data of 0.2 mg/mL wild-type TK binding to TPP at 9 mM Mg$^{2+}$, fitted to a double-Hill function.

**TK$_{high}$ and TK$_{low}$ are two distinct forms of transketolase**

The co-existence of two distinct forms of TK, with different affinities for TPP, but with no previous structural evidence to suggest how they differ, prompted me to look for chemical changes by mass spectrometry.
a) Detection of transketolase post-translational modification by mass spectrometry

Liquid chromatography electrospray ionisation mass spectrometry (LC-ESI-MS) of purified wild-type transketolase revealed a major unmodified transketolase species at the predicted molecular weight of 73,035 Da, and two defined, higher molecular weight species with smaller peak areas (Figure 32A). At even higher molecular weights, other yet smaller peaks were present, though not well resolved. The peaks were fitted to the sum of multiple Gaussian functions to determine the peak area and hence abundance of the unmodified transketolase species, $\text{TK}_{\text{unmodified}}$, relative to the first two, best defined, higher molecular weight species, defined as $\text{TK}_{\text{modified}}$ (i.e. peak two and three combined). The $\%\text{TK}_{\text{modified}}$ was $31.0 \pm 1.7\%$, comparable to the $\%\text{TK}_{\text{high}}$ ($33.6 \pm 2.9\%$) obtained from the global $\%B_{\text{max(high)}}$, implying that $\text{TK}_{\text{low}}$ was the unmodified transketolase species, whilst $\text{TK}_{\text{high}}$ formed as the result of post-translational oxidation during fermentation or purification. The average difference in molecular weight between each of the first three transketolase peaks was 15.8 Da, the equivalent of an oxygen atom. Furthermore, over-oxidation of sulfenic acid leads to the formation of sulfinic and sulfonic acids, each leading to further 16 Da increases in molecular weight per oxidation.

Figure 32: The mass spectra of wild-type transketolase in the absence and presence of 0.005% CHP added half-way through an eight-hour fermentation, and its impact on $\%\text{TK}_{\text{high}}$, TPP-binding and
activity. The mass spectra of purified wild-type transketolase expressed in A) the absence and B) the presence of 0.005% CHP added half-way through an eight-hour fermentation, fitted to the sum of multiple Gaussian functions. The major peak corresponds to unmodified transketolase, while the next two peaks correspond to modified transketolase. Higher molecular-weight peaks correspond to inactive, over-oxidised TK. C) Experimental TPP-binding data of 0.05 mg/mL wild-type TK expressed in the presence (black) and absence (red) of 0.005% CHP, with 9 mM Mg^{2+}, fitted to a double-Hill function. Dotted lines represent the contribution from \( TK_{\text{high}} \) and hence the \( \%B_{\text{max}}(\text{high}) \). D) Activity data of purified 0.067 mg/mL wild-type transketolase; black: 50 µM TPP, 9 mM Mg^{2+}; red: 2.4 mM TPP, 9 mM Mg^{2+}; blue: wild-type TK expressed the presence of 0.005% CHP, with 2.4 mM TPP, 9 mM Mg^{2+}.

The mass spectrometry data alone is insufficient to categorically rule out \( TK_{\text{modified}} \) as a collection of singly-oxidised species, modified at different residues located near the TPP-binding site (e.g. Met153, Cys157, Met158, Met159, Cys167). Equally, the more-oxidised, inactive species with the highest molecular weight peaks are potentially a combination of over-oxidised states at these residues. I therefore attempted to pinpoint the oxidation modification by re-analysis of previously-published X-ray crystal structures.

**b) Crystallographic evidence of hydroxylation at Cys157**

It was indicated previously from X-ray crystal structure data (29) that Cys157, located near the TPP-binding site, was present in an unusual hydroxylated form, which would have an increased molecular weight of 16 Da relative to unmodified cysteine (127). Later work found that direct air-oxidation *in vitro* was deactivating for TK, though only in the presence of TPP (127), and the authors suggested that the crystallographic evidence for a hydroxylated Cys157 may have been an artefact of the crystallization process. The mass spectrometry data indicated that in fact the hydroxylation of Cys157 was not an artefact of crystallization, but rather occurred during expression or purification of the enzyme. Interestingly, at least one of the oxidised states was more active than the unmodified TK, which means that the inactivated product of air-oxidation observed previously by Mitra et al. (127), was not \( TK_{\text{high}} \), and that the two oxidation mechanisms are different, or perhaps that air-oxidation led to lower activity through over-oxidation at the same site.

The detection of transketolase hydroxylation via mass spectrometry prompted me to re-inspect the published electron density maps for the crystal structure of wild-type *E. coli* holo-transketolase (1QGD). In the published structure, Cys157 was retained as unmodified sulphydryl.
However, closer inspection of the $2F_0F_C$ and $F_0F_C$ electron density difference maps around residue Cys157 in subunit A and B, at 3σ omit level (Figure 33), suggested a substantial amount of electron density that has not been accounted for in the structural model. This difference in electron density can be explained by the existence of populations of singly and doubly-oxidised Cys157 that form sulfenic and sulfinic acids, respectively, as also detected by LC-ESI-MS. Cys157 is conserved across many bacteria, yeast, protozoa and plants but not in animal species such as *Homo sapiens* or *Mus musculus*. It is therefore likely that many domains of life upregulate transketolase via oxidation of Cys157.

**Figure 33:** The $2F_0F_C$ (green) and $F_0F_C$ (orange) electron density difference maps around residue Cys157 in subunit A and B of wild-type *E. coli* TK (1QGD) at 3σ omit level. The $2F_0F_C$ maps illustrate the fitted electron density of the solved structural model. The $F_0F_C$ maps illustrate the electron density that has not been accounted for in the structural model.

c) **Oxidative stress during fermentation increased %**$TK_{modified}$, %$TK_{high}$ **and enzyme activity**

The effect of increasing cellular oxidative stress on the %$TK_{modified}$, %$B_{max(high)}$ and specific activity was investigated by addition of 0.005% cumene hydroperoxide (CHP) to *E. coli* cells half-way through an eight-hour fermentation. LC-ESI-MS, TPP-binding and specific activity data revealed a proportional increase in the %$TK_{modified}$, %$B_{max(high)}$ and enzyme activity by approximately 37% (Figure 32; Table 7). These findings not only imply that $TK_{low}$ and $TK_{high}$ are equivalent to $TK_{unmodified}$ and $TK_{modified}$, respectively, but provides a strong causative link where the more active $TK_{high}$ is formed as a direct response to oxidative stress during fermentation.
**Table 7**: Summary of the $\%B_{\text{max(high)}}$, $\%\text{TK}_{\text{modified}}$ and specific activity of wild-type transketolase in the presence and absence of 0.005% CHP. Associated errors are the fitting error when fitting to the respective functions.

<table>
<thead>
<tr>
<th>TK variant</th>
<th>$%B_{\text{max(high)}}$ ±</th>
<th>$%\text{TK}_{\text{modified}}$ ±</th>
<th>Specific activity (µmol/mg/min) ±</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type</td>
<td>33.6% 2.9%</td>
<td>31.0% 1.7%</td>
<td>0.31 0.01</td>
</tr>
<tr>
<td>Wild-type ± 0.005% CHP</td>
<td>45.8% 3.8%</td>
<td>44.1% 1.5%</td>
<td>0.41 0.01</td>
</tr>
</tbody>
</table>

**d) Detection and location of sulfenic acid formation via dimedone labelling**

Reaction with dimedone specifically traps sulfenic acids without reacting with thiol groups and over-oxidised states such as sulfones, and was hence used to detect a) the existence of sulfenic acid(s) through intact protein LC-ESI-MS, and b) the location of sulfenic acid formation as a response to oxidative stress via peptide digest followed by mass spectrometry.

Apo-dimeric wild-type transketolase was initially reacted with dimedone in 50 mM Tris-HCl. No shift in molecular weight was observed, however, presumably because the sulfenic acid was not solvent-exposed in correctly-folded apo-dimeric transketolase (Figure 34). The reaction was subsequently repeated in 3.8 M urea, the highest concentration transketolase can tolerate without irreversible denaturation (41), to expose the sulfenic acid modification, and 2 mM iodoacetamide, to prevent the oxidation of exposed sulfhydryls to sulfenic acids. Under these conditions, sulfenic acids were trapped by reaction with dimedone at up to two locations, while one sulfhydryl residue was trapped by reaction with iodoacetamide. Dimedone labelling was accompanied by the significant reduction of the first oxidised peak of wild-type transketolase, which corresponded to the sulfenic acid of TK$_{\text{high}}$, but not the second oxidised peak, which, therefore, likely corresponded to a sulfonic form of TK$_{\text{high}}$ rather than a second sulfenic acid at a second location.
Figure 34: Evidence for existence and location of sulfenic acid modifications. A) Mass spectra of wild-type transketolase and 2 mM dimedone in the absence (black) and presence (red) of 3.8 M urea and 2 mM iodoacetamide (IAM), normalised for total peak area. The spectra show the shift in molecular weight as a result of reaction between dimedone molecules and protein sulfenic acids. B) The predicted peptide modifications between G141-A170 after reaction of 2 mM dimedone, 2 mM IAM in 3.8 M urea, trypsin digestion and mass spectrometry. Additional supplementary data regarding the peptide digest and mass spectrometry analysis are provided in Appendix Three.

No peak-shift was observed that corresponded to the reaction with two dimedone and one iodoacetamide molecules, which implied the peak that corresponded to reaction with two dimedone molecules may in fact be an artefact from oxidation of a sulfhydryl to a sulfenic acid, which was subsequently trapped by a dimedone molecule.

The dimedone modification at trapped sulfenic acid(s) was utilised to locate the oxidised residue of TK_{high} via reaction with dimedone followed by trypsin digestion and mass spectrometry. The experiment and subsequent analysis was carried out by Clive Metcalfe at NIBSC, who is aware of the inclusion of his work in this thesis. Analysis of the peptide fragments indicated no dimedone modifications were present, but instead identified potential partial oxidation of Met153, Cys157,
Met158, Met159 and Cys167, all of which are located near the TPP binding site (Figure 34). The exact location of oxidation remains unclear, as the mass spectrometer had a low resolution MS2 scan and struggled to assign the correct position of modification within a single peptide. However, the above mass spectrometry data, combined with the new crystallographic evidence (Figure 33), suggested Cys157 may have been oxidised in TK\textsubscript{high}. Nevertheless, any of these modifications could, in theory, complete an inter-subunit proton wire between the two active sites of the dimer. Of the 34 peptides that covered the TPP-binding region of the protein sequence, 38% had only one oxidation modification, while 21%, 12% and 3% had only 2, 3 and 4 modifications. These ratios are unrepresentative of those of the native structure, however.

The detection of dimedone modifications in only intact mass spectra was unexpected, and may have been because the peptide fragment analysis software was unable to assign the dimedone modification, the labelling strategy required further optimisation, or because the proportion of labelled dimedone was too small. The experiment was, therefore, repeated with DCP-Bio1, a biotinylated dimedone molecule that was used to enrich the dimedone-labelled population via purification using a monomeric avidin column. Intact mass spectra of the resulting soluble sample had no peaks corresponding to the TK-DCP-Bio1 species, nor were DCP-Bio1 modifications detected after peptide digest of both the soluble and insoluble sample fractions after labelling (Appendix Three).

In summary, I was unable to definitively pin-point the location of the dimedone modification(s) to a particular residue. However, analysis of oxidation and carbamidomethylation sites, supplemented with crystallographic evidence, suggested oxidation likely occurred at least at Cys157. The dimedone-labelling strategy used in this study was not optimised, however, and future optimisation may facilitate the search for the location of sulfenic acid modification(s).

**TK\textsubscript{low} is a low-activity form of transketolase**

The activity of wild-type transketolase towards 50 mM glycolaldehyde (GA) and 50 mM HPA at 2.4 mM TPP and 9 mM Mg\textsuperscript{2+} (TK\textsubscript{high} and TK\textsubscript{low} saturated) increased only 9.0% relative to the activity at 50 μM TPP (only TK\textsubscript{high} saturated) (Figure 32D). These data, combined with the fractional saturation of TK\textsubscript{high} and TK\textsubscript{low} from the TPP binding data, was used to determine that the activity of TK\textsubscript{low} towards GA was only 4.5% relative to that of TK\textsubscript{high}, implying that the TK\textsubscript{low} dimer was effectively inactive, and that the oxidation of TK to form TK\textsubscript{high} therefore resulted independently in increased dimer formation, higher affinity for TPP, and also higher activity.
Ruling out alternative possible origins of $\%B_{\text{max(\text{high})}}$

While, the mass differences correlated directly to the observed ratio of $\text{TK}_{\text{high}}$:$\text{TK}_{\text{low}}$ from the $\%B_{\text{max(high)}}$ ratio, several other potential mechanisms might be hypothesised to explain the observed TPP-binding behaviour. These include: i) allostery; ii) asymmetric TPP binding to non-identical active sites in all homodimers; iii) population of intermediate states during TPP binding. However, none of these mechanisms fit the observations made in this work, and I have addressed each of them in detail below.

a) Allostery

Allosteric-activator binding of TPP to transketolase is unlikely given the importance of TPP to the reaction mechanism, the scarcity of TPP in the cell, and the lack of kinetic or structural evidence of such binding outside of the two active-sites. It is therefore likely that the low-affinity binding site is located within the same active-site binding pocket as the high-affinity binding site. Indeed, all X-ray crystal structures of holo-transketolase showed only two TPP molecules per dimer (18,29).

b) Asymmetric TPP binding to non-identical active sites in all homodimers

The presence of two TPP-binding events with significantly different affinities may at first glance support the hypothesis that two TPP molecules bind the two binding sites of a single transketolase dimer asymmetrically; one with high affinity, one with low affinity. This trail of thought is dispelled by several observations. Firstly, two independent binding events were determined, both of which were cooperative. The observed binding events must therefore be more than just cooperativity between two sites. Furthermore, the dissociation constant of the low-affinity TPP binding site detected in this study was 42-fold and 600-fold greater than those of the two non-equivalent, negatively cooperative active sites reported previously (101), suggesting the low-affinity site reported here is an entirely different binding site. Indeed, early work on transketolase TPP binding detected an additional binding site with a far higher dissociation constant than the two non-equivalent ‘high-affinity’ binding sites also detected, but no dissociation constant was obtained nor characterised (117). On the other hand, the dissociation constant of the high affinity site reported in here was between those of the two non-equivalent sites reported previously (101), as would be expected if two dissociation constants are analysed as a single dissociation constant using the Hill function.

c) Formation of an intermediate state during TPP binding

Transketolase cofactor-binding and activation has been shown to occur in at least two general steps. The first stage is fast and readily reversible and results in the formation of the catalytically
inactive TK---TPP primary complex. Subsequent conformational changes convert the primary complex into catalytically active holo-transketolase, TK*-TPP. The second step is quasi-irreversible and slow in nature (112,117).

In theory, it is possible that TK\textsubscript{high} and TK\textsubscript{low} may be synonymous with TK---TPP and TK*-TPP, and that the TK\textsubscript{low} population is unable to undergo the second, quasi-irreversible step to TK\textsubscript{high}. However, the maximum TPP concentration used in the previous kinetic experiments using yeast transketolase was 100 µM and therefore only saturating for TK\textsubscript{high}. As such, the two-step activation of transketolase was observed only in the formation of TK\textsubscript{high}, and may well be different for TK\textsubscript{low}. Additional experiments at higher [TPP] would be required to elucidate the mechanism of cofactor-binding and activation of TK\textsubscript{low}.

I also examined whether a lack of TPP during overexpression might provide the conditions under which TK differentiates into the TK\textsubscript{high} and TK\textsubscript{low} forms. \textit{E. coli} is known to form approximately 6% holoTK from overexpressed TK, by using the available cellularly synthesised TPP (128). However, supplementation of the fermentation culture with 0.5 mM thiamine, which \textit{E. coli} can import and subsequently convert to TPP, was found to have no significant effect on the \%\textit{B}_{\text{max(high)}} obtained (Figure 35).
Figure 35: Experimental data of 0.05 mg/mL wild-type transketolase binding to TPP at 9 mM Mg$^{2+}$, purified from a fermentation supplemented with 0.5 mM thiamine. Experimental data-points were fitted to the double-Hill function (black). The % TK$_{\text{high}}$ was 34.5 ± 15.8%, comparable to that of transketolase purified from un-supplemented fermentations.
Figure 36: A schematic diagram of the Two-Species Model of TK activation. The model is based on the combined TPP-binding, AUC, mass spectrometry and enzyme activity data. L (light blue) represents a TK\textsubscript{low} monomer and H (navy blue) a TK\textsubscript{high} monomer. TK\textsubscript{low} is converted to TK\textsubscript{high} via oxidation (brown) of an active site methionine or cysteine. The most likely modification site is Cys157.

DISCUSSION

It is unclear why TK\textsubscript{low} was not detected in other cofactor-binding studies, but this could have been one, some or all of the following reasons: a) the larger range of cofactor concentrations used in this study; b) the larger datasets taken at each [Mg\textsuperscript{2+}]; c) there may be no inducible absorption band at 320 nm upon cofactor binding to TK\textsubscript{low}; and d) TK\textsubscript{low} activity is negligible relative to TK\textsubscript{high} and so invisible to measurements of cofactor affinity based on activity measurements. The presence of a low-affinity transketolase binding event has been suggested previously for yeast transketolase (117) but was perhaps presumed to be irrelevant or unimportant and so never further characterised. Therefore, for decades it has been assumed that transketolase expressed and assembled always as a single native conformation, when in fact two monomeric species may often have been formed. Crystallography studies would not have revealed the second species simply through partial TPP
CHAPTER FIVE

occupancy as cofactors have always been added to saturation, even of the low-affinity site (18,23,26,29).

The available crystal structures of transketolase have not been able to reveal a structural basis for the non-equivalence of the two active sites in terms of TPP-binding cooperativity, or alternating sites reactivity. In the one case where this has been observed in a TPP-dependent enzyme, it manifested as subtle differences in the mobility of structural regions around the active site of E1 PDC (111). Such differences are difficult to observe, are potentially subdued by crystal packing, and could readily average out rather than resolving into distinct conformers. Furthermore, non-equivalence may originate from dynamics, rather than underlying structural differences, between the active sites and would therefore only be detected in kinetic-based assays. Similar differences between TK_{high} and TK_{low} would also be challenging to resolve. Even the structural differences resulting from the addition of only one or more oxygen atoms as detected by mass spectrometry, would be difficult to resolve, or easily overlooked by crystallography. Additionally, the oxidation state and hence formation of TK_{high} may yet be found to vary considerably with the range of expression systems and bioreactor conditions used in different studies, as these would lead to different levels of oxidative stress in vivo. Thus, many crystal structures and studies of TK may have been carried out on samples with much lower levels of TK_{high}.

The detection of TK_{low} also highlights a major challenge with measuring the specific activity or \( k_{cat} \) of overexpressed enzymes. It has always been assumed that saturation with cofactors led to maximum activity. However, the TK_{low} species accounts for roughly two-thirds of the purified transketolase in solution, but only 9% of the overall catalytic activity. Therefore, the true catalytic potential, in particular \( k_{cat} \) for the active transketolase species, TK_{high}, has been wildly underestimated. My research also showed that the standard concentrations of 2.4 mM TPP and 9 mM Mg\(^{2+}\) are unnecessarily high for studying TK_{high}, but have been likely optimised to achieve maximum activity through cofactor-saturation, in samples that were predominantly or even entirely the un-oxidised TK_{low} form. Going forward, industrial applications and experimental work may need to re-optimise the fermentation conditions to maximise TK_{high}, such as by maximising oxidative stress, and then adopt a working concentration of only 70 µM TPP to give a TK_{high} saturation of 99% at 9 mM Mg\(^{2+}\).

The observed oxidation of TK caused a dramatic 100-fold improvement in TPP-affinity and a 20-fold increase in the inherent catalytic activity of the enzyme. Therefore, the oxidation to TK_{high} is very likely to be physiologically important, either as a necessary post-translational modification
for normal cellular function, or as a rapid response mechanism to cope with the increased metabolic demands under oxidative stress. Metabolically, this is compatible with the location of transketolase in the PPP, and transketolase may indeed act as a redox-sensitive regulatory mechanism that increases metabolic flux through the PPP during oxidative stress. Furthermore, the location and potential role of the identified sulfenic acids in the formation of a proton wire between active sites suggests a second potential function. However, while induction of oxidative stress during fermentation can lead to an increase in $\text{TK}_{\text{high}}$, further oxidation during purification cannot be ruled out at this stage. Nevertheless, a 20-fold oxidative-stress induced improvement in activity relative to that under milder oxidative conditions is unusual, and could potentially be exploited deliberately for biocatalysis.

While $\text{TK}_{\text{low}}$ is unmodified, and assumed in all previous literature to be the physiologically relevant form, it is unclear whether only the overexpression in $E. \text{coli}$ under oxidative conditions designed to maximise cell growth, has led to the partial oxidation to form $\text{TK}_{\text{high}}$, or whether or not this form exists naturally. It is known that air-oxidation of native holoTK actually leads to loss of activity via an unknown mechanism, but one that must be either at a different site (129), or resulting from over-oxidation at the same site. Therefore, the activating oxidation in $\text{TK}_{\text{high}}$ must be more site-selective or controlled, or even occurring only at an intermediate stage of translation and folding.

The complex trends in affinity and cooperativity of $\text{TK}_{\text{high}}$ binding to TPP and $\text{Mg}^{2+}$ are of a form that is often found to be physiologically important. $\text{TK}_{\text{high}}$ binds TPP with the highest cooperativity at 4 mM $\text{Mg}^{2+}$ which matches the intracellular concentration of free $\text{Mg}^{2+}$ in $E. \text{coli}$ of approximately 1-5 mM (125). Therefore, $E. \text{coli}$ transketolase has evolved to utilise $\text{Mg}^{2+}$ to a) enhance the affinity of TK for TPP; and b) maximise binding cooperativity at physiologically relevant [$\text{Mg}^{2+}$], with the ultimate goal of increasing TK saturation by TPP at lower [TPP]. It is therefore likely that the decrease in cooperativity at > 4 mM $\text{Mg}^{2+}$ is not physiologically relevant. Nevertheless, these trends remain acutely relevant to in vitro biocatalysis which often pushes enzymes far beyond their natural means, and for which maximizing the activity from overexpressed enzymes is desirable.

Given the potential tripling in activity offered by simply converting $\text{TK}_{\text{low}}$ to $\text{TK}_{\text{high}}$, further research into the origin of the Two-Species Model of transketolase interconversion is warranted, as would research into methods that can enhance the population that is correctly oxidised to $\text{TK}_{\text{high}}$, given that this form was the most active. It would also be interesting to analyse donor substrate binding and inhibition, and the unusual phenomenon of heat-activation of transketolase (51), with respect to $\text{TK}_{\text{high}}$ and $\text{TK}_{\text{low}}$.
SUMMARY

This chapter utilised a novel fluorescence-quenching transketolase TPP-binding assay to detect and characterise a previously-unknown low-affinity (TPP) and low-activity transketolase sub-species, \( \text{TK}_{\text{low}} \), that was 20-fold less active than the previously-characterised high-affinity (TPP) and high-activity \( \text{TK}_{\text{high}} \) sub-species, over the largest range of [TPP] and [Mg\(^{2+}\)] undertaken so far. It explored the origin and interconversion of the two sub-populations via oxidation of Cys157 and potentially other active-site cysteines/methionines, and proposed a novel Two-Species Model of transketolase activation, and commented on the mechanistic and physiological insights it provides into transketolase function and regulation. This research has discovered a novel mechanism of TK redox regulation, and the first within the non-oxidative phase of the pentose phosphate pathway.
CHAPTER SIX: Transketolase donor substrate-binding, inhibition and heat-activation to refine the Two-Species Model of transketolase interconversion

The discovery and characterisation of the Two-Species Model of transketolase interconversion raised several interesting questions with respect to unexplained, transketolase-specific phenomena, such as donor substrate binding and inhibition, heat-activation and the potential existence of a three-species dimeric system consisting of TK\textsubscript{high}-TK\textsubscript{high}, TK\textsubscript{high}-TK\textsubscript{low}, and TK\textsubscript{low}-TK\textsubscript{low}. As the fluorescence-quenching-based TPP-binding assay was potentially adaptable to donor substrate-binding and subsequent formation of a stable TPP-substrate and TPP-DHE intermediates, HPA- and pyruvate-binding to transketolase were next investigated with respect to the aforementioned phenomena.

INTRODUCTION

In biocatalysis, β-hydroxypyruvate (HPA) is often used as the donor substrate due to the irreversible, concomitant release of CO\textsubscript{2} as a by-product. Strong substrate inhibition has been observed above 25 mM HPA with an inhibition constant of around 42 mM (77,130). The cause of this substrate inhibition is currently unknown and is addressed in this study.

The inactive, apo-form of transketolase is in a monomer-dimer equilibrium that is dependent on protein concentration. Upon cofactor binding, both the inactive apo-monomer and -dimer are converted into the catalytically active, dimeric holo-form of seemingly structurally identical subunits, with two active-sites per homodimer, located at the subunit interface. Each active site is comprised of one divalent cation, such as Mg\textsuperscript{2+}, and one TPP molecule.

Transketolase displays ping-pong kinetics and catalyses two sequential half-reactions; formation of the dihydroxyethyl-TPP (DHE-TPP) carbanion intermediate, followed by transfer of this two-carbon ketol group from the carbanion intermediate to an acceptor aldehyde, thus returning the enzyme to its starting state. It is thought that there is considerable communication between active sites in order to coordinate the ping-pong kinetics between active sites, such that catalysis alternates
between active sites, giving rise to the ‘half-of-the-sites reactivity’ phenomenon that has been observed in transketolase (110).

There is evidence that in the E1 subunit of the pyruvate dehydrogenase complex (PDC) (EC 1.2.4.1), the two TPP-containing active-sites communicate via a 20 Å proton wire that shuttles a proton between active sites, enabling the cofactors to operate reciprocally as general acid-base catalytic moieties, thus synchronising the ping-pong mechanism across active sites (111). It was suggested by the authors that many thiamine-dependent enzymes may function in a similar way.

Transketolase is unusual in that incubation at between 40-55 °C for 1 h increased the residual activity significantly, when measured after re-cooling the samples to 25 °C. For example, incubation at 42 °C for 1 hr increased activity by 50% (51,53). It was postulated that this curious phenomenon may be the result of an inactive form of transketolase being physically altered or activated by temperature, although no such species had been detected at that time.

Until this PhD, it was thought that purified transketolase existed in a single form with a high affinity for TPP, which is essential for catalytic activity. However, in Chapter Five, extensive evidence was provided that indicated transketolase exists as two distinct subunit species in purified samples, TK\textsubscript{high} and TK\textsubscript{low}, with over 200-fold different affinities for TPP at high [Mg\textsuperscript{2+}]. The TK\textsubscript{low} subunit, which has only 4.5% of the activity of TK\textsubscript{high} in the presence of saturating concentrations of TPP, was also found to have disrupted cooperativity between TPP-binding sites found in TK\textsubscript{high}. The two distinct monomeric subunits were found to combine into at least two distinct dimer forms, the TK\textsubscript{high}-TK\textsubscript{high} and TK\textsubscript{low}-TK\textsubscript{low} dimers, but the potential to form also the TK\textsubscript{high}-TK\textsubscript{low} mixed dimer was not ruled out.

The $\%B_{\text{max(high)}}^{\text{TPP}}$ (33.6 ± 2.9%), i.e. the fraction of TPP that can bind to TK\textsubscript{high} relative to TK\textsubscript{low}, was found to be invariant to changes in [Mg\textsuperscript{2+}] and [TK], while addition of [thiamine] during fermentation to increase cellular [TPP] also had no effect. It was hypothesized that oxidative stress may play a role in determining the ratio of TK\textsubscript{high}:TK\textsubscript{low}, and detection of post-translational oxidations at Cys157 by mass spectrometry supported this hypothesis.

The pentose phosphate pathway runs parallel to glycolysis and has several cellular functions, including the generation of pentose sugars as well as ribose-5-phosphate, the latter the precursor for nucleotide biosynthesis. It is therefore a key branch-point in the diversion of metabolic flux towards
biosynthetic pathways. However, one of the PPP’s most important utilities is to respond to and negate oxidative stress, often due to reactive oxygen species (ROS), through production of an anti-oxidant, NADPH (98).

Intracellular NADPH is continuously used as a reducing agent to replenish the reduced glutathione pool to protect against oxidative stress and to maintain a stable cellular redox potential (131–133). Recently, it has been demonstrated that in the short-term, oxidative stress can be mediated by redox-sensitive enzymes in lower glycolysis (96,97,134,135). In these instances, oxidative post-translational modifications (PTMs) can provide regulation with a rapid response time by cysteine oxidation and subsequent diversion of glycolytic flux through the PPP to generate NADPH and nucleotide precursors for DNA repair.

While a few control points have also been identified in the oxidative phase of the PPP (98), to the authors’ knowledge, redox regulation of enzymes in the non-oxidative phase of the PPP are yet to be identified or fully characterised. In Chapter Five, data was provided that supported the hypothesis that transketolase may be such a regulatory control point.

Here, the binding of substrate to the transketolase species populations are examined directly, for both wild type and the variant S385Y/D469T/R520Q, and provide further evidence that transketolase is a redox-regulated enzyme that likely plays a major role in the control of flux through the PPP during oxidative stress. The new substrate inhibition and heat-activation data indicated a mixed dimer species, TK\text{high}_\text{TK}_\text{low}, may be the cause of donor-substrate inhibition and heat-activation. Finally, an updated Two-Species Model for transketolase activation, regulation and inhibition is proposed that includes explanations of the origin of \%B_{\text{max}(\text{high})}, HPA inhibition, active-site cooperativity, redox control and heat-shock activation. The identification of such regulatory mechanisms may also inform studies into the role of transketolase in tumour progression, given its increased activity found in many cancer cells (136).

**EXPERIMENTAL PROCEDURES**

*Enzyme preparation*

ApoTK was prepared as described in Experimental Procedures, Chapter Five. For TPP-binding assays, series of 2x concentrated cofactor solutions were prepared and purified TK was added to a final concentration of 0.05 mg/mL. The samples were incubated at 22 °C for 45 minutes to allow
TK-TPP binding to reach equilibrium. For HPA-binding assays, 2x concentrated, purified holoTK was prepared 2x concentrated at 0.1 mg/mL TK, 18 mM Mg$^{2+}$ and 0.6 mM TPP and incubated at 22 °C for 45 minutes to allow TK-TPP binding to reach equilibrium. The purified holoTK was added to a series of 2x concentrated HPA solutions and incubated at 22 °C for 10 minutes to allow holoTK-HPA binding to reach equilibrium. For heat-inactivation studies, TK samples were incubated at 42 °C for 1 hour and subsequently re-equilibrated at 4 °C for 30 minutes and at 22 °C for 30 minutes prior to assays.

**Fluorescence assay to detect TPP binding**

TPP-binding was measured using a Fluoromax-4 (Horiba, UK) spectrofluorometer ($\lambda_{ex}=240$ nm; $\lambda_{em}=330$ nm; integration time = 0.1 s; slit width = 8 nm), as described in Chapter Five.

**Transketolase activity assay**

Purified, dialysed apo-transketolase (0.2 mg/mL) was incubated with 7.2 mM TPP and 27 mM Mg$^{2+}$ for 45 minutes at 22 °C (3x concentrated). 50 µL was added to 100 µL 150 mM GA, 150 mM HPA, giving final substrate concentrations of 50 mM. The reaction was performed in triplicate at 22 °C in a 96 well plate with shaking at 300 rpm using a Thermomixer Comfort shaker. 10 µL of the reaction was quenched with 190 µL 0.1% trifluoroacetic acid (TFA) after 3, 5, 10, 15, 20, 30, and 40 minutes. Samples were subsequently analysed by a Dionex HPLC system (Camberley, UK) with a Bio-Rad Aminex HPX-87H reverse phase column (300 x 7.8 mm$^2$) (Bio-Rad Labs., Richmond, CA, USA), via Chromeleon client 6.60 software, to separate and analyse the change in the concentration of substrate (GA) and product (Ery) over the course of the reaction using the method described previously (123).

**Fluorescence assay to detect HPA binding**

The same methodology was used as the TPP-binding fluorescence assay, except cofactor concentrations were kept constant at 9 mM Mg$^{2+}$ and 0.3 mm TPP, and an IFE CF was generated between 0-80 mM HPA.

**Mass spectrometry (LC-ESI-MS)**

LC-ESI-MS was performed as described previously in General Experimental Procedures.
RESULTS

The two-species transketolase model is persistent across variants

In Chapter Five, the TPP-binding parameters of wild-type transketolase were measured over a range of cofactor concentrations. Here, TPP-binding parameters for S385Y/D469T/R520Q, a variant that was previously engineered for activity towards aromatic aldehydes (48), are reported. The rationale behind these measurements was to a) confirm the existence of both TK\textsubscript{high} and TK\textsubscript{low} across variants; and b) to demonstrate a direct correlation between % substrate inhibition, %B\textsubscript{max(high)}, and %TK\textsubscript{modified}. The TPP-binding parameters of TK\textsubscript{high} and TK\textsubscript{low} between 0-18 mM Mg\textsuperscript{2+} are summarised below (Figure 37 and 38; Table 8 and 9). The double-Hill function was again utilised to determine the independent TPP-binding parameters of TK\textsubscript{high} and TK\textsubscript{low}. At 0 mM Mg\textsuperscript{2+}, however, the dissociation constants of TK\textsubscript{high} and TK\textsubscript{low} were too similar to deconvolve with accuracy, therefore the dissociation constants and Hill coefficients were constrained to equal each other within the double-Hill function.

**Figure 37:** Experimental data of 0.05 mg/mL TK S385Y/D469T/R520Q binding to TPP at 0 mM (black), 1 mM (red), 4.5 (blue), 9 mM (magenta) and 18 mM (green) Mg\textsuperscript{2+}. Experimental data-points A) at all [Mg\textsuperscript{2+}] fitted to the double-Hill function. The contributions of the high (red) and low affinity
(blue) binding sites of the double-Hill function are also shown; B) at 9 mM Mg$^{2+}$ (double-Hill function) with the contributions of the high and low affinity binding sites shown as dashed lines; normalised contributions to double-Hill functions, of the C) high affinity and D) low affinity binding sites at each [Mg$^{2+}$].

Figure 38: Dependence of TPP-dissociation constants and Hill coefficients, of the high and low affinity binding sites for 0.05 mg/mL TK (S385Y/D469T/R520Q), at 0 mM, 4.5 mM, 9 mM and 18 mM Mg$^{2+}$. A) & B) – the dissociation constants of the high and low affinity binding sites, respectively. C) & D) – the Hill coefficients of the high and low affinity binding sites, respectively. Solid lines represent quantified, fitted graphs; dashed lines represent clear, unquantified trends; dotted lines represent possible trends that fit the data.
Table 8: Summary of the TPP-binding parameters of the high affinity binding sites of the TK$_{\text{high}}$ species of S385Y/D469T/R520Q. A TK concentration of 0.05 mg/mL was used in each binding assay. Associated errors are the fitting error when fitting to the double-Hill function.

<table>
<thead>
<tr>
<th>[Mg$^{2+}$] (mM)</th>
<th>$K_{d\text{high}}^{TPP}$ (µM)</th>
<th>±</th>
<th>$n_{\text{high}}^{TPP}$</th>
<th>±</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>206</td>
<td>14</td>
<td>1.34</td>
<td>0.07</td>
</tr>
<tr>
<td>4.5</td>
<td>8.03</td>
<td>1.59</td>
<td>1.20</td>
<td>0.19</td>
</tr>
<tr>
<td>9</td>
<td>6.60</td>
<td>1.64</td>
<td>1.03</td>
<td>0.17</td>
</tr>
<tr>
<td>18</td>
<td>5.96</td>
<td>1.52</td>
<td>0.94</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Table 9: Summary of the TPP-binding parameters of the low affinity binding sites of the TK$_{\text{low}}$ species of S385Y/D469T/R520Q. A TK concentration of 0.05 mg/mL was used in each binding assay. Associated errors are the fitting error when fitting to the double-Hill function.

<table>
<thead>
<tr>
<th>[Mg$^{2+}$] (mM)</th>
<th>$K_{d\text{low}}^{TPP}$ (µM)</th>
<th>±</th>
<th>$n_{\text{low}}^{TPP}$</th>
<th>±</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>206</td>
<td>14</td>
<td>1.34</td>
<td>0.07</td>
</tr>
<tr>
<td>4.5</td>
<td>8.03</td>
<td>42</td>
<td>1.22</td>
<td>0.17</td>
</tr>
<tr>
<td>9</td>
<td>6.60</td>
<td>57</td>
<td>1.11</td>
<td>0.14</td>
</tr>
<tr>
<td>18</td>
<td>5.96</td>
<td>57</td>
<td>1.12</td>
<td>0.13</td>
</tr>
</tbody>
</table>

The trends in the TPP-binding parameters of TK$_{\text{high}}$ and TK$_{\text{low}}$ of S385Y/D469T/R520Q were similar to that of wild-type transketolase (Chapter Five). A significant decrease and comparatively small increase in the dissociation constants of TK$_{\text{high}}$ and TK$_{\text{low}}$ was observed, respectively. However, TK$_{\text{high}}$ and TK$_{\text{low}}$ of S385/D469T/R520Q bound to TPP with a lower affinity than that of wild-type (4.5-fold and 1.6-fold lower at 18 mM Mg$^{2+}$, respectively), and TPP-binding to TK$_{\text{high}}$ was less cooperative in S385/D469T/R520Q. Furthermore, the binding cooperativity changed from being positive, to non-cooperative as [Mg$^{2+}$] was increased, but no longer peaked at the physiologically relevant [Mg$^{2+}$] of 4 mM as observed in wild-type TK (Chapter Five). The trends were most likely non-linear in nature like wild-type, but a linear relationship could not be ruled out (Figure 38C and D). These results indicated that while changes in maximum affinity may change between variants, the response of TK$_{\text{high}}$ and TK$_{\text{low}}$ to Mg$^{2+}$ was persistent across variants, although the cooperativity was impacted quite considerably.
**Adaption of the fluorescence quenching-based TPP-binding assay for donor substrate binding**

Similar to the previous fluorescence quenching-based cofactor-binding assay (Chapter Five), binding of HPA to holo-transketolase was shown to further quench the intrinsic fluorescence of holo-transketolase ($\lambda_{ex}= 240$ nm; $\lambda_{em}= 330$ nm) (Figure 39). Like TPP, HPA absorbed relatively strongly at 240 nm; the sample signal was therefore corrected for the inner filter effect (IFE) by generating a correction factor, determined empirically from the change in fluorescence intensity of free HPA and TPP and 50 mM Tris-HCl buffer (Figure 40).

![Figure 39](image)

**Figure 39**: Evidence of fluorescence quenching upon HPA-binding. The observed (black) and corrected (red) fluorescence intensity of 0.05 mg/mL TK, 0.3 mM TPP, 9 mM Mg$^{2+}$ and 0-80 mM HPA in Tris-HCl pH 7.0 ($\lambda_{ex} = 240$ nm; $\lambda_{em} = 330$ nm). Fluorescence intensity was corrected for the inner filter effect (IFE) according to MacDonald *et al.* (120).
The fluorescence intensity of free HPA in 50 mM Tris-HCl and 0.3 mM TPP pH 7.0 ($\lambda_{ex} = 240$ nm; $\lambda_{em} = 330$ nm), fitted to the correction function described previously (120).

The successful adaption of the TPP-binding assay to donor substrates prompted me to expand the assay’s repertoire to pyruvate. Recently, two variants had been evolved to accept pyruvate as a donor substrate (52). The adapted assay was utilised to investigate pyruvate-binding to wild-type, S385/D469T/R520Q and the two pyruvate-accepting variants (Appendix Four).

**Characterisation of E. coli transketolase binding to HPA**

The dissociation constant of HPA-binding to transketolase, $K_d^{HPA}$, has never been previously determined directly, although the closely-related Michaelis-Menten constant of HPA-binding to wild-type transketolase, $K_m^{HPA}$, has been reported as $5.5 \pm 0.5$ mM (137) and 5.3 mM (40), whilst substrate inhibition by HPA, $K_i^{HPA}$, has also been measured as 42.2 mM and 43 mM (77,130), providing suitable benchmarks for comparison. The double-Hill function (Chapter Five) was again used to determine the HPA-binding parameters for both wild-type transketolase and S385Y/D469T/R520Q, as the fit to the data was superior to a single Hill function (Figure 41; Table 10), as reported previously (Chapter Five).
Figure 41: Experimental data of 0.05 mg/mL A) wild-type and B) S385Y/D469T/R520Q transketolase binding to HPA in the presence of 9 mM Mg$^{2+}$ and 0.3 mM TPP in 50 mM Tris-HCl buffer. Experimental data was fitted to the double-Hill function (Chapter Five). The change in binding was reported as the ‘relative fluorescence quenching signal’ in B) because of the change in signal from fluorescence quenching to increased fluorescence at above 20 mM HPA. As such, normalisation to give fractional saturation is not possible in B).

<table>
<thead>
<tr>
<th>Variant</th>
<th>$K_{d\text{[high]}}^{HPA}$ mM</th>
<th>$K_{d\text{[low]}}^{HPA}$ mM</th>
<th>$n_{\text{[high]}}^{HPA}$</th>
<th>$n_{\text{[low]}}^{HPA}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type</td>
<td>3.42 ± 0.45</td>
<td>39.3 ± 0.7</td>
<td>1.39 ± 0.19</td>
<td>3.14 ± 0.18</td>
</tr>
<tr>
<td>S385Y/D469T/R520Q</td>
<td>3.61 ± 0.31</td>
<td>43.1 ± 0.4</td>
<td>1.15 ± 0.06</td>
<td>4.60 ± 0.21</td>
</tr>
</tbody>
</table>

Table 10: Summary of the binding parameters of the high and low affinity binding sites of TK$_{\text{high}}$ and TK$_{\text{low}}$ of wild-type and S385Y/D469T/R520Q transketolase. Each binding assay used 0.05 mg/mL transketolase, 9 mM MgCl$_2$ and 0.3 mM TPP in 50 mM Tris-HCl buffer. Associated errors are the fitting error when fitting to the double-Hill function.
Two distinct HPA-binding events, $\text{TK}_{\text{high}}$ and $\text{TK}_{\text{low}}$, were detected, with an 11-fold difference in affinity, supporting the previously postulated Two-Species Model of transketolase interconversion. The three mutations within the S385Y/D469T/R520Q variant appeared to have no significant effect on the affinity of HPA binding, but significantly decreased the cooperativity of both TPP- and HPA-binding in $\text{TK}_{\text{high}}$, and increased the cooperativity of HPA-binding in $\text{TK}_{\text{low}}$ (Table 10), potentially reflecting an increase in the total number of possible orientations within the active site. In addition, binding of HPA to $\text{TK}_{\text{low}}$ resulted in a decrease in fluorescence quenching in S385Y/D469T/R520Q. This was presumably related to the introduction of the fluorescent tyrosine at residue 385. Importantly, the proportion of HPA bound to $\text{TK}_{\text{high}}$ relative to all TK at saturation, $\%B_{\text{max(high)}}^{\text{HPA}}$, matched that for TPP bound to $\text{TK}_{\text{high}}$, $\%B_{\text{max(high)}}^{\text{TPP}}$, for each variant (Table 11), indicating that the structural difference between $\text{TK}_{\text{high}}$ and $\text{TK}_{\text{low}}$ impacted both TPP and HPA binding in the same way, and confirmed that the two species remained distinct from each other. Furthermore, the $\%B_{\text{max(high)}}^{\text{HPA}}$ and $\%B_{\text{max(high)}}^{\text{TPP}}$ also matched the $\%\text{TK}_{\text{modified}}$, determined from the mass spectra of wild-type transketolase (Chapter Five), and the variant S385Y/D469T/R520Q (Figure 42; Table 11).
**Figure 42:** The impact of heat-activation on %TK<sub>high</sub>, TPP-binding and activity. The mass spectra of purified A) S385/D469T/R520Q B) heat-activated wild-type TK (1 hr at 42 °C). The major peak corresponded to unmodified TK, while the next two peaks corresponded to modified TK. C) TPP-binding to 0.05 mg/mL heat-activated wild-type TK (1 hr, 42 °C), 9 mM Mg$^{2+}$. D) Activity data of purified 0.067 mg/mL wild-type TK with 50 mM GA and 50 mM HPA, pre-incubated with 9 mM Mg$^{2+}$ and 50 µM TPP before (black) and after (red) heat-activation (1 hr, 42 °C).

<table>
<thead>
<tr>
<th>Variant</th>
<th>%&lt;i&gt;B&lt;/i&gt;&lt;sub&gt;max(high)&lt;/sub&gt;&lt;sup&gt;TPP&lt;/sup&gt; ±</th>
<th>%&lt;i&gt;B&lt;/i&gt;&lt;sub&gt;max(high)&lt;/sub&gt;&lt;sup&gt;HPA&lt;/sup&gt; ±</th>
<th>%TK&lt;sub&gt;modified&lt;/sub&gt; ±</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type</td>
<td>33.6% ± 2.9%</td>
<td>33.7% ± 3.0%</td>
<td>31.0% ± 1.7%</td>
</tr>
<tr>
<td>S385Y/D469T/R520Q</td>
<td>27.1% ± 5.3%</td>
<td>23.8% ± 5.0%</td>
<td>30.0% ± 0.1%</td>
</tr>
</tbody>
</table>

Table 11: The %<i>B</i><sub>max(high)</sub><sup>TPP</sup> and %<i>B</i><sub>max(high)</sub><sup>HPA</sup> and %TK<sub>modified</sub> of wild-type transketolase (From Chapter Five) and variant S385Y/D469T/R520Q, determined by TPP-binding, HPA-binding and mass spectrometry data.

The $K_d$<sup>HPA</sup> of wild-type TK<sub>high</sub>, $K_d$<sup>high</sup><sup>HPA</sup>, was slightly lower than the previously reported $K_m$<sup>HPA</sup> (40,137). This was expected because $K_m$ is a chemical pseudo-equilibrium which is convoluted with an additional forward reaction and the formation of product, in other words substrate turnover ($k_{cat}$). The $K_d$<sup>HPA</sup> of wild-type TK<sub>low</sub>, $K_d$<sup>low</sup><sup>HPA</sup>, was very similar to the previously reported inhibition constant of HPA, $K_i$<sup>HPA</sup> (77,130), which suggested a relationship between TK<sub>low</sub> binding to HPA and substrate inhibition by HPA, despite the fact that the $K_i$<sup>HPA</sup> was obtained with no knowledge of the Two-Species Model of transketolase. HPA appeared to bind to the wild-type TK<sub>high</sub> active-sites with slight positive cooperativity ($n_{high} = 1.38 ± 0.19$), while HPA binding to TK<sub>low</sub> was highly cooperative with a Hill coefficient of 3.10 ± 0.16. This high Hill coefficient suggests an important role of TK<sub>low</sub> inhibition by HPA in the regulation of TK activity.

**Heat-induced activation and conversion of TK<sub>low</sub> to TK<sub>high</sub>**

The previous study into heat-activation of TK used cofactor concentrations of 0.5 mM Mg$^{2+}$ and 50 µM TPP, rather than 9 mM Mg$^{2+}$ (51). At these concentrations, the [TPP] was semi-saturating for TK<sub>high</sub> (approximately 60% saturated) but not TK<sub>low</sub> (approximately 20% saturated) (Chapter Five). In other words, the majority of TK<sub>high</sub> was in the catalytically active holo-form while TK<sub>low</sub> was mostly in the inactive apo-form. It was hypothesised that heat exposure may convert TK<sub>low</sub> to TK<sub>high</sub>, hence increasing [holoTK] and overall activity of the sample.
The hypothesis was tested by taking fluorescence quenching measurements of 0.05 mg/mL TK, 9 mM Mg$^{2+}$ and a range of [TPP] before (Chapter Five) and after incubation of at 42 °C for 1 hour (Figure 42; Table 12). As hypothesised, the $%B_{\text{max}(\text{high})}$ increased from 33.7% to 51.7% after heat-activation. In addition, the affinity of both TK$_{\text{high}}$ and TK$_{\text{low}}$ increased significantly. Performing the same heat activation for a second time on the same sample increased the $%B_{\text{max}(\text{high})}$ only slightly to 53.5% (Figure 43), indicating that no further change could be induced.

<table>
<thead>
<tr>
<th>Sample</th>
<th>$K_{d(\text{high})}$</th>
<th>±</th>
<th>$K_{d(\text{low})}$</th>
<th>±</th>
<th>$%B_{\text{max}(\text{high})}$</th>
<th>±</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type</td>
<td>2.28 ± 0.23</td>
<td></td>
<td>276 ± 23</td>
<td></td>
<td>33.6% ± 2.9%</td>
<td></td>
</tr>
<tr>
<td>Heat-activated wild-type</td>
<td>1.75 ± 0.26</td>
<td></td>
<td>157 ± 28</td>
<td></td>
<td>51.7% ± 5.1%</td>
<td></td>
</tr>
</tbody>
</table>

Table 12: Summary of the TPP-dissociation constants of wild-type TK$_{\text{low}}$ and TK$_{\text{high}}$ pre- and post-incubation at 42 °C for 1 h. A TK concentration of 0.05 mg/mL was used in each binding assay. Associated errors are the fitting error when fitting to the double-Hill equation.

Figure 43: Experimental data of 0.05 mg/mL transketolase variant binding to TPP at 9 mM Mg$^{2+}$, after two consecutive rounds of heat activation at 42 °C followed by re-equilibration at 22°C for 30 minutes. Experimental data-points were fitted to the double-Hill function.
CHAPTER SIX

The activity of 0.05 mg/mL TK, 9 mM Mg$^{2+}$ and 50 μM TPP towards 50 mM GA and 50 mM HPA pre- and post-incubation was determined to calculate the activity of heat-activated TK$_{\text{high}}$ relative to pre-incubated TK$_{\text{high}}$ (Figure 42). Overall, a 24.6% increase in transketolase activity was observed after heating, approximately 50% of that expected from the $\%B_{\text{max(high)}}$ increase. This may indicate lower activity in the heat-induced TK$_{\text{high}}$-like state compared to the oxidised TK$_{\text{high}}$, or partial unfolding and inactivation during heating. These and other possible mechanisms were not investigated further.

**The existence of a TK$_{\text{high}}$-TK$_{\text{low}}$ mixed dimer species that mediates HPA substrate inhibition**

The equivalence between the $K_{d(\text{low})}^{\text{HPA}}$ and the $K_{i}^{\text{HPA}}$ measured in previous activity assays suggested that binding of HPA to TK$_{\text{low}}$ gave rise to the observed overall inhibition of transketolase activity. However, the activity of TK$_{\text{low}}$ was already only 4.5% that of TK$_{\text{high}}$ (Chapter Five), implying the TK$_{\text{low}}$-TK$_{\text{low}}$ dimer would be effectively inactive already, and so binding of HPA to that dimer species could not contribute significantly to the observed reaction inhibition. Therefore, the interaction between HPA and TK$_{\text{low}}$ must inhibit the TK$_{\text{high}}$ activity, which in turn suggested that inhibition can only occur within a TK$_{\text{high}}$-TK$_{\text{low}}$ mixed dimer. The relative proportions of the three dimeric species, TK$_{\text{high}}$-TK$_{\text{high}}$, TK$_{\text{high}}$-TK$_{\text{low}}$, and TK$_{\text{low}}$-TK$_{\text{low}}$, were therefore estimated from the heat-activation, HPA-binding, and enzyme activity data.

**a) Estimation of the % dimer from heat-activation data**

It is highly possible that partial heat-activation to a TK$_{\text{high}}$-like state occurred because only the proportion of TK$_{\text{low}}$ that is within the TK$_{\text{high}}$-TK$_{\text{low}}$ mixed dimer was converted to a TK$_{\text{high}}$-like conformation. Having assumed all TK$_{\text{low}}$ subunits of the mixed dimer were completely converted to a TK$_{\text{high}}$-like state, the relative proportions of each dimeric species before heat-activation were calculated from the change in $\%B_{\text{max(high)}}$ upon heat-activation:

(A) $\%\text{TK}_{\text{high}}$-TK$_{\text{low}}$ = ($\%B_{\text{max(high)}}^{\text{heated}} - \%B_{\text{max(high)}}^{\text{unheated}}$)$^2$ = (51.7% - 33.6%) * 2 = 36.2%;
(B) $\%\text{TK}_{\text{high}}$-TK$_{\text{high}}$ = $\%B_{\text{max(high)}}^{\text{unheated}} - (A) / 2 = 33.6% - (36.2 / 2) = 15.5%$;
(C) $\%\text{TK}_{\text{low}}$-TK$_{\text{low}}$ = 100% - (A) – (B) = 100% - 36.2% - 15.5% = 48.3%.
b) Estimation of the % dimer from HPA-inhibition data

Previous studies into the inhibition of transketolase determined the inhibition constants and the maximum inhibition of wild-type transketolase activity by HPA, $\%I_{\text{max}}$, for wild-type ($K_{\text{HPA}} = 43$ mM; $\%I_{\text{max}} = 48.1 \pm 5.1\%$) (77) and also the D469T variant of transketolase ($K_{\text{HPA}} = 40$ mM; $\%I_{\text{max}} = 46.8 \pm 11.5\%$) (138).

Making the assumption that HPA-binding to the TK$_{\text{low}}$ subunit of the mixed dimer resulted in total inhibition of the TK$_{\text{high}}$ subunit of the mixed dimer, and that TK$_{\text{high}}$ accounted for 91% of total activity, the relative proportions of the three dimeric species were predicted as follows:

\[
\begin{align*}
(A) \ %\text{TK}_{\text{high}}\text{-TK}_{\text{low}} &= (\%B_{\text{max(high)}}^{\text{unheated}} \times (\%I_{\text{max}} / 91\%)) \times 2 = (33.6\% \times (48.1\%/91\%)) \times 2 = 35.5\% \\
(B) \ %\text{TK}_{\text{high}}\text{-TK}_{\text{high}} &= \%B_{\text{max(high)}}^{\text{unheated}} - (A) / 2 = 15.8\%; \\
(C) \ %\text{TK}_{\text{low}}\text{-TK}_{\text{low}} &= 100\% - (A) - (B) = 100\% - 35.5\% - 15.8\% = 48.7\%.
\end{align*}
\]

The similarity between the relative proportions of the three dimer species, calculated from the $\%I_{\text{max}}$ obtained through activity data, and from TPP-binding data after heat-activation, provided evidence that HPA binding to the TK$_{\text{low}}$ subunit within the mixed dimer resulted in inhibition of the associated TK$_{\text{high}}$ subunit of the mixed dimer. It also implied heat-activation may have occurred via relief of donor substrate inhibition by conversion of the TK$_{\text{low}}$ subunit of the mixed dimer to a TK$_{\text{high}}$-like state. The fact that inhibition via the mixed dimer species persisted into a single-mutant TK variant, and with a different acceptor substrate, also suggested that this phenomenon is fundamental to the TK structure and mechanism when utilising HPA as the donor substrate.

<table>
<thead>
<tr>
<th>Dimer species</th>
<th>% Species (heat-activation data)</th>
<th>% Species (inhibition data)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TK$<em>{\text{high}}$-TK$</em>{\text{high}}$</td>
<td>15.5%</td>
<td>15.8%</td>
</tr>
<tr>
<td>TK$<em>{\text{high}}$-TK$</em>{\text{low}}$</td>
<td>36.2%</td>
<td>35.5%</td>
</tr>
<tr>
<td>TK$<em>{\text{low}}$-TK$</em>{\text{low}}$</td>
<td>48.3%</td>
<td>48.7%</td>
</tr>
</tbody>
</table>

**Table 13:** Summary of the predicted % dimer of TK$_{\text{high}}$-TK$_{\text{high}}$, TK$_{\text{high}}$-TK$_{\text{low}}$ and TK$_{\text{low}}$-TK$_{\text{low}}$, calculated from TPP-binding data after heat activation and HPA-inhibition data. Note that the % species from the HPA inhibition data accounts for TK$_{\text{high}}$ having only 91% of total activity.
The unified Two-Species Model of transketolase activation, regulation and inhibition

The development of assays capable of detecting both TK\text{high} and TK\text{low} has facilitated investigations into donor substrate inhibition, active site cooperativity and heat-activation, and finally lead to the discovery of a novel mechanism of transketolase regulation. This can be summarised in the unified Two-Species Model of transketolase activation, regulation and inhibition (Figure 44). In summary:

- Transketolase existed as two distinct transketolase species; TK\text{high} and TK\text{low}.
- Inactive TK\text{low} was the reduced, unmodified form of transketolase that was converted to TK\text{high} via oxidation of Cys157 and potentially other active-site methionine and cysteine residues, in response to cellular oxidative stress.
- TK\text{high} was not only significantly more active but also had a 35-fold increased affinity for TPP at physiologically relevant [Mg\textsuperscript{2+}].
- Oxidised Cys157 improved coordination (i.e. cooperativity) between active sites, possibly by providing the final residue of the proton wire between active sites.
- Redox regulation of transketolase provides an important mechanism of control in diverting flux from glycolysis to the PPP during oxidative stress.
- Heat-activation of transketolase occurred through conversion of the TK\text{low} subunit of the mixed dimer to a TK\text{high}-like state, which relieved substrate inhibition of the TK\text{high} subunit.
- Heat-activation may offer cells a degree of heat shock protection by activating transketolase - a key enzyme in central metabolism - without the energetic or time-cost of protein biosynthesis.
- The mixed dimer is responsible for substrate inhibition by HPA; this is not physiologically relevant.
**DISCUSSION**

The aim of this study was to resolve several unanswered questions remaining from my previous study into transketolase activation (Chapter Five); does the two-species phenomenon persist across variants and from cofactor to donor substrate; what is the cause of substrate inhibition; and what is the origin and physiological relevance of heat-activation?

Though wild-type, D469T and S385Y/D469T/R520Q all exist as a mix of TK$_{\text{high}}$ and TK$_{\text{low}}$, the susceptibility of TK$_{\text{low}}$ to oxidation could potentially be enhanced through protein mutations in order to maximise [TK$_{\text{high}}$] and hence activity. Alternatively, controlled oxidation of purified transketolase samples may be possible, but this may require careful control to avoid over-oxidation of Cys157, or at other sites that cause inactivation.

---

**Figure 44:** The Two-Species Model of transketolase regulation, activation and inhibition. The model is based on the combined TPP binding and AUC data before and after heat activation, as well as activity data. L (light blue) represents an inactive TK$_{\text{low}}$ monomer, H (navy blue) an active TK$_{\text{high}}$ monomer, and H (red) a TK$_{\text{high}}$-like monomer post-heat activation. The mechanism of conversion of TK$_{\text{low}}$ to TK$_{\text{high}}$ occurs through oxidation of Cys157 (brown).
It is interesting that HPA binding to TK\textsubscript{low} had such a high degree of positive cooperativity, and may indeed result from multiple HPA molecules binding to a single active-site, or in multiple inhibitory binding modes, to inhibit activity in that active site. Furthermore, it may also inhibit active-site synchronisation by inhibiting the shuttling of protons along the proton wire. The existence of the TK\textsubscript{high}\text{-}TK\textsubscript{low} mixed dimer potentially complicates further the analysis of Hill coefficients for TK\textsubscript{high} and TK\textsubscript{low}. However, the true Hill-coefficient of TK\textsubscript{high} is likely to be higher than the reported apparent value, whereas the true Hill-coefficient of TK\textsubscript{low} is presumably lower.

The heat-activation of TK appeared to increase activity through a slightly different mechanism to that of oxidation of TK\textsubscript{low} to TK\textsubscript{high}. Conformational rearrangements might result from heat-activation and give rise to a TK\textsubscript{high}-like conformational state that apparently relieves HPA inhibition of the TK\textsubscript{high} subunit of the mixed dimer. The heat-activation of transketolase may even be part of the cellular response to thermal stress, resulting in the upregulation of the PPP and hence NADPH production, while also increasing flux through critical biosynthetic pathways such as nucleotide biosynthesis. The heat-sensitivity of transketolase therefore has a potential role in facilitating rapid global changes in metabolism without the additional energetic burden of protein synthesis.

Finally, the discovery of a redox-sensitive regulatory system that activates transketolase during oxidative stress could have important implications for future advances in cancer treatments. Cancer cells are often exposed to higher levels of oxidative stress than normal cells, and elevated transketolase activity has been reported in a number of cancer types in order to reduce ultimately catastrophic damage from high oxidative stress (139). The redox-sensitivity of transketolase not only makes transketolase itself an antioxidant, but also an important mediator of the biosynthesis of a second anti-oxidant, NADPH. Reversal or prevention of transketolase oxidation at residue Cys157 by drug delivery or gene therapy may offer an effective way to prevent antioxidant production and hence proliferation in cancer cells.

**SUMMARY**

This chapter further investigated the preservation of the TK\textsubscript{high}\text{-}TK\textsubscript{low} ratio across variants and from cofactor (TPP) to donor substrate via adaption of the fluorescence-quenching assay to a novel HPA-binding assay. The assay provided the first $K_d$ value ($3.4 \pm 0.5$ mM) for TK\textsubscript{high} binding to HPA and was consistent with previously determined $K_m$ values (5.5 mM and 5.3 mM (40,137)). The
$K_d$ of $\text{TK}_{\text{low}}$ (39.3 ± 0.7 mM) was consistent with previously determined HPA inhibition constants (42.2 mM and 43 mM (77,130)), and hence implicated $\text{TK}_{\text{low}}$ in the mechanism of substrate inhibition and hence the existence of three dimeric species, $\text{TK}_{\text{high}}$-$\text{TK}_{\text{high}}$, $\text{TK}_{\text{low}}$-$\text{TK}_{\text{low}}$, and the mixed dimer $\text{TK}_{\text{high}}$-$\text{TK}_{\text{low}}$. Formation of the mixed dimer was responsible for the previously unexplained phenomenon of donor substrate inhibition and heat-activation, and the proportion of each dimeric species were estimated from data related to these phenomena. The model was updated to the Two-Species Model of transketolase regulation, activation and inhibition. The model will have a significant impact on future transketolase research, and is yet another example of redox regulation through active site cysteines.
CHAPTER SEVEN: A novel cell-lysate-compatible protein stability assay using bio-orthogonal FRET-labelling of a non-canonical amino acid (ncAA) reveals the impact of macromolecular crowding on local structural stability of *E. coli* transketolase

The final goal of my PhD was to exploit ncAAs to develop a novel, fluorescence-based stability assay via bio-orthogonal labelling of the ncAA with bio-orthogonal fluorescent dyes. This study contrasts to Chapter Four in that the ncAA facilitates the development of the stability assay itself, rather than improving a characteristic of the enzyme, and demonstrates the wide applicability of ncAA-incorporation across numerous fields of research. The bio-orthogonal nature of certain ncAA functional groups, and the bio-orthogonal dyes you can therefore covalently attach in a site-specific manner, opens the door to bio-orthogonal, lysate-based stability screens that may be compatible with ultra-high-throughput screening techniques for improved protein stability.

INTRODUCTION

Several methods are available for measuring the conformational stability and aggregation of proteins including by intrinsic fluorescence, size-exclusion chromatography, differential scanning fluorimetry of GFP-tagged proteins (DSF-GTP) (140), hydrophobic fluorescent probes (e.g. SPYRO Orange) (141), far-UV circular dichroism, and certain NMR-detected amide proton exchange assays (142). The non-specific nature of many of these protein stability and aggregation assays is both their strength and their weakness. They are widely applicable to almost all proteins, but are limited to purified, homogeneous samples and thus cannot be readily used in a cellular context, such as in cell lysates. For example, intrinsic fluorescence intensity measures global protein unfolding by detecting the change in the fluorescence of all tryptophan residues, either from the intensity at 340 nm, or the ratio of intensities at 350 nm and 330 nm, as their local environment changes during unfolding. When applied to complex lysates, the signal from tryptophan residues from the protein of interest is further convoluted with that from all other proteins.
The above methods are generally constrained to purified protein and are therefore incompatible with macromolecular crowding studies in cell-like conditions, or with multi-protein systems, and add considerable time, effort and cost to high-throughput screening of protein stability and aggregation. While innovative solutions have been developed, such as NMR-detected amide proton exchange between target proteins and simple crowding agents such as Ficoll-70 (143), and a colony filtration (CoFi) blot stability screen (144), they both have their limitations. For example, the former is unable to detect stability at very high or low protein and/or crowding agent concentrations, is constrained by typical NMR-related molecular weight ranges, and is incompatible with ultra-high-throughput approaches such as microdroplet arrays. Secondly, the latter methodology may perturb native protein structure, requires strong overexpression, and gives little information regarding the impacts of macromolecular crowding on protein stability and aggregation. FRET-based reporter systems have been developed to track in cellulo protein translation and folding visually (145), and to analyse the folding equilibrium between folded and unfolded states at standard growth temperatures using a GFP-based reporter system (146). These methodologies provided vital insights into the special and temporal mechanism and regulation of protein translation and folding in cellular conditions, but are less-applicable to the study of the impact of macromolecular crowding on global and local thermal degradation.

There is a growing realisation that the vast majority of protein stability and aggregation studies performed in vitro at highly dilute concentrations bear little resemblance to the complex, concentrated macromolecular conditions found in the cell, which account for at least 30% of the cell volume and can reach macromolecular concentrations of 200-300 g/L (147). Thus, a major current aim is to step beyond the previous reductionist approaches to protein stability and aggregation, and to now understand, and ideally model, the impact that the crowded cellular environment has on different proteins and cellular mechanisms, through macromolecular crowding effects, and transient interactions.

The macromolecular crowding effect is dominated by a combination of hard-core repulsions and soft (i.e. chemical) interactions. Hard-core repulsions, in which the protein does not form any interactions with molecules in the environment, are inherently stabilising and are increasingly significant at high macromolecular concentrations because the effective volume available to the protein of interest is decreased by the high concentration of macromolecules in solution. Repulsions are therefore entirely entropic and shift the unfolding/aggregation equilibria towards the most compact state, which is usually the native structure (143).
Conversely, soft (i.e. chemical) interactions can be attractive (destabilising) or repulsive (stabilising) (143). Repulsive soft interactions are stabilising for the same entropic reason as hard-core repulsions. Attractive interactions can either stabilise proteins through specific interactions with the native structure surface only, or destabilise proteins through non-specific interactions with the protein backbone that shifts the equilibria towards the greater backbone exposure in the unfolded state. Attractive chemical interactions have been shown to have an enthalpic component and hence can destabilise proteins at high macromolecular concentrations (148).

The original theory of macromolecular crowding predicted hard-core repulsions to dominate at high macromolecular concentrations (149). Much subsequent work has therefore focused on mimicking the volume exclusion effects of hard-core repulsions while limiting protein-macromolecular chemical interactions through the use of inert macromolecular crowding agents such as Ficoll-70 (150–152). Use of such agents has advantages, such as simplifying the analysis of thermal stability to a single protein species. However, crowding agents are unrepresentative of the cellular environment, which is composed of thousands of diverse macromolecules with irregular geometries and functional group arrangements, and various abilities to form weak or strong interactions with the protein of interest. Thus, recent work has shown that the contribution of chemical interactions has been greatly underestimated, and in many cases even dominates over the contribution of hard-core repulsions (143).

All TPP-dependent enzymes consist of two highly conserved domains; the pyrophosphate (PP)-binding domain (2-322 aa) and the pyrimidine (Pyr)-binding domain (323-539 aa). The third, C-terminal domain is less conserved and has an unknown biological function, but has been proposed as a regulatory binding site (18). The C-terminal domain is relatively flexible (54) and is considered a source of global instability, at least in vitro.

Transketolase has a single cofactor-binding site located at the interface of the PP and Pyr domains, hence has one active-site per monomeric unit (18,29). Yeast and E. coli transketolases function as homodimers of apparently structurally-identical subunits, and are active when bound by up to two TPP molecules per homodimer, one per active site (101). The affinity and cooperativity of TPP binding are highly dependent on the concentration of divalent cations, such as Mg$^{2+}$ (Chapter Five).

The global thermal denaturation profile of E. coli transketolase is complex. Between 5-55 °C, wild-type transketolase can undergo heat-activation, where there is a 10% non-cooperative loss
of secondary structure due to a subtle conformational rearrangement (51). Recent evidence suggests that the structural change is sufficient to convert the inactive form of holo-transketolase, TK\text{low}, into a catalytically active TK\text{high}-like form of transketolase (Chapter Six). Circular dichroism (CD) spectra indicated a sharp, highly cooperative transition with a mid-point at 58.3 °C, largely due to aggregation above 58 °C, as detected by dynamic light scattering (DLS). Meanwhile, the intrinsic fluorescence ratio, a measure of global thermal denaturation, had a less cooperative transition with a thermal transition midpoint, $T_m$, of 65.6 °C. It therefore appears that the unfolding and aggregation transitions of wild-type transketolase are convoluted.

Taken at face value, it is perhaps unsurprising that thermal unfolding and aggregation may overlap given that an unfolded protein is more likely to aggregate. However, it raises some interesting questions at the level of local structural dynamics during thermal denaturation. Do certain local structural motifs initiate global unfolding and/or aggregation at different temperatures? Which motifs are the major driving forces behind unfolding and aggregation? Does aggregation occur before or after dimer dissociation? Do any local unfolding events lead to aggregation? What impact does protein concentration have on local unfolding and/or aggregation?

Answers to each of these questions would provide invaluable information about the local structural triggers of unfolding and aggregation, for both purified systems and within more complex environments, and would help guide protein engineering strategies, including the design of directed evolution libraries, for generating more thermostable, and less aggregation-prone variants. As discussed above, even armed with this knowledge, current techniques for the directed evolution of protein stability often involve purification of tens or hundreds of variants, which is time- and cost-inefficient.

The aim in this chapter was to develop a FRET-based assay that can detect local protein unfolding and/or aggregation of dimeric proteins both as purified protein and in cell lysates (Figure 45). The technique is based around the single site-specific incorporation of a non-canonical amino acid (ncAA), para-azidophenylalanine ($p$AzF), using a genetically encoded orthogonal Methanococcus jannaschii tRNA/synthetase pair, and subsequent bio-orthogonal labelling of each monomeric unit with Alexafluor-488 (AF488) and Alexafluor-594 (AF594) to form an inter-dimeric FRET pair (Figure 45). While the ncAA $p$AzF and the two Alexafluor dyes have been used previously to interrogate protein structure and stability under different conditions (153), to this author’s
knowledge this is the first time doubly-incorporated and labelled pAzF has been used to study unfolding in real time via FRET, both in purified protein and in a lysate background.

**Figure 45:** A schematic diagram summarising site-specific incorporation & labelling (azide-alkyne click chemistry) of the UAA pAzF into transketolase & the detection of unfolding and/or inter-subunit aggregation via an increase in FRET distance.

In this study, the utilisation of the assay in purified protein was validated by investigating the impact of [TK] on local structural stability. The versatility of the assay and its potential in cell lysate-based stability assays was demonstrated by assessing the impact of macromolecular crowding on local unfolding and aggregation. Finally, the future development and wider applications of this system is discussed.

**EXPERIMENTAL PROCEDURES**

*Mutagenesis at residue 603 and 105*

Quikchange site-directed mutagenesis was performed using the manufacturers protocol (Stratagene, Cambridge, UK) and the following primers and their reverse complement, designed for specific mutations at residue 603 and 105:

5’- GTACTGCGTACGCGGTACTGCAC-3’  Mutant 603pAzF
5’- GGAAGTGCTTGGTGGACGCTGGTGTTGAAACC-3’  Mutant 105pAzF

The dpnI-digested PCR product was transformed into XL10-gold competent cells and the plasmid subsequently isolated using a Qiagen miniprep kit (Stratagene, Cambridge, UK).
Preparation of and co-transformation into competent C321. ΔA.exp “amberless”
cells

Competent C321.ΔA.exp “amberless” cells were prepared as described in Experimental
Procedures, Chapter Three.

Enzyme preparation and labelling

Variants of the transketolase mutant S385Y/D469T/R520Q were co-expressed with the
ncAA-incorporation machinery from the pUltra plasmid for eight hours in C321. ΔA.exp cells in the
presence of 1 mM pAzF and 1 mM IPTG. The resulting cell pellet was lysed, purified, ultrafiltrated
and quantified, as described previously in this thesis.

1 mg/mL protein was labelled with a 7-fold molar excess of both AF488 and AF594 in a
single reaction at room temperature overnight. Labelled protein was ultrafiltrated four times using an
Amicon Ultra-4 10k MWCO centrifugal filter to remove excess dye. The protein concentration was
determined by absorbance at 280 nm in water, taking into account absorbance due to the dyes using
the following equation:

\[
[\text{TK} - \text{AF488/594}] \ (M) = \left[ A_{280} - (A_{494} \times 0.11) - (A_{590} \times 0.56) \right] / \epsilon_{\text{TK}}
\]

where TK-AF488/594 is doubly-labelled transketolase, 0.11 and 0.56 are correction factors to
account for the absorption of AF488 and AF594 at 280 nm, respectively, and \( \epsilon_{\text{TK}} \) is the extinction
coefficient of the transketolase variant.

Mass spectrometry

LC-ESI-MS was performed as described previously in General Experimental Procedures.

SDS-PAGE analysis

SDS-PAGE was performed as described previously in General Experimental Procedures.
**Thermal denaturation and Static light scattering (SLS) measurements using the UNIt**

The $T_m$-values of TK variants were measured in the UNIt (Unchained Laboratories, Wetherby, UK) via their intrinsic fluorescence emission ratio (350/330 nm). Simultaneously, static light scattering (SLS) measurements were taken at 266 and 473 nm. The microcuvette arrays were loaded with 9 µL of 0.5 mg/mL sample, 2.4 mM TPP and 9 mM Mg$^{2+}$ and excited with a 266 nm laser. The fluorescence was measured as a function of temperature in the range of 30–90 ºC with steps of 1 ºC and an equilibration time of 30 s at each temperature. Thermal denaturation and SLS curves were analysed by fitting the baseline and single transition to a two-state model (88–90). Triplicate measurements were performed at each [TK] for each variant.

**Global thermal denaturation measurements using the Fluoromax-4 via intrinsic fluorescence (350/330)**

The temperature-ramp was controlled by a Thermo Scientific sc150 water bath system that heated the cuvette inside the Fluoromax-4 cuvette holder. The temperature inside the cuvette during the temperature-ramp was calibrated in triplicate using a Pico Technology TC-08 Thermocouple data logger thermometer inside a 1.5 mm x 1.5 mm quartz cuvette holding only PBS. The $T_m$-values of TK variants were measured via their intrinsic fluorescence emission ratio (350/330 nm). The quartz cuvette was loaded with 16 µL 0.04–0.5 mg/mL TK, 2.4 mm TPP and 9 mM Mg$^{2+}$ and excited at 280 nm. The fluorescence was measured in triplicate as a function of temperature in the range of 30–80 ºC with steps of 4 ºC/min. Thermal denaturation curves were analysed by fitting the baseline and single transition to a two-state model (88–90).

**Local thermal denaturation measurements using the Fluoromax-4 via the AF488-594 FRET system**

The local $T_m$- and $T_{agg}$ values of TK variants were measured via the change in FRET signal ($\lambda_{ex} = 488$ nm; $\lambda_{em} = 620$ nm). The quartz cuvette was loaded with 16 µL 0.04–0.5 mg/mL TK-AF488/594, 2.4 mm TPP and 9 mM Mg$^{2+}$. The fluorescence was measured in triplicate as a function of temperature in the range of 30–80 ºC with steps of 4 ºC/min. Thermal denaturation curves were analysed by fitting the baseline and single transition to a two-state (88–90) or three-state (154) model.
Local thermal denaturation measurements of TK in cell lysate using a Fluoromax-4 via the AF488-594 FRET system

Purified TK-AF488/594, TPP and Mg\textsuperscript{2+} was added back to 0-13.8 mg/mL clarified lysate to final concentrations of 0.16 mg/mL, 2.4 mM and 9 mM and left to equilibrate for 30 minutes. The lysate was clarified from XL10-Gold \textit{E. coli} expressing the control plasmid pUC-18. The local \(T_m\) and \(T_{agg}\) values of TK variants were measured in triplicate via the change in FRET signal, as described above.

RESULTS

Site-specific incorporation and labelling of para-azidophenylalanine (pAzF)

The ncAA pAzF was incorporated into transketolase using the strategy developed in Chapter Three; \textit{pUltra}, encoding the IPTG-inducible ncAA-incorporation system (74), and \textit{pQR791}, encoding a constitutively expressed transketolase variant with an engineered in-frame stop codon at residue Y105 or K603, were co-expressed in an “amberless” \textit{E. coli} strain, \textit{C321.AA.exp} (71), in the presence of 1 mM pAzF. Holo-transketolase exists as a homodimer, therefore only one amber stop codon was introduced to the gene sequence to form a dimer with one pAzF residue per subunit.

While a number of loci were considered as regions of interest, two locations were chosen in particular for the proof-of-concept study, at the C-terminal domain (K603) and a PP-binding domain loop (Y105) within the PP domain (Figure 2). The C-terminal domain of transketolase (540-663 aa) is less compact, more flexible and hence a potential source of instability (54). Local denaturation at residue 603 may therefore be expected to report on the overall domain stability. Conversely, loop 105 of the PP-binding domain (2-322 aa) is relatively stable in wild-type transketolase but highly flexible in the variant S385/D469T/R520Q which was evolved to accept aromatic substrates (48,54,155). Local denaturation at residue 105 is therefore likely to be loop-specific, highly localised and destabilising. Within the two regions of interest, residues K603 and Y105 were chosen as specific incorporation sites because they are solvent-exposed surface residues and have inter-subunit Förster resonance energy transfer (FRET) distances of 53 Å and 19 Å, respectively, which are well below the maximum Förster distances (20 nm) that FRET can report efficiently. Therefore, both sites reported on dimer dissociation, as well as domain and local unfolding, and aggregation. While the high solvent exposure of K603 makes mutagenesis unlikely to cause structural problems, the tyrosine at residue 105 was also considered acceptable as it would minimise structural perturbations when replaced with the structurally-similar pAzF. pAzF-incorporated transketolase was purified and then
simultaneously labelled by a ‘click chemistry’ reaction between the azide of pAzF and the alkyne of a donor fluorophore, Alexafluor-488 (AF488), or an acceptor fluorophore, Alexafluor-594 (AF594), yielding the covalent 1,5-disubstituted 1,2,3-triazole products. Both dyes were added to transketolase at 7:1 molar ratio of dyes:protein, in a single reaction overnight at 22 °C, which resulted in stoichiometric labelling of monomeric transketolase units and hence produced a population of doubly-labelled AF488/AF594 homodimers as a mixture of both the same and different dyes in the two monomers.

**Figure 46:** The site-specific incorporation and labelling of pAzF at residue 603 and 105. A) The structure of a S385/D469T/R520Q transketolase homodimer with residues 603 (blue) and 105 (green) highlighted. The PP, Py and C-terminal domains of the two subunits are shown in dark grey/pink, light grey/red and white/brick red, respectively. B) A summary of the efficiency and fidelity of incorporation, %TK$_{\text{high}}$ and the labelling efficiency at residue 603 and 105 relative to S385/D469T/R520Q. C) & D) The mass spectra of unlabelled mutants 603$_{\text{unlabelled}}$ and 105$_{\text{unlabelled}}$. Each mass spectrum (black) is fitted to the sum of multiple Gaussian functions (smaller individual peaks). The cumulative fit of all peaks is shown in red. Higher-molecular weight peaks to the right of the ncAA-incorporated transketolase represent oxidised forms of the pAzF-incorporated species,
as described in Chapter Five and Six. E) & F) the mass spectra of mutants 603-AF488/594 and 105-AF488/594 labelled with both AF488 and AF594 in a single reaction.

As outlined in Chapter Three, low but significant levels of misincorporation of glutamine, phenylalanine, and tyrosine, in addition to oxidised forms of transketolase (Chapter Five and Six), were detected by liquid chromatography electrospray ionisation mass spectrometry (LC-ESI-MS) in both unlabelled and labelled transketolase variants (Figure 46; Table 14). The labelling efficiency (i.e. the percent of labelled protein relative to unlabelled) of pAzF was significantly higher at residue 105 relative to 603, which was a surprising result given residue K603 was more solvent exposed than Y105. At 1 mg/mL transketolase, the concentration used during labelling, apo-transketolase existed exclusively as a dimer (Chapter Five). The proportion of transketolase in its oxidised, active form, TK\textsubscript{high}, calculated from the peak areas of singly (+16 Da), and doubly (+32 Da) hydroxylated molecular weight peaks relative to unmodified peak area, were roughly the same for each variant, but was less than that observed for other transketolase variants. The reason for this is unclear, but would suggest that either both variants are less sensitive to oxidative stress, or that addition of pAzF to the fermentation broth reduces oxidative stress.

<table>
<thead>
<tr>
<th>Variant</th>
<th>Efficiency of incorporation ±</th>
<th>Fidelity of incorporation ±</th>
<th>%TK\textsubscript{high} ±</th>
<th>Labelling efficiency ±</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mutant 603</td>
<td>66.6 % 7.0 %</td>
<td>82.2 % 5.4 %</td>
<td>27.6 % 1.8 %</td>
<td>14.5 %</td>
</tr>
<tr>
<td>Mutant 105</td>
<td>69.2 % 9.0 %</td>
<td>80.8 % 5.1 %</td>
<td>26.4 % 1.3 %</td>
<td>76.2 %</td>
</tr>
</tbody>
</table>

Table 14: Summary of the efficiency and fidelity of incorporation, the %TK\textsubscript{high} and the labelling efficiency when incorporating at residue 603 or 105. The efficiency of incorporation (i.e. yield of mutant protein relative to S385Y/D469T/R520Q) was determined by SDS-PAGE; the fidelity of incorporation (i.e. % pAzF-incorporated species relative to misincorporated species), labelling efficiency and %TK\textsubscript{high} were determined by ESI-LC-MS.

**Incorporation and labelling of pAzF minimally perturbs global stability**

A major aim of this study was to better-understand both the local and global denaturation of transketolase over a wide range of concentrations. It was first confirmed that incorporation and labelling of pAzF had minimal impact on the global thermal stability of transketolase. The intrinsic fluorescence ratio (350/330) and static light scattering (SLS) at 266 nm were determined for S385Y/D469T/R520Q, labelled mutant 603 (K603 to pAzF-AF488/594), labelled mutant 105 (Y105 to pAzF-AF488/594), unlabelled mutant 603\textsubscript{unlabelled} (K603 to pAzF), and unlabelled mutant...
105\textsubscript{unlabelled} (Y105 to pAzF) (Figure 47; Table 15) as a function of temperature using the UNit in order to assess the impact of ncAA incorporation and labelling on global unfolding and aggregation, respectively.

Figure 47: Experimental data of A) the normalised intrinsic fluorescence ratio (350/330) and B) SLS data (266 nm) of TK variants S385Y/D469T/R520Q (black), 603\textsubscript{unlabelled}, (red), 603-AF488/594 (blue), 105\textsubscript{unlabelled}, (magenta), and 105-AF488/594 (green), measured using the UNit. Data was fitted to the two- and three- state models of protein denaturation (88–90,154).
Table 15: Summary of the thermal stability of mutants S385Y/D469T/R520Q, 603unlabelled, 105unlabelled, 603-AF488/594 and 105-AF488/594. Thermal stability of 0.5 mg/ml TK, 2.4 mM TPP, 9 mM Mg, in 50 mM PBS, pH 7.0, in terms of their thermal transition midpoint, Tm, van’t Hoff enthalpy change, ΔHvh, the aggregation transition midpoint, Tagg, and the temperature at which 10% of the sample was aggregated, Ton. The first two measurements were determined from the intrinsic fluorescence ratio (330/500 nm) and the second two from static light scattering (SLS) measurements at 266 nm. Measurements were taken using the UNit.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Tm (ºC)</th>
<th>±</th>
<th>ΔHvh (kcal/mol)</th>
<th>±</th>
<th>Tagg (ºC)</th>
<th>±</th>
<th>Ton (ºC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S385Y/D469T/R520Q</td>
<td>63.1</td>
<td>0.5</td>
<td>87</td>
<td>8</td>
<td>65.4</td>
<td>2.0</td>
<td>62.5</td>
</tr>
<tr>
<td>603unlabelled</td>
<td>62.8</td>
<td>0.2</td>
<td>154</td>
<td>17</td>
<td>65.5</td>
<td>0.2</td>
<td>58.5</td>
</tr>
<tr>
<td>105unlabelled</td>
<td>62.6</td>
<td>0.3</td>
<td>105</td>
<td>12</td>
<td>64.4</td>
<td>1.6</td>
<td>60.6</td>
</tr>
<tr>
<td>603-AF488/594</td>
<td>61.9</td>
<td>0.2</td>
<td>144</td>
<td>17</td>
<td>63.9</td>
<td>1.7</td>
<td>61.3</td>
</tr>
<tr>
<td>105-AF488/594</td>
<td>62.7</td>
<td>0.7</td>
<td>82</td>
<td>15</td>
<td>66.7</td>
<td>3.9</td>
<td>60.6</td>
</tr>
</tbody>
</table>

Incorporation of pAzF at residue 603 and 105 had little impact on the thermal transition midpoint, Tm, relative to S385Y/D469T/R520Q, and, rather surprisingly, increased the van’t Hoff enthalpy for unfolding, ΔHvh, for 603unlabelled and 105unlabelled by 77% and 21%, respectively, indicating improved overall packing. Labelling at residue 105 had minimal impact on the Tm and ΔHvh, while labelling at residue 603 appeared to slightly destabilise the variant, as shown by the 1.2 ºC decrease in Tm.

The aggregation transition midpoints, Tagg, derived from the SLS for each labelled and unlabelled mutant, were all within error of 65 ºC. The temperature at which 10% of the sample was aggregated, Ton, was also similar in each labelled and unlabelled variant, although qualitatively, it appears the onset of aggregation may occur at a slightly higher temperature in S385Y/D469T/R520Q.

Overall, incorporation and labelling of pAzF at residue 603 and 105 had little impact on the global stability and aggregation relative to S385Y/D469T/R520Q. Furthermore, thermal unfolding, indicated by the Tm derived from the fluorescence intensity ratio data, occurred at approximately 2 ºC lower than Tagg, for each unlabelled and labelled variant, indicating that thermal aggregation continued to be a direct result of global thermal unfolding in these experiments.
Effect of [TK] on global stability for S385Y/D469T/R520Q, 603-AF488/594 and 105-AF488/594

As the UNit used above cannot measure the FRET signal, the global stability of S385Y/D469T/R520Q, 603-AF488/594 and 105-AF488/594 was additionally assessed by measuring the intrinsic fluorescence ratio (350/330) as a function of temperature using a Fluoromax-4 which can also measure the FRET signals. This ensured that the global stability of S385Y/D469T/R520Q, 603-AF488/594 and 105-AF488/594, indicated by the change in intrinsic fluorescence ratio (350/330), could be compared to the local stability of 603-AF488/594 and 105-AF488/594, as indicated by the change in distance and therefore FRET signal between Alexafluor dyes in 603-AF488/594 and 105-AF488/594, all as a function of temperature. The global stability was determined between 0.04-0.5 mg/mL to analyse the impact of [TK], and hence macromolecular crowding or aggregation propensity.

S385Y/D469T/R520Q and 105-AF488/594 were shown to have similar thermal transition midpoints, $T_m$, while 603-AF488/594 was slightly less stable at all protein concentrations (Figure 48 and 49). These results agreed with those measured using the Unit, although the absolute $T_m$ values are slightly lower when measured using the Fluoromax-4 because the heating rate (4 °C/min) was double that with the UNit (2 °C/min). In general, all three variants had the same trend in $T_m$ values as a function of protein concentration. Initially, as the concentration was increased from 0.04 mg/mL to 0.3 mg/mL, the entropic hard-core repulsions between homodimers appeared to stabilise holo-transketolase by the formation of a more compact dimer that was also less likely to dissociate into less-stable monomeric subunits before unfolding, as indicated by the increased van’t Hoff enthalpy for unfolding, $\Delta H_{v\theta}$. At higher concentrations, the global thermal stability of transketolase decreased as unfavourable chemical interactions between homodimers began to dominate and caused a decrease in $\Delta H_{v\theta}$. It is likely that the decrease in thermal stability reflected an increased contribution from aggregation. The increased heating rate when using the Fluoromax-4 compared to the UNit at 0.5 mg/mL TK significantly increased the $\Delta H_{v\theta}$ of S385Y/D469T/R520Q, and indicated a greater contribution from stabilising, hard-core repulsions compared to unfavourable chemical interactions between homodimers at higher heating rates. In contrast, the $\Delta H_{v\theta}$ values of 603-AF488/594 and 105-AF488/594 were relatively unchanged at each heating rate.
Figure 48: The global (350/330) and local (FRET) denaturation curves of purified TK variants S385Y/D469T/R520Q, 603-AF488/594 and 105-AF488/594, measured using the Fluoromax-4. A) C) and E) the global denaturation curves of S385Y/D469T/R520Q, 603-AF488/594 and 105-AF488/594, respectively, from the change in the normalised intrinsic fluorescence ratio as a function of temperature. B) No local stability data was available for S385Y/D469T/R520Q without pAzF incorporation and labelling. D) and F) the local denaturation curves of S385Y/D469T/R520Q, 603-AF488/594 and 105-AF488/594, respectively, from the change in the normalised FRET intensity as a function of temperature. All samples consisted of 0.04 mg/mL (black), 0.08 mg/mL (red), 0.16 mg/mL (blue), 0.25 mg/mL (orange), 0.32 mg/mL (magenta), or 0.5 mg/mL (green) with 9 mM Mg$^{2+}$ and 2.4 mM TPP. Data was fitted to the two- and three- state models of protein denaturation (88–90,154).
Figure 49: The change in global and local thermal stability of S385Y/D469T/R520Q, 603-AF488/594 and 105-AF488/594 as a function of [holoTK]. A) and B) – The change in global stability parameters (350/330) of S385Y/D469T/R520Q (black), 603-AF488/594 (red) and 105-AF488/594 (blue), as a function of [TK]. C) and D) Examples of the change in the normalised FRET signal intensity of 603-AF488/594 and 105-AF488/594, respectively, at 0.16 mg/mL as a function of temperature, fitted to a function describing a two-state and three-state denaturation transition, respectively. E) and F) A comparison between the change in global (350/330; black) and local (FRET; red) stability parameters of 603-AF488/594, as a function of [TK]. G) and H) A comparison between the change in global (350/330; black) and local (FRET; blue and magenta) stability parameters of 105-AF488/594, as a function of [TK]. Global (350/330) and local (FRET) stability parameters were determined by the change in intrinsic fluorescence ratio and FRET signal, respectively, at 0.04-0.5 mg/ml TK, 2.4 mM TPP, 9 mM Mg, in 50 mM PBS, pH 7.0, using a Fluoromax-4.

**Utilisation of the AF488/AF594 FRET system to probe the effects of macromolecular crowding on the local thermal stability of purified transketolase**

The pAzF-labelled AF488/AF594 FRET system was used to probe for any local thermal unfolding and/or intermolecular interactions at residues 603 and 105 of purified 603-AF488/594 and
105-AF488/594, respectively. The FRET signal during thermal infolding was compared to the profile from the intrinsic fluorescence ratio (350/330), to identify any deviations from the global denaturation. The change in FRET signal ($\lambda_{ex} = 488\text{ nm}; \lambda_{em} = 620\text{ nm}$) of 603-AF488/594 and 105-AF488/594 was determined as a function of temperature using the Fluoromax-4 at between 0.04-0.5 mg/mL TK, 9 mM Mg$^{2+}$ and 2.4 mM TPP (Figure 48 and 49). As the FRET signal intensity is dependent on inter-subunit FRET distance, a decrease in signal was interpreted as protein unfolding or subunit dissociation, while an increase in signal was interpreted as intermolecular aggregation.

The local thermal denaturation pathway of the C-terminal domain was predominantly via unfolding, given the observed decrease in FRET intensity of 603-AF488/594 as a function of temperature (Figure 48D and 49C). The aggregation known to occur as a result of global unfolding did not lead to an increase in FRET for 603-AF488/594, indicating that the aggregates did not bring two 603 residues from different homodimers into close proximity (<100 Å) (Figure 50). In other words, aggregation does not appear to be driven via contacts between C-terminal domains.

**Figure 50:** Schematic diagram of the aggregation of two transketolase molecules. The two subunits of each dimer are shown in grey and red, with the location of residue 603 and 105 shown as blue and green circles, respectively. The striped green circle represents a residue hidden from view. The intra- and inter-subunit distances are shown in yellow and black, respectively.
The FRET-measured local $T_m$ of the C-terminal domain of 603-AF488/594 was the same as the global $T_m$ at 0.04 mg/mL TK (Figure 49E), which suggested macromolecular crowding had little impact on both global and C-terminal domain denaturation at low [TK]. As [TK] was increased, the global $T_m$ increased at a greater rate than the local $T_m$, which indicated the C-terminal domain was only one of many local regions that contributed to increased global stability. These local regions, including the C-terminal domain, were most-likely stabilised as hard-core repulsions, under increasing macromolecular crowding, forced them into a more compact structure, as also indicated by the increase in $\Delta H_{vh}$. At high [TK], a relatively constant local $T_m$ and only a slight decrease in the local $\Delta H_{vh}$ (Figure 49F) contrasted to a decrease in both global $T_m$ and $\Delta H_{vh}$, which indicated the C-terminal domain contributed little to global destabilisation at high [TK]. Despite this, the modest discrepancy between the local and global $T_m$ between 0.04-0.5 mg/mL TK implied that the C-terminal domain stability was a relatively good indicator of global stability, and highlighted the importance of C-terminal domain stability on global stability. The shape of the FRET-measured transitions indicated no obvious influence of aggregation (i.e. a positive thermal transition) (Figure 51A). Furthermore, the normalised FRET magnitude of the unfolding transition, defined as the unfolded FRET signal relative to the initial folded FRET signal, was relatively stable between 0-0.32 mg/mL TK (Figure 51B), but decreased at 0.5 mg/mL, which suggested the kinetic equilibrium between unfolding and aggregation of the C-terminal domain had shifted towards the former.
Figure 51: The local denaturation (FRET) profiles and change in the normalised FRET magnitude or fractional aggregation of 603-AF488/594 and 105-AF488/594 as a function of [holoTK]. The change in FRET signal of A) purified 603-AF488/594 and C) purified 105-AF488/594 at 0-0.5 mg/mL TK, 9 mM Mg\(^{2+}\) and 2.4 mM TPP, measured using a Fluoromax-4. B) The normalised FRET magnitude of purified 603-AF488/594, defined as the unfolded FRET intensity relative to the folded FRET intensity. D) The fractional aggregation of purified 105-AF488/594, defined as the change in fluorescence intensity of transition two relative to the total combined magnitude of transition one and two.

The local thermal denaturation pathway of loop 105, again measured using the FRET assay, was more complex. It consisted of two convoluted thermal transitions, forming a distinctive ‘V-shaped’ curve (Figure 49D). The fluorescence intensity decreased during the first transition, which indicated local unfolding or increased flexibility of loop 105 at a temperature 8 °C lower than the global \(T_m\). The first transition was interrupted mid-transition by the second transition, which increased the fluorescence intensity as a result of intermolecular aggregation. Aggregates in this case created new FRET pairs by bringing residue 105 in close proximity (<100 Å) to another residue 105 from a different homodimer (Figure 50). The convolution of the two transitions was interpreted as evidence
that local unfolding and/or increased flexibility of loop 105 at the lower temperature transition resulted in the formation of an intermediate native-like structure, prior to the aggregation at the higher temperature, and that the flexibility at or close to loop 105 may play a role in the formation of aggregates.

The local $T_m$ of the first transition of 105-AF488/594, measured via FRET (Figure 49G), was approximately equal to the onset temperature of the global unfolding transition, measured via the intrinsic fluorescence ratio. The local $T_{agg}$ of loop 105 of 105-AF488/594 (Figure 49G), was at a temperature comparable to the global $T_m$ of 105-AF488/594 (Figure 49G). Together, these results implied that local unfolding of loop 105 at approximately 54 ºC may have initiated both global unfolding and aggregation between approximately 54-70 ºC. This is in agreement with measurements taken using the Unit above, and previous experiments (51), which have demonstrated a significant overlap between thermal aggregation and unfolding of transketolase.

Both the local $T_m$ and the higher local $T_{agg}$ of 105-AF488/594 were invariant to [TK] between 0.04-0.5mg/mL (Figure 49G), which suggested macromolecular crowding had no impact on the stability of loop 105 at these concentrations. Conversely, the global $T_m$ of mutant 105-AF488/594 followed the same trend as the other variants – an increase at low [TK] followed by a decrease at high [TK]. The global unfolding $\Delta H_{vh}$ profile of 105-AF488/594, determined by the intrinsic fluorescence ratio (350/330), suggested that there is an overall compaction of the protein as the TK concentration is increased. However, the FRET analysis for this variant suggests that the change in $\Delta H_{vh}$ is not contributed to by the 105 loop region, as the fits to the thermal denaturation profiles measured by FRET at residue 105 do not reveal any change in $\Delta H_{vh}$ as a function of TK concentration.

The local structure of loop 105 may have responded significantly less to macromolecular crowding at these concentrations compared to the C-terminal domain because of their relative sizes. Larger structural motifs, such as entire domains, are forced into more compact structures at lower concentrations than small, flexible loops. The total compaction of the global structure may also arise from the suppression of some but not all locally flexible regions, dependent on their relative local stability and solvent accessibility.

Despite no change in the thermodynamics of unfolding and aggregation, macromolecular crowding impacted the kinetics of local thermal denaturation of loop 105. The fractional aggregation, defined as the change in fluorescence intensity of transition two relative to the total combined
magnitude of transition one and two, increased with transketolase concentration and hence shifted the kinetic pathway of denaturation towards aggregation (Figure 51C and D).

The above study validated the use of the pAzF-labelled AF488/AF594 FRET system for the detection and deconvolution of local thermal unfolding and aggregation. It indicated that TK aggregation is driven through partial unfolding events at residue 105 and/or at other nearby residues in the PP-domain of TK, whereas the C-terminal domain contributes to global unfolding but does not directly interact with other C-terminal domains in aggregates. This shows that further labelling studies in different locations could be used to pin-point other regions that locally unfold and/or cross-interact during aggregation.

**Local denaturation profiles provide information on the drivers of global unfolding and aggregation**

The present study focused on only two loci of a relatively large protein of 73 kDa per subunit (146 kDa per dimer), yet a comparison between global unfolding and local unfolding/aggregation at residues 603 and 105 has already provided important structural insights into the impact local structural motifs have on the choice of thermal denaturation pathway (unfolding vs aggregation).

As explained previously, the global denaturation profile of transketolase can be divided into two phases; i) stabilisation by hard-core repulsions at low transketolase concentrations; and ii) destabilisation/aggregation by attractive chemical interactions at higher transketolase concentrations. A strong correlation between the global and local denaturation profiles of 603-AF488/594 during the first stabilisation phase suggests the C-terminal domain is a major driver of global stabilisation through the macromolecular crowding effect at low transketolase concentrations. Or put another way, the inherent flexibility of the C-terminal domain at low transketolase concentrations drives both local and global unfolding. Similarly, the impact of loop 105 may be greatest at high [TK], where the global $T_m$ of mutant 105-AF488/594 is similar in both trend and value to the local $T_{agg}$ value.

By repeating the present study at more loci to better understand local unfolding throughout transketolase, additional drivers of global instability and aggregation could be identified. This information can then be used to design semi-rational directed evolution libraries to improve global stability and aggregation.
Utilisation of the AF488/AF594 FRET system as a reporter of local thermal stability in cell-like conditions (cell lysate)

The bio-orthogonal nature of the stability assay offers a potential compatibility with cell lysate samples and hence hypothetically many more applications, including high-throughput stability screens. The pAzF-labelled AF488/AF594 FRET system was adapted to cell lysate and used to determine the effect of macromolecular crowding on local thermal stability, again using the labelled mutants 603-AF488/594 and 105-AF488/594, within cell-like conditions over a range of lysate concentrations, by labelling purified TK and subsequently adding it back to cell lysate at a TK concentration of 0.16 mg/mL. As with purified transketolase, transketolase unfolding and aggregation were successfully detected in a cell lysate background, with two- and three-state transitions observed in 603-AF488/594 (Figure 52A) and 105-AF488/594 (Figure 52C), respectively, and were repeated over a range of lysate concentrations (Figure 53).

Figure 52: The local denaturation (FRET) profiles and change in the normalised FRET magnitude or fractional aggregation of 603-AF488/594 and 105-AF488/594 as a function of [lysate]. The change in FRET signal of 0.16 mg/mL A) purified 603-AF488/594 and C) purified 105-AF488/594, incubated with at 0-13.6 mg/mL clarified lysate, 9 mM Mg\(^{2+}\) and 2.4 mM TPP, measured using a
Fluoromax-4. B) The normalised FRET magnitude of purified 603-AF488/594, defined as the unfolded FRET intensity relative to the folded FRET intensity. D) The fractional aggregation of purified 105-AF488/594, defined as the change in fluorescence intensity of transition two relative to the total combined magnitude of transition one and two.

**Figure 53**: The change in local thermal stability (FRET) of residue 603 and 105 as a function of [lysate]. The change in A) $T_m$ and B) $\Delta H_{vh}$ of 603-AF488/594; the change in C) $T_m$ (blue) and $T_{agg}$ (magenta) and D) $\Delta H_{vh}$ of 105-AF488/594, at 0.16 mg/mL TK, 2.4 mM TPP, 9 mM Mg$^{2+}$ and 0-13.6 mg/mL clarified lysate.

The local $T_m$ of the C-terminal domain of 603-AF488/594, measured using the FRET assay, increased as a function of lysate concentration in an isotherm-like trend (**Figure 53A**) and the maximum local $T_m$ observed was 3 °C higher in lysate compared to purified TK. Furthermore, like at high purified [TK], no decrease in $T_m$ was observed at high [Lysate]. This was reflected by an increase in $\Delta H_{vh}$ at high [lysate] (**Figure 53B**), which suggested compaction of the C-terminal domain in cell-like conditions, but contrasted to the slight decrease in $\Delta H_{vh}$ observed at high purified [TK].
The normalised FRET magnitude, an indicator of the kinetic equilibrium between unfolding and aggregation denaturation pathways, decreased in an isotherm-like trend (Figure 52B), which suggested a shift towards unfolding denaturation pathways of the C-terminal domain. It was unclear if the increase in FRET signal, observed above 340 K at only 3.0 mg/mL lysate (Figure 52A), was evidence of the emergence of a second, positive transition at high [lysate] and hence a shift towards C-terminal domain aggregation rather than unfolding pathways. Further evidence at higher [lysate] is required to confirm this shift.

Both the thermodynamics and kinetics of the local thermal denaturation of 105-AF488/594, measured using the FRET assay, were altered by the presence of lysate. Both the \( T_m \) and \( T_{agg} \) of mutant 105 increased at low lysate concentrations as loop 105 was compacted by the lysate macromolecules (Figure 53A and B). At higher lysate concentrations, not only did both thermodynamic constants decrease, but the three-state thermal denaturation profile reverted to a single transition with only two states. Subsequent analysis of the fractional aggregation as a function of lysate concentration revealed partial and eventually total inhibition of inter-subunit aggregation at loop 105 (Figure 52C and D), which explained the disappearance of \( T_{agg} \) at 14 mg/mL lysate. The local denaturation pathway of loop 105 in cell lysate is therefore dictated by both kinetics and thermodynamics. It is unclear if the shift in denaturation pathway is towards unfolding or simply another aggregation pathway, such as aggregation between transketolase and other proteins within the cell lysate.

**DISCUSSION**

The aim of this work was to demonstrate an assay that could measure protein stability in terms of local unfolding, and local intermolecular interactions within aggregates, with the ultimate goal of analysis within complex cellular milieu. Therefore, it was important that ncAA incorporation and subsequent labelling didn’t significantly impact overall protein stability. While this was certainly true of mutant 105-AF488/594, it was also modest for 603-AF488/594, with a 1.2 ºC decrease in \( T_m \) relative to S385Y/D469T/R520Q. The decrease in \( T_m \) was a result of labelling rather than ncAA incorporation, and hence may be due to a slight perturbation of local structure due to the presence of the label.

In this study, the investigation was limited to two locations; in reality, a multiplicity of local structural motifs will impact global stability in different ways. Increased coverage across the protein surface could, in future, build a more complete picture of the local drivers of global stability and
instability, while simultaneously providing insights into the regions that should be targeted by directed evolution libraries. This assay would be particularly synergistic in combination with molecular dynamic simulations that identify particular regions to be more flexible than others (54).

I have demonstrated the potential that this assay has to measure stability in cell-like conditions by introducing labelled purified transketolase back into cell lysate. The lysate concentrations used in this study were well below those experienced in cellulo, but were not limited by the assay and I am confident that detection of unfolding and aggregation at cell-like macromolecular concentrations can be achieved. However, labelling and subsequent removal of free dye in cell lysate has yet to be demonstrated. Preliminary experiments have indicated background noise due to non-specific interactions between dye molecules and other cellular proteins may be an issue moving forward. However, this challenge can be overcome by optimising the labelling conditions, changing the fluorophore and/or choosing alternative available click chemistries.

With such developments, this assay could in theory, be adapted to ultra-high-throughput systems such as microdroplet systems, which can screen $>10^8$ variants in 10 hours (156), and hence utilised as a direct, bio-orthogonal stability screen for the directed evolution of any protein, including vaccines, therapeutic proteins and biocatalysts.

**SUMMARY**

This chapter developed and implemented a novel FRET-based stability assay to quantify the dynamic unfolding and aggregation of proteins during a temperature ramp. While the ncAA pAzF and the two Alexafluor dyes have been used previously to interrogate protein structure and stability under different conditions (153), to this author’s knowledge this is the first time double ncAA- incorporation and labelling have been used to study unfolding in real time via FRET, both in purified protein and in a lysate background. The local stability (via FRET measurements) was compared to global stability (via intrinsic fluorescence ratio measurements), providing insights into the impact of particular local structures on global stability. This methodology could be complimented by in silico techniques to understand local structure and ultimately to evolve or engineer protein stability. This assay can, in theory, be applied universally to any protein, including biocatalysts, vaccines and therapeutic proteins.
FINAL SUMMARY, DISCUSSION & FUTURE WORK

FINAL SUMMARY

Research output:

1. Two fluorescence-quenching-based TPP- and donor-substrate-binding assays for transketolase
2. The Two-Species Model of transketolase activation, regulation and inhibition
3. A ncAA-incorporation platform that will be adapted to other proteins in the Dalby lab
4. A methodology of utilising the efficiency and fidelity of incorporation to extract true kinetic and stability parameters for the ncAA-incorporated species
5. Highly stable and catalytically active aromatic aldehyde-accepting transketolase variants through ncAA-incorporation saturation mutagenesis
6. A FRET-based stability assay that will be adapted to other proteins in the Dalby lab

Summary - Chapter Three:

- Established the methodology of ncAA incorporation.
- First application of pUltra/C321 in transketolase research.
- To my knowledge a novel in-depth quantitative analysis of the fidelity of incorporation by fitting mass spectra to the sum of multiple Gaussian functions.
- Analysis of the efficiency and fidelity of incorporation of four para-substituted phenylalanine derivatives and the creation of a new, inducible (tacI) transketolase expression plasmid with improved protein yield but, unfortunately, similar fidelities of incorporation.

Summary - Chapter Four:

- Utilised the ncAA incorporation platform established in Chapter Three to engineer the active site of a previously-evolved transketolase variant, S385Y/D469T/R520Q.
- Used a smart, small library of 5 para-substituted phenylalanine derivatives, three of which were ncAAs.
- A novel function (the Modified Michaelis-Menten function) was derived to describe and quantify the extensive substrate inhibition observed at high [3-HBA].
- A novel approach to accounting for misincorporation, and sub-species activity in the determination of the true catalytic/stability parameters of ncAA-incorporated species.
The best all-round performer was pAMF, which had a 2.6-fold, 2.4-fold and 1.1-fold improvement in catalytic efficiency, $K_m$ and $k_{cat}$, respectively; and a 5.4 °C increase in $T_m$.

To my knowledge, pAMF is the first example of an ncAA-incorporated variant with both enhanced activity and stability at 22 °C, and demonstrates the benefits of including ncAAs in site-specific, directed evolution libraries.

Exemplifies the power of protein engineering with ncAAs in order to augment catalytic performance and/or stability.

May serve as a useful template outlining one approach to protein engineering via rational active site mutagenesis and subsequent analysis of structure-function relationships.

Summary - Chapter Five:

- Developed a novel fluorescence-quenching transketolase TPP-binding assay.
- Detected and characterised a previously-unknown low-affinity (TPP) and low-activity transketolase sub-species, TK$_{low}$, and the previously characterised TK$_{high}$.
- TK$_{low}$ was 20-fold less-active than TK$_{high}$.
- It explored the origin and interconversion of the two sub-populations via oxidation of Cys157 and potentially other active-site cysteines/methionines.
- Proposed a novel Two-Species Model of transketolase activation, and commented on the mechanistic and physiological insights it provided into transketolase function and regulation.
- Discovered a novel mechanism of redox regulation, and the first within the non-oxidative phase of the pentose phosphate pathway.

Summary - Chapter Six:

- Further investigated the preservation of TK$_{high}$ and TK$_{low}$ across variants and from cofactor (TPP) to donor substrate.
- Adapted the fluorescence-quenching assay to a novel HPA-binding assay.
- The $K_d$ of HPA binding to wild-type transketolase was consistent with previously determined $K_m$ and $K_i$ values, respectively, and implicated TK$_{low}$ in substrate inhibition.
- Provided evidence of the existence of three dimeric species, TK$_{high}$-TK$_{high}$, TK$_{low}$-TK$_{low}$, and the mixed dimer TK$_{high}$-TK$_{low}$, and estimated their proportions.
- Formation of the mixed dimer was responsible for the previously unexplained phenomenon of donor substrate inhibition and heat-activation.
- The proportion of each dimeric species were estimated from data related to these phenomena.
• Proposed the updated Two-Species Model of transketolase regulation, activation and inhibition.
• The model will have a significant impact on future transketolase research, and is yet another example of redox regulation through active site cysteines.

Summary - Chapter Seven:
• Developed and implemented a novel FRET-based stability assay for quantification of the dynamic unfolding and aggregation of proteins during a temperature ramp.
• While the ncAA pAzF and the two Alexafluor dyes have been used previously to interrogate protein structure and stability under different conditions, to this author’s knowledge this is the first time double ncAA-incorporation and labelling have been used to study unfolding in real time via FRET, both in purified protein and in a lysate background.
• The local stability (via FRET measurements) was compared to global stability (via intrinsic fluorescence ratio measurements), providing insights into the impact of particular local structures on global stability.
• This methodology could be complimented by in silico techniques to understand local structure and ultimately to evolve or engineer protein stability.
• This assay can, in theory, be applied universally to any protein, including biocatalysts, vaccines and therapeutic proteins.

FINAL DISCUSSION & FUTURE WORK
The Two-Species Model of transketolase activation, regulation and inhibition should have a significant impact on transketolase research, since the detection and characterisation of TK\text{low} will influence future calculations of true catalytic performance of transketolase variants. The importance of the model is self-evident in its re-emergence in each chapter. Furthermore, since transketolase activity can be roughly tripled simply by maximising the \%TK\text{high}, efforts should be made to exclusively express TK\text{high} or a TK\text{high}-like form of transketolase. For example, mutation at Cys157 to Thr, Ser or Asp, in an attempt to mimic the polarity or charge of the Cys157 sulfenic acid, may yield TK\text{high}-like variants with improved TPP-binding and activity, and reduced inhibition. Exploration at Cys157 with ncAAs, potentially even directly incorporating a cysteine sulfenic acid, may yield further-improved TK\text{high} or TK\text{high-like} variants. In addition, fermentation conditions should be optimised to maximise oxidation of TK\text{low} to TK\text{high}.
The locating of the sulfenic acid at Cys157 may have implications in the analysis of future crystal structures, and may even present an opportunity to compare the crystal structures of pure TK\textsubscript{low}, TK\textsubscript{high} and heat-activated TK. Furthermore, the potential role of the sulfenic acid in completion of the proton wire between the active sites of the holo-dimer should be investigated further, as should the physiological relevance of heat-activation and half-of-the-sites reactivity.

The study into active site ncAA incorporation provides a success story that will hopefully pave the way for many more similar studies. Indeed, this PhD has only scratched the surface; the concept can be taken much further, and the breadth of ncAA structures that can now be incorporated is vast, and the possibilities endless. For example, further evolution around an incorporated active site ncAA is possible, as is incorporation of further ncAAs at more positions in the active site.

It is hoped that the FRET-based stability assay can be adapted to virtually any protein. It’s potential compatibility with cell lysate, or even \textit{in cellulo} labelling, means that this research has strong potential to lead the field in the directed evolution of protein stability.


88. Santoro MM, Bolen DW. Unfolding Free Energy Changes Determined by the Linear


BIBLIOGRAPHY


131. Pais EF, Schulz GE. The Catalytic Mechanism of Glutathione Reductase as Derived from X-ray Diffraction Analyses of Reaction Intermediates.


APPENDIX ONE: CPEC schematic diagram
APPENDIX TWO: *p*CNF as a genetically-encoded fluorescent probe

**Introduction**

Fluorescence is a highly sensitive analytical and diagnostic tool that occurs on the nanosecond timescale and is sensitive to changes in protein environment at the nanometre level, and is hence an ideal nano-scale probe. It is common practice to use fluorophores in biochemistry and biophysical chemistry to probe protein structure and dynamics, including ligand-induced conformational changes, protein-protein and protein-nucleic acid interactions, protein trafficking, and enzyme activity (157). GFP-fusion proteins, smaller fluorescent tags attached to the protein surface and fluorophores conjugated to reactive amino acids such as cysteine have been widely utilised for these investigations. However, their versatility is often limited by their large size, which can interfere with native interactions, function or location; and the fact that they must generally be located on the protein surface (157). Consequently, extrinsic fluorescent probes are rarely introduced into protein active sites because of the resulting structural perturbations and hence reduced enzyme activity.

The nitrile group of *p*CNF makes it a useful IR probe for local environment (158–161). Furthermore, *p*CNF is, in theory, an excellent fluorescent probe for protein binding, conformational change and protein folding. It is similar in structure to Phe/Tyr and, when replacing these amino acids in a protein sequence, are likely to have little impact on local and global protein structure. It can stably reside in both the hydrophobic core or the solvent-exposed protein surface, since its polarity is between an amide group and a methylene. The absorption spectrum of *p*CNF in water has three maxima in the deep-UV region (233, 274 and 280 nm) ([Figure A1](#)) and the fluorescence emission wavelength (295 nm) is independent of the excitation wavelength (e.g. 233 or 280 nm) (162). The overlap of the emission and excitation spectra of *p*CNF and Trp, respectively, can result in Förster resonance energy transfer (FRET) from *p*CNF to Trp at Förster distances of \(\leq 16\) Å and hence has been utilised in the study of peptide conformational (94) and unfolding (95) studies.
Figure A1: Absorption spectra of $p$CNF in different solvents. [$p$CNF] was approximately 25 µM in water, ethanol and tetrahydrofuran, and approximately 370 µM in water (162).

The fluorescence intensity (i.e. quantum yield) of $p$CNF is highly sensitive to its local environment and is mediated by the strength of H-bonding between the nitrile group of $p$CNF and H-bond donors; the stronger the H-bond (e.g. in different solvents; Figure A2), the greater the fluorescence (162). Moreover, like benzonitrile formation in the gas phase, $p$CNF may also exist in a bent conformation with a ‘dark’ S1 state that is nearly isoenergetic with the linear conformation in the solution phase. This state would present a major non-radiative decay channel of the excited state population and therefore hindrance of this bending motion by H-bonding may limit the non-radiative decay rate, thus increasing fluorescence lifetime (162).
Figure A2: Fluorescence emission spectra of $p$CNF in different solvents. $[p$CNF$]$ was approximately 370 µM and the excitation wavelength was 275 nm (162).

In theory, $p$CNF has the potential to be a superior fluorescent probe to Phe, Tyr and Trp because it a) has a five-fold greater quantum yield than Phe when excited at 240 nm; b) can be selectively introduced at a single specific site of the protein of interest; and c) can be selectively excited in the presence of other fluorescent aromatic amino acids because of its red-shifted absorption spectrum. For example, the molar absorptivity of $p$CNF at 240 nm is about 7 times greater than that of Trp, 50 times greater than that of Tyr and more than 400 times greater than that of Phe (94).

However, in practice, $p$CNF fluorescence is often quenched by surrounding amino acid side chains over short distances (163), which potentially reduces the effectiveness of $p$CNF as a fluorescence probe. The relative extent of quenching of $p$CNF fluorescence by each amino acid side chain was determined as follows: $Y > H^° > M > > C > H^+ > N > R = K^+$ (where $H^°$ denotes a neutral...
His side chain, and K⁺ and H⁺ denotes a positively charged Lys and His side chain, respectively) (163).

In Chapter Three, the ncAA para-cyanophenylalanine (pCNF) was efficiently and selectively incorporated into the active site of an E. coli transketolase variant (S385Y/D469T/R520Q) at residue 385. It was hypothesised that incorporation of pCNF may potentially offer an alternative method to assess relative enzymatic performance, substrate preference, active site protein dynamics, cofactor binding and/or protein unfolding. Such assays may even be compatible with cell lysate rather than purified protein given the selective excitation of pCNF over Trp, Phe and Tyr at 240 nm. The fluorescent properties of the apo- and holo-forms of variants S385Y/D469T/R520Q (mutant Y) and S385pCNF/D469T/R520Q (mutant pCNF) were therefore characterised and compared to assess the potential use of pCNF as an active-site fluorescent probe.

Results

The fluorescence spectra of apoTK

The fluorescence emission spectra of the apo- forms of each variant were compared at excitation wavelengths of 240 nm and 285 nm (Figure A3). Mutant Y and pCNF had similar emission peaks at approximately 326 nm. As expected, excitation at 240 nm reduced the maximum fluorescence intensity of both variants to 40% of that when excited at 285 nm because of the decreased quantum yield of tryptophan at 240 nm (162). However, no emission peak at 295 nm was observed in mutant pCNF when excited at 240 nm, indicating either pCNF formed a FRET pair with a tryptophan residue within the active site, the pCNF signal was quenched by neighbouring residues, or both. The fluorescence intensity of mutant pCNF was significantly less than mutant Y at 0.08 mg/mL TK, therefore the pCNF fluorescence signal was quenched in apoTK. However, the absence of a direct pCNF-induced peak at 295 nm suggested a potentially highly-complex system that may have limited applications as an effective active-site probe.
Figure A3: Fluorescence spectra of the apo-form of mutant Y and mutant pCNF at 0.08 mg/mL TK, 9 mM Mg\(^{2+}\) and 2.4 mM TPP in 50 mM Tris-HCl and pH 7.0. Samples were excited at 240 nm and 280 nm, bandwidth 8 nm and integration time 0.2 s.

**Structural evidence for FRET and quenching of pCNF fluorescence**

Structural analysis of the energy-minimised structure of the holo-form of mutant pCNF (Chapter Four) identified four possible tryptophan residues, as well as TPP, that may form a FRET pair with 385pCNF (Figure A4). TPP was only 8.4 Å from 385pCNF and had an absorption spectrum that significantly overlaps with the fluorescence emission spectrum of pCNF, and hence was likely to form the strongest FRET interaction in holoTK. In the absence of TPP in apoTK, the strongest FRET interaction was likely to have occurred with W390, however, the four possible candidates created an extremely complex network of interactions that would be difficult to interpret in terms of probing local dynamics. pCNF fluorescence may also be quenched by numerous active-site histidine residues, in addition to Mg\(^{2+}\), which potentially explains the fluorescence quenching observed in apoTK.
**Figure A4:** The energy-minimised structure (see Chapter Four) of mutant pCNF. The possible FRET-distances between 385pCNF and TPP (anti-clockwise from top right: W491, W503, W390 and W196; salmon red) or bound TPP (magenta) are depicted as dotted yellow lines.

**TPP absorption spectra at pH 6.5-7.8**

A previous study on yeast transketolase showed the intrinsic fluorescence intensity of transketolase decreased upon the addition of cofactor (164). However, the study did not consider the impact of the inner filter effect (IFE) that arises from strong absorption of a proportion of the incident light by free ligand (TPP) before it can excite the sample (‘primary’ IFE), decreasing the observed fluorescence (165).

The absorption spectrum of Free TPP had strong absorbance at both 240 nm and 285 nm (Figure A5), especially in comparison to apoTK (Table A1), which indicated the observed fluorescence intensity of holoTK will almost certainly be distorted at high [TPP]. In addition, the absorbance spectra were independent of pH at between 6.5-7.8 and significantly overlapped with the emission spectrum of pCNF (Figure A2). A pCNF-TPP FRET interaction is therefore likely to occur in holoTK.
### Table 16: The absorbance of 0.8 mM TPP and 0.2 mg/mL apoTK (S385Y/D469T/R520Q) at 240 nm and 285 nm.

<table>
<thead>
<tr>
<th>Wavelength</th>
<th>TPP (0.8 mM)</th>
<th>ApoTK (0.2 mg/mL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>240 nm</td>
<td>0.735</td>
<td>0.037</td>
</tr>
<tr>
<td>285 nm</td>
<td>0.351</td>
<td>0.025</td>
</tr>
</tbody>
</table>

**Figure A5:** Absorption spectra of 0.8 mM TPP in solution at pH 6.5, 6.8, 7.0, 7.2, 7.5 and 7.8.

**The fluorescence spectrum of holoTK**

The fluorescence intensity of both variants significantly decreased by over 99% upon the addition of 9 mM Mg$^{2+}$ and 2.4 mM TPP ([Figure A6](#)), and the spectra resembled that of free TPP, which again indicated a strong IFE at high concentrations of free TPP. Again, no emission peak was observed at 295 nm. However, in contrast to apoTK, the fluorescence signal of mutant $\rho$CNF was 50% higher than that of mutant Y. TPP-binding to the active site, only 8.4 Å from $\rho$CNF, a) reduced quenching from surrounding amino acids, and b) established a FRET interaction between $\rho$CNF and TPP, and hence increased the overall fluorescence signal.
Figure A6: Fluorescence spectra of 2.4 mM TPP (green), and the holo-forms of mutant Y (black) and mutant pCNF (red) at 0.08 mg/mL TK, 9 mM Mg\(^{2+}\) and 2.4 mM TPP in 50 mM Tris-HCl and pH 7.0. Samples were excited at 240 nm bandwidth 8 nm and integration time 0.2 s.

In summary, pCNF was an ineffective active site probe in transketolase because of the complex network of fluorescent interactions within the active site. However, the results indicated that fluorescence quenching, when corrected for the IFE, could potentially be used to determine the binding parameters of TPP, and potentially also HPA, independently of enzyme activity. The potential for the development of a protein intrinsic fluorescence-quenching-based assay for cofactor- and donor substrate-binding was investigated further in Chapters Five and Six.
APPENDIX THREE: Trypsin digest and mass spectrometry supplementary material

*Labelling with dimedone:*

<table>
<thead>
<tr>
<th>Protein Group</th>
<th>Protein ID</th>
<th>Accession</th>
<th>Score (%)</th>
<th>-10logP</th>
<th>Coverage (%)</th>
<th>#Peptides</th>
<th>#Unique</th>
<th>PTM</th>
<th>Avg. Mass</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>32609</td>
<td>P27302</td>
<td>TK1_ECOLI</td>
<td>99.2</td>
<td>515.94</td>
<td>87</td>
<td>350</td>
<td>334</td>
<td>Y</td>
<td>Transketolase 1 G5=Escherichia coli (strain K12) GN=k5 TA PE=1 SV=5</td>
</tr>
<tr>
<td>3</td>
<td>32614</td>
<td>B1IRG0</td>
<td>DNAK_ECOLC</td>
<td>99.2</td>
<td>400.54</td>
<td>72</td>
<td>68</td>
<td>68</td>
<td>Y</td>
<td>Chaperone protein Dnak O S=Escherichia coli (strain ATCC 8739 / DSM 1578 / Crook) GN=dnak PE=3 SV=1</td>
</tr>
<tr>
<td>3</td>
<td>32615</td>
<td>A7ZV7V</td>
<td>DNAK_ECOHS</td>
<td>99.2</td>
<td>400.54</td>
<td>72</td>
<td>68</td>
<td>68</td>
<td>Y</td>
<td>Chaperone protein Dnak O S=Escherichia coli O9:H4 (strain HS) GN=dnak PE=3 SV=1</td>
</tr>
<tr>
<td>3</td>
<td>32616</td>
<td>A7ZHA4</td>
<td>DNAK_ECO24</td>
<td>99.2</td>
<td>400.54</td>
<td>72</td>
<td>68</td>
<td>68</td>
<td>Y</td>
<td>Chaperone protein Dnak O S=Escherichia coli O139:H28 (strain E2437 7A / ETEC) G N=dnak PE=3 SV=1</td>
</tr>
<tr>
<td>3</td>
<td>32617</td>
<td>P0A6Y9</td>
<td>DNAK_ECOL6</td>
<td>99.2</td>
<td>400.54</td>
<td>72</td>
<td>68</td>
<td>68</td>
<td>Y</td>
<td>Chaperone protein Dnak O S=Escherichia coli O6:H1 (strain CPT073 / ATCC 700928 / UPEC) GN=dnak PE=3 SV=2</td>
</tr>
<tr>
<td>3</td>
<td>32618</td>
<td>Q0TLX5</td>
<td>DNAK_ECOL5</td>
<td>99.2</td>
<td>400.54</td>
<td>72</td>
<td>68</td>
<td>68</td>
<td>Y</td>
<td>Chaperone protein Dnak O S=Escherichia coli O6:K15:H31 (strain 536 / UPEC) GN=dnak PE=3 SV=1</td>
</tr>
<tr>
<td>3</td>
<td>32611</td>
<td>A1A766</td>
<td>DNAK_ECOX1</td>
<td>99.2</td>
<td>400.54</td>
<td>72</td>
<td>68</td>
<td>68</td>
<td>Y</td>
<td>Chaperone protein Dnak O S=Escherichia coli O1:K1 / APEC GN=dnak PE=3 SV=1</td>
</tr>
<tr>
<td>3</td>
<td>32612</td>
<td>Q1RGI8</td>
<td>DNAK_ECOUT</td>
<td>99.2</td>
<td>400.54</td>
<td>72</td>
<td>68</td>
<td>68</td>
<td>Y</td>
<td>Chaperone protein Dnak O S=Escherichia coli (strain UT 189 / UPEC) G N=dnak PE=3 SV=1</td>
</tr>
<tr>
<td>3</td>
<td>32613</td>
<td>P0A6Z9</td>
<td>DNAK_ECOS7</td>
<td>99.2</td>
<td>400.54</td>
<td>72</td>
<td>68</td>
<td>68</td>
<td>Y</td>
<td>Chaperone protein Dnak O S=Escherichia coli O157:H7 GN=dnak PE=3 SV=2</td>
</tr>
</tbody>
</table>
1. Notes

2. Result Statistics

Figure 1. False discovery rate (FDR) curve. X axis is the number of peptide-spectrum matches (PSM) being kept. Y axis is the corresponding FDR.

Figure 2. PSM score distribution. (a) Distribution of PEAKS peptide score; (b) Scatterplot of PEAKS peptide score versus precursor mass error.

Figure 3. De novo result validation. Distribution of residue local confidence: (a) Residues in de novo sequences validated by confident database peptide assignment; (b) Residues in "de novo only" sequences.

Table 1. Statistics of data.
| # of MS scans | 18538 |
| # of MS/MS scans | 31659 |

Table 2. Result filtration parameters.

| Peptide -10 logP | ≥15 |
| Protein -10 logP | ≥20 |

Table 4. PTM profile.

<table>
<thead>
<tr>
<th>Name</th>
<th>ΔMass</th>
<th>Position</th>
<th>#PSM</th>
<th>-10 logP</th>
<th>Area</th>
<th>AScore</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oxidation</td>
<td>15.99</td>
<td>M</td>
<td>2043</td>
<td>94.66</td>
<td>4.9E8</td>
<td>1000.00</td>
</tr>
<tr>
<td>Deamidation</td>
<td>98</td>
<td>NQ</td>
<td>1203</td>
<td>99.15</td>
<td>1.4E6</td>
<td>185.60</td>
</tr>
<tr>
<td>Carbamidomethyl</td>
<td>57.02</td>
<td>C</td>
<td>815</td>
<td>103.33</td>
<td>2.22E7</td>
<td>1000.00</td>
</tr>
</tbody>
</table>
3. Experiment Control

Figure 4. Precursor mass error of peptide-spectrum matches (PSM) in filtered result. (a) Distribution of precursor mass error in ppm; (b) Scatterplot of precursor m/z versus precursor mass error in ppm.

<table>
<thead>
<tr>
<th>Missed Cleavages</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4+</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018_11_28 OXalised TC</td>
<td>3655</td>
<td>713</td>
<td>67</td>
<td>8</td>
<td>0</td>
</tr>
</tbody>
</table>

4. Other Information

Table 6. Search parameters.
Search Engine Name: PEAKS
Parent Mass Error Tolerance: 15.0 ppm
Fragment Mass Error Tolerance: 0.6 Da
Precursor Mass Search Type: monoisotopic
Enzyme: Specified by each sample
Max Miss Cleavages: 3
Non-specific Cleavage: one
Variable Modifications:
- Oxidation (M): 15.99
- Deamidation (NQ): 0.98
- Oxidation (C): 15.99
- Carbamidomethylation: 57.02
- Dimedone: 138.08
- Dimedone (2): 139.08
- Dimedone (3): 140.08
Max Variable PTM Per Peptide: 5
Database: Uniprot/Swissprot
Taxon: Escherichia coli
Searched Entry: 23011
FDR Estimation: Enabled
Merge Options: no merge
Precursor Options: corrected

Table 7. Instrument parameters.
Fractron: 20181120_TK-oxidisedCys_01.mw, 20181120_TK-oxidisedCys_02.mw
Ion Source: ESI(nano-spray)
Fragmentation Mode: high energy CID (y and b ions)
MS Scan Mode: FT/ICROrbitrap
MS/MS Scan Mode: Linear Ion Trap
<table>
<thead>
<tr>
<th>Protein Group</th>
<th>Protein ID</th>
<th>Accession</th>
<th>Score (%)</th>
<th>-10LogP</th>
<th>Coverage (%)</th>
<th>#Peptides</th>
<th>#Unique</th>
<th>PTM</th>
<th>Avg. Mass</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>P27302</td>
<td>99.2</td>
<td>329.87</td>
<td>83</td>
<td>171</td>
<td>132</td>
<td>Y</td>
<td>72212</td>
<td>Transketolase</td>
</tr>
<tr>
<td></td>
<td></td>
<td>XT1_ECOLI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1 OS=Escherichia coli (strain K12) GN=sk1 TA PE=1 SV=5</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>PGA9Q8</td>
<td>99.2</td>
<td>241.86</td>
<td>55</td>
<td>33</td>
<td>33</td>
<td>Y</td>
<td>96127</td>
<td>Alcohol-alcohol dehydrogenase OS=Escherichia coli G 157387 GN=na dHE PE=3 SV=2</td>
</tr>
<tr>
<td>12</td>
<td>5</td>
<td>PGA9Q7</td>
<td>99.2</td>
<td>241.86</td>
<td>55</td>
<td>33</td>
<td>33</td>
<td>Y</td>
<td>96127</td>
<td>Alcohol-alcohol dehydrogenase OS=Escherichia coli (strain K12) GN=adHE PE=1 SV=2</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>Q833F9</td>
<td>99.2</td>
<td>239.64</td>
<td>54</td>
<td>38</td>
<td>20</td>
<td>Y</td>
<td>96408</td>
<td>Translation initiation factor</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IF2_SH1FL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>F-2 OS=Shigellla flexneri GN=infIB PE=3 SV=1 V=1</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>Q3I1W47</td>
<td>99.2</td>
<td>239.64</td>
<td>54</td>
<td>38</td>
<td>20</td>
<td>Y</td>
<td>96408</td>
<td>Translation initiation factor</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IF2_SH1BS5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>F-2 OS=Shigellla boydii serotype 4 (strain SB227) GN=infIB PE=3 SV=1</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>Q3YX73</td>
<td>99.2</td>
<td>239.64</td>
<td>53</td>
<td>38</td>
<td>20</td>
<td>Y</td>
<td>97331</td>
<td>Translation initiation factor</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IF2_SH1SS5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>F-2 OS=Shigellla sonnei (strain Sa046) GN=infIB PE=3 SV=1</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
<td>B1IQV3</td>
<td>99.2</td>
<td>239.64</td>
<td>53</td>
<td>38</td>
<td>20</td>
<td>Y</td>
<td>97350</td>
<td>Translation initiation factor</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IF2_ECOLC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>F-2 OS=Escherichia coli (strain ATCC 8739 / DSM 1576 / Crooks) GN=infIB PE=3 SV=1</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>B7M076</td>
<td>99.2</td>
<td>239.64</td>
<td>53</td>
<td>38</td>
<td>20</td>
<td>Y</td>
<td>97350</td>
<td>Translation initiation factor</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IF2_ECOBA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>F-2 OS=Escherichia coli O8 (strain AI2) GN=infIB PE=3 SV=1</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>PGA705</td>
<td>99.2</td>
<td>239.64</td>
<td>53</td>
<td>38</td>
<td>20</td>
<td>Y</td>
<td>97350</td>
<td>Translation initiation factor</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IF2_ECOLI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>F-2 OS=Escherichia coli (strain K12) GN=infIB PE=1 SV=1</td>
</tr>
</tbody>
</table>

*total 4500 proteins*
1. Notes

2. Result Statistics

Figure 1. False discovery rate (FDR) curve. X axis is the number of peptide-spectrum matches (PSM) being kept. Y axis is the corresponding FDR.

Figure 2. PSM score distribution. (a) Distribution of PEAKS peptide score; (b) Scatterplot of PEAKS peptide score versus precursor mass error.

Figure 3. De novo result validation. Distribution of residue local confidence: (a) Residues in de novo sequences validated by confident database peptide assignment; (b) Residues in "de novo only" sequences.

Table 1. Statistics of data.

| # of MS scans | 8434 |
| # of MS/MS scans | 18390 |

Table 2. Result filtration parameters.

| Peptide - \(-10\lg P\) | ≥15 |
| Protein - \(-10\lg P\) | ≥20 |

Table 4. PTM profile.

<table>
<thead>
<tr>
<th>Name</th>
<th>ΔMass</th>
<th>Position</th>
<th>#PSM</th>
<th>-10lgP</th>
<th>Area</th>
<th>AScore</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbamidomethyl</td>
<td>57.02</td>
<td>C</td>
<td>1131</td>
<td>73.12</td>
<td>2.26E6</td>
<td>1000.0</td>
</tr>
<tr>
<td>Oxidation</td>
<td>15.99</td>
<td>M</td>
<td>948</td>
<td>67.43</td>
<td>9.4E3</td>
<td>1000.0</td>
</tr>
<tr>
<td>Deamidation</td>
<td>.98</td>
<td>NQ</td>
<td>327</td>
<td>66.52</td>
<td>9.46E7</td>
<td>28.70</td>
</tr>
</tbody>
</table>
3. Experiment Control

Figure 4. Precursor mass error of peptide-spectrum matches (PSM) in filtered result. (a) Distribution of precursor mass error in ppm; (b) Scatterplot of precursor m/z versus precursor mass error in ppm. %

4. Other Information

Table 6. Search parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search Engine Name:</td>
<td>PEAKS</td>
</tr>
<tr>
<td>Parent Mass Error Tolerance:</td>
<td>15.0 ppm</td>
</tr>
<tr>
<td>Fragment Mass Error Tolerance:</td>
<td>0.6 Da</td>
</tr>
<tr>
<td>Precursor Mass Search Type:</td>
<td>monoisotopic</td>
</tr>
<tr>
<td>Enzymes: Specified by each sample</td>
<td></td>
</tr>
<tr>
<td>Max Missed Cleavages:</td>
<td>3</td>
</tr>
<tr>
<td>Non-specific Cleavage:</td>
<td>one</td>
</tr>
<tr>
<td>Variable Modifications:</td>
<td>Carbamidomethylation: 37.02</td>
</tr>
<tr>
<td></td>
<td>Oxidation (M): 15.99</td>
</tr>
<tr>
<td></td>
<td>Deamidation (NQ): 0.98</td>
</tr>
<tr>
<td></td>
<td>DCP-Bio: 394.16</td>
</tr>
<tr>
<td></td>
<td>DCP-Bio: 395.16</td>
</tr>
<tr>
<td></td>
<td>DCP-Bio: 396.17</td>
</tr>
<tr>
<td>Max Variable PTM Per Peptide:</td>
<td>5</td>
</tr>
<tr>
<td>Database:</td>
<td>Uniprot/Swissprot</td>
</tr>
<tr>
<td>Taxon:</td>
<td>All</td>
</tr>
<tr>
<td>Searched Entry:</td>
<td>551690</td>
</tr>
<tr>
<td>FDR Estimation:</td>
<td>Enabled</td>
</tr>
<tr>
<td>Merge Options:</td>
<td>no merge</td>
</tr>
<tr>
<td>Precursor Options:</td>
<td>corrected</td>
</tr>
<tr>
<td>Charge Options:</td>
<td>no correction</td>
</tr>
</tbody>
</table>

Table 7. Instrument parameters.

<table>
<thead>
<tr>
<th>Fraction:</th>
<th>20190122_CM130_01.raw, 20190123_CM131_01.raw</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ion Source:</td>
<td>ESI(nano-spray)</td>
</tr>
<tr>
<td>Fragmentation Mode:</td>
<td>high energy CID (y and b ions)</td>
</tr>
<tr>
<td>MS Scan Mode:</td>
<td>FT-ICR/Orbitrap</td>
</tr>
<tr>
<td>MS/MS Scan Mode:</td>
<td>Linear Ion Trap</td>
</tr>
</tbody>
</table>

Table 3. Statistics of filtered result.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peptide-Spectrum Matches</td>
<td>10401</td>
</tr>
<tr>
<td>Peptide sequences</td>
<td>3153</td>
</tr>
<tr>
<td>Protein groups</td>
<td>513</td>
</tr>
<tr>
<td>Proteins</td>
<td>4500</td>
</tr>
<tr>
<td>Proteins (Unique Peptides)</td>
<td>2881</td>
</tr>
<tr>
<td>FDR (Peptide-Spectrum Matches)</td>
<td>0.5%</td>
</tr>
<tr>
<td>FDR (Peptide Sequences)</td>
<td>1.6%</td>
</tr>
<tr>
<td>FDR (Proteins)</td>
<td>0.2%</td>
</tr>
<tr>
<td>De Novo Only Spectra</td>
<td>1299</td>
</tr>
</tbody>
</table>
APPENDIX FOUR: Evaluation of the pyruvate-binding parameters of wild-type and pyruvate-accepting variants

Introduction

Previous work indicated that wild-type transketolase (Geobacillus stearothermophilus) can accept pyruvate as a donor substrate (166), but with low affinity and resulting catalytic turnover with glycolaldehyde (GA). More recently, E. coli transketolase was evolved to accept pyruvate with increased higher catalytic turnover (52). Two mutants were created; H192P/A282P/I365L/G506A/H100L/D469T (mutant P-F), which accepted pyruvate and 3-FBA; and H192P/A282P/I365L/G506A/H100L/D469E/R520Q (mutant P-P), which accepted pyruvate and propionaldehyde. While the kinetic parameters were determined for wild-type G stearothermophilus transketolase (166) and E. coli mutant P-F, they were not obtained for P-F because the product could not be quantified (52). Furthermore, no thermodynamic dissociation constant was determined for these mutants. The substrate-binding assay developed in Chapter Four was therefore utilised to investigate the HPA- and pyruvate-binding parameters of each mutant with respect to TK\text{high}, TK\text{low} and their $K_m$ and $k_{cat}$ values.

Results

HPA-Binding parameters of E. coli transketolase mutants P-F and P-P

HPA-binding was determined between 0-100 mM HPA, and, unlike the previous HPA-binding experiments with E. coli wild-type transketolase and S385Y/D469T/R520Q, only one binding event was observed. The binding curves were therefore fitted to a single Hill function. The HPA-binding affinity of mutants P-F and P-P decreased 5.1- and 9.9-fold relative to wild-type transketolase (Figure A7; Table A2), which suggested a significant shift in substrate specificity away from HPA. It is unclear if the difference between P-F and P-P was a result of mutations at D469E or R520Q. Unfortunately, no kinetic parameters were determined for mutants P-F and P-P with HPA as the donor substrate.
Figure A7: Experimental data of 0.05 mg/mL wild-type TK (black), mutant P-P (red) and mutant P-F (blue), 9 mM Mg\(^{2+}\) and 0.3 mM TPP binding to 0-80 mM HPA. Experimental data-points were fitted to either the Hill function (mutants P-P and P-F) and the modified Michaelis-Menten function (wild-type).

<table>
<thead>
<tr>
<th>Species</th>
<th>(K_m) (\pm) (k_{cat}) ((s^{-1}))</th>
<th>(K_{d(high)}) (\pm) (K_{d(low)}) mM</th>
<th>(n_{high}) (\pm) (n_{low})</th>
<th>(\pm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type</td>
<td>5.5 (\pm) 0.5 33 (\pm) 1</td>
<td>3.42 (\pm) 0.44 39.3 (\pm) 0.7</td>
<td>1.15 (\pm) 0.06 4.60 (\pm) 0.21</td>
<td></td>
</tr>
<tr>
<td>P-P</td>
<td>- - - -</td>
<td>33.8 (\pm) 4.3 - -</td>
<td>1.83 (\pm) 0.17 - -</td>
<td></td>
</tr>
<tr>
<td>P-F</td>
<td>- - - -</td>
<td>17.6 (\pm) 7.7 - -</td>
<td>0.94 (\pm) 0.19 - -</td>
<td></td>
</tr>
</tbody>
</table>

Table 17: Summary of the kinetic and binding parameters of wild-type, mutant P-P and mutant P-F using HPA as the donor substrate using 0.05 mg/mL TK, 9 mM Mg\(^{2+}\), 0.3 mM TPP and 0-80 mM HPA. \(^{1}\)Kinetic data from (137).

The \(K_{d(low)}\) of wild-type was approximately 10-fold higher than that of TK\(_{high}\). Therefore, assuming a similar ratio in mutants P-F and P-P, the \(K_{d(low)}\) would not be detected below 100 mM HPA. It is therefore likely that the observed binding event related to TK\(_{high}\), while a HPA-binding to TK\(_{low}\) occurred at concentrations of between approximately 150-350 mM HPA.
Pyruvate-Binding parameters of wild-type, mutant P-F and mutant P-P transketolase

Only a single binding event, to TK_{high}, was observed for each mutant up to 100 mM pyruvate. The dissociation constants of the two mutants were only slightly higher than that of *E. coli* wild-type transketolase (Figure A8; Table A3), which was significantly less variation than the difference observed for HPA. Furthermore, the $k_{cat}$ of mutant P-F (52) was 5.6-fold greater than that of *G. thermophilus* wild-type transketolase (166). The discrepancy between the change in $k_{cat}$ and $K_d$ between variants implied the mutations introduced had relatively little impact on pyruvate affinity, but instead re-orientated the pyruvate-TPP intermediate for a more-favourable attack by the respective acceptor substrates.

**Figure A8**: Experimental data of 0.05 mg/mL wild-type TK (black), mutant P-P (red) and mutant P-F (blue), 9 mM Mg$^{2+}$ and 0.3 mM TPP binding to 0-100 mM pyruvate. Experimental data-points were fitted to the Hill function.
<table>
<thead>
<tr>
<th>Species</th>
<th>$K_m \pm$</th>
<th>$k_{cat} \text{ (s}^{-1}) \pm$</th>
<th>$K_{d(high)} \text{ mM} \pm$</th>
<th>$n_{high} \pm$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type</td>
<td>40.2$^1$</td>
<td>3.5 0.039$^1$ 0.001</td>
<td>36.0$^2$ 0.7 1.92$^2$ 0.10</td>
<td></td>
</tr>
<tr>
<td>P-P</td>
<td>-</td>
<td>-    -</td>
<td>39.1 1.7 2.16 0.08</td>
<td></td>
</tr>
<tr>
<td>P-F</td>
<td>90.4$^3$</td>
<td>11   0.22$^3$ 0.01</td>
<td>49.1 3.1 1.81 0.09</td>
<td></td>
</tr>
</tbody>
</table>

Table 18: Summary of the kinetic and binding parameters of wild-type, mutant P-P and mutant P-F using HPA as the donor substrate using 0.05 mg/mL TK, 9 mM Mg$^{2+}$, 0.3 mM TPP and 0-100 mM pyruvate. $^1$Kinetic data from wild-type *G. stearothermophilus* transketolase (166). $^2$Pyruvate-binding data from wild-type *E. coli* transketolase. $^3$Kinetic data from (52).