Evolution of New Catalytic Mechanisms for Xenobiotic Hydrolysis

A thesis submitted for the degree of Doctor of Philosophy at the Australian National University

Australian National University

Elena Sugrue

October 2016

Declaration

Unless otherwise stated, this work has been carried out by the author in the laboratory of Associate Professor Colin Jackson during the period of February 2013 – October 2016. The work presented herein has not been submitted as part of any other degrees.

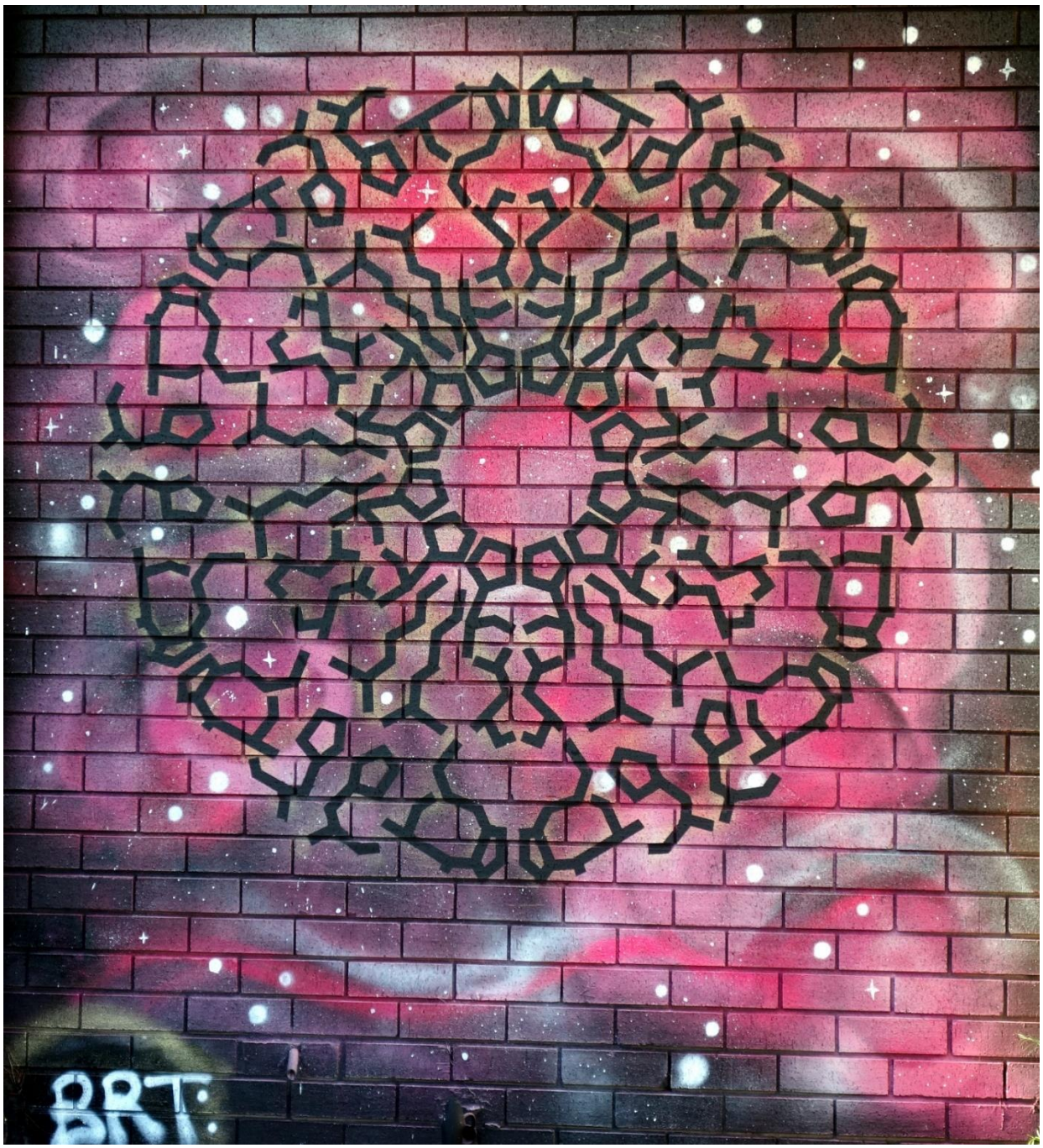
This thesis includes six original papers published in peer reviewed journals. The inclusion of co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research. In the case of chapters 1-6 my contribution to the work involved the following:

Thesis Chapter	Publication Status	Student contribution	Co-author name(s) and contributions
1	Published	80% I wrote the majority of the text and made all of the figures and tables	20% The primary investigator was Assoc. Prof. Colin J. Jackson. Dr. Carol Hartley and Dr. Colin Scott also made minor contributions
2	Published	30% My contributions were solving and analysing the two crystal structures presented	70% The engineering work shown was completed primarily by Dr. Tatheer Naqvi (50%) The contributions of the other authors (Dr. Andrew C. Warden, Dr. Nigel French, Dr. Paul D. Carr, Dr. Colin Jackson) amounted to ~20% of the effort. The primary investigator for this project was Dr. Colin Scott. Tatheer Naqvi, Andrew C. Warden, Dr. Colin Jackson and Dr Colin Scott wrote the manuscript.
3	Published	20% My contributions to this study involved testing and analysing the soluble expression levels of all variants	80% Dr. Christopher W. Coppin, Dr. Matthew Wilding were involved in the generation of mutants and screening with Dr

		under different conditions and time-points, as well as contributing to the structural analysis	Robyn Russell, Dr John Oakeshott and the co-primary author Dr Colin Scott. Dr. Paul D. Carr, Dr Janet Newman, Dr Martin Weik and Dr Tom Peat carried out the crystallization, collection and refinement of crystal structures with the co-primary author Assoc. Prof. Colin Jackson. Dr. Alexey Aleksandrov performed computational simulations under the supervision of Dr Martin Field. Mr Michael Paks and Ms Joanna Ubels were undergraduate project students that assisted with enzyme assays.
4	Published	70% I performed the majority of the experiments, I designed experiments, refined and analysed both crystal structures, performed extensive kinetic characterisations and all metal characterisations, performed all bioinformatics, made all the figures and contributed significantly to the preparation of the manuscript	30% Mr Nicholas J. Fraser performed preliminary work on MolA with Dr Colin Scott, resulting in the formation of the crystal that was later solved, Mr Davis H. Hopkins made and characterised mutants under my supervision in the 'Probing the structural determinants of MolA and PuhB substrate specificity' section of the manuscript, Dr. Paul D. Carr collected the crystal data and assisted with solving the structures, Dr. Jeevan L. Khurana prepared the crystal of PuhB that we later solved with Dr. John G. Oakeshott. The primary investigator for this project was Assoc. Prof. Colin Jackson
5	Published	70% I performed the majority of the experiments all; mutagenesis, enzyme	30% Dr. Paul D. Carr assisted with collection of data and solving of two of the crystal structures,

		purifications, preparation of crystals, kinetic and thermostability experiments, refining all of crystal structures, preparation of all the figures and contributing significantly to experiment design and manuscript preparation	Dr. Colin Scott assisted with the design of experiments and analysis, the primary investigator for this project was Assoc. Prof. Colin Jackson
6	Published	80% I performed all laboratory work, and contributed significantly to the preparation of the manuscript	20% The primary investigator for this work was Assoc. Prof. Colin J. Jackson. Dr. Colin Scott also made minor contributions

Elena Sugrue 11/10/2016



Co-Lab 2016

Britt Nichols/Elena Sugrue

Acknowledgements

I would first like to thank my supervisor, Dr. Colin Jackson, for the opportunity to work on this project and for his guidance and support. I would also like to acknowledge the RSC, ANU, and Australian government for their financial support.

I am extremely grateful for everyone in the Jackson lab, the great banter and terrible jokes made the long lab days not seem so long. Thank you to everyone in the group for the unique atmosphere and helpful discussions and ideas about my work. I particularly want to thank the '1st gen' members: Hafna, Will and Dongdi, for setting up the lab and the protocols we currently use, which was not a small feat. I am also very grateful for Dr. Paul Carr for his patience in all matters relating to protein crystallography.

Final thanks to my friends that have let me complain about very dull problems relating to metalloenzymes for over three years, and for their constant positivity. Finally, massive gratitude to my family for their overwhelming support, confidence and encouragement.

Abstract

Several different evolutionary processes allow enzymes to gain new catalytic activity, or enhance existing properties. Xenobiotic-degrading enzymes are often used to study the molecular basis for the evolution of new catalytic activity, as the substrates are often synthetic and comparatively new to the biosphere, the directionality of the evolutionary process is generally known and the evolved protein can often be traced back to a specific ancestral enzyme. However, the molecular basis behind the evolution of novel catalytic activities in different xenobiotic degrading enzymes is not well understood. In this work several different xenobiotic hydrolysing metalloenzymes were investigated to better understand new enzymatic reactions can evolve: the bacterial phosphotriesterases, triazine chlorohydrolases, and linuron/molinate hydrolases. These three enzyme families were studied using a combination of protein crystallography, enzyme kinetics, metal-ion dependence analysis, thermostability and solubility quantification, bioinformatics analyses, and critical analysis relevant literature. The conclusions that were drawn from these investigations cover fundamental concepts relating to enzymatic catalysis, including (1) the role of metal-ion cofactors in catalysis and the evolution of new function, (2) the desolvation of charged catalytic residues, the energetic cost of this desolvation and how this contributes to catalysis, and (3) the epistatic restrictions imposed on the evolution of new enzyme activity.

Various smaller milestones were accomplished as a result of these investigations. The molecular structures of hydrolases for the herbicides molinate and linuron (molinate hydrolase and phenylurea hydrolase B) have been solved for the first time and, in combination with extensive kinetic and metal analysis, a catalytic mechanism has been proposed. This information will be useful for optimization of both enzymes for the bioremediation of herbicide-contaminated sites. Detailed information outlining the structural and catalytic basis of triazine hydrolase activity and solubility was also obtained. Work towards understanding the structural basis underlying how triazine hydrolase can accumulate in a soluble, active, form and catalyse the hydrolysis of triazine based herbicides will aid the future protein engineering of more stable or active variants. The activity of phosphotriesterase was improved 5000-fold with the most widely used pesticide in the US, malathion. This phosphotriesterase variant has great potential for bioremediation applications and also provides a useful case study into the role that epistasis plays in the engineering of improved enzyme function. Finally, future research ideas are proposed, using results described in this work.

List of Abbreviations

a_3DH_3	Antiparallel three helix bundle
AHS	Aminodhydrolase Superfamily
AtzA	Atrazine Chlorohydrolase
DF1	Due-Ferri-1
DFT	Density functional theoretical
GpdQ	Glycerophosphodiesterase
GS	Ground State
Imp-1	β-lactamase
k_{cat}	Turnover number
k_{cat}/K_m	Turnover number/substrate concentration at half maximum velocity
K_m	Substrate concentration at half maximum velocity
MolA	Molinate hydrolase
MβL	Metallo-β-lactamase
NDM	New Delhi Metallo- β-lactamase-1
OpdA	Organophosphate-degrading enzyme
PLL	PTE-like lactonases
PON1	Paraoxonase 1
PTE	Phosphotriesterase
PuhA	Phenylurea hydrolase A
PuhB	Phenylurea hydrolase B
QM/MM	quantum mechanics/molecular mechanics
RS	Reactant state
SCCS	Supercoiled coil structure
T_m	Thermal stability
TriA	Melamine deaminase
TrzN	Triazine hydrolase
TS	Transition state
VIM	Verona integron-encoded metallo-β-lactamase

Table of contents

Chapter 1: Introduction and research objectives

Preface

The manuscript entitled 'The Evolution of New Catalytic Mechanisms for Xenobiotic Hydrolysis in Bacterial Metalloenzymes' gives a novel summary and analysis of different molecular mechanisms that have promoted the evolution of a diverse range of xenobiotic hydrolysing metalloenzymes. The fundamental biological (and inorganic chemistry) concepts that underlie how these enzymes are able to catalyse new chemical reactions with xenobiotics are also analysed. Finally, how the natural evolution of these metalloenzymes has been applied or mimicked in the field of artificial design is summarised. The overarching goal of my thesis has been to advance our understanding of how enzyme activity can be altered over evolutionary trajectories, in the context of xenobiotic hydrolysing metalloenzymes. This work summarises the current knowledge of the field, including some of the original work undertaken in my candidature.

Chapter 1

1.1 Research outline and objectives

1.1.1 Research objectives

The aim of this project was to use structural, biochemical and bioinformatic analyses of xenobiotic-hydrolysing metalloenzymes, to understand potential mechanisms of evolution that take place to facilitate new chemical reactions. Specifically, this work focuses on understanding what structural changes facilitate the associated catalytic mechanisms. It is hoped that this information may be useful in both applied and fundamental fields. The characterisation of xenobiotic hydrolysing enzymes is useful for the future application of these enzymes in contaminated sites and, fundamentally, this work provides an excellent case study into the mechanisms by which large functional evolutionary leaps can occur in response to anthropogenic environmental changes.

1.1.2 Research outline

This thesis consists of several separate investigations; each designed to study an aspect of enzyme evolution. Investigations are divided into the engineering (Chapters 2-3) and characterisation (Chapter 4-6) of recently discovered xenobiotic hydrolysing enzymes and detailed analysis of how they perform their chemical reactions, in the context of natural enzyme evolution.

Chapter 2. Changing the specificity of a phosphotriesterase through an epistatic reduction in steric hindrance

Chapter 3. Instability of apo-enzyme states influences the evolvability of metalloenzymes

Chapter 4. Loss of an ancestral metal ion coordination site enables xenobiotic hydrolysis in the molinate and diuron hydrolases

Chapter 5. Charged residue desolvation, thermostability tradeoffs and compensatory mutations in the evolution of a triazine hydrolase

Chapter 6. Epistasis and the evolution of bispecificity in triazine hydrolase

Chapter 7. Conclusions and future work

1.1.3 Experimental and computational methods

To investigate the evolution of different enzyme systems, a number of experimental and computational techniques were employed. X-ray crystallography was used to investigate the structures of molinate hydrolase, phenylurea hydrolase and triazine hydrolase variants. Metal ion occupancy and preference was tested using kinetic assays and inductively coupled plasma optical emission spectroscopy. The substrate specificity, pH-dependence of catalysis and computational ligand docking were used to investigate catalytic mechanisms. Bioinformatic analyses of the amidohydrolase superfamily were performed using sequence similarity networks, through the construction of phylogenetic trees and the ancestral reconstruction of selected nodes. Mutagenesis was used to determine the importance of residues to catalysis, substrate specificity and metal coordination. Finally, thermostability assays, circular dichroism and expression gels were used to investigate how modification of residues or cofactors changed the stability of the enzyme variants tested.

1.2 Publications

Several of the results of this thesis have been published in international journals. These publications are as follows:

Jackson, C. J., Coppin, C. W., Carr, P. D., Aleksandrov, A., Wilding, M., Sugrue, E., Ubels, J., Paks, M., Newman, J., and Peat, T. S. (2014) 300-Fold Increase in Production of the Zn2+-Dependent Dechlorinase TrzN in Soluble Form via Apoenzyme Stabilization, Applied and environmental microbiology 80, 4003-4011.

Naqvi, T., Warden, A. C., French, N., Sugrue, E., Carr, P. D., Jackson, C. J., and Scott, C. (2014) A 5000-fold increase in the specificity of a bacterial phosphotriesterase for malathion through combinatorial active site mutagenesis, PloS one 9, e94177.

Sugrue, E., Fraser, N. J., Hopkins, D. H., Carr, P. D., Khurana, J. L., Oakeshott, J. G., Scott, C., and Jackson, C. J. (2015) Evolutionary Expansion of the Amidohydrolase Superfamily in Bacteria in Response to the Synthetic Compounds Molinate and Diuron, Applied and environmental microbiology 81, 2612-2624.

Sugrue, E., Hartley, C., Scott C., Jackson,,C. J. (2016) The Evolution of New Catalytic Mechanisms for Xenobiotic Hydrolysis in Bacterial Metalloenzymes, Australian Journal of Chemistry.

Sugrue, E., Carr, P. D., Scott, C., Jackson, C. J. (2016) Active site desolvation and thermostability tradeoffs in the evolution of catalytically diverse triazine hydrolases, Biochemistry.

Sugrue, E., Scott, C., Jackson, C. J. (2017) Constrained evolution of a bispecific enzyme: lessons for biocatalyst design, Organic & biomolecular chemistry.

Chapter 2: Changing the specificity of a phosphotriesterase through an epistatic reduction in steric hindrance

Preface

The manuscript entitled ‘A 5000-Fold Increase in the Specificity of a Bacterial Phosphotriesterase for Malathion through Combinatorial Active Site Mutagenesis’ describes the use of saturation mutagenesis at the active site pocket of this enzyme, to improve activity with a comparatively poor substrate. In the work, the rationale behind the low turnover rate of this enzyme with malathion was investigated though computational docking and co-crystallisation of malathion with the most active variant. It was revealed that variants with the highest turnover rate with malathion had mutations at adjacent residues (Ser308Leu and Tyr309Ala) in the active site. Additionally, it was determined that if Tyr309 alone was mutated, the resultant proteins were insoluble. The contrasting impact of the Tyr309Ala mutation in different genetic contexts indicates that intramolecular epistasis is occurring. Epistasis is known to restrict evolutionary trajectories, which is often considered in natural evolution investigations (Chapter 6) but less frequently considered in protein engineering studies. The requirement for two adjacent mutations for both soluble protein and enhanced malathion activity highlights the importance of epistatic considerations in the engineering of enzymes.

Chapter 3. Instability of apo-enzyme states influence the evolvability of metalloenzymes

Preface

It is known that metal ions can contribute significantly to the stability and folding of a metalloenzyme. Even mutations that are remote from the coordination site can impact the stability, activity and respective metal ion affinities of metalloenzymes. As there are low levels of bioavailable metal ions present in the cytosol, factors that impact the ability of a metalloenzyme to bind to a metal ion can also restrict the maximum level of protein that can accumulate in the cell. In this work, three rounds of mutagenesis were performed to increase the soluble expression of triazine hydrolase. Structural analysis of the most soluble variant revealed that stabilisation of the unbound or apo-enzyme state facilitated the 300-fold increase in soluble expression. The limitations imposed on soluble protein expression by apo-enzyme state instability provides insight into the role of metal ions in protein folding and stability, and reveals an important target to consider in the engineering of metalloenzymes.

Chapter 3

Chapter 4: Loss of an ancestral metal ion coordination site enables xenobiotic hydrolysis in the molinate and diuron hydrolases

Preface

The manuscript entitled 'Evolutionary Expansion of the Amidohydrolase Superfamily in Bacteria in Response to the Synthetic Compounds Molinate and Diuron' describes the characterisation of two unique xenobiotic-hydrolysing metalloenzymes. By analysing the structures solved in this work, the kinetic properties of the enzymes, and their wider phylogenetic relationships, we have identified the novel and drastic loss of a metal binding site to enable new enzymatic function. It has previously been hypothesised that metal coordination transitions may have occurred in the diversification of metalloenzyme superfamilies, but this is the first example of such an occurrence. Examples like the molinate and diuron hydrolases provide insight into the breadth of modifications possible in natural evolution.

Supporting Information

Evolutionary Expansion of the Amidohydrolase Superfamily in Bacteria in Response to the Synthetic Compounds Molinate and Diuron

Elena Sugrue[†], Nicholas J Fraser[†], Davis H Hopkins[†], Paul D Carr[†], Jeevan L Khurana[‡], John G Oakeshott[‡], Colin Scott[‡], Colin J Jackson[#, †, ‡]

[†]Research School of Chemistry, Australian National University, Canberra, ACT, Australia; [‡]CSIRO Entomology, Canberra, ACT, Australia

#Corresponding author. Email: colin.jackson@anu.edu.au

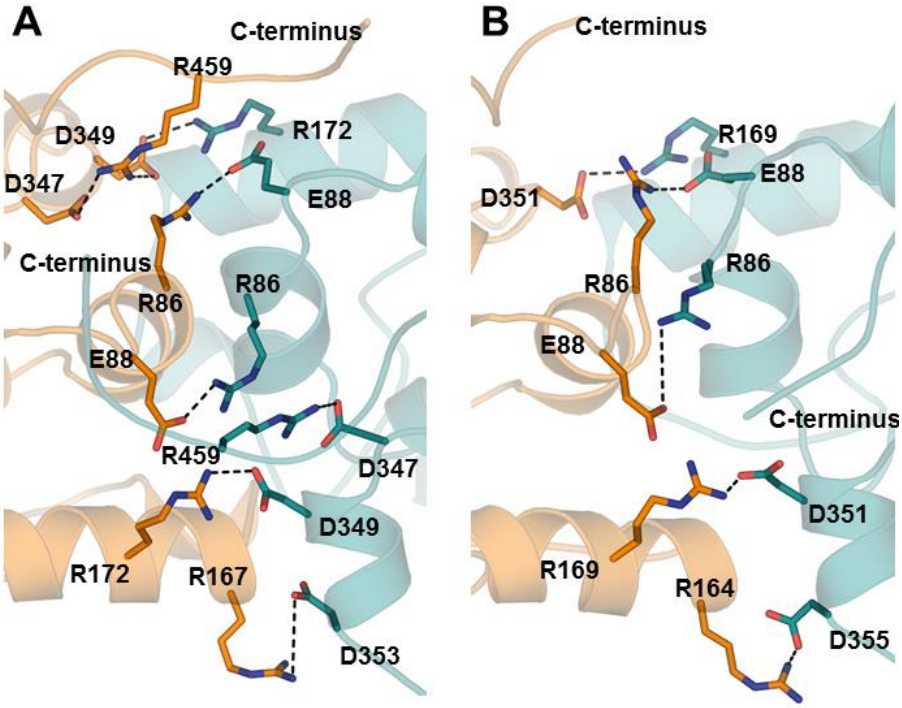


FIG. S1 Both MolA (**A**) and PuhB (**B**) have extensive salt bridge formation across the dimer-dimer interface (orange, blue). The extended length of the MolA C-terminus allows for additional salt-bridge formation *via* R459 (**A**).

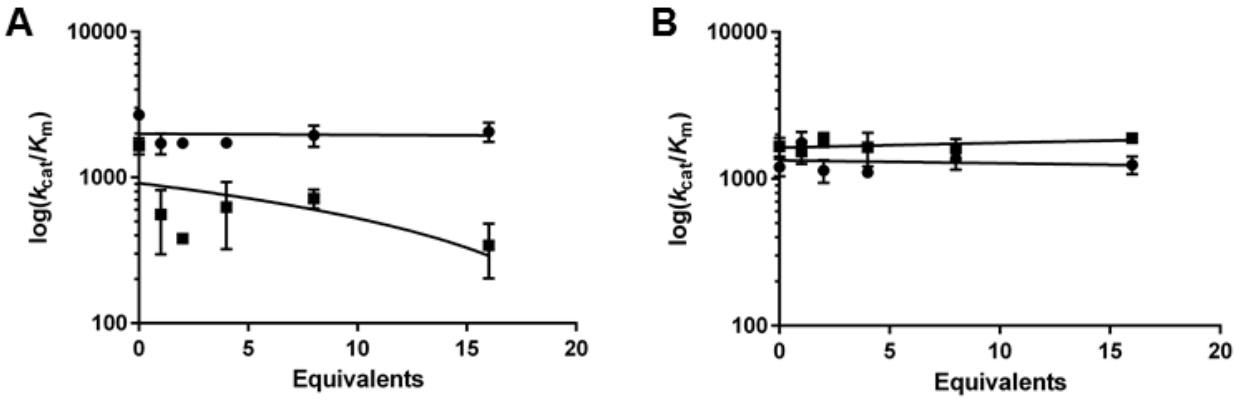


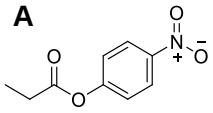
FIG. S2 (**A**) An exogenous supply of cobalt (●) and zinc (■) does not increase the activity of MolA. (**B**) Exogenous bicarbonate addition has no effect on Co-MolA (●) or Zn-MolA (■), in the presence of two equivalents of the respective metals. Error bars represent the standard deviation of two independent replicates. Equivalents refer to the relative concentration of exogenous metal or bicarbonate to protein concentration.

TABLE S1 Effect of the metal cation on molinate hydrolase (MolA) thermostability

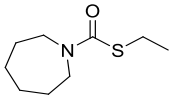
Enzyme	**Melting point (ºC)**
Zinc-MolA	46.0 ± 0.34
Apo-MolA	50.8 ± 0.31

FIGURE S3 (**A**) Substrates from which kinetic parameters could be calculated with molinate hydrolase. (**B**) Substrates with which there was detectable hydrolysis. (**C**) Substrates with which there was no detectable hydrolysis.

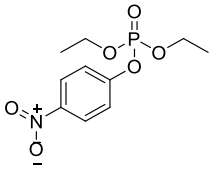
A



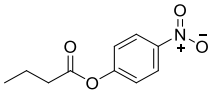
p-nitrophenyl propionate



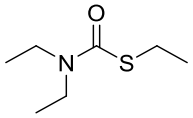
Molinate



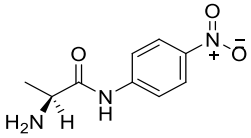
Paraoxon ethyl



p-nitrophenyl butyrate

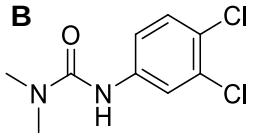


Ethiolate

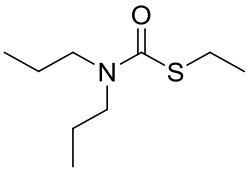


L-alanine *p*-nitroanilide

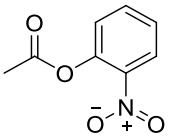
B



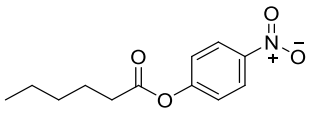
Diuron



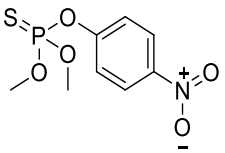
EPTC



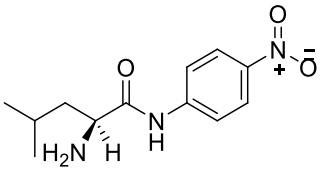
o-nitrophenyl acetate



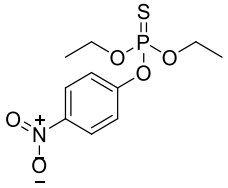
p-nitrophenyl hexanoate



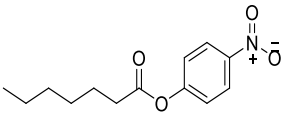
Parathion methyl



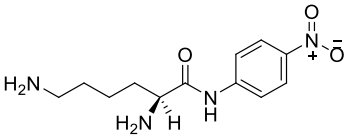
L-leucine *p*-nitroanilide



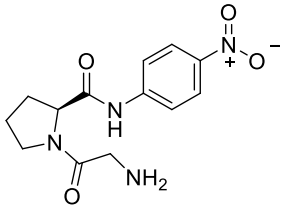
Parathion ethyl



p-nitrophenyl octanoate



L-lysine *p*-nitroanilide



Gly-Pro *p*-nitroanilide

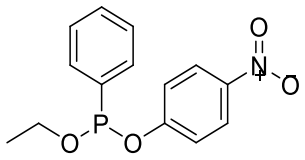
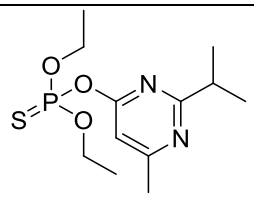
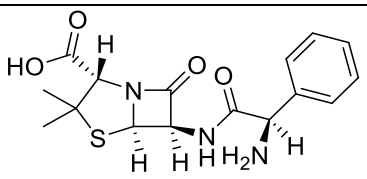
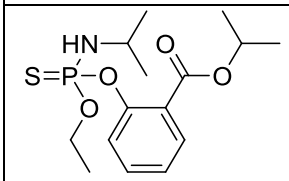
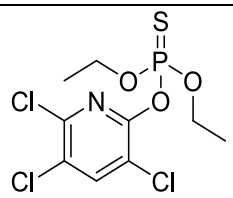
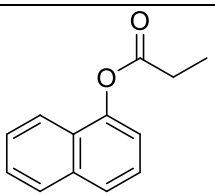
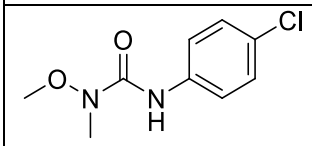
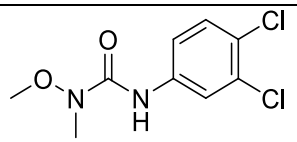
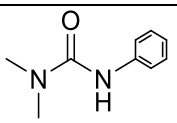
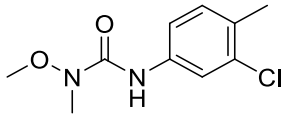
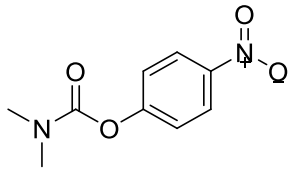
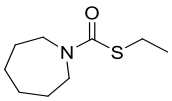
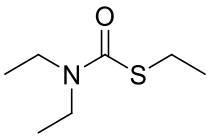
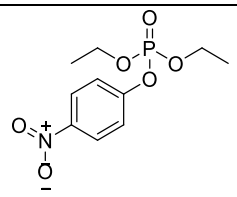
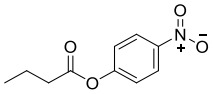
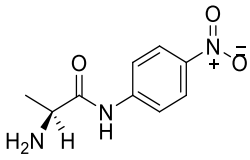
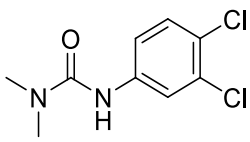
C  EPN	 Diazninon	 Ampicillin
 Isofenphos	 Chlorpyrifos	 1-Naphthylpropionate
 Monolinuron	 Linuron	 Fenuron
 Chlortoluron	 DMNPC	

TABLE S2 The kinetic parameters for wild-type MolA and variants with molinate, ethiolate, paraoxon ethyl, *p*-nitrophenyl butyrate, L-alanine *p*-nitroanilide and diuron. The k_{cat}/K_m of diuron was estimated using a single fixed substrate concentration well below the K_m. Values shown are the mean and 95% confidence interval of the calculated kinetics.

Substrate	**Variant**	K_m **(μM)**	k_{cat} **(10^{-3} s^{-1})**	k_{cat}/K_m **($M^{-1}s^{-1}$)**
Molinate 	Wild-type	7±2	8±1	1199
	D327E	27±1	12±0	449
	S319G	15±5	56±3	3763
	N69H	0	0	0
	H215N	0	0	0
	F318V	161±15	3±0	20
	I73M	17±5	12±1	1530
	V217C	59±16	12±0	197
	F318M	62±5	12±0	189
Ethiolate 	Wild-type	1976±322	67±2	34
	D327E	1965±723	29±3	15
	S319G	6267±1079	61±6	10
	F318V	2914±2206	10±2	4
	I73M	6100±2525	17±3	3
	V217C	6234±287	32 ±1	5
	F318M	7621±3023	24±4	3
Paraoxon ethyl	Wild-type	1299±132	21±2	16
	D327E	1756±139	47±6	27
	S319G	1592±74	60±3	38
	F318V	1965±605	20±5	10
	I73M	1931±834	19±5	10

	V217C	539±89	10±1	19
	F318M	581±433	4±0.2	6
p-Nitrophenyl butyrate 	Wild-type	407±18	1370±55	3355
	D327E	400±50	1670±102	4178
	S319G	465±29	8970±242	19299
	F318V	485±16	230±5	470
	I73M	475±28	650±18	1366
	V217C	345±48	828±56	2400
	F318M	433±12	1150±29	2658
L-Alanine *p*-nitroanilide 	Wild-type	1586±605	2±0	1
	D327E	1583±332	2±0	1
	S319G	1549±258	3±0	2
	F318V	0	0	0
	I73M	817±47	2±0	2
	V217C	696±18	2±0	2
	F318M	0	0	0
Diuron 	Wild-type	Could not be determined due to low solubility.		0
	D327E			2
	S319G			2
	F318V			1
	I73M			0
	V217C			0
	F318M			0

Chapter 5: Active site desolvation and thermostability tradeoffs in the evolution of catalytically diverse triazine hydrolases

Preface

The manuscript entitled ‘Active site desolvation and thermostability tradeoffs in the evolution of catalytically diverse triazine hydrolases’ presents a study of the evolutionary tradeoff between active site desolvation and thermostability of enzymes using triazine hydrolases as a model system. Triazine hydrolases typically utilise a glutamate residue as a proton donor, an atypical role for this residue as glutamate usually has a low p*K*a. The discovery of a triazine hydrolase lacking this catalytic glutamate residue, with a resultant inability to hydrolase one of the substrates, provided a unique opportunity to study a potentially inefficient ancestral triazine hydrolase. Tracing the potential evolutionary steps between the inefficient and efficient triazine hydrolase isoforms, through mutagenesis and structural, computational, and kinetic analysis, revealed the role of many fundamental catalysis concepts in the natural evolution of these enzymes. The chemical concepts governing many of these processes have been tested in model enzymes and utilised in computational enzyme design, but this is the first time they have been studied using naturally evolved variants. The solubly expressed variant of TrzN, engineered in Chapter 3, facilitated the study of these processes (Appendix 5.1).

Chapter 6: Epistasis and the evolution of bispecificity in triazine hydrolase

Preface

The following manuscript describes an investigation into the potential evolutionary trajectories of TrzN under different selection pressures, and what role intermolecular epistasis has in restricting these paths. Epistasis is known to constrain evolutionary paths and has been studied before in different systems, but there are a limited number of studies using naturally evolved enzymes. The presence of naturally evolved bispecific and monospecific TrzN isoforms with extremely similar sequences provides a unique test system in which to study possible evolutionary trajectories towards bispecificity. To gain understanding into the different evolutionary trajectories possible, sixteen variants that encapsulate the diversity between three TrzN isoforms were generated and analysed. The results provide insight into how epistasis can severely restrict the natural evolution of new enzymatic activity, and the trade-offs that can occur under different selection pressures, an important consideration in evolution and protein engineering.

Chapter 6

Supporting Information

Constrained evolution of a bispecific enzyme: lessons for biocatalyst design

E. Sugrue,[a] Colin Scott,[b] C. J. Jackson[a*]

[a]Research School of Chemistry, Australian National University, Canberra, Australia. E-mail: colin.jackson@anu.edu.au

Table 1 Impact of individual mutations on the thermostability and kinetic parameters of TrzN with atrazine and ametryn when compared to the *Nocardioides* sp. strain AN3 background.

Mutation	Background			Fold-change[a]		
				Atrazine hydrolysis	Ametryn hydrolysis[b]	Absolute change in T_m (°C)
Tyr67	-	-	-	-2	N/A	4.7
	Pro214	-	-		†	
	-	Tyr215		-1	N/A	4.4
	-	-	Glu241		†	
	Pro214	Tyr215	-	-1	N/A	2.1
	Pro214	-	Glu241		†	
	-	Tyr215	Glu241	-3	0	5.6
	Pro214	Tyr215	Glu241	-5	0	6.5
Pro214	-	-	-	-2	N/A	-1.6
	Tyr67	-	-		†	
	-	Tyr215	-	0	N/A	1.7
	-	-	Glu241		†	
	Tyr67	Tyr215	-	0	N/A	0.2
	Tyr67	-	Glu241		†	
	-	Tyr215	Glu241	1	2	-1.2
	Tyr67	Tyr215	Glu241	0	2	-0.3
Tyr215	-	-	-	-2	N/A	-2.3
	Tyr67	-	-	0	N/A	-2.6
	-	Pro214	-	0	N/A	2.6
	-	-	Glu241	-6	2	-1.5
	Tyr67	Pro214	-		†	
	Tyr67	-	Glu241		†	
	-	Pro214	Glu241		†	
	Tyr67	Pro214	Glu241		†	
Glu241	-	-	-	7	38000	-7.1
	Tyr67	-	-		†	
	-	Pro214	-		†	
	-	-	Tyr215	3	57000	-6.3
	Tyr67	Pro214	-		†	
	Tyr67	-	Tyr215	1	53000	-5.1
	-	Pro214	Tyr215	5	89000	-10.0
	Tyr67	Pro214	Tyr215	1	91000	-5.6

†One of the proteins does not express

[a]Fold-change when compared to the *Nocardioides* sp. strain AN3 background. Grey indicates negative change, yellow neutral, and green positive.

[b]Fold-change in activity when compared to the assay detection limit

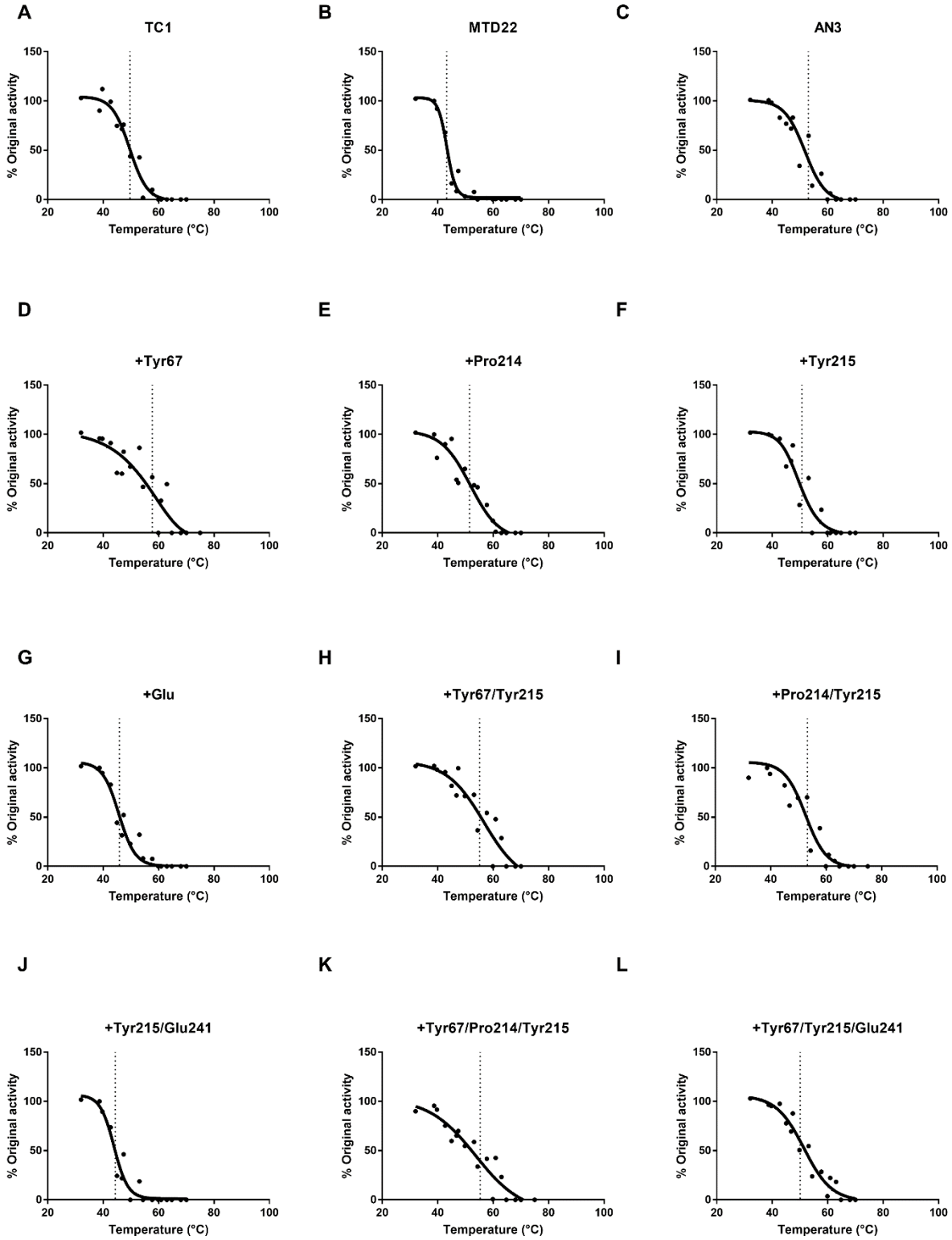


Figure 1 Representative replicates of thermostability assay curves used to calculate the T_m of variants. *Arthrobacter aurescens* strain TC1 TrzN, *Nocardioides* sp. strains MTD22 TrzN, and AN3 TrzN are shown from A-C, the remainder of soluble variants from D-L. The T_m (°C) calculated from three independent replicates is indicated with a dotted line.

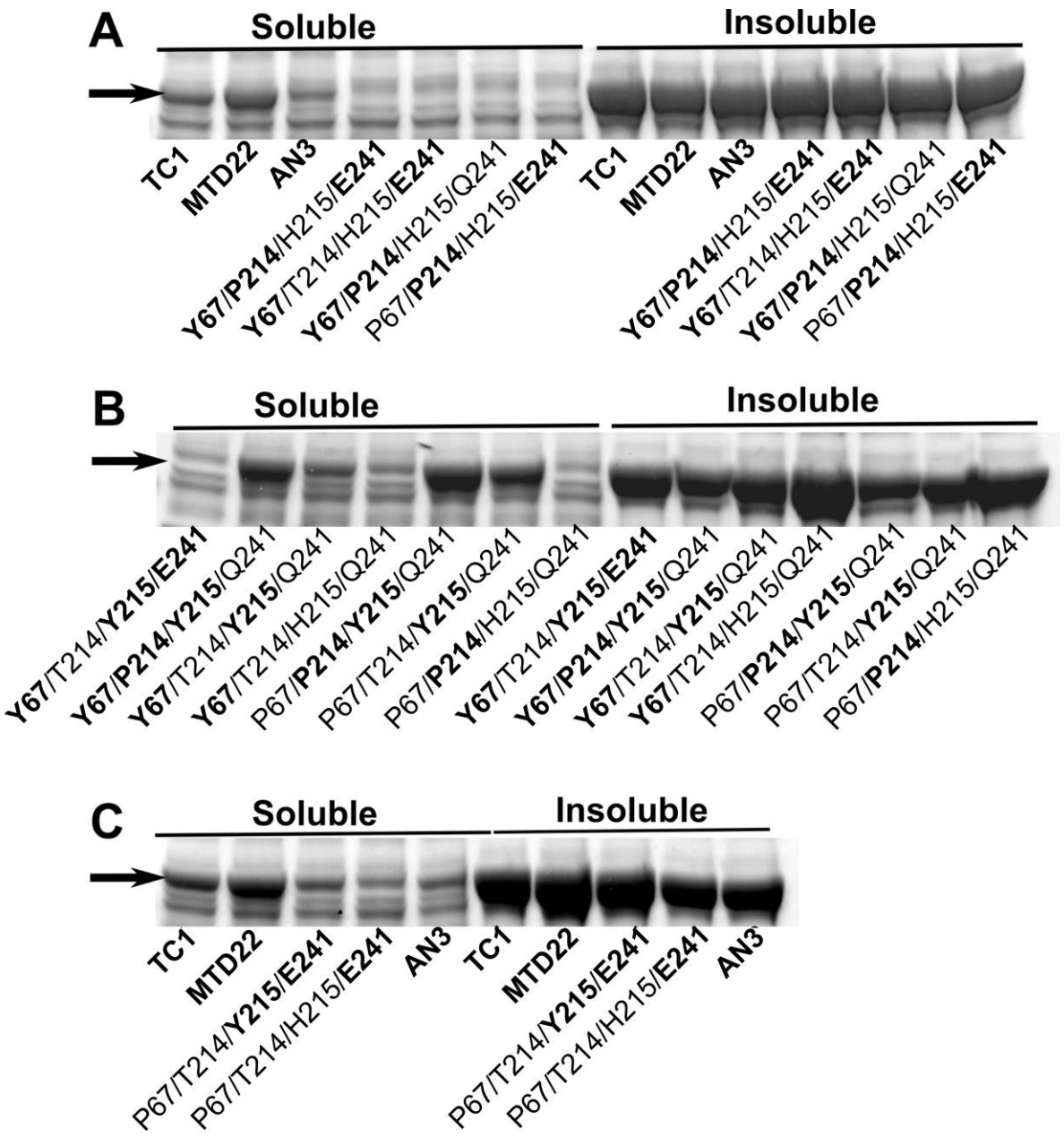


Figure 2 (A) Four variants were unable to be solubly expressed compared to *Arthrobacter aurescens* strain TC1 TrzN, *Nocardioides* sp. strains MTD22 TrzN, and AN3 TrzN, which exhibit soluble expression. **Tyr67**/**Pro214**/His215/**Glu241 TrzN, Tyr67**/Thr214/His215/**Glu241, Tyr67**/**Pro214**/His215/Gln241, and Phe67/**Pro214**/His215/**Glu241** show no soluble expression but substantial insoluble expression. (**B**) Soluble expression is much higher in the presence of Gln241. (**C**) Addition of Glu241 is observed to decrease expression.

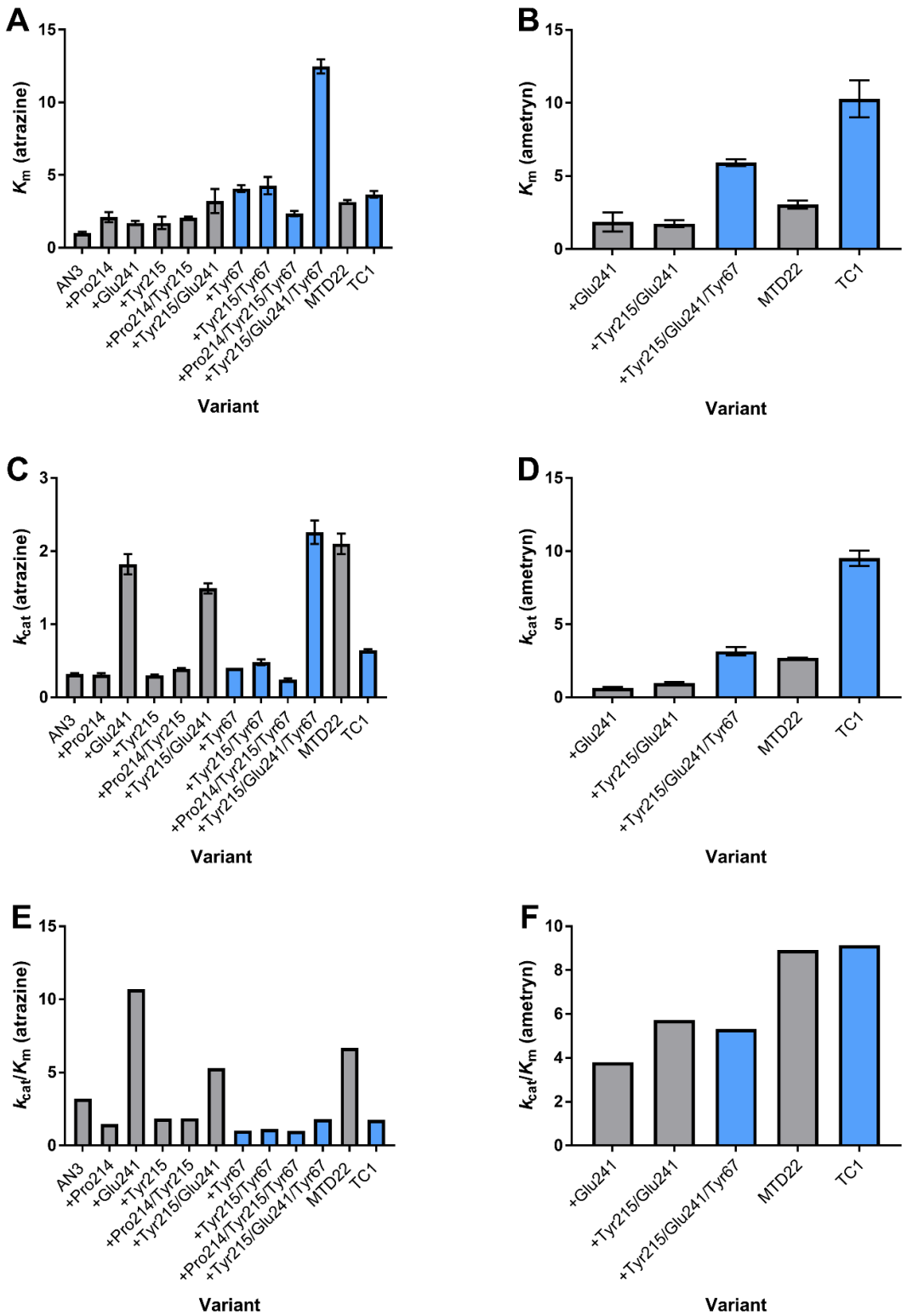


Figure 3 Modification of Phe67 to Tyr67 is observed to change the kinetic parameters of all TrzN variants. Modification of Phe67 to Tyr67 (blue bars) increases the K_m with both atrazine (**A**) and ametryn (**B**) independent of pre-existing mutations. Addition of Tyr67 is also observed to increase the k_{cat} with atrazine (**C**) and ametryn (**D**) in some genetic backgrounds. Overall, the k_{cat}/K_m with atrazine is observed to be reduced in Tyr67 containing variants (**E**) whereas the k_{cat}/K_m with ametryn is observed to be generally unchanged (**F**)

Chapter 7: Conclusions and future work

7.1 Conclusions

The central aim of this work was to gain greater understanding of potential molecular mechanisms of evolution that take place to facilitate new enzymatic reactions. Thorough analysis of PTE, MolA, PuhB and TrzN have provided additional demonstrations of established molecular mechanisms and novel evidence of unique pathways to xenobiotic hydrolysis.

Loop rearrangements, metal ion promiscuity and active site changes have been characterised in different systems before, as covered in the introduction, but findings from this work have provided additional insight into these processes. The structures solved of MolA and PuhB, when compared to the most similar ancestral enzyme, have enabled analysis into the loop differences between them. Identifying again that loop regions are very important for determining the substrate specificity of a metalloenzyme. The extensive promiscuity of metalloenzymes for different metal ions, and the role this has on substrate specificity was also identified in the MolA system. Despite its very constrained active site, MolA still displays promiscuity with diverse metal ions, which provides greater understanding of the innate metal ion promiscuity often observed in metalloenzymes. The known restrictive impact of epistasis on evolutionary trajectories was also identified in the TrzN system. The extremely constrained paths determined in this work contribute further understanding of epistasis. As the potential evolutionary trajectories were still heavily constrained by intramolecular epistasis, despite the already limited natural genetic variation. The structures of engineered PTE variants also revealed the importance of epistasis in the semi-rational design of enzymes for altered substrate specificities, a topic less frequently considered in applied studies.

Previously uncharacterised evolutionary strategies have also been identified in this work. It is known that metal coordination is tuned over evolutionary trajectories, but the discovery of the loss of a second metal binding site to facilitate the hydrolysis of molinate and diuron is the first example of such a drastic coordination change over evolution. The characterisation of a trade-off between the catalytically beneficial burial of an ionisable residue, and the stability cost of this burial in the TrzN system is also unique. This topic has mostly been studied in an abstract fashion, to better understand the fundamental chemistry underlying side chain ionisation and stability, but now has been shown to play

a key role in natural evolutionary processes. Both novel evolutionary strategies are important for both our fundamental understanding of enzyme evolution, and for the applied engineering of enzymes to perform desired reactions.

7.2 Future work

7.2.1 Computational investigations into TrzN catalysis

The results outlined in Chapter 5 and 6 describe the origins of ametryn hydrolysis in TrzN. Several different factors are shown to contribute to the moderately efficient catalysis of this enzyme, and it is speculated as to whether TrzN represents the first step of enzyme evolution, with GS destabilisation being the primary means to reduce the activation energy barrier, with TS optimisation likely to occur subsequently. As described in Chapter 5, clear examples of GS destabilisation in enzyme evolution are rare.[1-3] Accordingly, the role of GS destabilisation in enzyme evolution is often debated. Therefore, computational analysis of the TrzN system could provide great insight into this contested area of research. To investigate the impact of mutations on the catalytic reaction more thoroughly, a hybrid approach using quantum mechanics/molecular mechanics (QM/MM) molecular dynamics simulations and density functional theoretical (DFT) calculations could be used to gain detailed information into the reaction pathway.[4] Additionally, investigations into the activation free energies between mutants and the solvation energies between the reactant state (RS) and TS between variants could provide insight into whether they changed along the evolutionary trajectory.[1]

The experimental TrzN data presented here also provides an ideal opportunity to benchmark current methods for p*K*a determination. The experimental determination of p*K*a values of variants includes differences below 1 pH unit between variants, this is at the accuracy limit of many of the current methods for determining the p*K*a of residues within proteins. The TrzN system would provide great test data to determine the best estimating methods.[5-7] In addition, more thorough analysis of the p*K*a of Glu241 using more intensive QM/MM calculations, as well as assessing the nature of water coordination in the active site during the reaction in a more robust manner is important work to carry out to confirm the results of Chapter 6.[8]

7.2.2 Recreating natural evolution in the lab

We have proposed that different evolutionary pressures resulted in the novel loss of a metal ion for MolA and PuhB, and the burial of a catalytic glutamate residue in the case of TrzN. Replicating these evolutionary processes in the lab would provide further information. Directed evolution could be performed using a carboxypetidase starting point for MolA and PuhB respectively, and selecting for molinate and diuron hydrolysis over consecutive rounds.[9] The number of successive rounds required for catalysis, the order of mutations, whether any intermediate states are able to catalyse both ancestral and contemporary reactions, and differences in the catalytic trajectories between MolA and PuhB, would all provide information unavailable from examination of the already existing proteins. It would also be interesting to investigate the TrzN system via directed evolution, starting with the Gln241/His215/Thr214 background and selecting for enhanced ametryn hydrolysis over successive rounds. Whether the endpoint would be the same for TrzN under these conditions, or if a different solution was possible or further optimisation was possible from the Glu241/Tyr215/Pro214 TrzN scaffold would all be interesting investigations into the fitness landscape of TrzN.

7.2.3 Reversing the evolutionary process

Evolutionary trajectories are frequently investigated using directed evolution or ancestral protein reconstruction, but very few studies examine both the forward and reverse evolutionary trajectories of the system studied.[10] Analysing both the forward and reverse evolutionary trajectories provides insight into the role of epistasis, as it can differentially restrict both paths within the same system. MolA and PuhB have constrained active sites and substrate diffusion tunnels, which have been tuned for their respective substrates and reactions. Due to both the instability and constrained active sites of both MolA and PuhB, the reverse evolutionary trajectory towards a carboxypeptidase ancestor may require prior stabilising mutations to proceed. An initial directed evolution experiment, screening for increased thermostability/expression in MolA and PuhB may be required to allow destabilising functional mutations to be incorporated. A green fluorescent protein reporter system, or testing for activity under increased temperature could both be used for screening purposes.[11] Stabilising MolA and PuhB prior to performing the reverse trajectory would still provide information as to the nature of potential catalytic intermediate states, even if the stabilised starting point is slightly

artificial. The instability and constrained active sites of MolA and PuhB could provide a unique case study into the reversibility of enzyme evolution to an ancestral state."

In addition, the evolutionary trajectory of TrzN towards ametryn hydrolysis could be investigated, using the efficient Glu241/His215/Thr214 TrzN as a starting point. Selection for atrazine hydrolysis over successive rounds could potentially cause a reversion of Glu241 to Gln241 and loss of ametryn hydrolysis. Destabilising conditions may be required to force the reversion, which would also be useful information.

7.2.4 Engineering new enzyme function

The enzyme scaffolds characterised in this work can be further manipulated to test their true catalytic potential, using different protein engineering tools. For example, the order and structure of the polypeptide chain can be changed using circular permutation, which can produce a variety of novel outcomes inaccessible through point mutations.[12,13] The creation of non-native termini and subsequent reorganisation of the polypeptide chain can alter the quaternary and tertiary structure of the engineered proteins. At the oligomeric state level, MolA and PuhB have a unique cruciform assembly that could be changed through circular permutation, which would impact the stability, dynamics and possible entry sites to the active site.[14] In addition, circular permutation can potentially affect the position and dynamics of important loops involved in substrate specificity, this can create or improve promiscuous activity**.** Circular permutation experiments performed on TrzN would potentially impact solvent access to the active site, which could improve catalytic efficiency. Rearrangement of loops or alteration of the position and size of the tunnel would also change the microenvironment surrounding Glu241, which could also impact catalysis. Substrate specificity could also be modified through these changes, in a manner comparable to MolA and PuhB. Any changes to the structures and resultant activities of MolA, PuhB, and TrzN could subsequently be manipulated and optimised using techniques such as directed evolution or rational design.

As described in Chapter 1, atypical metal ions can also be used to promote novel enzymatic reactions or change substrate specificities. MolA has been shown to be promiscuous with diverse metal ions, but the alternatively metalated forms have not been screened extensively with different substrates. It is possible that a metal isoform of MolA, such as Cu-MolA, has unique activity or specificity. Testing of alternatively metalated isoforms can also be used in tandem with other techniques such as directed evolution,

ancestral protein reconstruction, or circular permutation. Screening the activity of the alternatively metalated variants is a viable option, as any modifications to the protein structure can drastically change metal ion coordination. Combining isoform screening with protein modification is also a viable option for enzymes that are not as promiscuous with metal ions as MolA, such as TrzN.

References

1. M. P. Frushicheva, J. Cao, Z. T. Chu and A. Warshel, *Proc. Natl. Acad. Sci. U.S.A.*, 2010, **107**, 16869-16874.
2. M. P. Frushicheva, J. Cao and A. Warshel, *Biochemistry*, 2011, **50**, 3849-3858.
3. E. A. Ruben, J. P. Schwans, M. Sonnett, A. Natarajan, A. Gonzalez, Y. Tsai and D. Herschlag, *Biochemistry*, 2013, **52**, 1074-1081.
4. H. Park, E. N. Brothers and K. M. Merz, *J. Am. Chem. Soc.*, 2005, **127**, 4232-4241.
5. P. Barth, T. Alber and P. Harbury, *Proc. Natl. Acad. Sci. U.S.A.*, 2007, **104**, 4898-4903.
6. J. E. Nielsen, M. Gunner and E. García-Moreno, *Proteins: Struct., Funct., Bioinf.*, 2011, **79**, 3249-3259.
7. T. Meyer and E.-W. Knapp, *J. Chem. Theory Comput.*, 2015, **11**, 2827-2840.
8. N. Ghosh, X. Prat-Resina, M. Gunner and Q. Cui, *Biochemistry*, 2009, **48**, 2468-2485.
9. O. Kuchner and F. H. Arnold, *Trends Biotechnol.*, 1997, **15**, 523-530.
10. M. Kaltenbach, C. J. Jackson, E. C. Campbell, F. Hollfelder and N. Tokuriki, *eLife*, 2015, **4**, e06492.
11. C. Roodveldt, A. Aharoni and D. S. Tawfik, *Curr. Opin. Struct. Biol.*, 2005, **15**, 50-56.
12. Z. Qian and S. Lutz, *J. Am. Chem. Soc.*, 2005, **127**, 13466-13467.
13. Y. Yu and S. Lutz, *Trends Biotechnol.*, 2011, **29**, 18-25.
14. Z. Qian, J. R. Horton, X. Cheng and S. Lutz, *J. Mol. Biol.*, 2009, **393**, 191-201.

Appendix

5.1: Use of TrzNG3

Chapter 3 outlines the optimisation of TrzN derived from *Arthrobacter aurescens* for enhanced soluble expression (*trzNG3).* The optimised genetic background of *trzNG3* was also used in Chapter 5, as it was required to obtain enough protein to perform crystallography and for the kinetic assays used. A sequence similarity of 99% is shared between *trzNG3* and both *Nocardiodies* MTD22 and AN3. In addition, the three mutations made to enhance soluble expression are not present in close proximity to glu241, the closest residue modified is 11 Å away. However, the modification of all three residues does alter the kinetic profile compared to wildtype *A. aurescens* TrzN. A raised K_m (53 µM compared to 20 µM in wildtype) combined with an increased k_{cat} (3.7 s^{-1} compared to 2.1 s^{-1}) slightly decreases the overall k_{cat}/K_m to 0.7 M^{-1} s^{-1} in *trzNG3* compared to 1.1 M^{-1} s^{-1} in wild-type. The changes to the kinetic profile may be relatively minor, but they do reduce the weight of any evolutionary implications. Accordingly, a different genetic background is used in Chapter 6 where evolution is the focus. The work of Chapter 5 focuses on the chemical nature of Glu241 and the direct microenvironment surrounding this residue, where the *trzNG3* genetic background should not impact the conclusions drawn.