MOLECULAR BASIS OF THE EVOLUTION OF NEW PROTEIN FUNCTIONS: KEMP ELIMINATION AND PHOSPHORYL TRANSFER REACTIONS

Nan-Sook Hong

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

Research School of Chemistry
Australian National University
Canberra, Australia

July 2016

DECLARATION BY AUTHOR

The contents of this thesis are the result of my original research under supervision of A/Prof. Colin J Jackson (Research School of Chemistry, ANU).

Prof. Michelle Coote and co-workers (Research School of Chemistry, ANU) provided the results of the transition state optimizations using the KE07 active site fragments (Figure 5.9, Chapter 5).

Prof. Chris Easton and co-workers synthesized the 3’D-5-nitrobenzisoxazole substrate (Figure 3.3, Chapter 3). Prof. Chris Easton and co-workers and William Zhang and I prepared 3’H-5-nitrobenzisoxazole substrate (Figure 3.2, Chapter 3).

I declare that the results presented in this thesis are my own original work, except as acknowledged in the text above, and that the material has not been submitted elsewhere.

Nan-Sook Hong

July 2016

Research School of Chemistry
Australian National University

ACKNOWLEDGEMENT

I would like to thank my supervisor, A/Prof. Colin Jackson, for giving me the valuable opportunity to work on this interesting project, for his guidance to the PhD completion, for his encouragement, and for his inspirational enthusiasm.

I am very grateful to A/Prof. Nobuhiko Tokuriki, for the interesting metallo-beta-lactamase super family project and his support and enthusiasm.

I am grateful to Prof. Michelle Coote and her group members, Dr. Ching-Yeh Lin and Dr. Ganna Gryn'ova, for training me on the use of Gaussian and Gaussrate, for giving me timely advice, and for providing the result of quantum cluster modelling.

I am grateful to Prof. Chris Easton and his group members, Dr. Timothy Altamore, Mr. Daniel Bartkus, and Ms. Charlotte Rouse, and William Zhang for providing and helping me to synthesize the substrates for the KE07 project.

I would like to thank Dr. Paul Carr, for training and helping me on the use of the X-ray diffractometer system, data processing, refinement, and PDB submission, for his timely advice on the whole process of X-ray crystallography.

I would like to thank Dr. Peter Mabbitt, for showing me tryptophan fluorescence measurement and analysis, for his extensive help on my writings about the KE07 project, and for timely advice and encouragement.

I am grateful to Mr. Benjamin Clifton and Mr. Galen Correy, for showing me on the use of AutoDock vina, Gromacs, and data analysis for the molecular dynamics simulations. I thank Mr. Galen Correy to generate the figure for the tryptophan angle analysis using Mathematica program.

I am grateful to Ms. Tracy Murray and Ms. Dongdi Li, for training and helping me on the cloning process.

I am grateful to Dr. Livnat Afriat-Jurnou, for showing me enzyme kinetics, data analysis, and micro-seeding technique.

I am grateful to the technical and administrative support provided by the staff at the Research School of Chemistry (RSC), ANU, throughout my PhD. I thank Ms. Alison Scott for her timely and kind help for admission, scholarship, and completing milestones.

I would like to acknowledge the funding from the Australian National University International student PhD Scholarship.

I am very grateful to Dr. Peter Mabbitt and Mr. Benjamin Clifton, for proofreading my thesis with insightful advice.

I thank the Jackson group, past and current members, for helping me make it to the end.

Last but not least, I am grateful to my husband, for his constant support, company, love, advice and encouragement throughout my PhD.

ABSTRACT

This is a series of investigations into the molecular basis of the evolution of new protein functions. The broad objective of this work was to determine exactly how a series of single amino acid mutations, typical of an evolutionary trajectory, can result in dramatic changes in catalytic activity, specificity and protein solubility. Various strategies were employed to achieve this aim, including analysis of existing literature concerning the various models and theories relating to molecular evolution, protein crystallography, extensive enzyme kinetics and thermodynamic analysis, theoretical analysis of catalytic mechanisms and computational simulation of protein dynamics. Three model systems were investigated: the *de novo* designed Kemp Eliminase (KE07), the metallo-β-lactamases NDM1 and VIM2, and the N-acyl-homoserine lactonase AiiA. Based on these studies, I was able to identify three clear phenomena that are important in molecular evolution: first, preorganization of the active sites residues is essential for efficient catalysis; second, remote mutations are capable of causing quite drastic rearrangements to the active site and substrate binding site by modulating the conformational landscape of a protein; third, intramolecular epistasis, the way that mutations interact with each other and the sequence background that they are introduced to, can constrain evolutionary trajectories and make the evolutionary potential of a protein contingent on its starting sequence.

In Chapters 3-5 I focus on KE07, performing detailed kinetic analysis of hydrogenated and deuterated substrate, which revealed entropy-enthalpy compensation in the improvement in activity as well as an unusual change in the kinetic properties in the middle of the evolutionary trajectory. This is followed by comprehensive structural analysis, which reveals the enzyme has evolved to adopt a completely unexpected active site configuration *via* remote mutations. Finally, using computational simulations and solution fluorescence spectroscopy, I confirm that the *in crystallo* and kinetic observations are consistent with the behaviour of the protein in solution. Chapter 6 consists of a manuscript that describes the effects of conformational tinkering on the N-acyl-homoserine lactonase AiiA, specifically how remote mutations can have dramatic effects on activity by modulating the conformation of the active site. My contribution to this work included crystal

structures and molecular dynamics simulations. Finally, Chapter 7 is a second manuscript that focuses on evolutionary contingency: by examining two related sub-families of the metallo-β-lactamases, NDM1 and VIM2 we show that the evolvability of each is constrained by intramolecular epistasis and contingent on the starting sequence. To achieve the same final goal (greater whole cell activity), NDM1 evolved higher activity, while VIM2 evolved greater solubility. The crystals structures that I solved revealed the structural basis for the enhanced activity in NDM1 and that enhanced solubility in VIM2 is a result of an unprecedented (for an enzyme) structural rearrangement where the two halves of the α/β sandwich metallo-β-lactamase protein fold have separated and rearranged in an domain-swapped dimer.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

CHAPTER 2: METHODS

CHAPTER 3: KINETICS ANALYSIS OF KE07 VARIANTS

CHAPTER 4: STRUCTURAL PRE-ORGANIZATION IN KE07 EVOLUTION

CHAPTER 5: SOLUTION AND COMPUTATIONAL ANALYSIS OF THE CONFORMATIONAL DYNAMICS OF KE07 AND VARIATNS

CHAPTER 6: CONFORMATIONAL TINKERING DRIVES EVOLUTION OF A PROMISCUOUS ACTIVITY THROUGH INDIRECT MUTATIONAL EFFECTS

CHAPTER 7: ENZYME EVOLABILITY IS CONTINGENT ON THE INITIAL SEQUENCE BACKGROUND

LIST OF TABLES

LIST OF FIGURES

MANUSCRIPT

* represents co-first authorship

Yang, G.*, **Hong, NS ***, Baier, F., Jackson, C.J., Tokuriki, N., 2016. Conformational tinkering drives evolution of a promiscuous activity through indirect mutational effects. *Submitted to Biochemistry.*

Baier, F., **Hong, NS**, Yang, G., Carr, P. D., Jackson, C. J., Tokuriki, N., Enzyme evolvability is contingent on the initial sequence background. *Submission ready.*

ABBREVIATIONS

AiiA	Acyl-homoserine lactonase
bp	Base pairs
DFT	Density functional theory
DNA	Deoxy nucleic acid
dNTP	2'-Deoxyribonucleotide triphosphate
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EtOH	Ethanol
FPLC	Fast protein liquid chromatography
GS	Ground state
HEPES	4-(2-hydroxyethyl)-piperazine-1-ethanesulfonic acid
IPTG	Isopropyl β-D-1-thiogalactopyranoside
KIE	Kinetic isotope effect
LB	Luria broth
LBA	Luria broth supplemented with ampicillin
LBK	Luria broth supplemented with kanamycine
MBL	Metallo-β-lactamase
MD	Molecular dynamics
MeOH	Methanol
MPH	Methyl parathion hydrolase
MPD	2-Methyl-2,4-pentanediol
NDM1	New Delhi metallo-β-lactamase-1
OP	Organophosphate
PBC	Periodic boundary conditions
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
PME	Particle mesh Ewald
PTE	Phosphotriesterase
QM	Qunatum dynamics
QQL	Quorum quenching lactonase
RMSD	Root mean square deviation
SAP	Shrimp alkaline phosphatase
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
ssDNA	Single-stranded DNA
TB	Terrific broth
TIM	Triosphosphate isomerase
TRIS	Tris (hydroxymethyl) aminomethane
TS	Transition state
VIM2	Verona integron-encoded metallo-β-lactamse-2

Chapter 1: Introduction

1.1 Preamble

Rational enzyme design has been studied intensely in the previous decade (reviewed in (1–3)), particularly to create catalysts for reactions involving synthetic compounds such as benzisoxazoles and organophosphates (4–7). However, the low precision of the sequence-based prediction of protein structure and dynamics makes creating enzymes with high k_{cat} difficult (6, 8, 9). In this chapter, achievements and limitations in the design of protein structure and dynamics are presented.

1.2 Enzyme design

Enzyme engineering has been used to generate enzymes with a wide range of beneficial applications in fuels biosynthesis, fructose biosynthesis, biodegradation of toxic synthetic compounds, protein vaccines, and so on (3, 5, 10–14). Routine production of enzymes for specific purposes based on rational enzyme design is ultimate goal of this field, since it will allow us to create useful enzymes that do not exist in nature (e.g. to decompose toxic nerve agents and disease related proteins (5, 15, 16), to repair disease related DNA (3)). Transition state theory (TST), which was first suggested by Eyring for chemical reactions (17) and by Pauling for enzymatic reactions (18), is widely used to understand enzyme mechanisms and for enzyme engineering. In particular, Pauling's concept of pre-organized enzyme active sites that stabilize transition states (19), developed further by Warshel and others (19–25), suggests that TS-stabilization reduces the activation energy for a reaction, and has been the foundation of many enzyme design experiments. The current method for creating enzymes, *de novo,* is first to find the transition state of a reaction and identify the optimal positioning of amino acid side chains to stabilize the TS, and then to combine the designed fragments with a stable scaffold (2).

1.3 Enzyme design and directed evolution

Before the emergence of computational design methods, monoclonal antibodies were created to transform synthetic compounds (26–28) and analysed (29–35), taking advantage of the adaptability of the antibody. These structural and theoretical studies of catalytic antibodies have provided molecular insight into various designed reaction mechanisms (6, 7, 9, 36–38). Subsequently, computationally designed active site fragments were grafted onto compatible and stable scaffolds, which were identified through the application of algorithms that have been developed by the groups of David Baker (39, 40), Stephen Mayo (41) and others (42). Although *de novo* enzymes for the retro-aldol, Diels-Alder, and Kemp elimination reactions have been created, their activities are significantly lower (k_{cat}/K_M ~10^{-1}-10^{3} $M^{-1}s^{-1}$) than efficient natural enzymes (e.g. triosephosphate isomerase (TIM) or beta lactamase: ~10^{8} $M^{-1}s^{-1}$) (43, 44). However, the weak activities of designed proteins can be significantly improved by directed evolution (6, 45–48).

Directed evolution is widely used to improve protein stability or activity using random mutagenesis, copying natural evolution. Three techniques for generating genetic diversity are used together for directed evolution: random mutagenesis *via* error-prone PCR, focused mutagenesis *via* site-saturation mutagenesis to target mutations to the active site, and gene recombination, for example, by recombination of entire sequences or exchange of binding loops between proteins (45, 46, 49, 50). Intense X-ray crystallography (8, 45–48, 51) and quantum mechanics/molecular mechanics (QM/MM) (1, 15, 52–57) studies to understand how the initially weak activities of designed enzymes could be improved by directed evolution have provided molecular insight into the evolutionary routes taken to improve activity. Some specific examples (retro-aldolases, Diels-Alderases, Kemp eliminases; Table 1.1) are discussed in the following sections.

Table 1. 1 | Computationally designed enzymes and antibodies for transformation of synthetic compounds, and evolved variants thereof.

Reaction	Catalyst	Fold (PDB ID)	k_{cat} (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)	Ref
Kemp elimination	Ab 34E4		0.66	5.5×10^3	(28)
	$\text{HG3}_{\text{design}}$	TIM barrel (1GOR)	3.0 ± 0.1	1.3×10^3	(45)
	$\text{HG3.17}_{\text{evolved}}$		700 ± 60	2.3×10^5	
	$\text{KE59}_{\text{design}}$	TIM barrel (1A53)	Nd	160	(6)
	$\text{KE59.13}_{\text{evolved}}$		9.5 ± 0.6	5.9×10^4	
	$\text{KE07}_{\text{design}}$	TIM barrel (1THF)	0.017 ± 0.001	12.2 ± 0.1	(6)
	$\text{KE07.7}_{\text{evolved}}$		1.37 ± 0.14	2590 ± 300	
	$\text{KE70}_{\text{design}}$	TIM barrel (1JCL)	0.14 ± 0.01	126 ± 4	(6)
	$\text{KE70.8}_{\text{evolved}}$		7.47 ± 0.32	53,100 ± 8,200	
Retro-Aldol	Ab 38C2		0.012 ± 0.001	480	(26)
	$\text{RA95.0}_{\text{design}}$	TIM barrel (1LBL)	0.01 ± 0.00002	0.19	(8)
	$\text{RA95.5-8}_{\text{evolved}}$		0.17 ± 0.02	850	(47)
Diels-Alder	Ab 7D4		5.83×10^{-5}	20.18	(58)
	$\text{DA20}_{\text{design}}$	β-propeller (1E1A)	$(2.78 \pm 0.56) \times 10^{-5}$	0.06	(9)
	$\text{DA20.10}_{\text{evolved}}$		$(59.2 \pm 6.7) \times 10^{-5}$	6.23	
	$\text{CE6}_{\text{designed}}$	Helix-turn-helix lid on to DA20.10	$(61.1 \pm 2.78) \times 10^{-5}$	87 ± 14	(46)
	$\text{CE20}_{\text{evolved}}$		$(300 \pm 14) \times 10^{-5}$	540 ± 70	

Retro-Aldolase RA95

Catalytic antibodies were created for the complex multistep retro-aldol reaction in 1997 (Figure 1.1) (26). *De novo* aldolases were created with multiple scaffolds and evolved (8, 36). The activity of the designed retro-aldolase RA95, which was created by insertion of a designed active site into an $(\beta/\alpha)_8$ barrel scaffold (PDB ID 1LBL) was subsequently enhanced 4,400-fold by directed evolution (47). The crystal structures of the intermediates in the directed evolution experiment showed changes in the location of the catalytic motif, substrate binding mode, and the mobility of loops during RA95 evolution (47). Another aldolase was constructed from the PLP-dependent racemase by a single point mutation (59).

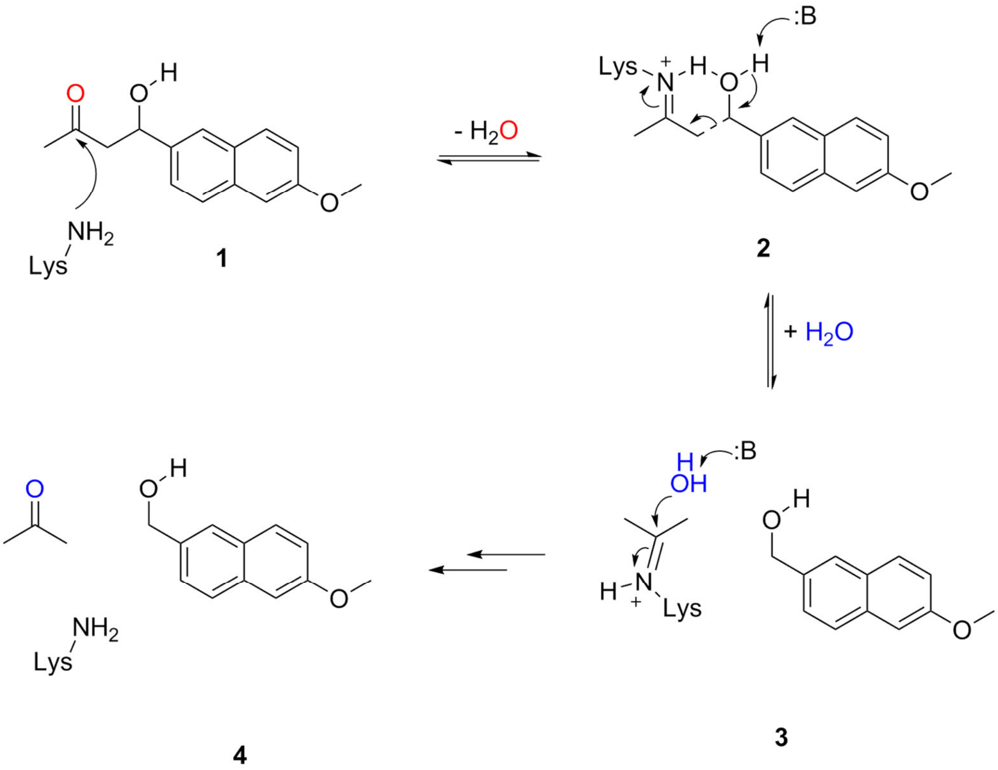


Figure 1. 1 | Reaction scheme for the Retro-aldol reaction catalysed by lysine and a base residue inside of the active site. The synthetic compound 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone (**1**) forms protonated Schiff base intermediates (**2**). Deprotonation of the beta alcohol causes C-C bond cleavage (**3**) and then the imine is hydrolysed (**3**) to form the final products 6-methoxy-2-naphthaldehyde and acetone (**4**).

Diels-Alderase

A catalytic antibody was created for the challenging C-C bond-forming cycloaddition Diels-Alder reaction in 1989 (Figure 1.2) (27). *De novo* Diels-Alderases were created with multiple scaffolds in 2010 (9). In the case of DA20.00, which was created by insertion of a designed active site into a six-bladed beta-propeller scaffold (PDB ID 1E1A), catalytic activity was enhanced 9,700-fold by installation of a lid loop and then through directed evolution (46). Crystal structures of intermediates in the directed evolution experiment showed that the active site was reshaped to stabilize the transition state and became more effectively pre-organized.

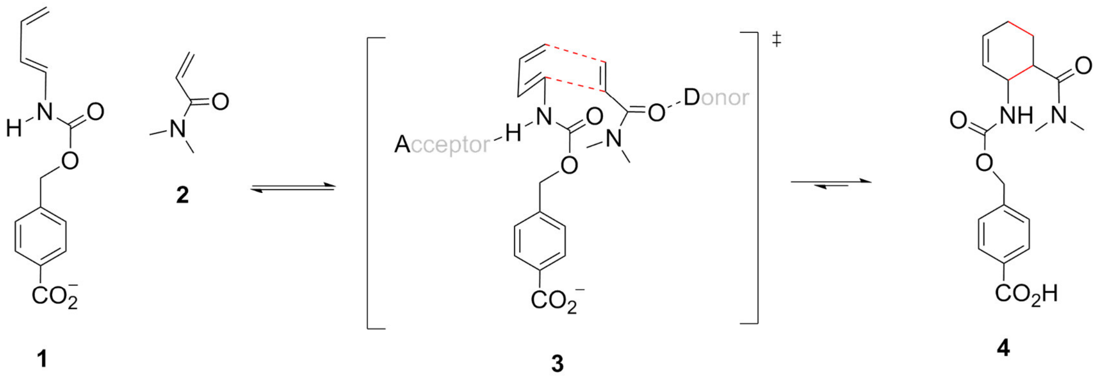
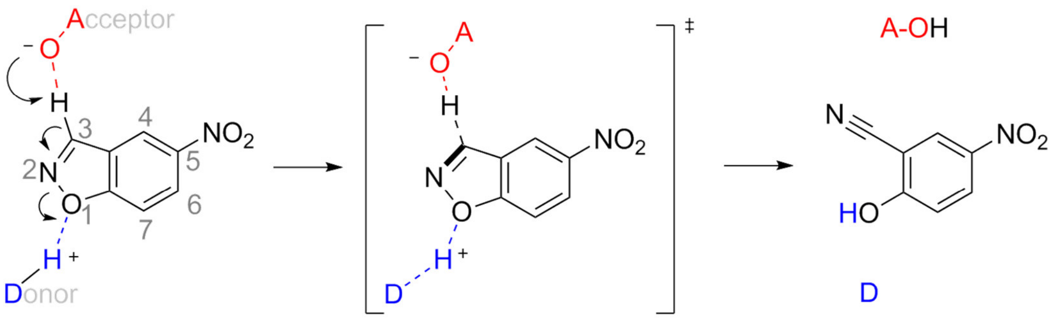


Figure 1. 2 | Reaction scheme for the Diels-Alder reaction. Diene 4-carboxybenzyl-*trans*-1,3-butadiene-1-carbamate (**1**) and dienophile *N,N*-dimethylacrylamide (**2**) are activated by donor and acceptor residues in the protein active site (**3**) and forms the final product 3R,4S endo cyclohexane (**4**).

1.4 *De novo* Kemp eliminases

Kemp elimination of 5-nitrobenzisoxazole

The Kemp elimination is an irreversible bimolecular ring opening of benzisoxazoles that occurs by a C-H activated H transfer mechanism (Figure 1.3). The substrates were first synthesized and the reaction was defined by D.S. Kemp (60). The relative stability of benzisoxazoles, due to aromaticity and various simple functionalizations has led to their usage in drug development (61). For example, benzisoxazoles are used as a starting material to synthesize bioactive molecules such as antipsychotic and anticancer drugs (62–70). Among benzisoxazoles, 5-nitrobenzisoxazole has been used for computational studies of the catalytic mechanism and for design of de-novo Kemp eliminases (6, 28, 31, 45). The slow, spontaneous decomposition of 5-nitrobenzisoxazole (**1**) to the 5-nitrosalicylonitrile anion can be catalysed by anions present in buffer. The reaction proceeds *via* an E2 reaction with a single transition state, so the C-H bond is weakened and a partial negative charge develops on the oxygen simultaneously (**2**). The isoxazole ring is opened by C-H bond cleavage, resulting in a negatively charged intermediate which is later neutralized by protonation of the oxyanion (**3**). The reaction rate is greatly affected by the charge of the base and is sensitive to the local environment (60).



1 **2** **3**

Figure 1. 3 | Reaction scheme for the Kemp elimination of 5-nitrobenzisoxazole. The nucleophilic oxygen atom of the H-acceptor (A) donates electrons to the electrophilic 3'-H of 5-nitrobenzisoxazole and the electronegative oxygen atom of the isoxazole group forms a hydrogen bond with a H-donor (D) (**1**), forming a transition state in which the C-H and N-O bonds are weakened (**2**). The removal of the 3'-H from the substrate leads to an anionic phenoxide intermediate, which is then protonated, forming the final product (**3**).

Kemp eliminases

The Kemp elimination reaction (60) has been a focus of the field of computational enzyme design (6, 7, 38, 45, 48, 71). The popularity of *de novo* Kemp eliminases is due to their well-defined, simple, but environmentally sensitive and biologically important mechanism (ring opening *via* C-H activation), the possibility of accurate spectroscopic monitoring of reaction kinetics, and previously well studied enzymatic mechanisms by computational modelling using catalytic antibodies (e.g. the catalytic antibody 34E4) (28–31, 34).

The combination of computational methods and directed evolution was successfully applied to create *de novo* the KE series of Kemp eliminases, including KE07, KE70 and KE59, using the Rosetta algorithm (39) in 2008 (6). Later, the HG series of Kemp eliminases were created *de novo* using a TIM barrel scaffold (PDB ID 1GOR) with the Phoenix algorithm (41), (45). Their activities ($M^{-1}s^{-1}$) were improved 200-400-fold through directed evolution (45). The designs and evolved variants were extensively analysed to improve computational algorithms (15, 45, 48, 52–57). To be specific, the evolutionary trajectories of KE07 (71), KE70 (48), KE59 (51), and HG3 (45) were analysed using X-ray crystallography and computational simulation. The analysis often showed emergence of an alternative conformation of the base (E, D, or a D-H dyad) and π-stacking residue (W or Y), changing the substrate binding mode (e.g. KE70, KE59, HG3). Computer simulations of Kemp eliminases (e.g. KE70, KE59, HG3, and so on) show that the most catalytically efficient substrate binding mode can be enriched through evolution by distant mutations to reduce reorganization energy (52, 54). AlleyCat is another Kemp eliminase, which was constructed from the non-catalytic binding protein calmodulin by a single point mutation (38).

Ab34E4

The most efficient antibody catalyst for the Kemp elimination reaction (34E4) was produced in 1995 (28). Subsequently, quantum mechanical calculations were performed, suggesting that the proper orientation of base and the introduction of an oxyanion binding residue affected the rate acceleration (30, 31). Later, crystal structures of 34E4 were solved and showed two conformations of the substrate π-stacking residue W91 in the apo-structure and one conformation in the substrate bound structure (32). The authors mentioned that future efforts to solidify the active conformation of W91 might increase the activity of the most efficient catalytic antibody 34E4 further by eliminating the slow conformational isomerization process of the inactive conformation of W91 (32).

KE59

KE59 was designed (6) using a TIM barrel scaffold (PDB ID 1A53, indole 3-glycerol phosphate synthase – (72)). This enzyme was the least thermostable of the KE series, but was active with a variety of benzisoxazoles (6). The activity of KE59 was enhanced 400-fold by directed evolution (51). The substrate binding mode was changed with conformational selection of the π-stacking residue W109, and the motion of the catalytic base E230 was decreased during evolution (51). The selected conformation of W109 formed a hydrogen bond with the base E230.

KE70

KE70 was designed using a TIM barrel scaffold (PDB ID 1JCL, deoxyribose phosphate aldolase) and its activity was increased 300-fold by directed evolution (6). Crystal structures showed that a narrower active site entry and stronger electrostatic network were developed during evolution, which improved substrate binding and increased precise substrate positioning (48).

HG3

HG3 was designed using a TIM barrel scaffold (PDB ID 1GOR, thermostable xylanase 1 from *Thermoascus aurantiacus*) by a focused scaffold search and a molecular dynamics (MD)-assisted design approach in 2012 (7). The activity was enhanced 200-fold by directed evolution, entirely by increases in k_{cat} (45). Although the variants showed low pH-stability and thermostability (T_{m} 40-60 ℃) with high K_{M} values (2.4-8.3 mM) the k_{cat} of most active variant (HG3.17) was the highest that has been reported among designed Kemp eliminases (700 s^{-1}). The substrate binding mode, alignment of the substrate with the catalytic base D127, and the identity of the oxyanion binder (from T265 at the bottom of active site (designed) to Q50 at the entry of active site) changed throughout the directed evolution trajectory. Two conformational states of the substrate stabilizing residue W44 were identified in the crystal structure of most active variant. In one conformation, the indole nitrogen of W44 forms a hydrogen bond with the catalytic base D127. The indole nitrogen of W44 in the other conformation forms a hydrogen bond with T265. Quantum Mechanics/Molecular Mechanics (QM/MM) simulations of the HG3 variants highlighted the increased flexibility of the active site (53) and decreased sampling of unproductive substrate binding modes, and decreased reorganization energy due to pre-organization of the active site during evolution (52).

KE07

KE07 was designed using a TIM barrel scaffold: the HisF protein from *Thermotoga maritima* (PDB ID 1THF), an enzyme involved in histidine biosynthesis (73–76)). This enzyme is the most thermostable (T_m >95 °C) among *de novo* design enzymes to date (6). Its activity was increased 200-fold through directed evolution (6, 71). Structures and biochemical analyses of the evolutionary end-points and intermediates highlighted the increased basicity of E101, which was caused by reorganization of the electrostatic network by the mutation I7D, and increased flexibility of the active site by conformational isomerism of R202, during evolution. QM/MM studies have been performed, showing reorganization of water molecules (1, 56, 57) and reduction of reorganization energy during evolution (55).

1.5 KE07: *de novo* design and directed evolution

Because this chapter is focused on understanding how the catalytic efficiency of KE07 increased through directed evolution, the following section will describe the design, evolution and biophysical/biochemical characterization of KE07 that has already been performed in greater detail.

Design

The catalytic residues included in the KE07 design were a catalytic base (E101) for C-H abstraction, a general acid (K222) to stabilize the oxyanion, and a π-stacking residue (W50) to polarize the substrate. This active site arrangement was calculated *via* high-level quantum chemical calculations to stabilize the TS in the Kemp elimination reaction (6). A scaffold compatible with the designed geometry of the residues was searched for using the RosettaMatch algorithm (39). The TIM barrel HisF (PDB ID 1THF) from *T. maritima* was selected as a scaffold for KE07 design (Figure 1.4). *T. maritima* was first isolated in 1986 from hot marine sediments in Vulcano, Italy (70–100 ℃, pH 6.5–7) (73). *T. maritima* is phylogenetically significant because it is a particularly slowly evolving bacterium (77). Its genome was sequenced in 1999, showing 24% similarity to archaeal genes (74). In addition, it is an ideal model system to understand biological systems because it has one of the smallest and most diverse genomes. The high thermostability (>95 ℃) and functional diversity of the *T. maritima* proteome has provided many stable scaffolds for enzyme engineering (78–81). The TIM-barrel scaffold is predominant in the central metabolic network of *T. maritima* (82). The archetypal TIM-barrel, triosephosphate isomerase (TIM), is known to catalyse its isomerisation reaction at rates up to the diffusion limit (10^8 $M^{-1}s^{-1}$). Notably, TIM uses a similar catalytic mechanism to KE07 (E165 as a catalytic base, and K12 and H95 to stabilize the enolate intermediate) (83).

Although the designed KE07 (Round 1, henceforth R1) accelerated the Kemp elimination 10^4-fold relative to the non-enzymatic reaction, with a reasonable K_M value (1.4 mM) and high thermostability (>95 ℃), its turn-over rate (0.018 s^{-1}) was many orders of magnitude lower than the most efficient natural enzymes (e.g. TIM:

4,300 s^{-1} for glyceraldehyde phosphate, β-lactamase: 2,400 s^{-1} for methyl penicillin) (43, 44).

Mutagenic characterisation of catalytic residues in the KE07 design suggests the involvement of other residues in addition to the designed catalytic base/acid residues (E101/K222) in a rate limiting pathway (6). To be specific, the k_{cat} of KE07 R1 was reduced only 20-fold by mutation of catalytic base to alanine (E101A), whereas other KE designs lose all activity with the mutation of the catalytic base to alanine (6). Furthermore, the k_{cat} of KE07 design was actually increased 2-fold by mutation of the putative catalytic acid (K222) to alanine.

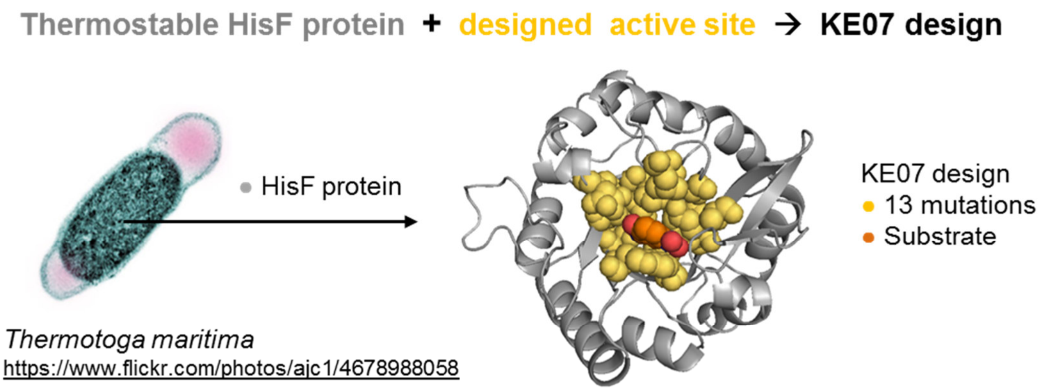


Figure 1. 4 | KE07 design using a TIM barrel scaffold from *Thermotoga maritima*. The microscopy image of the *T. maritima* is shown on the left side. The protein HisF, involved in the histidine biosynthesis pathway, was chosen as a scaffold to incorporate the designed transition state active site residues, shown in yellow spheres on cartoon representation. The designed position of the substrate 5-nitrobenzisoxazole is shown in orange spheres.

Evolution

The catalytic efficiency of the initial design of KE07 (R1, 12 $M^{-1}s^{-1}$) was enhanced 200-fold (2,590 $M^{-1}s^{-1}$), and its turnover rate (0.018 s^{-1}) was enhanced 100-fold (1.37 s^{-1}) (6). The most active variant KE07 R7-2 shows 2-fold higher turnover rate (1.4 s^{-1}) than the most efficient antibody catalyst 34E4 (0.66 s^{-1}) (28).

Although crystal structures of the designed KE07 (R1), and several evolutionary variants (R4, R6, R7), have been solved, the crystal structure of the most active variant (R7-2) has not been solved yet (71). The structure of KE07 R1 showed a salt bridge between E101 and K222, which reduces the basicity of E101. This salt bridge was broken in R4 by the mutation I7D, which forms a hydrogen bond with K222 and bridges K222 *via* an electrostatic network behind the active site (K222-D7-R5-E167-K99). The emergence of charged residues at the entry of active site (N224D, G202R) which are consistently present from R2 results in the formation of a new electrostatic network with H201, which was strengthened at R7. This network potentially interacts with substrate, facilitating catalysis. R202, positioned on the surface binding loop, showed two conformations in R7, suggesting increased flexibility of the active site during evolution. These changes are shown in Figure 1.5.

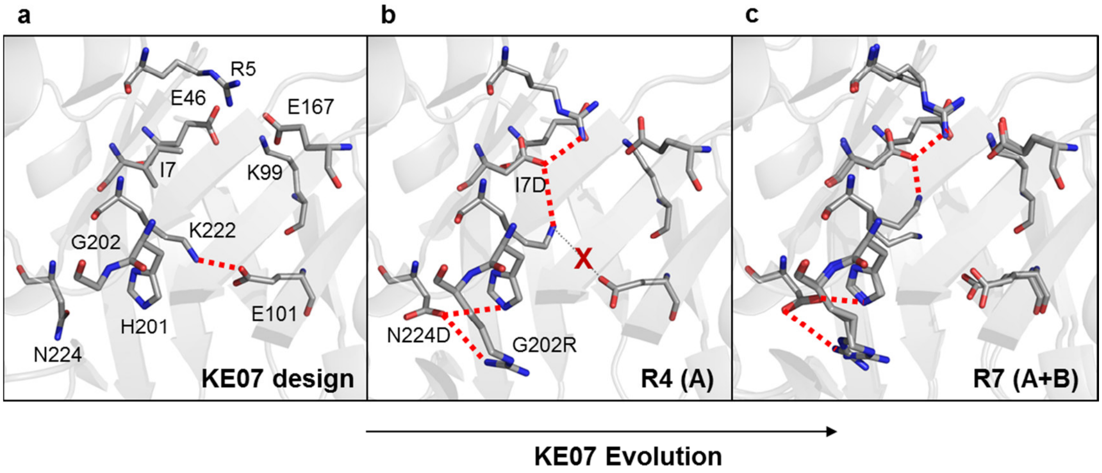


Figure 1. 5 | Crystal structures of KE07 variants. The salt-bridge between E101 and K222 and new electrostatic interactions during evolution are shown as red dotted lines. The broken salt-bridge at R4 crystal structure is highlighted by **X** on the grey dotted line. (A) Chain A shown. (A+B) Superposition of chains A and B shown. PDB ID for the structures: 2RKX (**a**), 3IIO (**b**), 3IIV (**c**).

Computational Evaluation

The designs and evolutionary trajectories of KE07 have been evaluated by computer simulations to learn from the improvements and enhance the prediction of the structures and energetics of future designs (1). For example, the free energy surfaces of KE07, along with other enzymes in the KE series (KE10, KE15, KE16), were calculated to predict the initial activity of the designed enzymes (15). Later, KE07 variants, other KE designs, and Ab34E4 were evaluated with QM/MM simulations (57). More water molecules were located near the catalytic base (<3.2 Å) of KE07 in the evolved variant (R7, 3.9) compared to the initial variants (R1, 2), even though solvent accessibility was reduced. KE07 evolution was also studied by Warshel and co-workers with molecular orbital-combined quantum mechanics/ molecular mechanics (MO-QM/MM) and empirical valance bond (EVB) simulations (56). In contrast to the previous finding, fewer waters were found near the catalytic base (E101) in the evolved variant (R6) compared to R1, showing that the increase in catalytic activity during evolution was caused by reduction of solvation rather than transition state stabilization. Another study using the EVB method in 2013 showed that the reorganization energy was reduced by distant mutations (V12L, F77I, M207T) (55).

1.6 Overview of thesis

The development of design methods to create tailored enzymes for synthetic molecular reactions has been intensely studied over the last decade. The *de novo* designed enzyme KE07 is one of the first and most successful designs for catalysis of the synthetic Kemp elimination reaction. The KE07 design had a reasonable K_M but very low k_cat, compared to naturally efficient enzymes, which is a problem associated with most designs. The k_cat of KE07 was gradually increased through 7 rounds of directed evolution with about one mutation accumulated per round. This evolutionary trajectory has been studied by crystallography and computational simulations. However, the structure of the most active variant has not been solved (R7-2) and no detailed kinetic characterization (e.g. determination of thermodynamic parameters from Arrhenius plots or kinetic isotope effects) has been performed.

Here, the evolutionary trajectory of KE07 was analysed by enzyme kinetics, spectroscopy, X-ray crystallography, and computer simulation to explain how the accumulation of mutations during evolution changes energetics, the active site conformation, and the substrate-binding mode, which are not well understood in this system. Firstly, enthalpic and entropic contributions to the free energy barrier and primary kinetic isotope effects in KE07 variants were quantified using Arrhenius plots (Chapter 3). Next, I crystallized the initial design (R1) and all variants in later rounds, including the most active variant R7-2 (R5, R6, R7, R7-2). The crystals were then soaked with substrate to observe conformational changes in the presence of substrate (Chapter 4). Unexpectedly, the π-stacking residue, tryptophan, at position 50 exhibited a dramatically altered position in the intermediate variants (>90° rotation comparing to the designed position), which was stabilized in the most active variant (R7-2). Therefore, we further analysed KE07 evolution using molecular dynamics simulations to observe movement of the side chain of W50, tryptophan fluorescence to evaluate the conformational states in solution, and quantum mechanical calculations using the initial and final configurations of W50 (Chapter 5).

1.7 References

1. Kiss G, Çelebi-Ölçüm N, Moretti R, Baker D, Houk KN (2013) Computational enzyme design. *Angew Chemie - Int Ed* 52(22):5700–5725.
2. Kries H, Blomberg R, Hilvert D (2013) De novo enzymes by computational design. *Curr Opin Chem Biol* 17(2):221–228.
3. Wilson CJ (2015) Rational protein design: developing next-generation biological therapeutics and nanobiotechnological tools. *Wiley Interdiscip Rev Nanomedicine Nanobiotechnology* 7(3):330–341.
4. Hemmert AC, et al. (2011) Nerve agent hydrolysis activity designed into a human drug metabolism enzyme. *PLoS One* 6(3):e17441.
5. Cherny I, et al. (2013) Engineering V-type nerve agents detoxifying enzymes using computationally focused libraries. *ACS Chem Biol* 8(11):2394–2403.
6. Röthlisberger D, et al. (2008) Kemp elimination catalysts by computational enzyme design. *Nature* 453(7192):190–195.
7. Privett HK, et al. (2012) Iterative approach to computational enzyme design. *Proc Natl Acad Sci* 109(10):3790–3795.
8. Althoff EA, et al. (2012) Robust design and optimization of retroaldol enzymes. *Protein Sci* 21(5):717–726.
9. Siegel JB, et al. (2010) Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science* 329(5989):309–13.
10. Steen EJ, et al. (2010) Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* 463(7280):559–562.
11. McKenna R, Nielsen DR (2011) Styrene biosynthesis from glucose by engineered E. coli. *Metab Eng* 13(5):544–554.
12. Sollid LM (2002) Coeliac disease: dissecting a complex inflammatory disorder. *Nat Rev Immunol* 2(9):647–55.
13. Perez EE, et al. (2008) Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat Biotechnol* 26(7):808–16.
14. Bornscheuer UT, et al. (2012) Engineering the third wave of biocatalysis. *Nature* 485(7397):185–194.
15. Alexandrova AN, Röthlisberger D, Baker D, Jorgensen WL (2008) Catalytic mechanism and performance of computationally designed enzymes for Kemp elimination. *J Am Chem Soc* 130(47):15907–15915.
16. Gordon SR, et al. (2012) Computational design of an α-gliadin peptidase. *J Am Chem Soc* 134(50):20513–20520.
17. Eyring H (1935) The activated complex and the absolute rate of chemical reactions. *Chem Rev* 17(1):65–77.
18. Pauling L (1946) Molecular architecture and biological reactions. *Chem Eng News* 24(10):1375–1377.
19. Villa J, Warshel A (2001) Energetics and dynamics of enzymatic reactions. *J Phys Chem B* 105(33):7887–7907.
20. Warshel A (2014) Multiscale modeling of biological functions: From enzymes to molecular machines (Nobel Lecture). *Angew Chemie Int Ed* 53(38):10020–10031.
21. Warshel A (1998) Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites. *J Biol Chem* 273(42):27035–27038.
22. Ram Prasad B, Warshel A (2011) Prechemistry versus preorganization in DNA replication fidelity. *Proteins: Struct, Funct, Bioinf* 79(10):2900–2919.
23. Olsson MHM, Parson WW, Warshel A (2006) Dynamical contributions to enzyme catalysis: Critical tests of a popular hypothesis. *Chem Rev* 106(5):1737–1756.
24. Liu H, Warshel A (2007) The catalytic effect of dihydrofolate reductase and its mutants is determined by reorganization energies. *Biochemistry* 46(20):6011–6025.
25. Hwang JK, King G, Creighton S, Warshel A (1988) Simulation of free energy relationships and dynamics of SN2 reactions in aqueous solution. *J Am Chem Soc* 110(16):5297–5311.
26. Barbas CF, et al. (1997) Immune versus natural selection: antibody aldolases with enzymic rates but broader scope. *Science* 278(December):2085–2092.
27. Hilvert D, Hill KW, Nared KD, Auditor M-TM (1989) Antibody Catalysis of a Diels-Alder Reaction. *J Am Chem Soc* 111(12):9262–9263.
28. Thorn SN, Daniels RG, Auditor M-TM, Hilvert D (1995) Large rate accelerations in antibody

catalysis by strategic use of haptenic charge. *Nature* 373(6511):228–30.

29. Hilvert D (2000) Critical analysis of antibody catalysis. *Annu Rev Biochem* 69(1):751–793.
30. Na J, Houk KN, Hilvert D (1996) Transition state of the base-promoted ring-opening of isoxazoles. Theoretical prediction of catalytic functionalities and design of haptens for antibody production. *J Am Chem Soc* 118(27):6462–6471.
31. Hu Y, Houk KN, Kikuchi K, Hotta K, Hilvert D (2004) Nonspecific medium effects versus specific group positioning in the antibody and albumin catalysis of the base-promoted ring-opening reactions of benzisoxazoles. *J Am Chem Soc* 126(26):8197–8205.
32. Debler EW, Müller R, Hilvert D, Wilson IA (2008) Conformational isomerism can limit antibody catalysis. *J Biol Chem* 283(24):16554–16560.
33. Cauerhff A, Goldbaum F a, Braden BC (2004) Structural mechanism for affinity maturation of an anti-lysozyme antibody. *Proc Natl Acad Sci U S A* 101(10):3539–3544.
34. Foote J, Milstein C (1994) Conformational isomerism and the diversity of antibodies. *Proc Natl Acad Sci U S A* 91(22):10370–4.
35. Wilson I a, Stanfield RL (1994) Antibody-antigen interactions: new structures and new conformational changes. *Curr Opin Struct Biol* 4(6):857–867.
36. Jiang L, et al. (2008) De novo computational design of retro-aldol enzymes. *Science* 319(5868):1387–1391.
37. Fleishman SJ, et al. (2011) Computational Design of Proteins Targeting the Conserved Stem Region of Influenza Hemagglutinin. *Science* 332(6031):816–821.
38. Korendovych I V, et al. (2011) Design of a switchable eliminase. *Proc Natl Acad Sci* 108(17):6823–6827.
39. Zanghellini A, et al. (2006) New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci* 15(12):2785–2794.
40. Richter F, Leaver-Fay A, Khare SD, Bjelic S, Baker D (2011) De novo enzyme design using Rosetta3. *PLoS One* 6(5):e19230.
41. Dahiyat BI, Gordon DB, Mayo SL (1997) Automated design of the surface positions of protein helices. *Protein Sci* 6(6):1333–7.
42. Nosrati GR, Houk KN (2012) SABER: A computational method for identifying active sites for new reactions. *Protein Sci* 21(5):697–706.
43. Putman SJ, Coulson AFW, Farley IRT, Riddleston B, Knowles JR (1972) Specificity and kinetics of triose phosphate isomerase from chicken muscle. *Biochem J* 129:301–310.
44. Christensen H, Martin MT, Waley SG (1990) Beta-lactamases as fully efficient enzymes. Determination of all the rate constants in the acyl-enzyme mechanism. *Biochem J* 266(3):853–861.
45. Blomberg R, et al. (2013) Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature* 503(7476):418–421.
46. Preiswerk N, et al. (2014) Impact of scaffold rigidity on the design and evolution of an artificial Diels-Alderase. *Proc Natl Acad Sci* 111(22):8013–8018.
47. Giger L, et al. (2013) Evolution of a designed retro-aldolase leads to complete active site remodeling. *Nat Chem Biol* 9(8):494–498.
48. Khersonsky O, et al. (2011) Optimization of the in-silico-designed Kemp eliminase KE70 by computational design and directed evolution. *J Mol Biol* 407(3):391–412.
49. Stemmer WP (1994) DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc Natl Acad Sci U S A* 91(22):10747–10751.
50. Park H-S, et al. (2006) Design and evolution of new catalytic activity with an existing protein scaffold. *Science* 311(5760):535–8.
51. Khersonsky O, et al. (2012) Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed Kemp eliminase KE59. *Proc Natl Acad Sci* 109(26):10358–10363.
52. Osuna S, Jiménez-Osés G, Noey EL, Houk KN (2015) Molecular dynamics explorations of active site structure in designed and evolved enzymes. *Acc Chem Res* 48(4):1080–1089.
53. Świderek K, Tuñón I, Moliner V, Bertran J (2015) Protein flexibility and preorganization in the design of enzymes. The Kemp elimination catalyzed by HG3.17. *ACS Catal* 5(4):2587–2595.
54. Frushicheva MP, Cao J, Warshel A (2011) Challenges and advances in validating enzyme design proposals: The case of kemp eliminase catalysis. *Biochemistry* 50(18):3849–3858.
55. Labas A, Szabo E, Mones L, Fuxreiter M (2013) Optimization of reorganization energy drives evolution of the designed Kemp eliminase KE07. *Biochim Biophys Acta* 1834(5):908–17.
56. Frushicheva MP, Cao J, Chu ZT, Warshel A (2010) Exploring challenges in rational enzyme

design by simulating the catalysis in artificial kemp eliminase. *Proc Natl Acad Sci U S A* 107(39):16869–16874.

57. Kiss G, Röthlisberger D, Baker D, Houk KN (2010) Evaluation and ranking of enzyme designs. *Protein Sci* 19(9):1760–1773.
58. Yli-Kauhaluoma JT, et al. (1995) Anti-Metallocene Antibodies: A New Approach to Enantioselective Catalysis of the Diels-Alder Reaction. *J Am Chem Soc* 117(27):7041–7047.
59. Seebeck FP, Hilvert D (2003) Conversion of a PLP-dependent racemase into an aldolase by a single active site mutation. *J Am Chem Soc* 125(34):10158–10159.
60. Casey ML, Kemp DS, Paul KG, Cox DD (1973) Physical organic chemistry of benzisoxazoles. I. Mechanism of the base-catalyzed decomposition of benzisoxazoles. *J Org Chem* 58(I):33–34.
61. Dubrovskiy A V, Larock RC, Chem RWJM (2010) Synthesis of benzisoxazoles by the [3+ 2] cycloaddition of in situ generated nitrile oxides and arynes. *Org Lett* 12(6):1180–1183.
62. Davis L, et al. (1992) 3-substituted-1, 2-benzisoxazoles: novel antipsychotic agents. *Drug Des Discov* 8(3):225–240.
63. Gopalsamy A, et al. (2008) Discovery of benzisoxazoles as potent inhibitors of chaperone heat shock protein 90. *J Med Chem* 51(3):373–375.
64. Jain M, Kwon C (2003) 1, 2-Benzisoxazole phosphorodiamidates as novel anticancer prodrugs requiring bioreductive activation. *J Med Chem* 46(25):5428–5436.
65. Strupczewski JT, et al. (1985) Synthesis and Neuroleptic Activity of 3-(1 -Substituted-4-piperidinyl)-1,2-benzisoxazoles. *Am Chem Soc* 28:761–769.
66. Oldiges H, et al. (1979) Derivatives and Their Anticonvulsant Activities. 22(1976):180–183.
67. Villalobos A, et al. (1994) Novel benzisoxazole derivatives as potent and selective inhibitors of acetylcholinesterase. *J Med Chem* 37(17):2721–2734.
68. Strupczewski JT, et al. (1995) 3-[[(Aryloxy) alkyllpiperidinyl]-l, 2-Benzisoxazoles as D2/5-HT2 antagonists with potential atypical antipsychotic activity: Antipsychotic profile of iloperidone. *J Med Chem* 38:1119–1131.
69. Janssen PA, et al. (1988) Pharmacology of risperidone (R 64 766), a new antipsychotic with serotonin-S2 and dopamine-D2 antagonistic properties. *J Pharmacol Exp Ther* 244(2):685–693.
70. Priya BS, Swamy SN, Rangappa KS (2005) Synthesis and characterization of novel 6-fluoro-4-piperidinyl-1, 2-benzisoxazole amides and 6-fluoro-chroman-2-carboxamides: antimicrobial studies. 13:2623–2628.
71. Khersonsky O, et al. (2010) Evolutionary optimization of computationally designed enzymes: Kemp eliminases of the KE07 series. *J Mol Biol* 396:1025–1042.
72. Hennig M, Darimont BD, Jansonius JN, Kirschner K (2002) The catalytic mechanism of indole-3-glycerol phosphate synthase: crystal structures of complexes of the enzyme from Sulfolobus solfataricus with substrate analogue, substrate, and product. *J Mol Biol* 319(3):757–766.
73. Huber R, et al. (1986) Thermotoga maritima sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C. *Arch Microbiol* 144(1976):324–333.
74. Nelson KE, et al. (1999) Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of Thermotoga maritima. *Nature* 399(6734):323–329.
75. Lang D, Thoma R, Henn-Sax M, Sterner R, Wilmanns M (2000) Structural evidence for evolution of the β/α barrel scaffold by gene duplication and fusion. *Science* 289(5484):1546–1550.
76. Canaves JM, Page R, Wilson IA, Stevens RC (2004) Protein biophysical properties that correlate with crystallization success in Thermotoga maritima: Maximum clustering strategy for structural genomics. *J Mol Biol* 344(4):977–991.
77. Achenbach-Richter L, Gupta R, Stetter KO, Woese CR (1987) Were the original eubacteria thermophiles? *Syst Appl Microbiol* 9(1):34–39.
78. Schröder C, et al. (1994) Glucose fermentation to acetate, CO2 and H2 in the anaerobic hyperthermophilic eubacterium Thermotoga maritima: involvement of the Embden-Meyerhof pathway. *Arch Microbiol* 161(6):460–470.
79. Wrba A, Schweiger A, Schultes V, Jaenicke R (1990) Extremely Thermostable D-Glyceraldehyde-3-phosphate Dehydrogenase from the Eubacterium Thermotoga maritima. *Biochemistry* 29:7584–7592.
80. Schut GJ, Adams MWW (2009) The iron-hydrogenase of Thermotoga maritima utilizes ferredoxin and NADH synergistically: a new perspective on anaerobic hydrogen production. *J Bacteriol* 191(13):4451–4457.

81. Winterhalter C, Liebl W (1995) Two extremely thermostable xylanases of the hyperthermophilic bacterium Thermotoga maritima MSB8. *Appl Environ Microbiol* 61(5):1810–1815.
82. Zhang Y, et al. (2009) Three-dimensional structural view of the central metabolic network of Thermotoga maritima. *Science* 325:1544–1549.
83. Knowles JR (1991) Enzyme catalysis: not different, just better. *Nature* 350(6314):121–124.

Chapter 2: Methods

2.1 Preamble

This chapter describes the general methods used in this thesis. First, the methods involving DNA manipulation including cloning, purification, sequencing, and transformation are explained. Second, the methods involving protein production including expression and purification procedures are described. The His-tag removal process, using the sequence specific Tobacco Etch Virus (TEV) protease, including the method to produce the TEV protease is included. Finally, the methods to obtain crystals and structures of proteins are described.

Experiments that were sensitive to contamination were performed with autoclaved glassware, tips and reagents. All buffers and reagents were made with fresh Milli-Q water. DNA work used nuclease free water (Sigma). Specific details, including the suppliers of chemicals and reagents, are presented in the appendices at the end of the thesis.

2.2 DNA

First, the details of vectors, transformation, and purification of DNA used and performed in this thesis are explained. Second, the details of site-directed mutagenesis using the Gibson assembly method (1) to make the W50A of KE07 (Chapter 4, 5) are described.

2.2.1 Vectors

All genes were inserted into the T7 based pET system (Novagen), which originated from pBR322 (2, 3). Recognition sites for restriction endonucleases and a six or ten histidine-tag (His-tag) were located between the T7 promoter and T7 terminator

TEV protease (Tobacco Etch Virus nuclear-inclusion-a endopeptidase isolated from Tobacco Etch Virus) (4–7) was cloned into a pET-17b vector (Novagen) with an N-terminal His_6-tag. This vector is ~3.3 kb and includes an ampicillin resistant gene and a T7 promoter.

KE07 variants were cloned into the pET-29b(+) vector (Novagen) with a C-terminal His_6-tag. This vector is ~5.4 kb and includes a kanamycin resistance gene and the *lacI* promotor.

Genes of metallo β-lactamase superfamily including acyl homo serine lactonase (AiiA) variants, New Delhi metallo-β-lactamase-1 (NDM1) variants, and Verona Integron-encoded metallo-β-lactamase-2 (VIM2) variants were cloned into the pET-28a (Novagen) with an N-terminal His_{10}-tag. A TEV recognition site (GAA AAC CTG TAT TTT CAG GGC) was included between the His_{10}-tag and the inserted gene to facilitate the removal of the His-tag from the target protein before crystallization screens were undertaken. This vector is ~5.4 kb and includes a kanamycin resistance gene and the *lacI* promotor.

2.2.2 DNA amplification and isolation

DNA was isolated using AxyPrep Plasmid Miniprep Kit following the manufacturers protocol (Axygen). A 5 ml overnight LB culture at 37 ℃ inoculated with a single colony of transformed *Escherichia coli* TOP10 (electro competent cells, Invitrogen) was pelleted by centrifugation at 4,400 rpm for 15 minutes on a benchtop centrifuge (Eppendorf). DNA was isolated using the AxyPrep Plasmid Miniprep Kit following the manufacturers protocol (Axygen). Bound DNA was eluted with 30 μl of nuclease-free water.

2.2.3 Transformation and sequencing

DNA (10 ng) was transformed into 50 μl of *E.coli* TOP10 electro-competent cells (Invitrogen) with a brief 2.5 kV electric shock using a Micropulser electrophorator (Bio-Rad). The bacterial cells containing the target DNA were inoculated into 1 ml LB and incubated at 37 ℃ for 1 hour. 100 μl of the resulting LB culture was spread onto the LBA or LBK-agarose gel plate (recipe in appendix A1) and incubated at 37 ℃ overnight to obtain colonies. One colony was picked and grown at 37 ℃ for 17 hours. DNA was purified by the method described in section 2.2.2. The purified plasmid DNA (100 ng/μl) was sent to the John Curtin School of Medical Research at Australian National University for sequencing. The resulting chromatograms were examined for all sequences to confirm the sequence reliability and stored at the website of biomolecular resource facility of John Curtin School of Medical Research (http://dnalims.jcs.anu.edu.au/cgi-bin/dna).

2.2.4 Site-directed mutagenesis

This section describes vector preparation, primer design, PCR, Gibson assembly, and DNA purification used for the site-directed mutagenesis of KE07 variants (W50A in KE07 design and R7-2, Chapter 4 and Chapter 5).

Vector preparation

The FastDigest restriction endonuclease (Thermo scientific) system showed sufficient DNA cleavage with both enzyme systems (Figure 2.1). Thus, we used FastDigest enzymes to prepare the empty vector for cloning. Linear vector without insert was prepared by digestion of plasmid DNA (pET-29b(+)) of KE07 design using NcoI and XhoI FastDigest enzymes (Thermo). A total volume of 50 µl mixture of plasmid DNA (5 µg), FastDigest Green Buffer (10X, 5 µl), FastDigest enzymes (NcoI 2.5 µl and XhoI 2.5 µl) were incubated at 37 °C for 2 hours followed by enzyme inactivation at 65 °C for 15 minutes. Shrimp Alkaline Phosphatase (NEB, 1 µl) and nuclease free water (10 µl) was added to the chilled mixture (10 minutes at room temperature) and incubated for 1 hour at 37 °C to remove phosphate from the end of the plasmid to prevent self-ligation. The mixture was heated at 80 °C for 5 min to inactivate the enzymes and was cooled down at a room temperature for 10 minutes. The linear empty vector was visualized using a Blue-Light Transilluminator (470 nm, home-made) and isolated with a sterilized cutter (red box in Figure 2.1), followed by gel-purification as described in the following section.

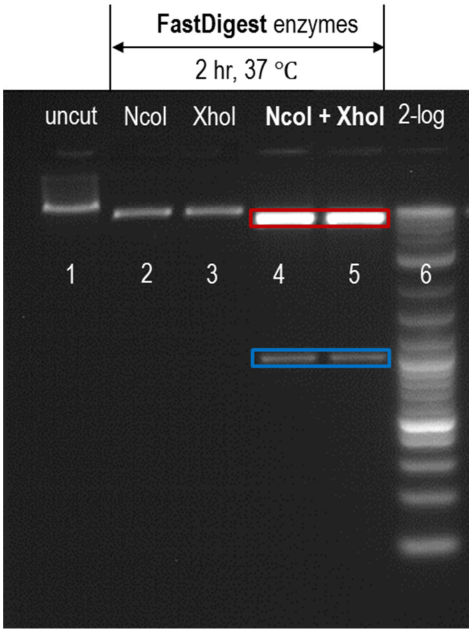


Figure 2. 1 | Vector preparation using restriction enzymes. Intact KE07 plasmid (**1**), NcoI only (**2**), XhoI only (**3**), both NcoI and XhoI (**4, 5**) using FastDigest enzymes and FastDigest green buffer (Thermo) are shown left side of 2-log DNA ladder (**6**). The linear empty vector is highlighted by red box and cleaved insert is highlighted by blue box.

DNA purification

DNA mixed with bromophenol blue loading dye (recipe in appendix A1) was loaded onto a 1% agarose gel with 1 × RedSafe nucleic acid staining solution (20,000 ×, Intron Biotechnology Ltd.). A 2-log DNA ladder (0.1-10 kb, NEB) was used as a size marker. DNA was separated using a Bio-Rad Mini Sub cell GT with 100-120 V (Thermo Scientific Electrophoresis Power Supply (EC300XL2)) for 1-2 hours in borate buffer (pH 8.0, recipe in appendix A1). The DNA was visualized using a Blue-Light Transilluminator (470 nm) instead of UV light to prevent DNA damage and isolated using sterilized cutter. Subsequently, the isolated agarose gel slice containing target DNA was purified using the Wizard SV Gel and PCR Clean-Up System (Promega) following the manufacturer's protocol.

Primer design

Primers were designed to have approximately 25 bp overlap in the 5' and 3' direction from the mutated residue. The DNA sequences corresponding to the amino acid sequences were determined based on codon preference in *E.coli* (8). Primer properties including Tm (melting temperature), GC content, and possible dimerization were analysed by OligoAnalyser at the IDT website (https://sg.idtdna.com/calc/analyzer). Designed primers were evaluated by the 'Run a PCR' program of SerialCloner 2.5 before application. The same primers were used for the W50A mutation of KE07 design and most evolved R7-2 because there were no other mutations around the residue at position 50 during KE07 evolution. Details of the primers for W50A of KE07 are shown below. The residue A50 is shown in red in the primers below.

Forward primer: Tm 70.3 °C, GC content 53.1%
5'-GGCATCGACGAGCTTAGTTTT **GCG** GACATTACCGCTTCGGTGGAAAAGC-3'

Reverse primer: Tm 70.3 °C, GC content 53.1%
5'-GCTTTTCCACCGAAGCGGTAATGTC **CGC** AAAACTAAGCTCGTCGATGCC-3'

PCR

The designed primers (10 μM) were mixed with stock solutions including Thermo Scientific Phusion Hot Start II High-Fidelity DNA Polymerase (F-549S/L) and Phusion GC buffer (5X), T7-primer (10 μM), dNTPs (10 mM), nuclease free water and plasmid (10 ng/μl) according to the manufacturer's instructions as shown in Table 2.1.

Table 2. 1 | PCR reaction conditions for W50A mutation.

	I_N	I_C	B (control)
Phusion GC buffer (5X)	10	10	10
dNTP (10 mM)	1	1	1
Plasmid (10 ng/µl)	1 (10 ng)	1 (10 ng)	0
Primer forward (10 µM)	T7 forward 2.5	W50A forward 2.5	T7 forward 2.5
Primer reverse (10 µM)	W50A reverse 2.5	T7 reverse 2.5	W50A reverse 2.5
Nuclease free water	32.5	32.5	33.5
Polymerase (2 U/µl)	0.5	0.5	0.5
Total volume	50 µl	50 µl	50 µl

The mixture for N' fragment (I_N) and C' fragment (I_C) with W50A mutation were prepared using plasmid DNA of KE07 pET-29b(+). Thermo Scientific Phusion Hot Start II High-Fidelity DNA Polymerase (F-549S/L) and 5X Phusion GC buffer were used following manufacturer's protocol. Details of T7 primers: T7 forward primer: 5'-TAA TAC GAC TCA CTA TAG GG-3' (Tm 47.5 °C) and T7 reverse primer: 5'-GCT AGT TAT TGC TCA GCG-3' (Tm 50.8 °C). Units for the columns are µl.

A temperature gradient (55-72 °C) was used to determine the optimal annealing temperature for the T7 primer and the designed primer combination (Figure 2.2**a**). The PCR efficiency was best at an annealing temperature of 55 °C (Figure 2.2**a**). Thus, the mixtures underwent thermal cycles (Figure 2.2**b**) with an annealing temperature of 55 °C using a MJ Mini Personal Thermal Cycler (Bio-rad) to amplify the DNA, following the phusion polymerase manufacturer's instructions (Thermo). The PCR products were gel-purified as described in DNA purification section above. The DNA was isolated from the gel under a Blue-Light Transilluminator (470 nm) and purified using the Wizard SV Gel and PCR Clean-Up System (Promega). Bound DNA was eluted with 30 µl of nuclease free water.

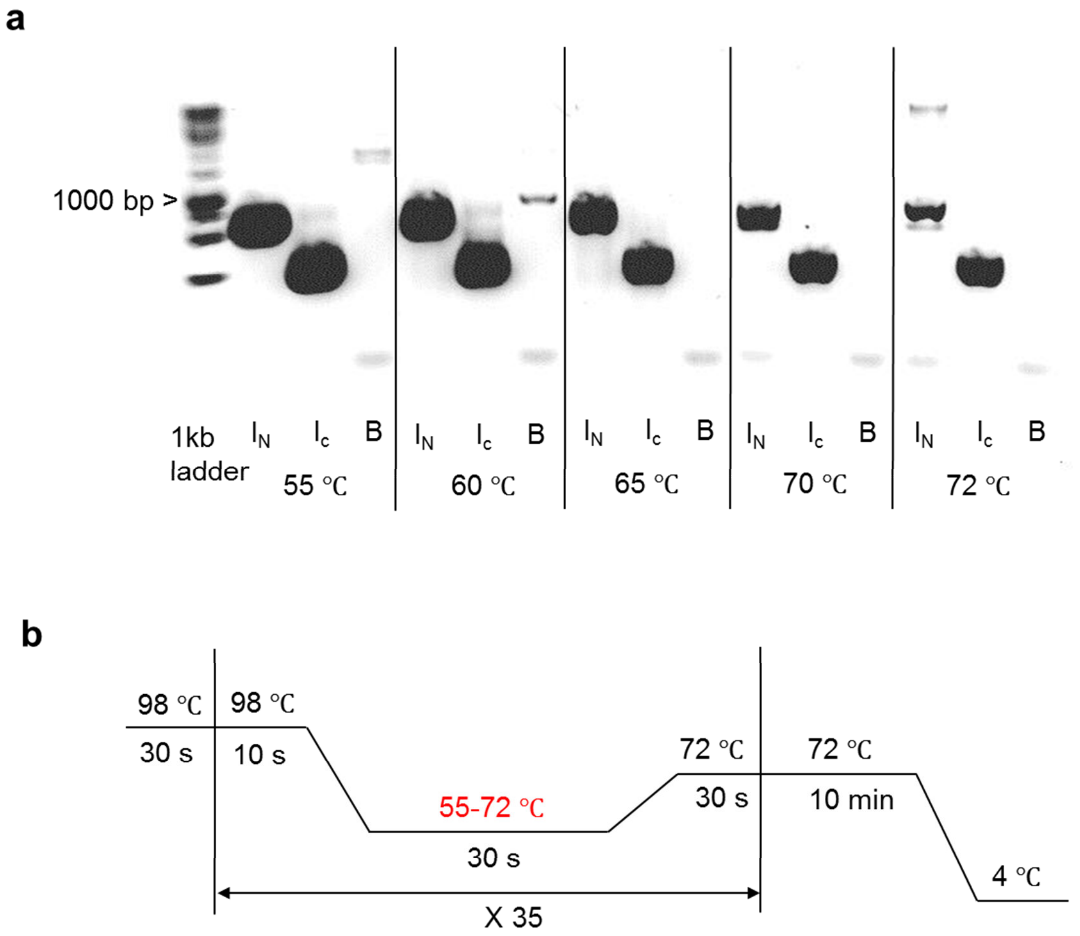


Figure 2. 2 | Temperature gradient to determine the optimal annealing temperature. The PCR products with different annealing temperatures (55-72 °C) for N' fragment (I_N), C' fragment (I_C), and control (B, no template DNA) with point mutation were loaded on a 1% agarose gel (**a**). Thermo Scientific GeneRuler 1kb DNA ladder was used as a size marker. The detailed PCR cycling (**b**) was based on the manufacturer's instruction provided with the Thermo Scientific Phusion Hot Start II High-Fidelity DNA Polymerase (F-549S/L).

Gibson assembly

50 ng (0.0142 pmol) of the linear empty vector and molar equivalents of insert DNAs (0.0142 pmol of I_N and I_C, calculation scheme is shown in Figure 2.3) were mixed with the Gibson Assembly Master Mix (NEB) (1, 9). The mixture (20 µl) was incubated at 50 °C for 1 hour, followed by purification using the PCR Clean-Up System (Promega). Bound DNA was eluted with 30 µl nuclease free water. 1 µl of the purified sample was transformed into *E.coli* TOP10 electro competent cells (Invitrogen) and inoculated into 1 ml LB and incubated at 37 °C for an hour. The colonies were grown overnight on the LBK plate at 37 °C. Two colonies were picked and DNA was purified and sequenced by following the procedures in section 2.2.3.

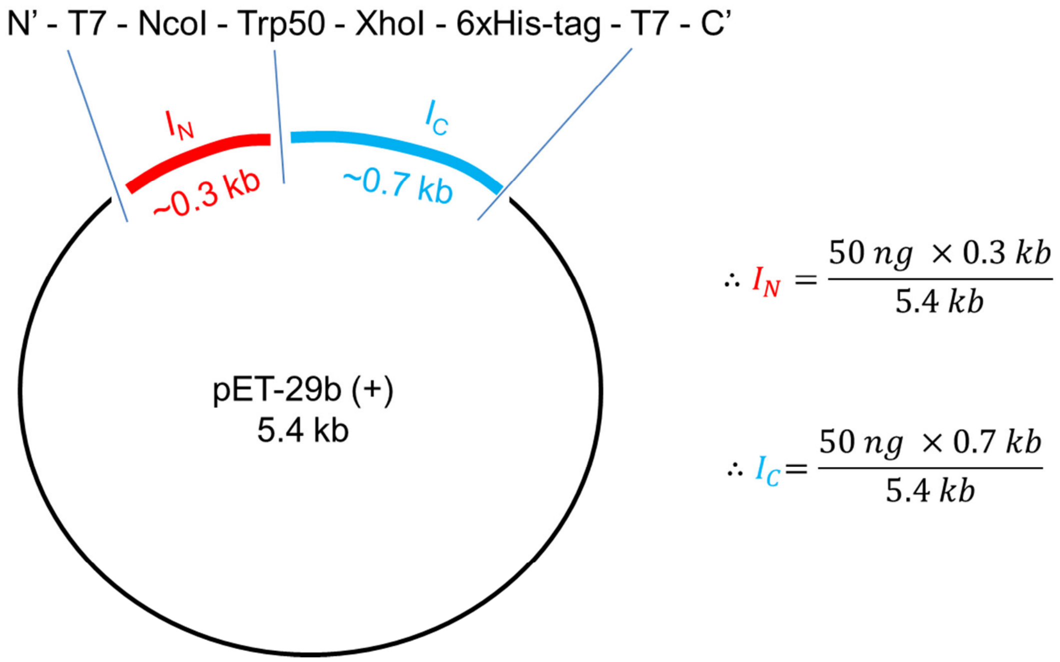


Figure 2. 3 | Scheme for ligation of three DNA fragments for W50A mutation of KE07. The inserts I_N and I_C using T7 polymerase specific primers and designed primers containing point mutation (W50A) were ligated with empty linear vector (pET-29b). The overlaps between each DNA fragments were now shown in the scheme (e.g. 25 pb overlap between I_N and I_C in actual reaction). The equations used to calculate same moles of I_N and I_C for 50 ng of vector (0.014 pmol) are presented on right side.

2.3 Protein

2.3.1 Protein expression

Proteins were expressed by auto-induction (10) or IPTG-induction. One colony of *E.coli* BL21(DE3) containing DNA of the target protein was inoculated into 5-10 ml LB supplemented with antibiotics (35-100 μg/ml) and incubated overnight (~17 hours) at 37 °C. 1-10 ml of the LB seed culture was added to 400-1000 ml LB or TB large culture supplemented with antibiotics. 200 μM $ZnCl_2$ was included throughout the purification process for metalloenzymes (e.g. AiiA, NDM1, and VIM2).

For the auto-induction, TB cultures supplemented with antibiotics and glycerol (0.4-1%) were incubated at 37 °C for 24 hours or 30 °C for 6 hours, followed by room temperature (22 °C) for 16 hours. For the IPTG induction, the LB large culture supplemented with antibiotics was incubated at 37 °C for 3-6 hours until the cell density OD_{600} was around 0.8-1. IPTG (final concentration 1 mM) was added to the large culture and the growth incubated overnight at 30 °C (~16 hours). All cultures were incubated with 100-200 rpm orbital shaking using an Innova44 incubator shaker (Eppendorf).

2.3.2 Lysis

The cells were pelleted by centrifugation at 5,000 rpm (4620 × *g*) for 15 min using Hitachi High-Speed Refrigerated Centrifuge (R9A rotor) at 4 °C. The collected cells were lysed immediately or stored at - 80 °C. The cells were resuspended in buffer A (20-25 mM Imidazole, 50 mM HEPES or Tris-HCl pH 7.25-8.0, 150-500 mM NaCl, and 200 μM $ZnCl_2$ for metalloenzymes) and 0.5 μl turbonuclease from *Serratia marcescens* (200 U/μl, Sigma, T4330) was added to reduce viscosity and degrade RNA and DNA in the sample. Subsequently, the mixture was lysed two times using a French Press (Thermo, 12,000 psi) or Ultrasonic homogenizer (OMNI sonic ruptor 400, 50% power, 6 minutes). The cell debris was removed by centrifugation at 13,500 rpm (29,070 × *g*) for 60 minutes using a Hitachi High-Speed Refrigerated Centrifuge (R15A rotor) at 4 °C. The supernatant was filtered (0.45 μm) to remove bacteria and particles before loaded onto nickel affinity column (5 ml, Qiagen or GE Healthcare).

Part of the culture (1 ml) was pelleted by centrifugation and lysed using Bugbuster protein extraction reagent (Novagen). The mixture was incubated for about 10-20 minutes at room temperature. The cell debris was removed by centrifugation at 12,000 rpm for 10 minutes using a benchtop centrifuge (Eppendorf model 5424). The supernatant (1-10 μl) was mixed with bromophenol blue loading dye (recipe in Appendix A1) and analysed by SDS-PAGE gel electrophoresis (see 2.3.3).

2.3.3 SDS-PAGE

20 μl of mixtures of 0.5-15 μl of protein and 5 μl of bromophenol blue protein loading dye (4X, recipe in Appendix A1) were heated at 95 ℃ for 10 minutes. The protein mixtures were loaded onto Bolt Bis-Tris Plus polyacrylamide gel (4-12%, 15 wells, Novex). Low molecular weight (LMW) protein ladder (NEB) was used as a size marker. The proteins were separated using 1X MES or 1X MOPS buffer (recipe in Appendix A1) with constant voltage (140-150 V) for 40-60 min using Bolt Mini Gel Tank (Novex). The gel was stained by Coomasie blue protein staining solution (recipe in Appendix A1) for an hour and then de-stained using distilled water overnight.

2.3.4 Nickel affinity purification

Nickel affinity column (5 ml, Ni-NTA column from Qiagen or Histrap FF column from GE Healthcare) was equilibrated before sample loading with 50 ml buffer A, 50 ml buffer B, and 50 ml buffer A, in that order. Buffer A was composed of 20-25 mM imidazole in a buffer containing 50 mM HEPES or Tris-HCl pH 7.5-8.0, 150-500 mM NaCl, and 200 μM $ZnCl_2$ for metallo-enzymes. Buffer B was composed of 250-500 mM imidazole in a buffer containing same reagents as buffer A. The protein was loaded on the column and washed with 100 ml buffer A. The bound protein was eluted with buffer B using an AKTA Pure HPLC purification system (GE Healthcare). After use, the column was washed with 30 ml buffer A, 50 ml water, 50 ml 0.2 M NaOH, 50 ml water, and 20% EtOH in order and stored at 4 ℃.

2.3.5 Dialysis and desalting

The samples in buffer B were exchanged to other buffers using SnakeSkin Dialysis Tubing (10K MWCO, Thermo) or using HiPrep 26/10 Desalting column (GE Healthcare).

Dialysis

Protein samples were transferred SnakeSkin Dialysis Tubing (Thermo) and equilibrated into 2,000-3,000 ml buffer with 200-500 rpm stirring using a magnetic stirrer (VWR) at 4 ℃. The buffer was exchanged to fresh buffer and equilibrated for a further 3-5 hours to remove imidazole further. The buffer was exchanged to fresh buffer and equilibrated for a final 2-3 hours.

Desalting

A HiPrep 26/10 desalting column (GE Healthcare) was equilibrated with 2 column volumes (CV) of the elution buffer before use. The sample (1-15 ml) was loaded onto desalting column using a peristaltic pump (Gilson) with a flow rate of 5-10 ml/min. The protein fractions were eluted using an AKTA FPLC purification system (GE Healthcare) and protein concentration was measured by NanoDrop ND-1000 Spectrophotometer (NanoDrop, A_{280} nm) using an extinctuion coefficient calculated by ProtParam (http://web.expasy.org/protparam/) and the purity of the protein was estimated by SDS-PAGE (see Appendix A4 for extinction coefficients of all proteins used in this thesis).

2.3.6 TEV protease

Expression and purification of TEV protease

The plasmid DNA for TEV protease (10 ng, pET-17b) was transformed into 50 µl of *E.coli* BL21(DE3) electrocompetent cells. The cells were transferred to 1 ml LB and incubated for an hour at 37 °C with 200 rpm orbital shaking (VWR). 50-100 µl of the LB culture was inoculated onto an LBA plate and incubated overnight at 37 °C. A colony was transferred to 10 ml of LB supplemented with ampicillin 100 µg/ml and incubated overnight at 37 °C at 200 rpm orbital shaking. The LB seed culture (10 ml) was inoculated into 1000 ml LB supplemented with ampicillin (final concentration 100 µg/ml) and incubated for 3-5 hours at 37 °C at 200 rpm. IPTG (final concentration 1 mM) was added to the LB culture when an OD_{600} of 0.6-0.8 was reached. Further incubation was performed overnight at 30 °C at 200 rpm. The culture was centrifuged at 4 °C at 5,000 rpm for 15 minutes. The cell pellets were resuspended in buffer A (25 mM sodium phosphate pH 8.0, 10% (v/v) glycerol, 0.2 M NaCl, and 25 mM imidazole) and 0.5 µl of turbonuclease (Sigma) was added and incubated for 30 minutes at 4 °C. Cells were lysed using a French Press (Thermo) or sonication at 50% power for 6 min twice using a Ultrasonic homogenizer (OMNI sonic ruptor 400). The lysate was centrifuged at 13,500 rpm for 60 min using Hitachi High-Speed Refrigerated Centrifuge (R15A rotor). The nickel affinity column (5 ml, Qiagen or GE health care) was washed with buffer B and equilibrated with 50-100 ml buffer A before use. The lysate was loaded on the column with at a 5 ml/min flow rate using a peristaltic pump (Gilson). The column was washed using 100 ml of buffer A and the bound protein was eluted using buffer B (25 mM sodium phosphate pH 8.0, 10% (v/v) glycerol, 0.2 M NaCl, 500 mM imidazole). Buffer B, containing TEV protein, was exchanged to TEV storage buffer (25 mM sodium phosphate pH 8.0, 10% (v/v) glycerol, 5 mM β-mercaptoethanol, 0.2 M NaCl) using a HiPrep 26/10 desalting column (GE Healthcare). The final sample purity containing TEV protein was evaluated using SDS-PAGE and flash-frozen using a mixture of dry ice and ethanol (70-100% purity, Merk) and stored at - 80℃.

His-tag removal using TEV protease

Engineered TEV protease with a His$_6$-tag (4) was produced and used to cleave the His-tag from target proteins. TEV protease is highly specific for the amino acid sequence ENLYFQ↓G/S (4, 6, 7). The proteins purified by nickel affinity chromatography were buffer exchanged to the TEV protease reaction buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM EDTA, 0.5 mM DTT) using a desalting column. The His-tag cleavage was performed by adding 1-2 mg of TEV protease to 10 mg of the proteins. Reactions were carried out at 4 °C for 5 days. The efficiency of the His-tag cleavage was determined by SDS-PAGE. The proteins, if more than 80% of the His-tag was cleaved, was exchanged to buffer A. The sample in buffer A was loaded onto the pre-equilibrated nickel affinity column and the flowthrough was collected. The purity of the target protein in the flowthrough was estimated by SDS-PAGE. The example of the protein purification steps using TEV protease (NDM1-R10 case) is shown in Figure 2.4.

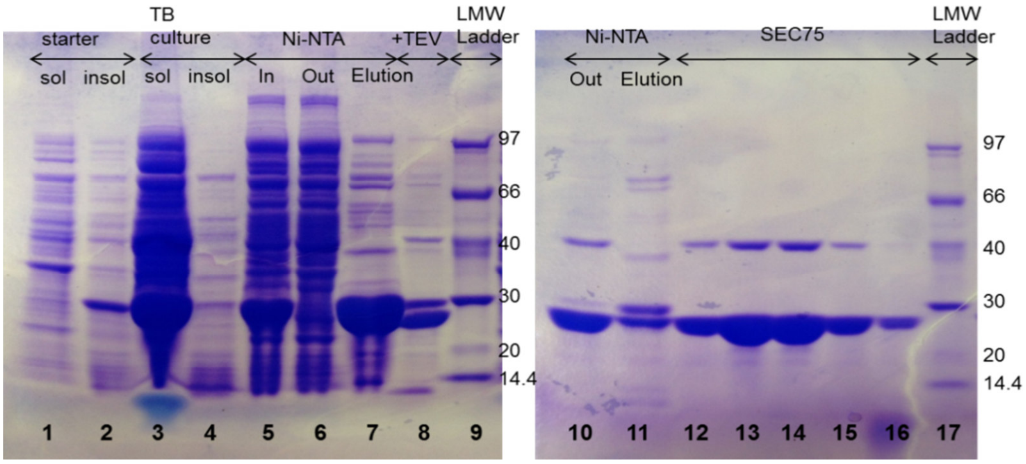


Figure 2. 4 | Example of protein purification steps including a TEV cleavage step in the case of NDM1-R10 purification. One colony of *E.coli* BL21(DE3) bacteria, containing NDM1-R10 protein, was inoculated into 10 ml LB supplemented with kanamycin antibiotics and incubated at 37 °C overnight and a 1 ml fraction was taken to check the expression and the solubility (**1, 2**). 'Sol' indicates the soluble fraction and 'insol' indicates the insoluble fraction. 1 ml of the LB culture was inoculated into 1000 ml of TB, supplemented with kanamycin antibiotics (35 µg/ml) and 1% glycerol, and incubated at 30 °C for 6 hours and then 22 °C for 16 hours (**3, 4**). The cells were sonicated and centrifuged. The filtered lysate (**5**, In) was loaded on His Trap FF column (5 ml, GE Healthcare). The flow-through was monitored (**6**, Out) to measure the efficiency of the column. The bound protein was eluted with 250 mM imidazole (**7**) followed by buffer exchange to TEV protease reaction buffer using Hi-Prep Desalting column (GE Healthcare). 20% TEV protease (w/w) relative to the protein was added to the protein solution and incubated at 4 °C for 5 days (**8**, ~80% cleaved). The protein solution, including TEV protease was buffer exchanged to the nickel affinity column binding buffer. The flow-through from the nickel affinity column (**10**) was collected for size exclusion chromatography using a HiLoad 16/600 Superdex 75 pg column (GE Healthcare, **12-16**). Bound TEV protease and protein fragments, attached to the His-tag, were eluted (**11**). Low molecular weight (LMW) ladder (NEB) was used as a size marker (**9, 17**).

2.3.7 Size exclusion chromatography

A HiLoad 16/600 Superdex 75 pg column (GE Healthcare) was equilibrated with 1.5 column volumes (CV) of the elution buffer (detailed composition of each elution buffer for specific protein is presented in each chapter) using an AKTA FPLC purification system (GE Healthcare). Proteins were concentrated to 0.5-2 ml using Amicon Ultra-15 centrifugal filter 10KDa (EMD Millipore) and loaded onto the column using a peristaltic pump (Gilson). The protein concentration and purity were determined by NanoDrop at 280 nm (ND-1000 spectrophotometer) and SDS-PAGE gel electrophoresis (Novex). Extinction coefficients of proteins were calculated using ProtParam website (http://web.expasy.org/protparam/). Extinction coefficients of proteins are presented in Appendix A4.

2.4 Crystallography

This section explains the general protein crystallization protocols used in this thesis. Milli-Q water was used for protein purification and crystal screens. The purified proteins were eluted in crystallization buffer (100-200 μM $ZnCl_2$ was included for metalloenzymes such as AiiA, NDM1, and VIM2 variants). Detailed crystallization buffer compositions are presented in the method section of individual chapters in this thesis. The concentration of protein used for crystal screening was 2.5-25 mg/ml depending on the property of the proteins (e.g. solubility). The protein samples were centrifuged at 12,000 rpm for more than 10 minutes to remove possible amorphous protein and particles before using in crystal screen. Protein purity was estimated by SDS-PAGE which showed more than 95% purity for all proteins used for crystallization screens.

2.4.1 Crystallization and data collection

To find initial crystallization conditions of proteins, we used 96 reagent crystallization screen kits from Hampton Research. 50 μl of the crystallization reagents of Grid Screen Salt HT, Crystal Screen HT, Index, PEG/Ion HT, PEGRx HT, and SaltRx HT were transferred to the reservoir of the 96 well sitting drop vapor diffusion plate (Corning). The total volume of the reservoir was 50 µl. 0.5-2 µl of the reservoir solution was transferred to a sitting drop followed by the transfer of the 0.5-1 µl of protein samples (2.5-25 mg/ml) to the reservoir of the sitting drop. This transfer process was performed using an 8 channel pipette (Thermo Scientific F1-ClipTip Pipettes) or a Mantis liquid dispenser (Formulatrix). The crystallization plates were then sealed by a clear sealing tape. The plates were placed in a constant temperature room (4 ℃ or 18 ℃). Crystal growth was checked using an Olympus SZ40 stereo zoom microscope. Crystals appeared between 2 days-4 weeks.

Once a crystallization condition was found, the condition was optimized by gradually changing pH, PEG concentration and salt concentration using crystallization plates (EasyXtal 15-well tools or Crystalgen 24-well super clear plates). Further optimization was performed in some cases to find bigger crystals by varing the ratio of the reservoir and protein (reservoir: protein sample (µl) = 1:1, 2:1, 2:2, 4:2). Micro-seeding was used in some cases to produce better quality crystals. For microseeding, a single crystal or crystals with a drop (1-5 µl) were crushed by a pipette tip and serial diluted to 10^5-10^6 times using a reservoir solution. The solution containing crystal fragments was transferred to the crystallization drop of the crystal screen plate instead of using the reservoir solution. Then the protein sample (2.5-25 mg/ml) was transferred to the drop. A good example of micro-seeding to improve the quality of crystals is shown in Figure 2.5**a**.

Crystal soaking into substrate containing solution was performed to obtain product bound structures or to observe structural changes. Crystals were soaked into a crystallization solution including additional 1-15 mM substrate for 30 second-1 hour. Crystals grown in salt or low concentration of cryo-protecting reagents (<35%) were briefly soaked into a cryo-buffer solution containing additional cryo-protectant such as glycerol, 2-methyl-2,4-pentanediol, or polyethylene glycol (PEG) . Cryo-cooling of protein crystals during an X-ray diffraction experiment eliminates the diffusion of radiolytic species, such as free radicals from causing additional damage within the crystal (11). The crystals were quickly mounted on a MarµX goniometer (1.5418 Å), located at the Australian National University to collect two dimensional images (90-180) by 1° rotation per image. An image plate Mar scanner (345 mm plate) was used. In the case of using Synchrotron beam sources, the crystals were vitrified using liquid nitrogen to be delivered to Australian Synchrotron in Melbourne for data collection on the MX1 or MX2 beamlines (0.9537 Å) (12). CCD detector (ADSC QUANTUM 210r) was used. The simplified process of crystal screening, micro-seeding, and mounting of a crystal onto the machine is shown in Figure 2.5.

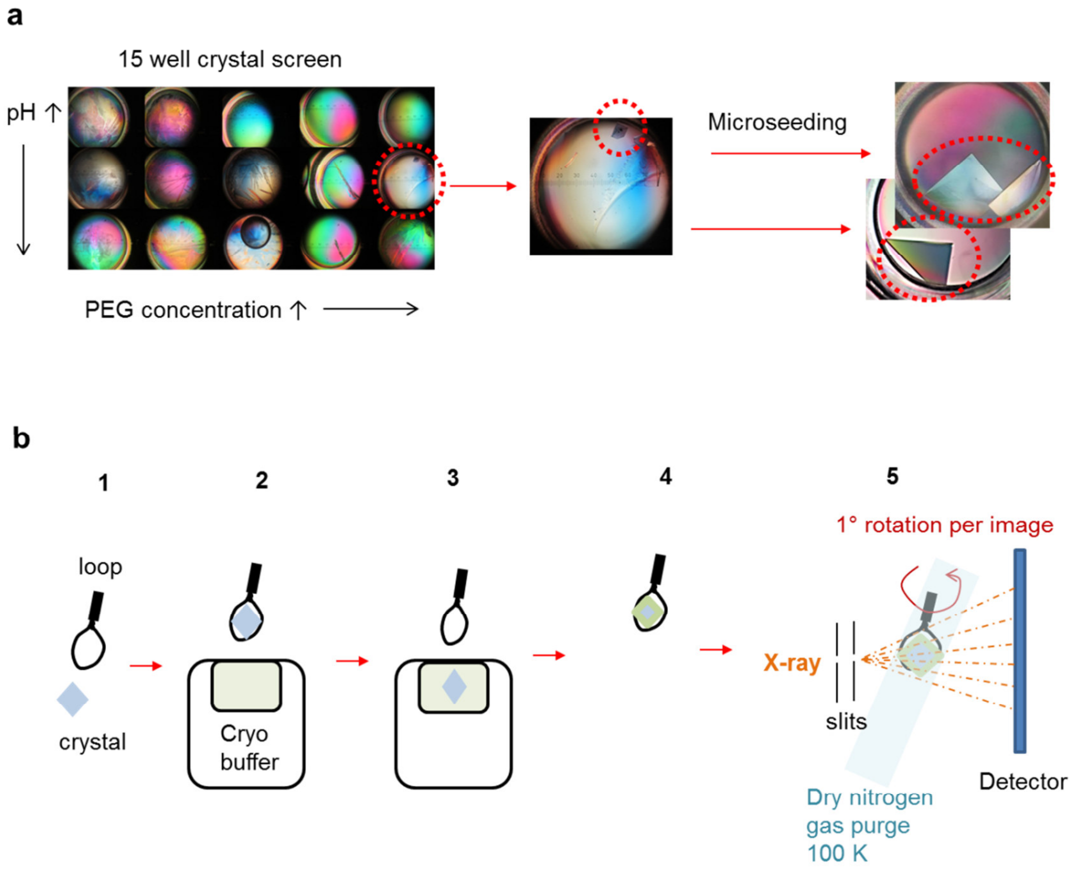


Figure 2. 5 | Example of micro-seeding and scheme of pre-treatment for X-ray diffraction. (**a**) The quality of the best crystal of MSMEG3380 from optimization screen was improved by micro-seeding. (**b**) A crystal was taken using a similar size loop (**1-2**) and briefly soaked into cryo-buffer (**3**) which contains additional cryo-protectant such as glycerol or PEG. Then, the crystal was taken again using the loop (**4**) before mounted to the X-ray diffraction machine (**5**). Diffraction images were taken at 1° intervals from 0°–180° and then merged to obtain a full dataset used for structure determination (Figure 2.6).

2.4.2 Data processing

The diffraction images were indexed and integrated by XDS (13) or iMosflm (14). The data were merged and scaled using Aimless in the CCP4 program suite (15). The resolution limit was decided using the overall half-dataset correlation CC(1/2) > 0.5 (16). Phases were calculated with the program MOLREP (17) in the CCP4 program suite using a model based on the sequence and previous structures. The model was manually improved using COOT 0.7 (18), based on the electron density map and then refined using Refmac5.7 in the CCP4 program suite (15), or phenix.refine (19). The simplified process of data processing is shown in Figure 2.6.

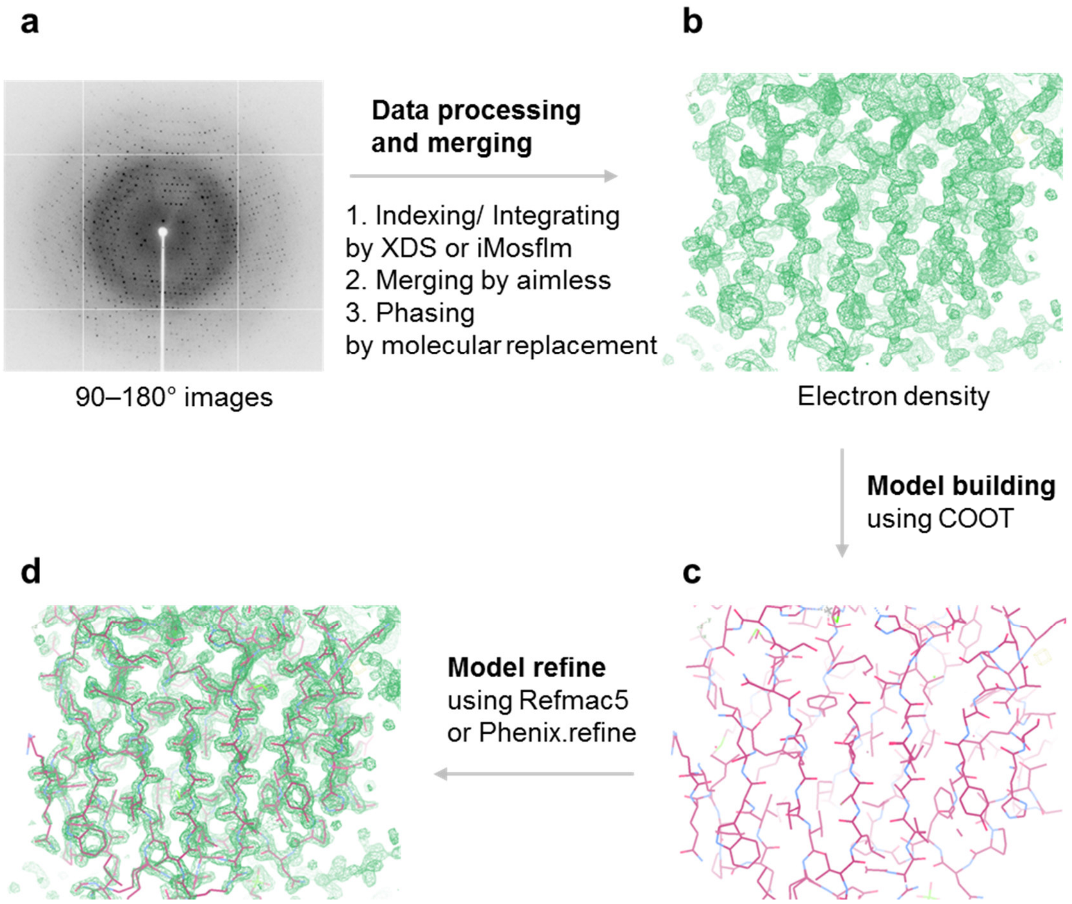


Figure 2. 6 | Example of data processing, model building and structure refinement in X-ray crystallography. X-ray diffraction images (**a**) of protein crystals were collected in different angles (1° rotation per image) using MarμX at Australian National University or using MX1 or MX2 beam sources at Australian Synchrotron (Melbourne). 90-180° images were indexed and integrated using XDS or iMosflm programs and merged using aimless (**b**). Previously existing structures were used to phase the data and obtain an initial electron density map. The resulting model structure was refined using Phenix and Refmac5 to fit the model into the electron density better (**c**-**d**).

2.5 References

1. Gibson DG, et al. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6(5):343–345.
2. Sutcliffe JG (1979) Complete nucleotide sequence of the Escherichia coli plasmid pBR322. *Cold Spring Harbor Symposia on Quantitative Biology* (Cold Spring Harbor Laboratory Press), pp 77–90.
3. Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW (1990) [6] Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol* 185:60–89.
4. Cabrita LD, et al. (2007) Enhancing the stability and solubility of TEV protease using in silico design. *Protein Sci* 16(11):2360–2367.
5. Phan J, et al. (2002) Structural basis for the substrate specificity of tobacco etch virus protease. *J Biol Chem* 277(52):50564–50572.
6. Parks TD, Howard ED, Wolpert TJ, Arp DJ, Dougherty WG (1995) Expression and purification of a recombinant tobacco etch virus NIa proteinase: biochemical analyses of the full-length and a naturally occurring truncated proteinase form. *Virology* 210(1):194–201.
7. Miladi B, et al. (2011) A new tagged-TEV protease: Construction, optimisation of production, purification and test activity. *Protein Expr Purif* 75(1):75–82.
8. Stock JB, Surette MG (1996) Escherichia coli and Salmonella: cellular and molecular biology. *Washington, DC Am Soc Microbiol*.
9. Gibson DG, Smith HO, Hutchison III CA, Venter JC, Merryman C (2010) Chemical synthesis of the mouse mitochondrial genome. *Nat Methods* 7(11):901–903.
10. Studier FW (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Expr Purif* 41:207–234.
11. Kmetko J, Warkentin M, Englich U, Thorne RE (2011) Can radiation damage to protein crystals be reduced using small-molecule compounds? *Acta Crystallogr Sect D Biol Crystallogr* 67(10):881–893.
12. McPhillips TM, et al. (2002) Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines. *J Synchrotron Radiat* 9(6):401–406.
13. Kabsch W (2010) Xds. *Acta Crystallogr Sect D Biol Crystallogr* 66(2):125–132.
14. Battye TGG, Kontogiannis L, Johnson O, Powell HR, Leslie AGW (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr Sect D Biol Crystallogr* 67(4):271–281.
15. Winn MD, et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr Sect D Biol Crystallogr* 67(4):235–242.
16. Karplus PA, Diederichs K (2012) Linking crystallographic model and data quality. *Science* 336(6084):1030–1033.
17. Vagin A, Teplyakov A (1997) MOLREP: an automated program for molecular replacement. *J Appl Crystallogr* 30(6):1022–1025.
18. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr Sect D Biol Crystallogr* 60(12):2126–2132.
19. Afonine P V., et al. (2012) Towards automated crystallographic structure refinement with phenix. refine. *Acta Crystallogr Sect D Biol Crystallogr* 68(4):352–367.

Chapter 3: Kinetic Analysis of KE07 Variants

3.1 Introduction

Enzyme design can be used medically and commercially in many ways (1–7). Some enzymes such as triosephosphate isomerase and beta-lactamases have the ability to catalyse reactions up to the diffusion limit (8, 9). The massive catalytic power of enzymes originates from pre-organisation of their active sites, which reduces reorganization energy compared to the non-enzymatic reactions (10, 11). In current design methods, the designed active site is based on the transition state and is then combined with a compatible and stable scaffold to create a new enzyme (12). The major problem of enzyme design is the low turnover rates of most designed enzymes. However, the low turnover rates of designed enzymes have been improved by directed evolution (13–17). These evolutionary trajectories provide detailed illustrations of the routes taken to increase the k_{cat} of the designed proteins *via* changes in their structure and dynamics (18). The *de novo* designed enzyme KE07 showed a gradual increase in k_{cat} throughout evolution, with about one mutation accumulated per round (13). Mutations in later rounds (R5-R7-2) in particular occurred outside the active site.

In this chapter, the relationship between the kinetic parameters and protein sequence is investigated by thermodynamic analysis and primary kinetic isotope effects of KE07 evolutionary variants, as determined by Arrhenius plots (19, 20).

Arrhenius parameters and energetics

The activation energies (E_{a}) and the pre-exponential factors (A) were obtained from Arrhenius plots. The plot of ln(k) against 1/T gives a straight line with slope –E_{a}/R and y-intercept ln(A), according to the linear form of the Arrhenius equation (Equation 1).

$$\ln(\boldsymbol{k}) = \ln(\boldsymbol{A}) - \frac{\boldsymbol{E_a}}{\mathbf{R}}\left(\frac{\mathbf{1}}{\mathbf{T}}\right) \tag{1}$$

where A, R, k, and T are pre-exponential factor, gas constant, reaction rate, and temperature in Kelvin, respectively. The activation energy E_a was calculated from the slope of the Arrhenius plot. A is calculated from the y-intercept of the plot. Because in reality, the plot is not linear outside the measured temperature range due to protein degradation at high or low temperatures, extrapolation is needed to obtain the collision factor A. According to Eyring's interpretation, A depends on the transmission coefficient κ and the activation entropy ($\Delta S^{\ddagger}$) (Equation 2) (20). Eyring states that "κ is just the ratio of systems crossing the barrier to systems reacting, and for many simple reactions has approximately the value unity" (20).

$$\boldsymbol{A = \kappa \frac{K_B}{h} T e^{\Delta S^{\ddagger}/R}} \qquad (2)$$

where K_B is the Boltzmann constant and h is the Planck constant. If we assume that κ is unity, the relationship between the Gibbs free energy $\Delta G^{\ddagger}$, activation enthalpy $\Delta H^{\ddagger}$ (~E_a), and activation entropy $\Delta S^{\ddagger}$ is given by Equation 3 (21).

$$\boldsymbol{k = \frac{k_B}{h} T e^{-\Delta G^{\ddagger}/RT} = \frac{k_B}{h} T e^{-\Delta H^{\ddagger}/RT} e^{\Delta S^{\ddagger}/R} \ (\because \kappa = 1)} \qquad (3)$$

Equations 2 and 3 show that higher ln$A$ from the Arrhenius equation is due to higher κ or $\Delta S^{\ddagger}$ and results in reduction of $\Delta G^{\ddagger}$.

To sum up, a lower activation energy (E_a), or higher pre-exponential factor (ln$A$), corresponds to a higher reaction rate or lower activation free energy ($\Delta G^{\ddagger}$). E_a is roughly equivalent to the potential energy of activation ($\Delta H^{\ddagger}$), and the pre-exponential factor (maximum frequency at infinite temperature) depends on the activation entropy ($\Delta S^{\ddagger}$, if κ = 1 by Equation 3). The case of thermophilic alcohol dehydrogenase provides an example of what information can be obtained from Arrhenius parameters. The Arrhenius parameters of this enzyme show decreasing activation energy at high temperature (23.6 kcal/mol at 303 K and 14.6 kcal/mol at 333 K) with temperature-independent activation free energy (22). The reduction of activation energy was interpreted as evidence for significant hydrogen tunnelling, which was thought to be a result of increased gating motions of the protein at high temperature. However, Villa and Warshel offered enthalpy-entropy compensation as an alternative explanation; that is, more fluctuations of the polar reactant state - compared to the TS at higher temperatures led to a decrease in activation entropy

($\Delta S^{\ddagger}$). Enthalpy-entropy compensation has been shown in non-enzymatic systems as well (23–25).

3.2 Methods

3.2.1 Purification

KE07 variants, cloned into the pET-29b(+) vector (Invitrogen), were expressed with C-terminal His_6-tags in *Escherichia coli* BL21(DE3) cells (Invitrogen). The cell pellet was resuspended in buffer A (50 mM Tris-HCl pH 8.0, 100 mM NaCl, 20 mM imidazole) and lysed by French pressure cell press (Thermo). The soluble fraction was loaded onto a Ni-NTA column (Qiagen) and elution was achieved with buffer B (50 mM Tris-HCl pH 8.0, 100 mM NaCl, 250 mM imidazole). After extensive dialysis against buffer (HEPES-NaOH 25 mM, pH 7.25, 100 mM NaCl), protein concentrations were determined using a NanoDrop spectrophotometer (Thermo) at 280 nm and SDS-PAGE gel electrophoresis (Figure 3.1). Extinction coefficients for each of the KE07 variants at 280 nm were calculated using ProtParam (26), and were listed in Appendix A4. The proteins were concentrated (10-30 mg/ml) and stored at 4 °C.

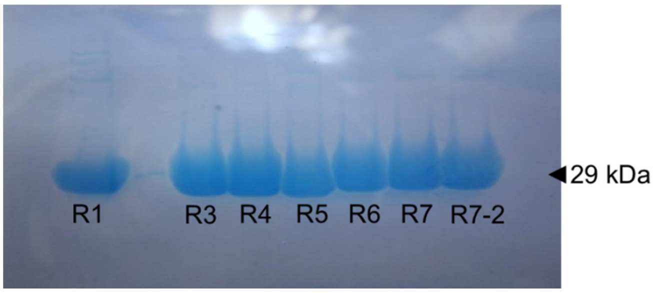


Figure 3. 1 | SDS-PAGE of KE07 variants. 6 µl of the purified proteins were mixed with 2 µl of sample reducing agent (10X) and 5 µl of SDS sample buffer (4X, Novex) and 7 µl of MQ water. 20 µl of the protein mixture was heated at 95 °C for 5 min and loaded onto the gel. The samples were run in 1X MES buffer with 100 mA current for 40 min. The gel was stained with Coomassie Blue dye and de-stained overnight with distilled water.

3.2.2 Preparation of the H/D substrates

5-Nitrobenzisoxazole was prepared from 1,2-benzisoxazole (Sigma, >99%) following the published protocol (27) with help from Prof. Chris Easton, Charlotte Rouse, Daniel Bartkus and William Zhang at ANU (Figure 3.2).

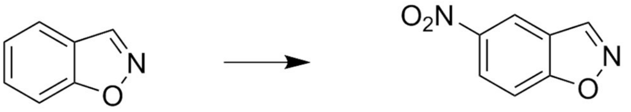


Figure 3. 2 | 5-Nitrobenzisoxazole synthesis from 1,2-benzisoxazole.

3-Deuterio-5-nitrobenzisoxazole was synthesized by Prof. Chris Easton and Dr. Timothy Altamore at ANU from 2-bromophenol following the published procedure (Figure 3.3) (28, 29).

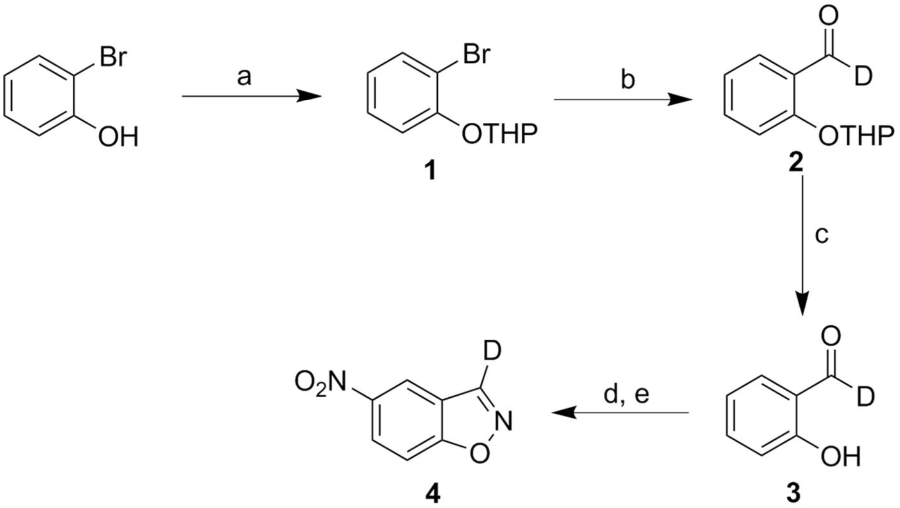


Figure 3. 3 | Synthesis of 3-deuterio-5-nitrobenzisoxazole from 2-bromophenol. *Reagents and conditions:* a) 2,3-Dihydropyran, pyridinium *p*-toluenesulfonate, CH_2Cl_2, *86%*; b) 1) 2.2M n-BuLi in hexanes, Et_2O, 0 °C, 2) [D^7]-DMF in Et_2O, rt, *94%*; c) 1M aq. DCl in D2O, THF, *75%*; d) 1) hydroxylamine-*O*-sulfonic acid, Et_2O, 2) aq. $NaHCO_3$; e) HNO_3/H_2SO_4, 0 °C, *28% over 2 steps*.
*** Figure 3.3 is from Prof. Chris Easton and Dr. Timothy Altamore, Research School of Chemistry, Australian National University.**

3.2.3 Kinetics

Kemp elimination of 5-nitrobenzisoxazole (PubChem CID: 142385) results in the formation of 2-cyano-4-nitrophenol (PubChem CID: 11116377). The product has maximum absorbance at 380 nm (molar absorption coefficient of 15,800 $M^{-1}cm^{-1}$ in water) (30). The rate of product formation was monitored *via* absorbance at 380 nm using a Cary 60 UV-Vis spectrophotometer (Agilent Technologies) or a SpectraMax-M2 Multi-mode microplate reader (Molecular Devices). Reaction rates were measured in buffer containing 25 mM HEPES pH 7.25, 100 mM NaCl at a range of temperatures (10, 20, 25, 30, 35, 40, 45, 50 °C). To increase substrate solubility 1.25% (v/v) glycerol was included in the assay buffer. Substrate stocks in acetonitrile (final concentration of acetonitrile 1-1.5% v/v) at 9-11 different concentrations (0.01-1.1 mM) were added to the enzyme-buffer mixture immediately before the measurement due to the water sensitivity of the substrates. The rate of the non-catalytic background reaction was subtracted before fitting the data for enzyme-catalysed reactions to a kinetic model. The calculated k_{cat} values at different temperatures were fitted to the linear form of the Arrhenius equation, Equation 1. The standard deviations of the slope ($-E_a/R$) and the y-intercept ($\ln A$) from the Arrhenius plots were calculated using LINEST command in the Excel software (version 2010). The standard deviations of the calculated values were estimated based on the error propagation rules, shown in Equation 4 and Equation 5.

$$\boldsymbol{X = a + b} \qquad \because \triangle \boldsymbol{X} = \sqrt{(\triangle \boldsymbol{a})^2 + (\triangle \boldsymbol{b})^2} \tag{4}$$

$$\boldsymbol{X = a \times b \;\; or \; X = \frac{b}{a}} \qquad \because \triangle \boldsymbol{X} = \boldsymbol{X} \times \sqrt{\left(\frac{\Delta a}{a}\right)^2 + \left(\frac{\Delta b}{b}\right)^2} \tag{5}$$

3.3 Results

3.3.1 Determination of the Arrhenius parameters

The 3’-H and 3’-D 5-nitrobenzisoxazole substrates (Figure 3.4) were used to measure k_H and k_D. Each measurement was repeated and the data were fitted to Michaelis-Menten curves to calculate k_{cat} and K_M. The initial velocity of the non-enzymatic reaction was linearly correlated with substrate concentration because the reactant supply (water) exceeds its consumption during the reaction. Thus, the pseudo-first order approximation was used for the non-enzymatic reaction (Figure 3.5).

The determined kinetic constants are shown in Table 3.1. $\ln k_H$ and $\ln k_D$ were plotted against 1/T, and the data were fitted to the Arrhenius equation (Eq.1) to calculate the activation energy (E_{a}) and pre-exponential factor (A) (Figure 3.6). The Arrhenius plots of k_H and k_D were overlaid for comparison and the kinetic isotope effects (k_H/k_D) were plotted next to the Arrhenius plot (Figure 3.7). Arrhenius parameters are shown in Table 3.2.

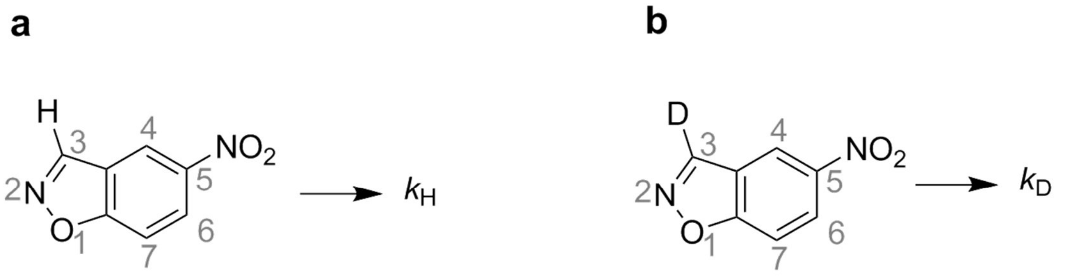


Figure 3. 4 | H/D substrates for kinetic measurements. 3’-H-5-nitrobenzisoxazole (**a**) was used for k_H measurement and 3’-D-5-nitrobenzisoxazole (**b**) was used for k_D measurement.

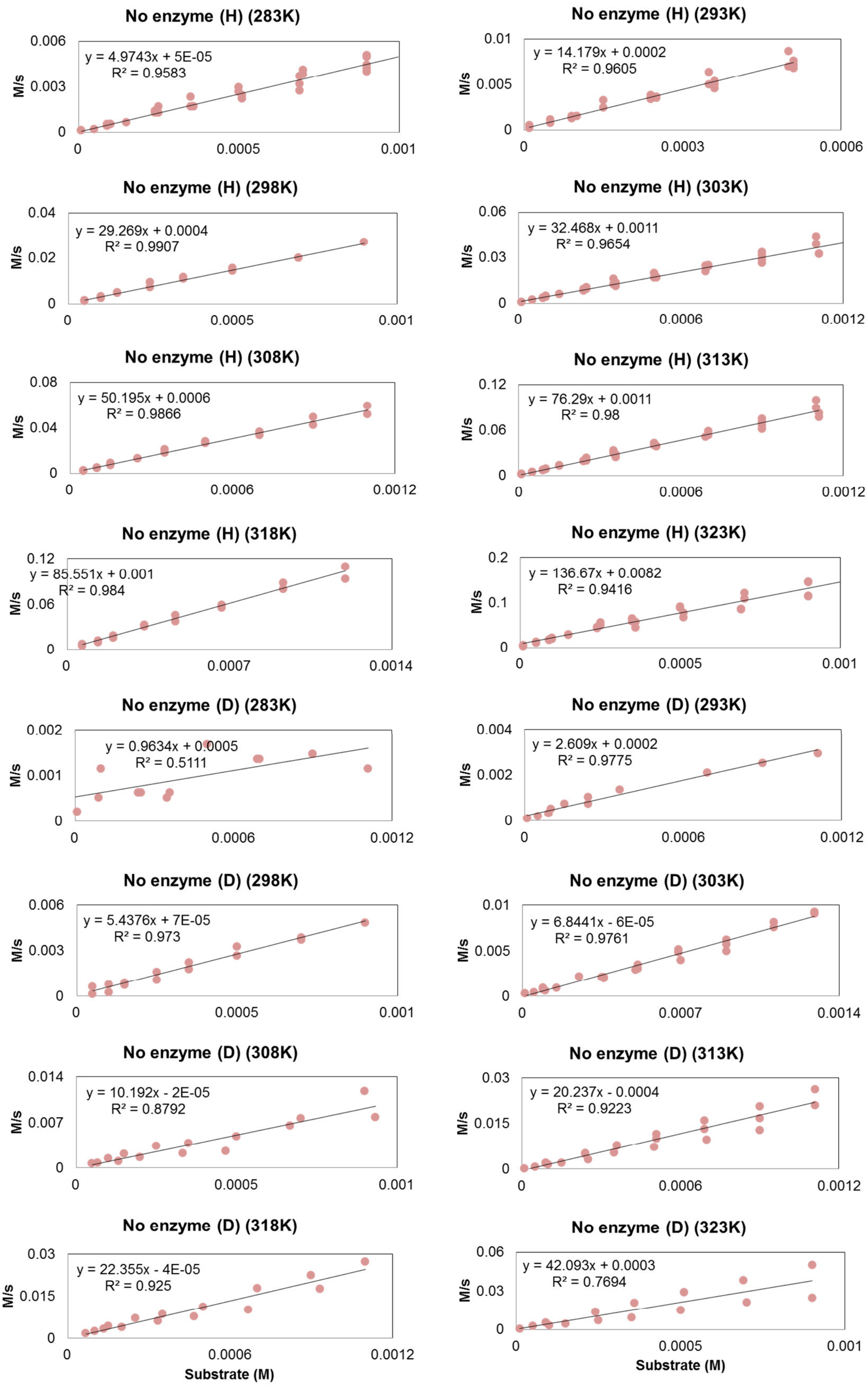


Figure 3. 5 | Pseudo-first order approximation for the non-enzymatic reactions with 3'-H/D-5-nitrobenzisoxazole in different temperatures. Each measurement was repeated 2–3 times in a temperature range 283–323 K. The data were fitted to linear regression curves to calculate k_{cat} and K_M. The equation of the fit and R^2 values for the fits are presented near each curve.

Table 3. 1-1 | Rate constants for conversion of 5-nitrobenzisoxazole to 2-cyano-4-nitrophenol for non-enzymatic and enzymatic reactions by KE07 variants (R1-R3)

	T (K)	k_H (s^{-1})	k_D (s^{-1})	K_M (H) (mM)	K_M (D) (mM)	k_{cat}/K_M (H) (s^{-1}·M^{-1})	k_{cat}/K_M (D) (s^{-1}·M^{-1})	KIE (k_H/k_D)[a]
B[b]	283	4.97E-06 ± 1.8E-07	9.63E-07 ± 2.6E-07					5.2 ± 1.4
	293	1.42E-05 ± 5.8E-07	2.61E-06 ± 1.25E-07					5.4 ± 0.3
	298	2.93E-05 ± 7.9E-07	5.44E-06 ± 2.5E-07					5.4 ± 0.3
	303	3.25E-05 ± 1.1E-06	6.84E-06 ± 2.3E-07					4.7 ± 0.2
	308	5.02E-05 ± 1.5E-06	1.02E-05 ± 1E-06					4.9 ± 0.5
	313	7.63E-05 ± 2E-06	2.02E-05 ± 1.3E-06					3.8 ± 0.3
	318	8.56E-05 ± 2.9E-06	2.24E-05 ± 1.8E-06					3.8 ± 0.3
	323	1.37E-04 ± 6.4E-06	4.21E-05 ± 6.4E-06					3.2 ± 0.5
R1	283	0.004 ± 0.0002	0.001 ± 0.0001	0.72 ± 0.05	0.50 ± 0.11	6.2 ± 0.5	1.1 ± 0.3	8.0 ± 0.9
	293	0.01 ± 0.0007	0.002 ± 0.00003	0.76 ± 0.09	0.49 ± 0.02	14.7 ± 1.9	3.5 ± 0.1	6.5 ± 0.4
	303	0.02 ± 0.0008	0.005 ± 0.002	0.96 ± 0.06	0.70 ± 0.06	24.5 ± 1.8	6.9 ± 0.7	4.9 ± 0.3
	313	0.03 ± 0.003	0.01 ± 0.0007	0.86 ± 0.12	0.58 ± 0.09	37.7 ± 6.1	14.1 ± 2.5	4.0 ± 0.5
	323	0.05 ± 0.005	0.02 ± 0.0009	0.84 ± 0.13	0.63 ± 0.07	59.3 ± 10.4	25.5 ± 3.0	3.1 ± 0.3
R3	283	0.21 ± 0.02	0.02 ± 0.004	0.33 ± 0.07	0.46 ± 0.18	634.3 ± 146	36.7 ± 17	12.3 ± 3.1
	293	0.39 ± 0.02	0.05 ± 0.004	0.35 ± 0.04	0.27 ± 0.06	1128.2 ± 145	176.6 ± 44	8.2 ± 0.8
	298	0.49 ± 0.04	0.07 ± 0.001	0.39 ± 0.08	0.32 ± 0.02	1262.1 ± 281	232.5 ± 14	6.6 ± 0.6
	303	0.77 ± 0.02	0.11 ± 0.01	0.38 ± 0.03	0.45 ± 0.06	2054.7 ± 172	251.8 ± 34	6.8 ± 0.4
	308	0.74 ± 0.03	0.12 ± 0.003	0.43 ± 0.04	0.39 ± 0.03	1691.2 ± 171	311.4 ± 23	6.1 ± 0.3
	313	1.24 ± 0.06	0.20 ± 0.01	0.45 ± 0.05	0.35 ± 0.04	2773.0 ± 338	563.7 ± 77	6.2 ± 0.5
	318	1.32 ± 0.11	0.26 ± 0.01	0.46 ± 0.09	0.50 ± 0.05	2900.1 ± 596	524.8 ± 60	5.0 ± 0.5
	323	1.67 ± 0.07	0.39 ± 0.03	0.50 ± 0.04	0.49 ± 0.08	3320.3 ± 298	798.1 ± 146	4.3 ± 0.4

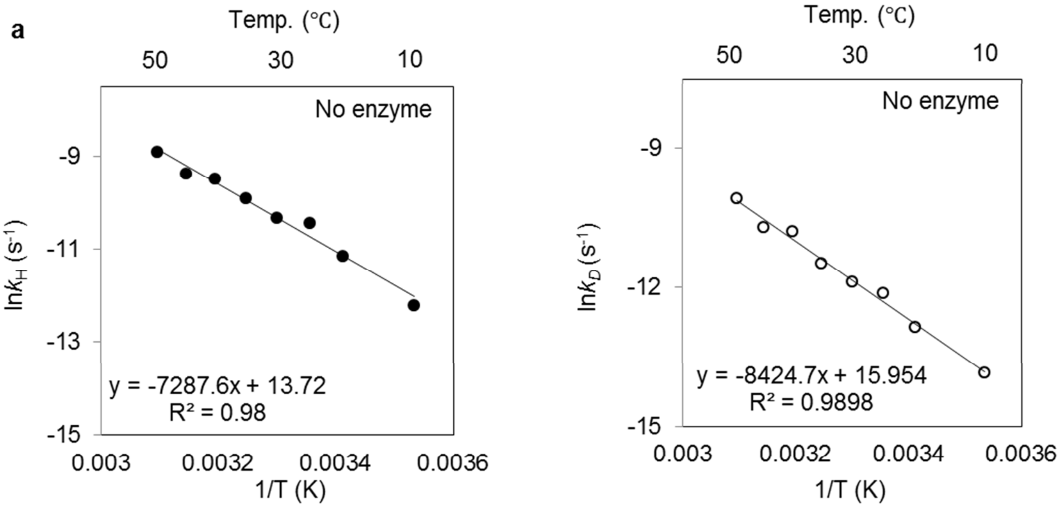
[a] KIE values were calculated directly from the rate constants from experimental values.
[b] B stands for the non-enzymatic reaction rate in buffer.

Table 3. 1-2 | Rate constants for conversion of 5-nitrobenzisoxazole to 2-cyano-4-nitrophenol for enzymatic reactions by KE07 variants (R4-R6)

	T (K)	k_H (s^{-1})	k_D (s^{-1})	K_M (H) (mM)	K_M (D) (mM)	k_{cat}/K_M (H) ($s^{-1}\cdot M^{-1}$)	k_{cat}/K_M (D) ($s^{-1}\cdot M^{-1}$)	KIE (k_H/k_D)[a]
R4	283	0.37 ± 0.05	0.04 ± 0.01	0.96 ± 0.23	0.79 ± 0.34	389.3 ± 109	53.0 ± 27	8.9 ± 2.6
	293	0.69 ± 0.05	0.12 ± 0.03	1.09 ± 0.14	1.38 ± 0.50	632.1 ± 92	85.0 ± 37	5.9 ± 1.5
	298	0.96 ± 0.13	0.13 ± 0.004	0.94 ± 0.19	1.02 ± 0.06	1075.2 ± 186	122.7 ± 8.5	6.9 ± 0.7
	303	0.96 ± 0.09	0.16 ± 0.02	1.10 ± 0.17	1.11 ± 0.23	876.6 ± 159	142.5 ± 35	6.1 ± 1.0
	308	1.07 ± 0.13	0.19 ± 0.02	0.99 ± 0.21	1.19 ± 0.14	1090.8 ± 268	162.2 ± 24	5.6 ± 0.8
	313	1.15 ± 0.07	0.24 ± 0.13	0.97 ± 0.08	1.10 ± 0.79	1187.5 ± 122	219.8 ± 195	4.7 ± 2.5
	323	1.34 ± 0.15	0.44 ± 0.14	1.10 ± 0.20	0.90 ± 0.46	1217.4 ± 257	483.6 ± 290	3.1 ± 1.0
R5	283	0.40 ± 0.03	0.05 ± 0.01	0.38 ± 0.06	0.26 ± 0.11	1071.4 ± 178	195.6 ± 86	7.8 ± 1.5
	293	0.83 ± 0.04	0.07 ± 0.004	0.44 ± 0.05	0.28 ± 0.04	1910.9 ± 243	262.7 ± 40	11.4 ± 0.8
	298	1.09 ± 0.17	0.13 ± 0.004	0.56 ± 0.15	0.28 ± 0.02	1931.8 ± 604	469.6 ± 38	8.2 ± 1.3
	303	1.43 ± 0.06	0.12 ± 0.005	0.53 ± 0.05	0.26 ± 0.02	2701.0 ± 257	464.8 ± 45	11.9 ± 0.6
	308	1.33 ± 0.12	0.17 ± 0.01	0.64 ± 0.10	0.30 ± 0.03	2080.2 ± 369	579.5 ± 61	7.6 ± 0.7
	313	2.08 ± 0.06	0.23 ± 0.02	0.57 ± 0.04	0.37 ± 0.06	3629.7 ± 249	618.2 ± 111	9.2 ± 0.7
	318	1.94 ± 0.26	–	0.68 ± 0.18	–	2844.0 ± 835	–	–
	323	2.39 ± 0.14	0.37 ± 0.03	0.62 ± 0.07	0.58 ± 0.10	3823.6 ± 505	645.4 ± 129	6.4 ± 0.7
R6	283	0.40 ± 0.04	0.05 ± 0.01	0.28 ± 0.07	0.53 ± 0.12	1431.7 ± 378	89.3 ± 22	8.4 ± 1.3
	293	0.86 ± 0.06	0.12 ± 0.01	0.40 ± 0.06	0.40 ± 0.07	2175.8 ± 359	299.3 ± 56	7.1 ± 0.7
	298	1.04 ± 0.05	0.16 ± 0.003	0.41 ± 0.04	0.42 ± 0.02	2533.8 ± 292	379.7 ± 17	6.5 ± 0.3
	303	1.56 ± 0.07	0.23 ± 0.01	0.50 ± 0.05	0.46 ± 0.06	3093.5 ± 332	500.9 ± 72	6.8 ± 0.5
	308	1.73 ± 0.14	0.28 ± 0.02	0.65 ± 0.10	0.63 ± 0.09	2678.7 ± 460	438.6 ± 71	6.2 ± 0.7
	313	2.46 ± 0.08	0.44 ± 0.01	0.61 ± 0.04	0.51 ± 0.03	4004.9 ± 311	855.4 ± 51	5.6 ± 0.2
	318	2.91 ± 0.22	0.49 ± 0.03	0.89 ± 0.14	0.68 ± 0.01	3260.6 ± 574	729.8 ± 42	5.9 ± 0.6
	323	3.21 ± 0.14	0.66 ± 0.04	0.66 ± 0.06	0.58 ± 0.08	4855.4 ± 465	1129.2 ± 172	4.9 ± 0.4

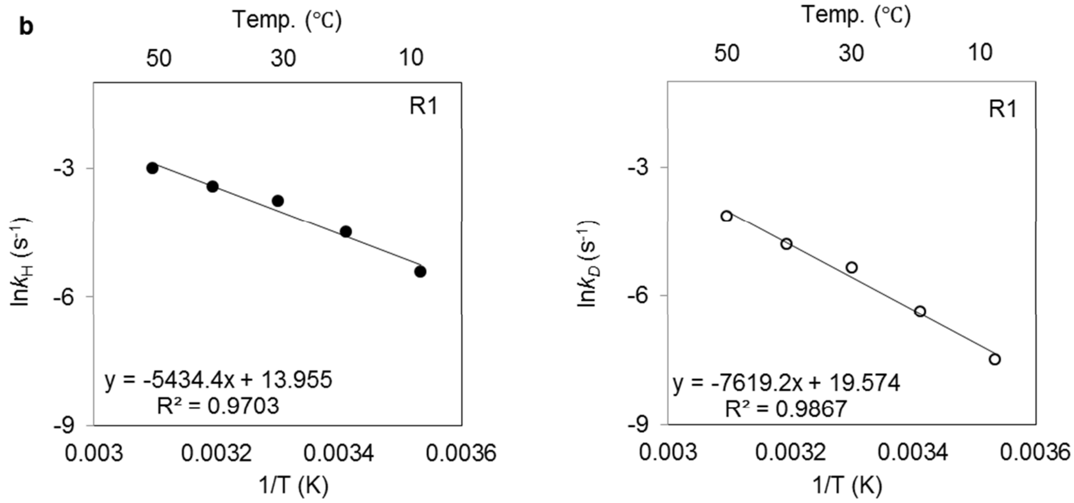
Table 3. 1-3 | Rate constants for conversion of 5-nitrobenzisoxazole to 2-cyano-4-nitrophenol for enzymatic reactions by KE07 variants (R7-R7-2)

	T (K)	k_H (s^{-1})	k_D (s^{-1})	K_M (H) (mM)	K_M (D) (mM)	k_{cat}/K_M (H) ($s^{-1}\cdot M^{-1}$)	k_{cat}/K_M (D) ($s^{-1}\cdot M^{-1}$)	KIE (k_H/k_D)[a]
R7	283	0.77 ± 0.05	0.06 ± 0.004	0.24 ± 0.05	0.23 ± 0.04	3166.6 ± 667	279.7 ± 58	12.2 ± 1.2
	293	1.54 ± 0.06	0.17 ± 0.01	0.45 ± 0.04	0.30 ± 0.04	3426.8 ± 339	581.5 ± 79	8.9 ± 0.5
	298	1.90 ± 0.11	0.29 ± 0.01	0.50 ± 0.06	0.44 ± 0.03	3793.9 ± 502	642.8 ± 44	6.7 ± 0.4
	303	2.51 ± 0.14	0.34 ± 0.02	0.58 ± 0.07	0.46 ± 0.07	4309.8 ± 568	741.1 ± 118	7.3 ± 0.6
	308	2.35 ± 0.09	0.40 ± 0.01	0.56 ± 0.05	0.56 ± 0.03	4184.6 ± 400	715.0 ± 41	5.9 ± 0.3
	313	3.10 ± 0.09	0.57 ± 0.03	0.61 ± 0.04	0.56 ± 0.06	5066.6 ± 360	1011.1 ± 114	5.4 ± 0.3
	318	3.64 ± 0.20	0.64 ± 0.02	0.81 ± 0.09	0.73 ± 0.05	4480.7 ± 533	881.9 ± 70	5.7 ± 0.4
	323	3.88 ± 0.24	0.82 ± 0.07	0.72 ± 0.09	0.78 ± 0.12	5409.8 ± 763	1044.9 ± 183	4.7 ± 0.5
R7-2	283	1.43 ± 0.06	0.11 ± 0.01	0.39 ± 0.04	0.20 ± 0.05	3714.3 ± 389	565.8 ± 170	12.9 ± 1.7
	293	2.47 ± 0.08	0.25 ± 0.01	0.40 ± 0.03	0.28 ± 0.02	6174.1 ± 520	911.0 ± 82	9.8 ± 0.4
	298	2.57 ± 0.12	0.38 ± 0.01	0.45 ± 0.04	0.40 ± 0.04	5769.7 ± 622	942.8 ± 89	6.8 ± 0.4
	303	3.83 ± 0.17	0.49 ± 0.02	0.55 ± 0.05	0.42 ± 0.03	6968.4 ± 669	1167.8 ± 102	7.9 ± 0.4
	308	3.92 ± 0.16	0.65 ± 0.02	0.57 ± 0.05	0.65 ± 0.05	6917.0 ± 690	992.3 ± 85	6.1 ± 0.3
	313	5.14 ± 0.21	0.81 ± 0.05	0.57 ± 0.05	0.51 ± 0.07	8967.5 ± 878	1577.3 ± 251	6.4 ± 0.5
	318	5.35 ± 0.26	0.99 ± 0.03	0.74 ± 0.07	0.79 ± 0.05	7236.1 ± 781	1257.5 ± 91	5.4 ± 0.3
	323	6.93 ± 0.37	1.56 ± 0.09	0.69 ± 0.07	0.90 ± 0.09	9992.8 ± 1188	1731.9 ± 194	4.4 ± 0.3



Slope [-$E_{a(H)}$/R] (K)	− 7287.6
Intercept [lnA_H]	13.72
$E_{a(H)}$ (kJ/mol)	60.6
A_H	908909

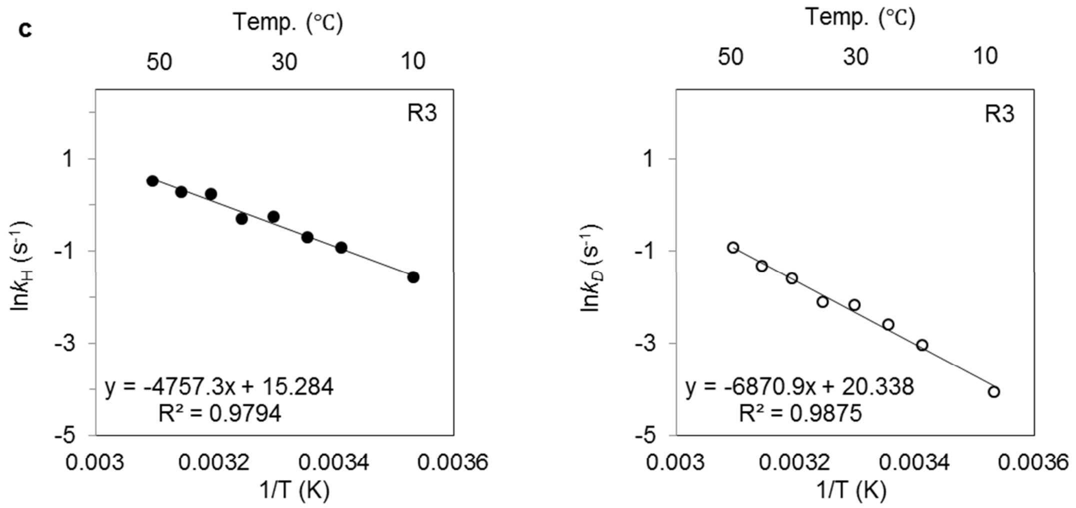
Slope [-$E_{a(D)}$/R] (K)	− 8424.7
Intercept [lnA_D]	15.954
$E_{a(D)}$ (kJ/mol)	70.0
A_D	8486608



Slope [-$E_{a(H)}$/R] (K)	− 5434.4
Intercept [lnA_H]	13.955
$E_{a(H)}$ (kJ/mol)	45.2
A_H	1149687

Slope [-$E_{a(D)}$/R] (K)	− 7619.2
Intercept [lnA_D]	19.574
$E_{a(D)}$ (kJ/mol)	63.3
A_D	316869318

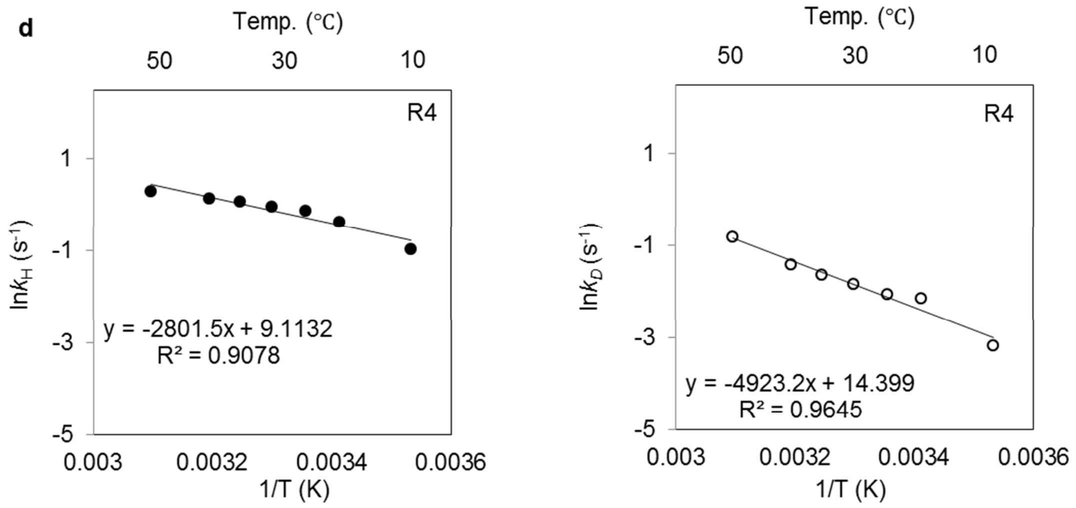
c



Slope [-$E_{a(H)}$/R] (K)	– 4757.3
Intercept [lnA_H]	15.284
$E_{a(H)}$ (kJ/mol)	39.6
A_H	4342670

Slope [-$E_{a(D)}$/R] (K)	– 6870.9
Intercept [lnA_D]	20.338
$E_{a(D)}$ (kJ/mol)	57.1
A_D	680269771

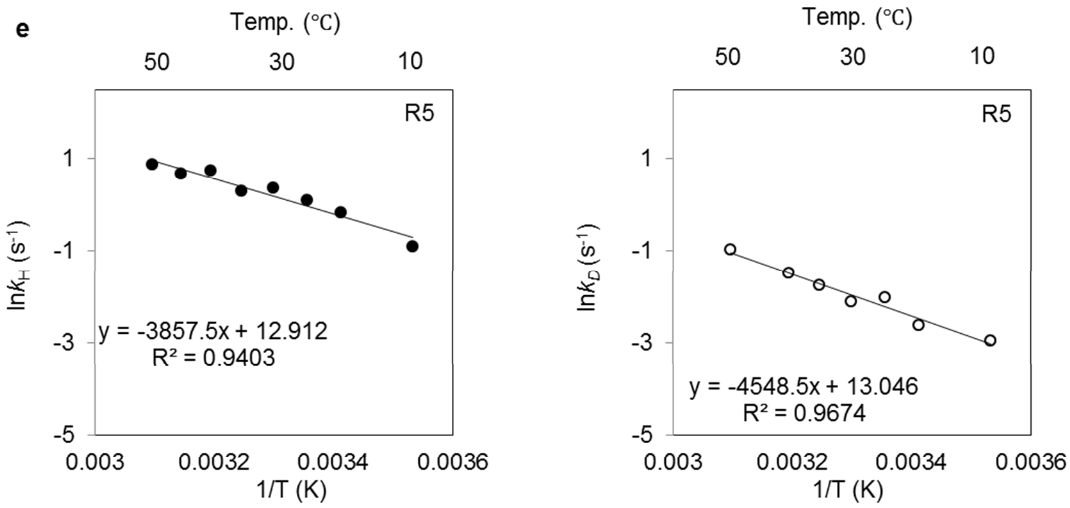
d



Slope [-$E_{a(H)}$/R] (K)	– 2801.5
Intercept [lnA_H]	9.1132
$E_{a(H)}$ (kJ/mol)	23.3
A_H	9074

Slope [-$E_{a(D)}$/R] (K)	– 4923.2
Intercept [lnA_D]	14.399
$E_{a(D)}$ (kJ/mol)	40.9
A_D	1792282

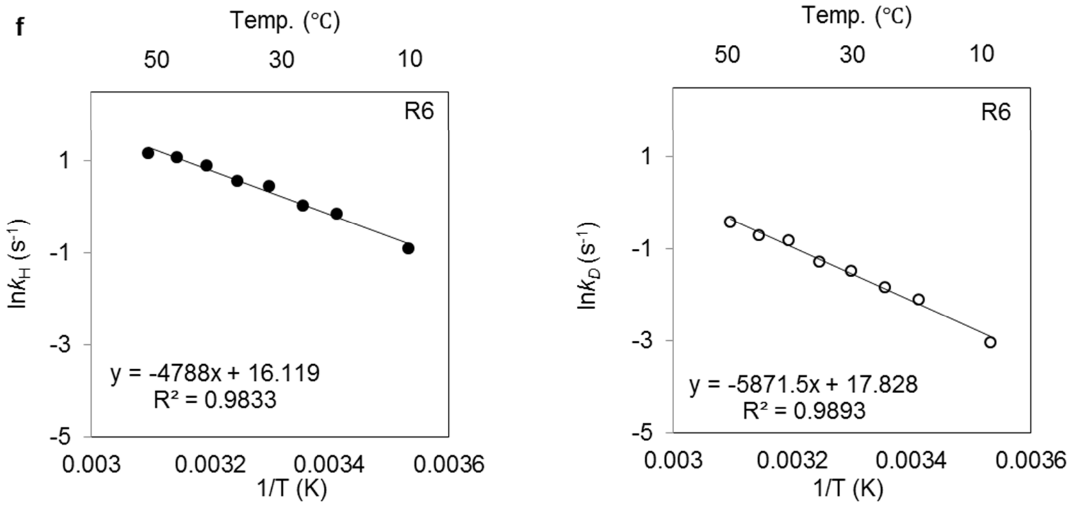
e



Slope [-$E_{a(H)}$/R] (K)	– 3857.5
Intercept [lnA_H]	12.912
$E_{a(H)}$ (kJ/mol)	32.1
A_H	405145

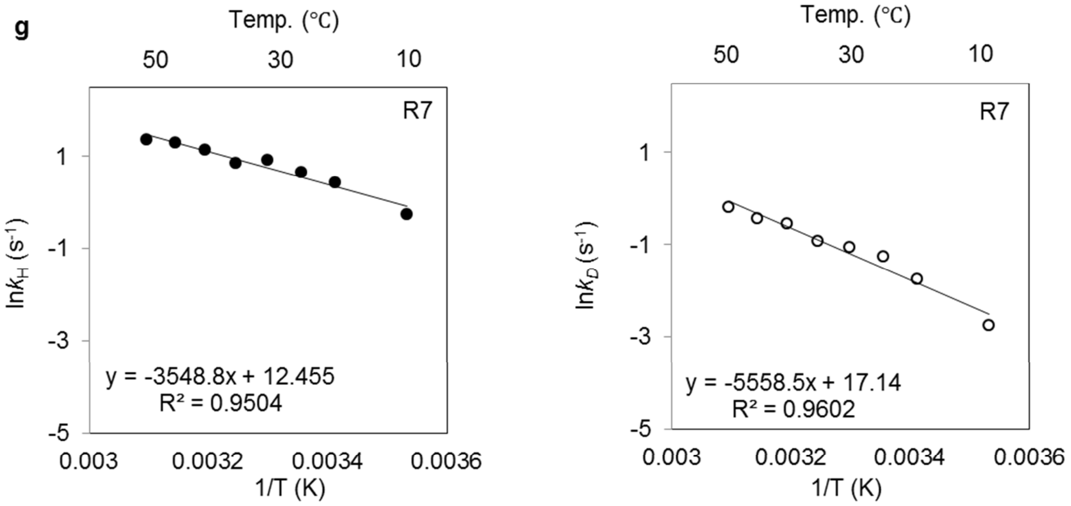
Slope [-$E_{a(D)}$/R] (K)	– 4548.5
Intercept [lnA_D]	13.046
$E_{a(D)}$ (kJ/mol)	37.8
A_D	463240

f



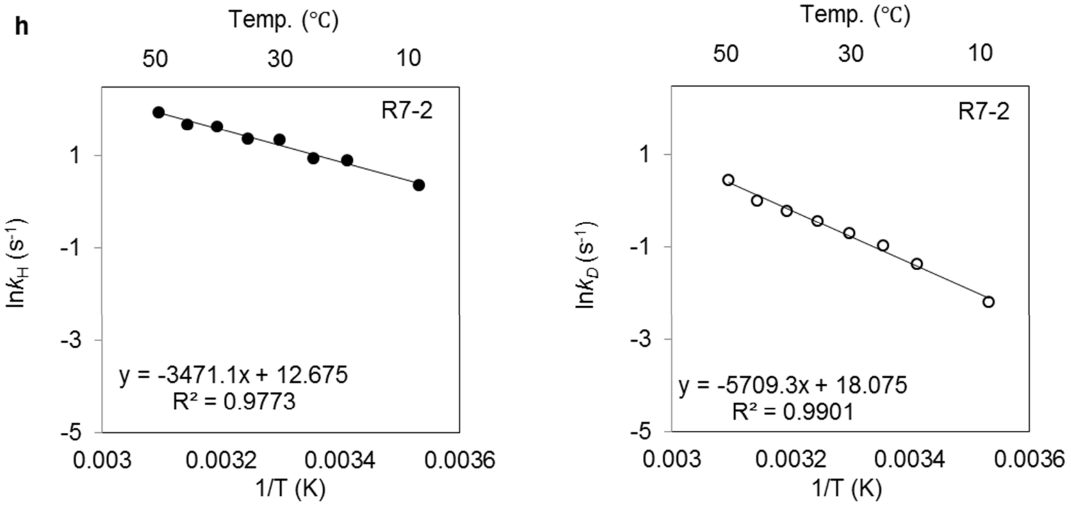
Slope [-$E_{a(H)}$/R] (K)	– 4788
Intercept [lnA_H]	16.119
$E_{a(H)}$ (kJ/mol)	39.8
A_H	10009048

Slope [-$E_{a(D)}$/R] (K)	– 5871.5
Intercept [lnA_D]	17.828
$E_{a(D)}$ (kJ/mol)	48.8
A_D	55284327



Slope [-$E_{a(H)}$/R] (K)	– 3548.8
Intercept [lnA_H]	12.455
$E_{a(H)}$ (kJ/mol)	29.5
A_H	256530

Slope [-$E_{a(D)}$/R] (K)	– 5558.5
Intercept [lnA_D]	17.14
$E_{a(D)}$ (kJ/mol)	46.2
A_D	27784809



Slope [-$E_{a(H)}$/R] (K)	– 3471.1
Intercept [lnA_H]	12.675
$E_{a(H)}$ (kJ/mol)	28.9
A_H	319656

Slope [-$E_{a(D)}$/R] (K)	– 5709.3
Intercept [lnA_D]	18.075
$E_{a(D)}$ (kJ/mol)	47.5
A_D	70773840

Figure 3. 6 | Arrhenius plots of KE07 variants - k_H and k_D. The KE07 R1 (**b**), R3 (**c**), R4 (**d**), R5 (**e**), R6 (**f**), R7 (**g**), R7-2 (**h**) reactions, and the non-catalytic-background reactions (**a**) are shown in filled and open circles, representing experimental measurements of k_H and k_D respectively. Fits to the Arrhenius equation are shown in solid lines. The equations and R^2 values for the Arrhenius fits are given for each variant. The slope (-E_a/R) and the y-intercept (ln*A*) of the fits were used to calculate activation energy (E_a) and the pre-exponential factor (*A*), which are shown under each graph.

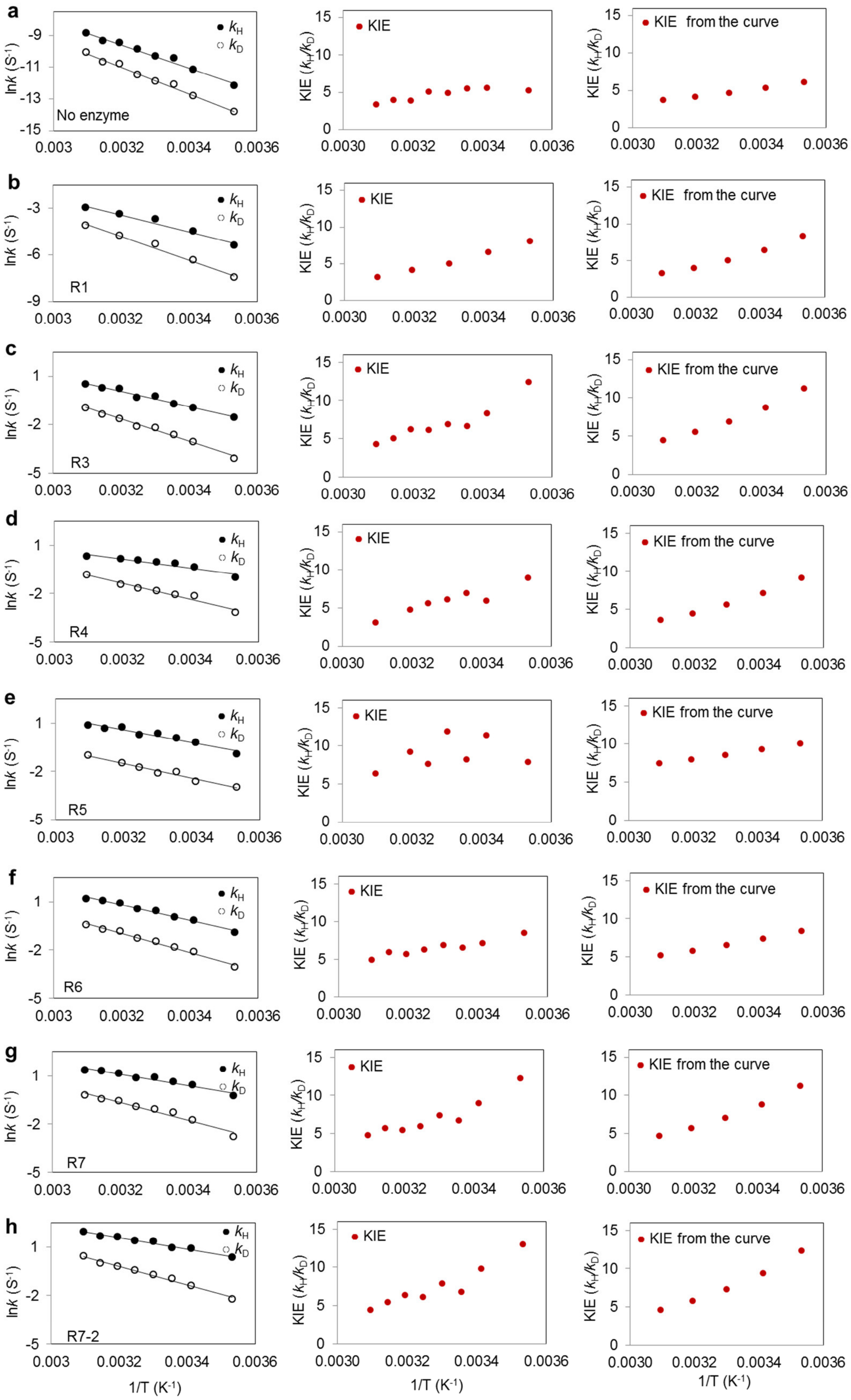


Figure 3. 7 | Arrhenius plots and KIEs. The Arrhenius plots of k_H (filled black circle) and k_D (open black circle) for each KE07 variant fitted to the Arrhenius equation (solid line). No enzyme (**a**), KE07 design R1 (**b**), R3 (**c**), R4 (**d**), R5 (**e**), R6 (**f**), R7 (**g**), and R7-2 (**h**) are shown. The KIEs directly calculated from measured points are shown in the middle and KIEs calculated from the Arrhenius equations (solid lines) are shown on the right side.

Table 3. 2 | Arrhenius parameters obtained from the curves

KE07 variant	$^{a}E_{a(H)}$	$^{a}E_{a(D)}$	$E_{a(D)} - E_{a(H)}$	$\ln A_H$	$\ln A_D$	A_H/A_D	$^{b}k_H/k_{D\ 298\ K}$
ᵇB	60.6 (3.5)	70.0 (2.9)	9.5 (4.6)	13.7 (1.4)	16.0 (1.1)	0.107 (0.01)	4.9 (0.1)
R1	45.2 (4.6)	63.3 (4.2)	18.2 (6.2)	14.0 (1.8)	19.6 (1.7)	0.004 (0.0006)	5.5 (0.2)
R3	39.6 (2.3)	57.1 (2.6)	17.6 (3.5)	15.3 (0.9)	20.3 (1.0)	0.006 (0.0005)	7.7 (0.1)
R4	23.3 (3.3)	40.9 (3.5)	17.6 (4.8)	9.1 (1.3)	14.4 (1.4)	0.005 (0.0009)	6.3 (0.2)
R5	32.1 (3.3)	37.8 (3.1)	5.7 (4.5)	12.9 (1.3)	13.0 (1.2)	0.875 (0.1)	8.9 (0.2)
R6	39.8 (2.1)	48.8 (2.1)	9.0 (3.0)	16.1 (0.8)	17.8 (0.8)	0.181 (0.01)	6.9 (0.1)
R7	29.5 (2.8)	46.2 (3.8)	16.7 (4.7)	12.5 (1.1)	17.1 (1.5)	0.009 (0.001)	7.8 (0.2)
R7-2	28.9 (1.8)	47.5 (1.9)	18.6 (4.0)	12.7 (0.7)	18.1 (0.8)	0.005 (0.0003)	8.3 (0.1)

[a] E_a is in kJ/mol, all values were calculated from Arrhenius equation from rate constant in a range of temperature 283-323 K.

[b] B stands for the non-enzymatic reaction rate in buffer.

3.3.2 Analysis of the mutations in KE07 evolution

To make sense of the kinetics, it is first necessary to provide an overview of the locations and effects at the mutations that accumulated along the trajectory based on previously published results (13, 31). The Kemp eliminase KE07 was designed using the Rosetta software (32) and random mutagenesis was performed using DNA shuffling and error-prone PCR. Active mutants were identified by lysate activity screening and used for subsequent rounds of mutagenesis. The activity of KE07 was increased 200-fold through seven rounds of the directed evolution and accumulated approximately one mutation in each round, which is summarized in Table 3.3. The potential effects of each mutation are analysed here to explain the effect of these mutations on the energetics and kinetic isotope effects of the reaction. Overall, the three mutations in the early rounds (R1-R4) are consistently present in later rounds of KE07 evolution. The side chains of those mutations (N224D, G202R, I7D) directly interact with the side chains of active site residues. Although those early mutations were described previously (31), the effects are explained here for clarity and to compare with mutations in later rounds (R5–R7-2). The side chains of the five mutations from the later rounds (V12M, K146T, F77I, F229S, I102F) have no direct interaction with the side chains of the active site.

The active site of KE07 is located on the beta barrel core surrounded by alpha helices of the $(\beta\alpha)_8$-barrel scaffold (Figure 3.8**a**). In the initial design of KE07, the substrate 5-nitrobenzisoxazole forms a π-stacking interaction with W50, and the proton at the 3' position of the substrate is transferred to the carboxylate anion of E101 in the reaction catalysed by KE07. K222 was designed to stabilize the oxyanion of the substrate (Figure 3.8**b**, see Chapter 1). However, K222 actually forms a salt bridge with E101, according to the previous crystal structure of KE07 R1 (Figure 3.8**b**) (13).

Table 3. 3 | Accumulation of mutations during KE07 evolution

No.	Resi.	R1	R2	R3	R4	R5	R6	R7	R7-2
1	**N224**		**D**	**D**	**D**	**D**	**D**	**D**	**D**
2	**G202**		**R**	**R**	**R**	**R**	**R**	**R**	**R**
3	**I7**			**Q**	**D**	**D**	**D**	**D**	**D**
4	**V12**					**M**		**L**	**M**
5	**K146**		T	T	E		**T**	**T**	**T**
6	**F77**							**I**	**I**
7	**F229**			S				**S**	**S**
8	**I102**								**F**
9	K19		E				E		
10	Q123		R						
11	F86			L					
12	H84							Y	
13	M207							T	

Mutations accumulated during evolution of KE07 are highlighted in yellow background. Mutations No.1-3 directly remodel the active site. Mutations No.4-8 indirectly regulate the active site. Inconsistent mutations are in grey background.

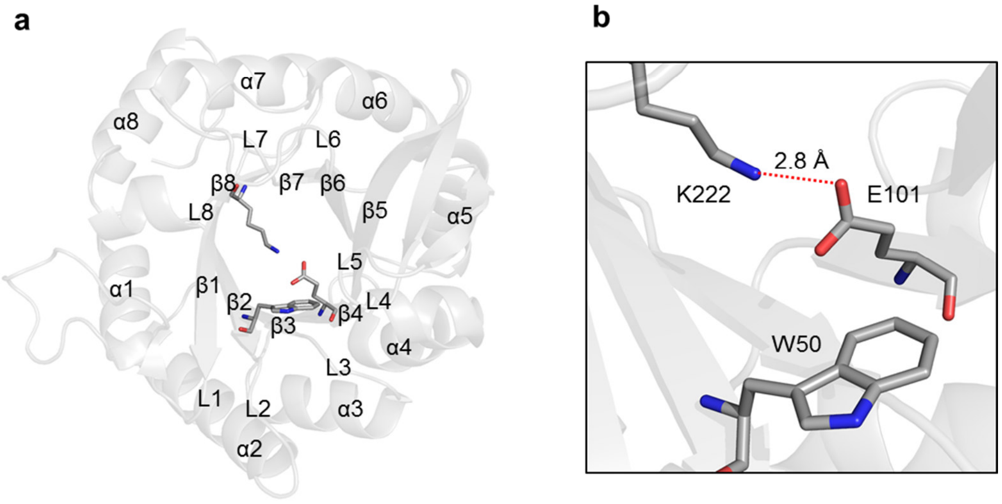


Figure 3. 8 | Crystal structure of KE07 design R1. The overall $(\beta\alpha)_8$–barrel scaffold of the KE07 design R1 (**a**) is shown in cartoon representation. The catalytic base residue E101, the catalytic acid residue K222, and the substrate stabilizing residue W50 are shown in stick representation (**a, b**). The salt bridge between E101 and K222 is shown as a red dotted line. PDB ID: 2RKX.

Initially, the entry to the active site was remodelled by the mutations N224D and G202R from R2. The mutated residues D224 and R202 formed a new electrostatic network with the side chains of H201 and Y128 (Figure 3.9, top panel). In addition, those mutations of uncharged residues to larger and charged residues reduce the size of the active site and introduced new charges into the active site (Figure 3.9, top). The next mutation was I7D from R4. The mutated residue D7 forms new hydrogen bonds with K222, resulting in a conformational change of K222 and breaking the salt bridge between K222 and the catalytic base, E101 (Figure 3.9, lower panel) (31). In summary, the mutations up to R4 occur in the active site and the side chains directly interact with active site side chains, which remodel the size, shape, and electrostatic networks of the active site, which became more charged and ordered (Figure 3.9). All mutations accumulated during the later rounds (R5–R7-2) occur outside of the active site and the side chains have no direct contact with side chains of active site residues.

To sum up, the early increase in activity of KE07 during directed evolution (R1-R4) is due to direct electrostatic optimization through active site mutations, while the later activity increase (from R5) is due to the indirect mutations, potentially through remote influences on the flexibility and conformation of the active site residues. Discussion of the crystal structures of R1-R7-2 is presented in Chapter 4.

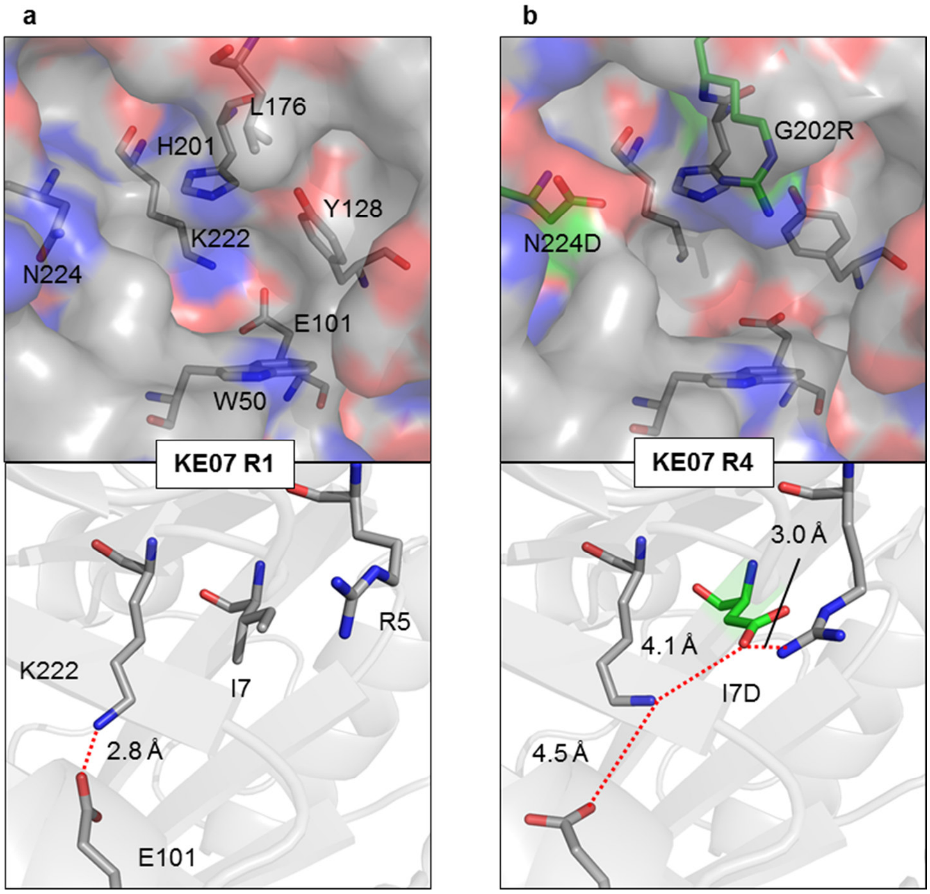


Figure 3. 9 | Mutations in early round (R1–R4) of KE07 evolution. The active site of the KE07 design R1 (**a**) and R4 (**b**) are shown in surface and stick representation for the entry of the active site and cartoon and stick style for the inside/behind active site. Mutated residues are colored in green. The change of electrostatic interactions due to I7D is shown in red dotted lines. The surface views of the active site entry of R1 (**a**) and R4 (**b**) show that the entry became smaller and charged by the N224D and G202R mutations. Red dotted lines show that I7D mutation forms new electrostatic interaction with K222 and R5. PDB ID: 2RKX (**a**), 3IIO (**b**).

3.3.3 Analysis of Arrhenius parameters (H)

Comparison between the Arrhenius parameters in non-enzymatic and enzymatic reactions

The activation energy (E_a) for the Kemp elimination in the absence of enzyme was 61 kJ/mol (Table 3.2). The activation energy was significantly reduced in all reactions catalysed by KE07 variants, by 23-45 kJ/mol. On the other hand, the pre-exponential factor (A) in enzymatic reactions was slightly higher or lower (ln$A$: 9.1-16.1) than in the non-enzymatic reaction (ln$A$: 13.7). Detailed Arrhenius parameters are shown in Table 3.2.

Changes in the Arrhenius parameters in KE07 evolution

The activation energy (E_a) decreased until R4, then increased in R5 and R6 and decreased again in the last round of KE07 evolution (Table 3.2, Figure 3.10). The overall trend in the pre-exponential factor (A) was a decrease in the early rounds (R1–R4), followed by an increase in R5 and R6 and a decrease in later rounds (R7, R7-2). The decrease in ln$A$ in R4 cancels out much of the decrease in E_a and is a typical enthalpy-entropy compensation. The increase in ln$A$ in R5-6 was sufficient to outweigh an increase in the activation energy. Interestingly, the middle rounds (R5, R6) were the points where remote mutations started to accumulate on the top of the active site mutations.

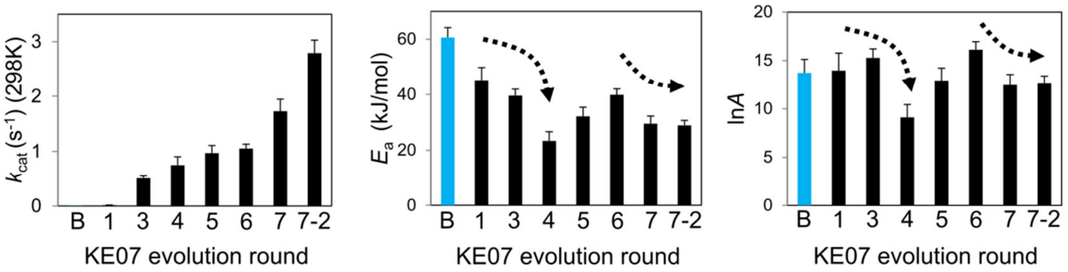


Figure 3. 10 | Arrhenius parameters (H) of KE07 variants. The blue column represents the non-enzymatic reaction and the black columns represent the enzymatic reactions catalyzed by KE07 variants. Arrows show the trends in early rounds (R1–R4) and later rounds (R6-R7-2).

3.3.4 Analysis of Arrhenius parameters (H/D)

Primary deuterium isotope effect in Kemp elimination

To determine primary deuterium isotope effects (1° KIE) in the Kemp elimination reaction, the 3'H was substituted with D, since the 3'H is transferred from C3 of 5-nitrobenzisoxazole to the carboxylate oxygen of E101. In this acid-base Kemp elimination reaction, the KIE is affected by the vibrational zero-point energy difference between C-H and C-D (33). The expected vibrational zero-point energy (ZPE) difference is $E_{a(D)} - E_{a(H)} < 5$ kJ/mol and $k_H/k_D < 6.9$ (298K) when the vibration of C-H is linear along the C-H-O coordinate (33). The KIE is higher when the vibration of C-H is non-linear along the C-H-O coordinate (34). In the extreme case, disappearance of the C-H bending frequency in the TS compared to the RS results in $k_H/k_D < 17.2$ (298K), $E_{a(D)} - E_{a(H)} < 9$ kJ/mol, and $A_H/A_D > 0.5$ (33). Deviations from these expected values have been explained by quantum tunnelling (33, 35–37).

Computational modelling predicts that KIEs and their temperature dependency increase as the donor-acceptor (D-A) distance increases (38). The D-A coordinate is affected by solvent composition in the non-enzymatic reaction, while it is affected by the preorganization of the active site in the enzymatic reaction (39, 40). Another computational modelling study predicted that alternative conformations can account for large KIEs and unusual temperature dependency in SLO-1, AADH, MADH, and DHFR (41).

To sum up, the KIE is sensitive to the p*K*a of the acid and base, vibrational linearity, and D-A distance, which are sensitive to the solvent composition, active site pre-organization, and sampling of alterative conformations.

Changes of Arrhenius parameters (H/D) in non-enzymatic and enzymatic reactions by KE07 variants

The activation energies, for both hydrogenated and deuterated substrate, were significantly reduced in the enzymatic reactions compared to the non-enzymatic reactions while the ln*A* values in the enzymatic reactions were slightly below or above the non-enzymatic values (Table 3.2, Figure 3.15). For example, the activation energies (E_a) were 61 kJ/mol for H and 70 kJ/mol for D in the non-enzymatic reaction. The activation energies were significantly lower for all enzymatic reactions catalysed by KE07 variants ($E_{a(H)}$: 23.3–45.2 kJ/mol); $E_{a(D)}$:37.8–63.3 kJ/mol) compared to the non-enzymatic reaction. On the other hand, differences between non-enzymatic and enzymatic values were marginal for the pre-exponential factor (< 10% difference). ln*A* values were 14 for H and 16 for D in the non-enzymatic reactions. The values of ln*A* in enzymatic reactions by KE07 variants were in between 9–16 (average 13) for H and in between 13–20 (average 17) for D.

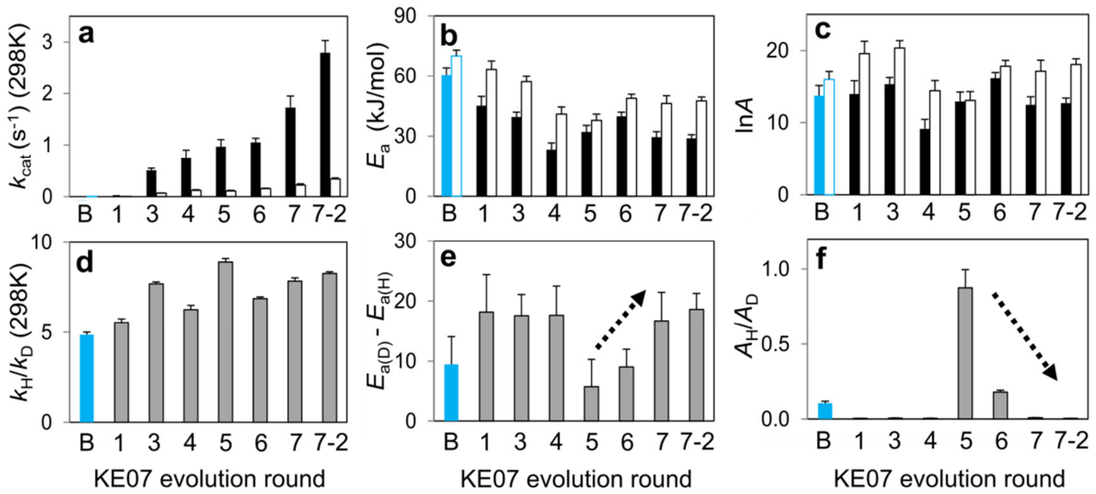


Figure 3. 11 | Arrhenius parameters (H/D) in KE07 evolution. Non-enzymatic values are shown in blue and enzymatic values are shown in black. Rate constants (s^{-1}), activation energies (*E*a, kJ/mol), and pre-exponential factors (*A*) are shown with black (H) and white (D) columns (**a-c**). The KIEs for each parameter are shown in **d-f**. The trend in the later rounds of the KE07 evolution is indicated by black dotted arrows.

Kinetic isotope effects of the Arrhenius parameters, including k_H/k_D, $E_{a(D)} - E_{a(H)}$, and A_H/A_D, were determined for the non-enzymatic and enzymatic reactions (Table 3.2, Figure 3.15). Although the overall kinetic isotope effect of $E_{a(D)} - E_{a(H)}$ was increased (~2-fold) and A_H/A_D was decreased (by 0.094) (without taking account R5 - R6) in the enzymatic reactions, compared to the non-enzymatic reactions, the values were close to the classical limitation in both non-enzymatic and enzymatic reactions. KIE classical limits are: $E_{a(D)} - E_{a(H)} > 5$ kJ/mol, $A_H/A_D < 0.5$, $k_H/k_D < 6.9$ (298K). To be specific, k_H/k_D (298 K) was 4.9 for the non-enzymatic reaction and 5.5–8.9 (average ~7) for the enzymatic reactions. $E_{a(D)} - E_{a(H)}$ was 9.5 kJ/mol for the non-enzymatic reaction, which is above the classical limit of 5 kJ/mol. $E_{a(D)} - E_{a(H)}$ values in R5-R6 (5.7-9.0 kJ/mol) were lower than the non-enzymatic reactions, while in the other variants, the $E_{a(D)} - E_{a(H)}$ values were higher (16.7–18.6 kJ/mol, average 18 kJ/mol). The pre-exponential factor ratio (A_H/A_D) was 0.1 for the non-enzymatic reaction, which is below the classical limit of 0.5. A_H/A_D values in R5-R6 (0.2-0.9) were higher than the non-enzymatic reactions while in the other variants, the A_H/A_D values were lower (0.004–0.01).

Changes of Arrhenius parameters (H/D) in KE07 evolution

These results show that the mutations that occurred in the active site (N224D, G202R, I7Q/D) in early rounds (R1–R4) did not change the kinetic isotope effect of the activation energy and pre-exponential factors. On the other hand, these values were significantly affected by remote mutations in the later rounds (R5–R7-2) (Figure 3.11, Table 3.2, Table 3.3). The kinetic isotope effect on the reaction rate fluctuated throughout evolution (k_H/k_D (298 K) = 6-9) and did not show a clear trend. The kinetic isotope effect on $E_{a(D)} - E_{a(H)}$ and A_H/A_D was constant in early rounds ($E_{a(D)} - E_{a(H)}$ = 18.2–17.6 kJ/mol, A_H/A_D = 0.004–0.006 in R1–R4). Then the values changed suddenly and were gradually restored to the early round values in the later rounds of KE07 evolution ($E_{a(D)} - E_{a(H)}$ = 5.7 → 18.6 kJ/mol, A_H/A_D = 0.875 → 0.005 in R5 → R7-2).

The kinetic isotope effect is sensitive to the D-A distance (38) and the pre-organized dipoles within the active site (10). For example, increasing $E_{a(D)} - E_{a(H)}$ (0.9 → 5.3) and decreasing A_H/A_D (18 → 0.027) were observed in the soybean lipoxygenase-1 (SLO-1) system due to mutation of the remote residue 553 (42). Crystal structures of SLO-1 wild type showed two conformations of the metal coordinating residue H499, which was solidified to one conformation upon mutations of residue 553 (42). This was explained by increasing D-A distance with decreasing gating motion along the coordinate due to stabilization of H499 by a remote mutation at position 553, resulting in a reduced hydrogen tunneling contribution (42). However, another computer simulation of the SLO-1 system was able to reproduce the experimental values with conventional transition state theory by adopting two conformational states rather than gating motion and tunneling (41), suggesting that changes to the preorganization of an active site due to remote mutations can alter the kinetic isotope effect. In the work presented here, the kinetic isotope effect was changed by remote mutations in later rounds of evolution of KE07 showing a similar trend to the SLO-1 system: increasing $E_{a(D)} - E_{a(H)}$ (5.7 → 18.6) and decreasing A_H/A_D (0.875 → 0.005) in R5-R7-2. This suggests that R5 and R6, which exhibit low $E_{a(D)} - E_{a(H)}$ and high A_H/A_D might be disordered, as seen in SLO-1 (further explanation in Chapters 4 and 5).

To sum up, we observed a sudden change in the kinetic isotope effects (activation energy and pre-exponential factor) in R5-R6, where remote mutations (V12M or K146T) are added to the active site mutations (N224D, G202R, I7Q/D). This happened at the same time as an increase in E_a and ln$A$ (Figure 3.11, Table 3.3). Additional remote mutations (F77I, F229S, I102F) in later rounds restored the KIEs to levels seen in R1–R4. This trend suggests a change in the preorganization of the active site by remote mutations in R5-6 of KE07 evolution, which possibly causes two reactant states as shown in SLO-1 (42) followed by stabilization of the active site in R7 and R7-2. To be specific, the values for the initial KE07 design and the last variant obtained by directed evolution (R7-2) were almost same ($E_{a(D)} - E_{a(H)} = 18.2$ kJ/mol at R1 and 18.6 kJ/mol at R7-2, $A_H/A_D = 0.004$ at R1 and 0.005 at R7-2).

3.4 Summary

Arrhenius parameters and primary KIEs for the Kemp elimination in solution and in KE07 variants were determined. The activation energy, but not the pre-exponential factor, was significantly decreased in the enzymatic Kemp elimination compared to the non-enzymatic reaction, demonstrating that KE07 enzymes reduce the activation free energy mainly by reduction of the enthalpic barrier. In addition, the kinetic isotope effects on activation energy and pre-exponential factor were smaller in the KE07-catalysed reactions compared to the non-enzymatic reactions, suggesting a potential D-A distance change or pre-organized dipole state change in the KE07 enzymes. The Arrhenius parameters in KE07 evolution show overall enthalpy-entropy compensation even though there was a sudden fluctuation in the middle of evolution.

The kinetic isotope effects on the Arrhenius parameters (i.e. $E_{a(D)} - E_{a(H)}$ and A_H/A_D) were constant in KE07 evolution, except for a sudden change in R5-R6. The decreased $E_{a(D)} - E_{a(H)}$ and increased A_H/A_D in R5-R6 indicates a significant structural change that affected the preorganization, as shown in SLO-1 (42). The observed increasing $E_{a(D)} - E_{a(H)}$ and decreasing A_H/A_D in later round of KE07 evolution could be related to a structural event such as the emergence of two reactant conformations, which are subsequently resolved in R7 and R7-2.

3.5 References

1. Wilson CJ (2015) Rational protein design: developing next-generation biological therapeutics and nanobiotechnological tools. *Wiley Interdiscip Rev Nanomedicine Nanobiotechnology* 7(3):330–341.
2. Bornscheuer UT, et al. (2012) Engineering the third wave of biocatalysis. *Nature* 485(7397):185–194.
3. Steen EJ, et al. (2010) Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* 463(7280):559–562.
4. McKenna R, Nielsen DR (2011) Styrene biosynthesis from glucose by engineered E. coli. *Metab Eng* 13(5):544–554.
5. Cherny I, et al. (2013) Engineering V-type nerve agents detoxifying enzymes using computationally focused libraries. *ACS Chem Biol* 8(11):2394–2403.
6. Sollid LM (2002) Coeliac disease: dissecting a complex inflammatory disorder. *Nat Rev Immunol* 2(9):647–55.
7. Perez EE, et al. (2008) Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat Biotechnol* 26(7):808–16.
8. Christensen H, Martin MT, Waley SG (1990) Beta-lactamases as fully efficient enzymes. Determination of all the rate constants in the acyl-enzyme mechanism. *Biochem J* 266(3):853–861.
9. Putman SJ, Coulson AFW, Farley IRT, Riddleston B, Knowles JR (1972) Specificity and kinetics of triose phosphate isomerase from chicken muscle. *Biochem J* 129:301–310.
10. Villa J, Warshel A (2001) Energetics and dynamics of enzymatic reactions. *J Phys Chem B* 105(33):7887–7907.
11. Warshel A (2014) Multiscale modeling of biological functions: From enzymes to molecular machines (Nobel Lecture). *Angew Chemie Int Ed* 53(38):10020–10031.
12. Kries H, Blomberg R, Hilvert D (2013) De novo enzymes by computational design. *Curr Opin Chem Biol* 17(2):221–228.
13. Röthlisberger D, et al. (2008) Kemp elimination catalysts by computational enzyme design. *Nature* 453(7192):190–195.
14. Blomberg R, et al. (2013) Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature* 503(7476):418–421.
15. Preiswerk N, et al. (2014) Impact of scaffold rigidity on the design and evolution of an artificial Diels-Alderase. *Proc Natl Acad Sci* 111(22):8013–8018.
16. Giger L, et al. (2013) Evolution of a designed retro-aldolase leads to complete active site remodeling. *Nat Chem Biol* 9(8):494–498.
17. Khersonsky O, et al. (2011) Optimization of the in-silico-designed Kemp eliminase KE70 by computational design and directed evolution. *J Mol Biol* 407(3):391–412.
18. Osuna S, Jiménez-Osés G, Noey EL, Houk KN (2015) Molecular dynamics explorations of active site structure in designed and evolved enzymes. *Acc Chem Res* 48(4):1080–1089.
19. Arrhenius S (1889) On the reaction velocity of the inversion of cane sugar by acids. *Zeitschrift für Phys Chemie* 4:226.
20. Eyring H (1935) The activated complex and the absolute rate of chemical reactions. *Chem Rev* 17(1):65–77.
21. Basran J, Sutcliffe MJ, Scrutton NS (1999) Enzymatic H-transfer requires vibration-driven extreme tunneling. *Biochemistry* 38(10):3218–3222.
22. Kohen A, Cannio R, Bartolucci S, Klinman JP (1999) Enzyme dynamics and hydrogen tunnelling in a thermophilic alcohol dehydrogenase. *Nature* 399(6735):496–499.
23. Hwang JK, King G, Creighton S, Warshel A (1988) Simulation of free energy relationships and dynamics of SN2 reactions in aqueous solution. *J Am Chem Soc* 110(16):5297–5311.
24. Tubergen MJ, Cable JR, Levy DH (1990) Substituent effects on the electronic spectroscopy of tryptophan derivatives in jet expansions. *J Chem Phys* 92(1):51.
25. Anderson JB (1995) Predicting rare events in molecular dynamics. *Adv Chem Phys* 91:381–432.
26. Gasteiger E, et al. (2005) *Protein identification and analysis tools on the ExPASy server* (Springer).
27. Hollfelder F, Kirby AJ, Tawfik DS, Kikuchi K, Hilvert D (2000) Characterization of proton-transfer catalysis by serum albumins. *J Am Chem Soc* 122(6):1022–1029.
28. Döbele M, Wiehn MS, Bräse S (2011) Traceless solid-phase synthesis of

trifluoromethylarenes. *Angew Chemie Int Ed* 50(48):11533–11535.

29. Manetsch R, Zheng L, Reymond MT, Woggon W, Reymond J (2004) A catalytic antibody against a tocopherol cyclase inhibitor. *Chem Eur J* 10(10):2487–2506.
30. Casey ML, Kemp DS, Paul KG, Cox DD (1973) Physical organic chemistry of benzisoxazoles. I. Mechanism of the base-catalyzed decomposition of benzisoxazoles. *J Org Chem* 58(I):33–34.
31. Khersonsky O, et al. (2010) Evolutionary optimization of computationally designed enzymes: Kemp eliminases of the KE07 series. *J Mol Biol* 396:1025–1042.
32. Richter F, Leaver-Fay A, Khare SD, Bjelic S, Baker D (2011) De novo enzyme design using Rosetta3. *PLoS One* 6(5):e19230.
33. Bell RP (1959) *The proton in chemistry* (Springer Science & Business Media).
34. Melander L (1960) *Isotope effects on reaction rates.*
35. Kim Y, Kreevoy MM (1992) The experimental manifestations of corner-cutting tunneling. *J Am Chem Soc* 114(18):7116–7123.
36. Kwart H (1982) Temperature dependence of the primary kinetic hydrogen isotope effect as a mechanistic criterion. *J Am Chem Soc* 265(4):401–408.
37. Cheng L, Doubleday C, Breslow R (2015) Evidence for tunneling in base-catalyzed isomerization of glyceraldehyde to dihydroxyacetone by hydride shift under formose conditions. *Proc Natl Acad Sci U S A* 112(14):4218–20.
38. Liu H, Warshel A (2007) Origin of the temperature dependence of isotope effects in enymatic reactions: The case of dihydrofolate reductase. *J Phys Chem B* 111(27):7852–7861.
39. Liu H, Warshel A (2007) The catalytic effect of dihydrofolate reductase and its mutants is determined by reorganization energies. *Biochemistry* 46(20):6011–6025.
40. Nagel ZD, Klinman JP (2009) A 21st century revisionist's view at a turning point in enzymology. *Nat Chem Biol* 5(8):543–550.
41. Glowacki DR, Harvey JN, Mulholland AJ (2012) Taking Ockham's razor to enzyme dynamics and catalysis. *Nat Chem* 4(3):169–176.
42. Meyer MP, Tomchick DR, Klinman JP (2008) Enzyme structure and dynamics affect hydrogen tunneling: The impact of a remote side chain (I553) in soybean lipoxygenase-1. *Proc Natl Acad Sci* 105(4):1146–1151.

Chapter 4: Structural Preorganization in KE07 Evolution

4.1 Introduction

In the previous chapter, kinetic isotope effects (KIEs) of the Arrhenius parameters changed dramatically at R5, i.e. the KIE $E_{a(D)} - E_{a(H)}$ decreased ~3-fold and the KIE A_H/A_D increased ~175-fold, relative to R1-4. These parameters were restored to the early round values in later rounds of KE07 evolution (R5 to R7-2; Chapter 3). The trend suggested a significant structural event at R5, which was refined/improved by R7. In later rounds of the KE07 evolution (R5–7-2) the only mutations that occur are remote from the active site. Defining the effects of mutations that are remote from the active site is one of the most challenging and crucial aspects in the field of enzyme design, as such mutations can change the preorganization and substrate binding mode of the enzyme indirectly and the effects of these mutations are difficult to predict (1–3). Although X-ray crystal structures of apo-R1, 4, 6, and 7 were previously published (4, 5), the X-ray crystal structures of R5 and the most active variant (R7-2) were not solved at the outset of this thesis. Moreover, the previous structures were all solved in different space groups, making comparison difficult in light of crystal packing artefacts. Here, I have solved the crystal structures of R5 and R7-2 for the first time. I have also solved the structures of these variants in the presence and absence of ligand.

Unexpectedly, in the course of KE07 evolution we observed an alteration in the conformations of the substrate π-stacking residue W50 and the catalytic residues E101 and K222. In particular, the alternate conformation of W50 was only observed from R5 onwards and results in a dramatic change in the active site. In the most active variant (R7-2), only the new, alternate, conformation of W50 was observed.

4.2 Methods

X-ray crystallography

Crystals of all KE07 variants (Figure 4.1; representative SDS-PAGE gels are presented in Chapter 3) were grown at 4 °C by hanging-drop vapour diffusion. Equal volumes of reservoir solution (12-25% PEG 3350, 0.1 mM Bis-Tris Propane pH 8.5) or (25 mM HEPES pH 7.25, 100 mM NaCl) were mixed with protein (5-30 mg/ml) and crystals reached maximum size after 7-60 days of incubation. Either 35% polyethylene glycol (PEG400) or 35% 2-methyl-2,4-pentanediol (MPD) was used as a cryo-protectant. Crystals were soaked in a solution containing 35% MPD and 0-7.5 mM 5-nitrobenzisoxazole for between 10 seconds and 40 minutes before vitrification at 100 K. Crystallographic data were collected at 100 K at the Australian National University (MarμX, 1.5418 Å) or Australian Synchrotron (MX1, 0.9537 Å). The obtained diffraction data were indexed and integrated with the program XDS (6). Resolution estimation and data truncation were performed on the basis of the datasets overall half-dataset correlation, a $CC_{1/2}$ value of 0.5 (7). All structures were solved by molecular replacement using the Molrep program in CCP4 (8) using either the structure deposited under PDB accession code 2RKX or 3IIV as a starting model. The models were refined using phenix.refine (9), and the model was subsequently optimized by iterative model building with the program COOT v0.7 (10). The alternative conformations were modelled based on *mFo-dFc* density and the occupancies and B-factors were determined using phenix.refine (9). The structures were then evaluated using MolProbity (11). Details of the refinement statistics were produced by Phenix v1.9 (12) and summarized in Table 4.1–4.3. Figures were made using COOT v0.7 (10) and PyMol v1.7 (13).

Site directed mutagenesis

The impact of the W50A mutation in R1 and R7-2 of KE07 was assessed by site-directed mutagenesis. Gibson assembly (14) based mutagenesis was used to introduce the W50A substitution into the plasmids encoding R1 and R7-2. The following primers were used to introduce the W50A substitution (forward primer 5'-GGC ATC GAC GAG CTT AGT TTT GCG GAC ATT ACC GCT TCG GTG GAA AAG C-3' and reverse primer 5'-GCT TTT CCA CCG AAG CGG TAA TGT CCG CAA AAC TAA GCT CGT CGA TGC C-3'). Full-length expression constructs were assembled using the following primers (T7-forward primer 5'-TAA TAC GAC TCA CTA TAG GG-3' and T7-reverse primer 5'-GCT AGT TAT TGC TCA GCG-3'). The gene inserts were sub-cloned into the empty pET-29b(+) vector using Gibson assembly master mix according to the manufacturer's directions (New England Biolabs). A detailed cloning process is described in Chapter 2.

4.3 Results

4.3.1 X-ray crystal structures of KE07 variants

To investigate the structural changes that underlie the KIEs observed in the later rounds of KE07 evolution (Chapter 3), KE07 variants (R1, R5, R6, R7, R7-2) were crystallized (Figure 4.1). Although crystal structures of R1, R4, R6, and R7 were previously published (5), R5, where the KIE changed significantly, and the most active variant (R7-2), are presented here for the first time. To investigate substrate binding, the crystals of R1, R5, R6, R7, R7-2 were soaked in substrate containing solution (5-nitrobenzisoxazole, 1.7–7.5 mM) for 10 seconds - 40 minutes. Structures obtained from the soaked crystals were compared with apo-structures (Figure 4.3). Data processing and refinement statistics for the eleven KE07 structures are given in Table 4.1-4.3.

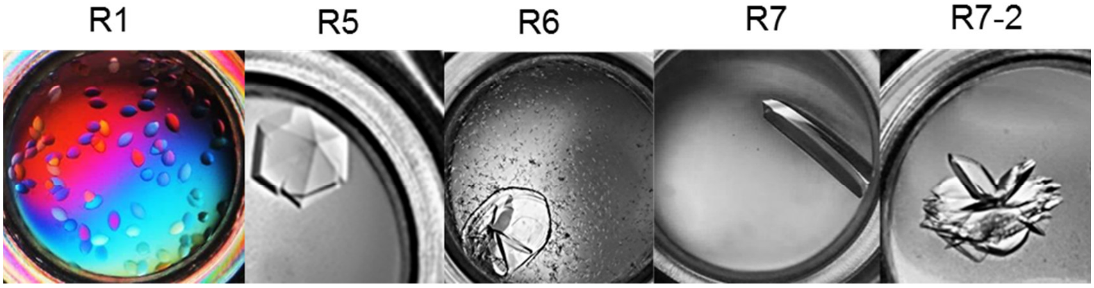


Figure 4. 1 | Crystals of KE07 variants. Crystals of R1, R5, R6, R7, and R7-2.

Table 4. 1 | Data processing and refinement statistics for R1 KE07.

Crystal No.	1	3	4	5
Protein	KE07 R1	KE07 R1	KE07 R1	KE07 R1
PDB ID	4ZO8	5D2T	5D2V	5D2W
Wavelength (Å)	0.9537	0.9537	1.5418	0.9537
Resolution range (Å)	40.65-1.80 (1.86-1.80)	46.01-1.87 (1.94-1.87)	46.05-2.02 (2.09-2.02)	42.06-1.66 (1.72-1.66)
Space group	P 6_1 2 2	P 6_1 2 2	P 6_1 2 2	P 6_1 2 2
Cell dimensions				
a, b, c (Å)	97.22 97.22 156.28	96.34 96.34 155.32	96.44 96.44 155.15	97.13 97.13 155.51
α, β, γ (°)	90.00 90.00 120.00	90.00 90.00 120.00	90.00 90.00 120.00	90.00 90.00 120.00
Total reflections	834696 (79320)	763137 (74658)	233462 (23282)	952950 (48532)
Unique reflections	41005 (3983)	35874 (3492)	28621 (2815)	51184 (4701)
Multiplicity	20.4 (19.9)	21.3 (21.4)	8.2 (8.3)	18.6 (10.3)
Completeness (%)	99.79 (98.93)	99.92 (99.52)	99.84 (100.00)	98.86 (93.03)
Mean I/sigma(I)	18.70 (0.86)	15.22 (1.38)	13.31 (1.57)	15.89 (1.51)
Wilson B-factor (Å^2)	27.88	23.71	31.15	20.36
R-merge[a]	0.13 (4.0)	0.22 (2.7)	0.11 (1.4)	0.13 (1.2)
R-meas[b]	0.13	0.23	0.12	0.13
CC1/2[c]	0.999 (0.626)	0.998 (0.687)	0.996 (0.569)	0.999 (0.637)
CC*[d]	1.00 (0.877)	1.00 (0.903)	0.999 (0.852)	1.00 (0.882)
R-work	0.222 (0.431)	0.191 (0.369)	0.190 (0.308)	0.184 (0.285)
R-free	0.267 (0.416)	0.232 (0.359)	0.237 (0.381)	0.216 (0.294)
Number of non-hydrogen atoms	2171	2225	2179	2200
macromolecules	2016	2019	1977	1987
ligands	-	12	36	16
water	155	194	166	201
Protein residues	253	251	250	250
RMS (bonds, Å)	0.006	0.007	0.005	0.016
RMS (angles, °)	1.15	1.00	1.05	1.51
Ramachandran preferred (%)	95.73	95.58	95.71	96.57
Ramachandran allowed (%)	3.42	3.10	3.00	2.58
Ramachandran outliers (%)	0.85	1.33	1.29	0.86
Clashscore	2.71	1.71	1.72	4.72
Average B-factor (Å^2)	37.1	28.6	35.8	26.1
macromolecules	36.8	27.7	34.7	25.3
ligands	-	32.2	63.2	59.7
solvent	40.3	37.5	42.5	32.3
Crystalization conditions	15% (w/v) PEG 3350, 0.1 M Bis-tris propane pH 8.5, 0.2 M NaF, 277.15K	19% (w/v) PEG 3350, 0.1 M Bis-tris propane pH 8.5, 0.2 M NaF, 277.15K	25 mM HEPES pH 7.25, 0.1 M NaCl, 277.15K	25 mM HEPES, pH 7.25, 0.1 M NaCl, 277.15K
Substrate conc. (mM)	-	1.7	-	7.5
Soaking time (min)	-	40	-	0.5
Cryo-protectant	35% MPD	35% PEG400	35% PEG400	35% MPD

* Highest-resolution shell is shown in parentheses.

[a] R-merge = $\sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$ (15).

[b] R-meas = $\sum_{hkl}\{N(hkl)/[N(hkl)-1]\}^{1/2} \times \sum_i |I_i(hkl) - \langle I(hkl)\rangle| \sum_i I_i(hkl)$ (16).

[c] CC1/2 = $\sum(x-\langle x\rangle)(y-\langle y\rangle)/[\sum(x-\langle x\rangle)^2 \sum(y-\langle y\rangle)^2]^{1/2}$ (7)(17).

[d] CC* = $\sqrt{\frac{2CC_{1/2}}{1+CC_{1/2}}}$ (7)(17).

Table 4. 2 | Data processing and refinement statistics for R5 KE07.

Crystal No.	6	7	9
Protein	KE07 R5	KE07 R5	KE07 R5
PDB ID	5D2X	5D2Y	5D30
Wavelength (Å)	0.9537	1.5418	0.9537
Resolution range (Å)	42.12-1.76 (1.82-1.76)	19.64-1.98 (2.06-1.98)	41.90-1.69 (1.75-1.69)
Space group	$P\ 6_1\ 2\ 2$	$P\ 6_1\ 2\ 2$	$P\ 6_1\ 2\ 2$
Cell dimensions			
a, b, c (Å)	97.26 97.26 156.28	97.88 97.88 156.82	96.80 96.80 155.41
α, β, γ (°)	90.00 90.00 120.00	90.00 90.00 120.00	90.00 90.00 120.00
Total reflections	903045 (78435)	634206 (63546)	951630 (53782)
Unique reflections	43891 (4267)	31217 (3072)	48768 (4772)
Multiplicity	20.6 (18.4)	20.3 (20.7)	19.5 (11.3)
Completeness (%)	99.89 (99.00)	99.51 (98.31)	99.99 (100.00)
Mean I/sigma(I)	20.64 (1.76)	24.47 (0.94)	17.38 (1.41)
Wilson B-factor	23.91	29.24	22.55
R-merge[a]	0.13 (2.0)	0.11 (3.3)	0.13 (1.5)
R-meas[b]	0.14	0.12	0.13
CC1/2[c]	0.999 (0.608)	0.999 (0.507)	0.999 (0.626)
CC*[d]	1.00 (0.870)	1.00 (0.820)	1.00 (0.878)
R-work	0.193 (0.299)	0.203 (0.269)	0.183 (0.283)
R-free	0.228 (0.328)	0.232 (0.302)	0.218 (0.316)
Number of non-hydrogen atoms	2259	2196	2261
macromolecules	2039	1979	2019
ligands	8	-	-
water	212	217	242
Protein residues	250	250	250
RMS (bonds)	0.008	0.008	0.017
RMS (angles)	1.06	1.04	1.69
Ramachandran preferred (%)	96.33	96.22	95.59
Ramachandran allowed (%)	2.75	2.94	2.64
Ramachandran outliers (%)	0.92	0.84	1.76
Clashscore	3.38	2.01	2.95
Average B-factor	27.5	33.5	26.4
macromolecules	26.6	32.8	25.5
ligands	37.8	-	-
solvent	35.5	39.9	33.8
Crystalization conditions	11% (w/v) PEG 3350, 0.1 M Bis-tris propane pH 8.5, 0.2 M NaF, 277.15K	25 mM HEPES pH 7.25, 0.1 M NaCl, 277.15K	25 mM HEPES pH 7.25, 0.1 M NaCl, 277.15K
Substrate conc. (mM)	-	-	7.5
Soaking time (min)	-	-	0.5
Cryo-protectant	35% MPD	35% MPD	35% MPD

* Highest-resolution shell is shown in parentheses.

[a] R-merge = $\sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$ (15).

[b] R-meas = $\sum_{hkl}\{N(hkl)/[N(hkl) - 1]\}^{1/2} \times \sum_i |I_i(hkl) - \langle I(hkl)\rangle| \sum_i I_i(hkl)$ (16).

[c] CC1/2 = $\sum(x - \langle x\rangle)(y - \langle y\rangle)/[\sum(x - \langle x\rangle)^2 \sum(y - \langle y\rangle)^2]^{1/2}$ (7)(17).

[d] CC* = $\sqrt{\frac{2CC_{1/2}}{1 + CC_{1/2}}}$ (7)(17).

Table 4. 3 | Data processing and refinement statistics for R7 and R7-2 KE07.

Crystal No.	11	12	16	17
Protein	KE07 R6	KE07 R7	KE07 R7	KE07 R7-2
PDB ID	5D32	5D33	5D37	5D38
Wavelength (Å)	0.9537	0.9537	1.5418	0.9537
Resolution range (Å)	18.96-2.10 (2.17-2.10)	44.24-1.59 (1.65-1.59)	42.83-2.04 (2.11-2.04)	23.41-1.43 (1.48-1.43)
Space group	$P\,2_1\,2_1\,2_1$	$P\,2_1\,2_1\,2_1$	$P\,2_1\,2_1\,2_1$	$P\,2_1\,2_1\,2_1$
Cell dimensions				
a, b, c (Å)	38.88 82.28 170.93	38.98 82.72 176.97	39.36 83.48 171.31	39.00 81.29 168.67
α, β, γ (°)	90.00 90.00 90.00	90.00 90.00 90.00	90.00 90.00 90.00	90.00 90.00 90.00
Total reflections	232526 (22637)	555153 (56169)	254526 (23731)	688545 (61602)
Unique reflections	32914 (3247)	76765 (7504)	36939 (3596)	100612 (9798)
Multiplicity	7.1 (7.0)	7.2 (7.5)	6.9 (6.6)	6.8 (6.3)
Completeness (%)	99.79 (99.51)	98.22 (97.12)	99.94 (99.92)	99.76 (98.63)
Mean I/sigma(I)	10.55 (1.48)	13.54 (1.66)	11.53 (2.39)	11.22 (3.39)
Wilson B-factor	29.56	19.56	25.59	13.06
R-merge[a]	0.17 (1.3)	0.11 (1.3)	0.12 (0.81)	0.11 (0.47)
R-meas[b]	0.18	0.12	0.13	0.12
CC1/2[c]	0.995 (0.550)	0.997 (0.538)	0.993 (0.607)	0.993 (0.835)
CC*[d]	0.999 (0.842)	0.999 (0.836)	0.998 (0.869)	0.998 (0.954)
R-work	0.200 (0.303)	0.199 (0.216)	0.234 (0.290)	0.194 (0.209)
R-free	0.249 (0.350)	0.229 (0.244)	0.318 (0.385)	0.226 (0.242)
Number of non-hydrogen atoms	4414	4516	4623	4901
macromolecules	4000	4088	4042	4159
ligands				
water	414	428	581	742
Protein residues	510	514	519	526
RMS (bonds)	0.007	0.007	0.008	0.006
RMS (angles)	1.03	1.03	1.1	1.09
Ramachandran preferred (%)	94.20	95.64	90.43	94.61
Ramachandran allowed (%)	3.80	2.70	8.01	3.39
Ramachandran outliers (%)	2.00	1.66	1.56	2.00
Clashscore	6.86	3.52	7.74	3.72
Average B-factor	33.5	25.6	27.2	19.2
macromolecules	33	24.8	26.3	17.4
ligands	-	-	-	-
solvent	38.2	32.5	33.5	29
Crystalization conditions	15% (w/v) PEG 3350, 0.1 M Bis-tris propane pH 8.5, 0.2 M NaF, 277.15K	11% (w/v) PEG 3350, 0.1 M Bis-tris propane pH 8.5, 0.2 M NaF, 277.15K	11% (w/v) PEG 3350, 0.1 M Bis-tris propane PH 8.5, 0.2 M NaF, 277.15K	21% (w/v) PEG 3350, 0.1 M Bis-tris propane PH 8.5, 0.2 M NaF, 277.15K
Substrate conc. (mM)	7.5	-	7.5	-
Soaking time (min)	0.5	-	1.5	-
Cryo-protectant	35% MPD	35% MPD	35% MPD	35% MPD

* Highest-resolution shell is shown in parentheses.

[a] R-merge $= \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$ (15).

[b] R-meas = $\sum_{hkl}\{N(hkl)/[N(hkl)-1]\}^{1/2} \times \sum_i |I_i(hkl) - \langle I(hkl)\rangle| \sum_i I_i(hkl)$ (16).

[c] CC1/2 = $\sum(x-\langle x\rangle)(y-\langle y\rangle)/[\sum(x-\langle x\rangle)^2 \sum(y-\langle y\rangle)^2]^{1/2}$ (7)(17).

[d] CC* = $\sqrt{\frac{2CC_{1/2}}{1+CC_{1/2}}}$ (7)(17).

Comparison of crystal structures with and without soaking

The KE07 R1 protein was crystallized at pH 7.25 (HEPES) and pH 8.5 (bis-tris propane). The C-terminal histidine (H263), part of the hexahistidine tag, of the adjacent protein within the crystal lattice is bound to the active site of KE07 in the structure of the R1 protein crystallized at pH 8.5 (Figure 4.2**a**). After soaking of R1 crystals (1.7 mM 5-nitrobenzisoxazole for 40 min) product density was observed (occupancy 0.3, B-factor ~30 Å^2) in addition to the histidine density (occupancy 0.7, B-factor ~30 Å^2, Figure 4.2**b**). The catalytic residues of R1 KE07 altered their conformations due to the presence of product after soaking. For example, both the E101 and K222 side chains adopted alternative conformations, with partial occupancies, after soaking (Figure 4.2**b**). The angle chi2 of E101 changed 25° from 165° to -170° in the alternative conformation (carboxyl carbon distance between two conformers: 1.3 Å). The side chain nitrogen of the alternative conformation of K222 was about 2.7 Å distance from the other conformation. The side chain of N224 also adopts an alternate position in the soaked crystal, the chi2 (CG-CB) angle is rotated about 60° creating a larger active-site entrance (Figure 4.2**a, b**).

In the structure of the R1 protein crystallized at pH 7.25, H263 was not present in the active site and the active site residues including E101, K222, and N224 adopted different conformations compared to the R1 structure at pH 8.5 (Figure 4.2**c**). In particular, the angle chi1 of N224 was 100° rotated towards substrate binding site (chi1 from -170° to -70°). Upon soaking into a solution containing 7.5 mM 5-nitrobenzisoxazole for 30 seconds, product was bound (occupancy 0.8, B-factor ~60 Å^2). The side chain nitrogen of N224 was 3.6 Å away from the oxyanion of the product and additionally adopted an alternative conformation which creates a larger active-site entrance (Figure 4.2**d**). The enlargement of the active-site entrance as a result of the alternative conformation of N224 (chi1: CA-CB, 100° rotation) possibly facilitates product release. Although the position of the product observed in the R1 crystal structure (Figure 4.2**d**) was similar to the previously calculated substrate position (4), K222 was positioned differently. K222 retained its salt-bridge with the

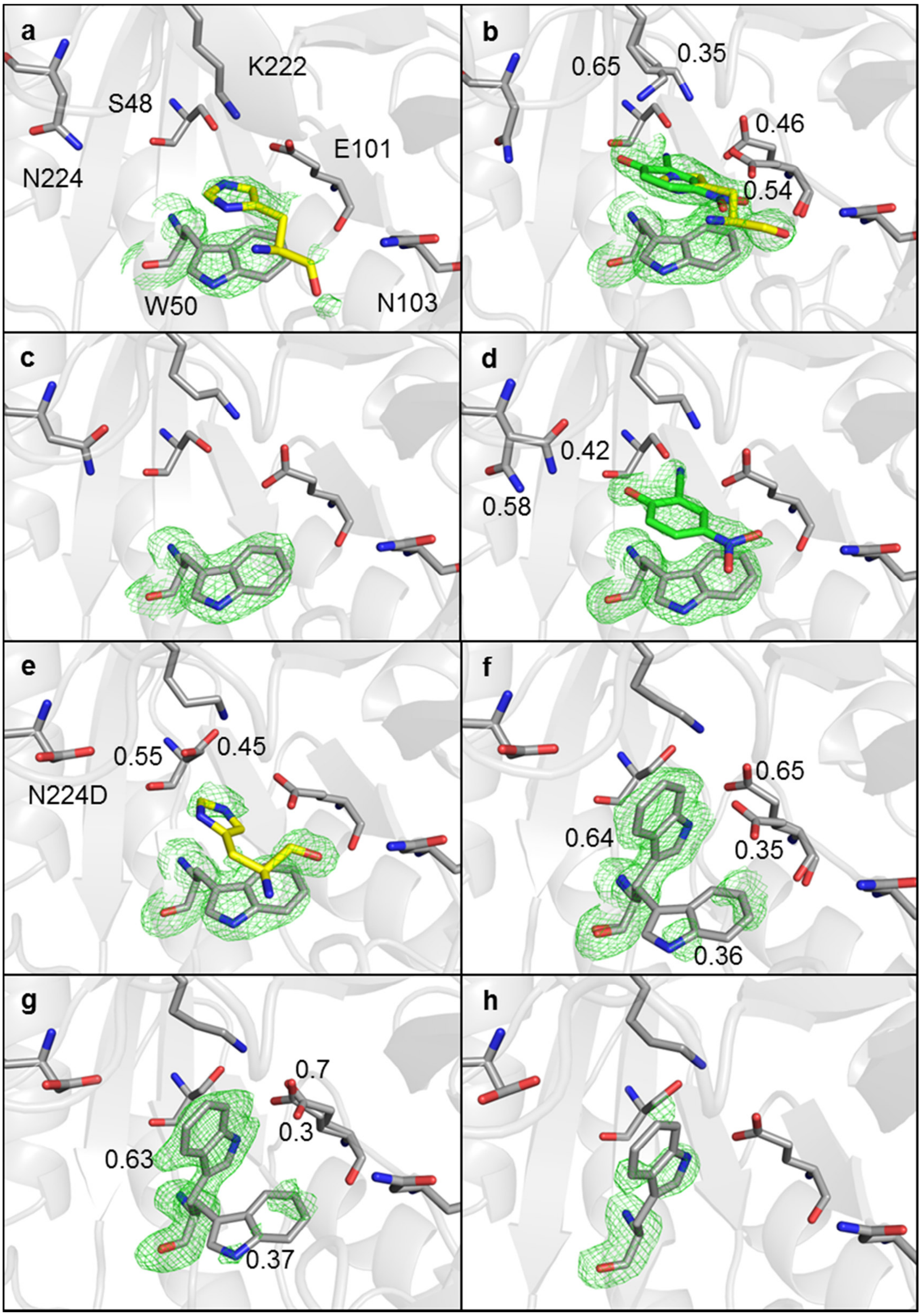


Figure 4. 2 | Comparison of crystal structures with and without soaking. The active site of R1 (pH 8.5) before (**a**) and after soaking (**b**), R1 (pH 7.25) before (**c**) and after soaking (**d**), R5 (pH 7.25) before (**e**) and after soaking (**f**), and R7 (pH 8.5) before (**g**) and after soaking (**h**) are shown in grey cartoons and sticks. H263 from the other subunit is shown in yellow stick, product in green stick and proximal residues changed its conformations by soaking and by evolution are shown in grey sticks. Occupancies of alternative conformations were determined by Phenix v1.9 and are shown next to the residues. Electron densities centered on W50, the histidine, and the product are shown. The F_o-F_c omit maps are contoured at 3.0σ (**a-f**) and at 2.5σ (**g-h**). PDB ID of the structures: **a** (4ZO8), **b** (5D2T), **c** (5D2V), **d** (5D2W), **e** (5D2Y), **f** (5D3O), **g** (5D33), **h** (5D37).

catalytic E101 upon soaking rather than moving to stabilize the oxyanion of the product and both E101 and K222 did not change position before and after soaking at pH 7.25, unlike at pH 8.5.

The KE07 R5 protein was crystallized at pH 7.25 (HEPES). The C-terminal histidine (H263) of another subunit was bound to the active site (occupancy 0.7, B-factor ~40 Å^2). In this structure, two alternate conformations of S48 were observed while one conformation of E101 was observed (Figure 4.2**e**). The bound histidine interacts with E101, potentially stabilizing the E101 conformation. To be specific, the distance between the imidazole nitrogen of H263 and the carboxylic oxygen of E101 is 3.3 Å. The crystal was soaked in substrate containing solution for 30 seconds. Unexpectedly, upon soaking W50 adopted an additional conformation, henceforth this conformation is referred to as position B (110° rotation of chi1: CA-CB, occupancy 0.6, B-factor 20 Å^2) relative to the previous conformation (henceforth, position A) despite the presence of histidine before soaking (Figure 4.2**f**). The pyrrole nitrogen of W50 in position B forms a hydrogen bond with the carboxylic oxygen of E101 (distance <3 Å). Upon soaking of the R5 crystal, a single S48 conformation was observed; this S48 conformation removes steric hindrance with the W50 in position B (the oxygen atom of S48 is 3.7 Å away from the benzene ring of W50 in position B) and forms a new hydrogen bond with D7 through a water molecule (Figure 4.4**b**). The E101 residue adopted an alternative conformation in addition to the previous position (carboxyl carbon distance of two conformers: 1.9 Å, angle chi2 difference: 25° (165° and -170°). When in position B, the carboxylic oxygen of E101 forms a hydrogen bond with the pyrrole nitrogen of W50 and the side chain nitrogen of K222. In this configuration, KE07 would be catalytically unproductive because both oxygens of E101 are hydrogen bonding with amines. We postulate that the high flexibility of E101 observed in R5 (described in detail in Chapter 5) allows these bonds to transiently break, permitting catalysis. This may account for the higher heterogeneity of the active site, relative to ligand bound structures and apo R1 structure, due to the high flexibility of the side chains (e.g. E101 and W50).

The side chains of S48 and K222 point away from the active site in R4 and R5, in comparison to R1, which is likely due to the I7D mutation (see Figure 1.5 for the I7D mutation in R4 in comparison to R1 and the view of the interactions of D7 in R5 is shown in the following section, effect of pH). These conformational changes in S48 and K222 remove some steric hindrance and allow W50 to sample position B. However, the conformational change of W50 appears to begin from R5 most directly due to the addition of the V12M; although the I7D mutation is present in both R4 and R5, the V12M mutation is only present in R5. The backbone amide of V12M on β1 forms hydrogen bond with the backbone oxygen of W50 on β2. In addition, V12M is surrounded by hydrophobic residues including F49, which is connected to W50 through peptide bond. For example, in the crystal structures, the distance between the side chains of the V12 and F49 was 4.2 Å at R4. The distance between the side chains of the bigger M12 and F49 was 3.6 Å at R5.

The R7 crystals formed in a different space group to R1-5 and included two protein chains in the asymmetric unit (Table 4.1-3). In chain B of R7, two conformations of E101 and W50 were observed before soaking. The occupancy of W50 in position B was 0.63 with a B-factor of 15 Å^2 (Figure 4.2**g**). The carbon atom distance of the carboxyl group of two conformers of E101 was ~0.8 Å and the chi2 angle difference was 15° (175° to -170°). After soaking with substrate, the side chains were observed in a single conformation (7.5 mM 5-nitrobenzisoxazole for 1.5 min, Figure 4.3**h**). No unambiguous electron density for the product or substrate was observed in the active site of chain B. The active site of chain A was occupied by H263 from chain B of a crystallographic symmetry mate. The active site side chains of the R7 chain A adopted a single conformation (W50 in position B), these conformations were unaltered upon crystal soaking. The mutations that occur at R7 and their connection to the catalytic residues are discussed in Chapter 3 (Figure 3.12). The active site side chains of both chain A and B of R7-2 adopted a single conformation (W50 in position B); these conformations were unaltered upon crystal soaking.

The changed catalytic configuration in the KE07 structures coincides with the KIEs shown in Chapter 3. The alternative conformation of W50 was first observed in R5, and dramatically rearranges the active site, affecting the position of the substrate and E101. E101 adopted one conformation in apo-R1 (Figure 4.2**c**) and two conformations with 25° angle chi2 difference (chi2 165 and -170) in apo-R5 (Figure

4.2**f**) and two conformations with 15° angle chi2 difference (chi2 175 and -170) in apo-R7 (Figure 4.2**g**). This structural heterogeneity coincides with the significant changes in the KIEs at R5. This correlation between structural heterogeneity and KIEs has been observed in other systems such as SLO-1 which is discussed in Chapter 3 (18). Thus, it is possible that the conformational heterogeneity of W50 in R5 leads to the sharp increase in A_H/A_D, and its stabilization in position B by R7-2 explains the reduction in A_H/A_D over the later rounds of KE07 evolution, as presented in Chapter 3.

An analagous substrate induced conformational change of a substrate π-stacking residue, such as W50, has been observed in another Kemp eliminase (34E4, an antibody catalyst) (19). In 34E4 the substrate π-stacking residue (W91) adopted two conformations in the apo-structure and a single conformation after soaking. The authors suggested that fixing W91 to the catalytically active conformation would produce a better antibody catalyst (19).

Effect of pH

The effect of pH on the structures of the KE07 variants was investigated. Interestingly, occupancy of position B of W50 was increased higher at elevated pH levels in later round variants (Figure 4.3). Although the R5 crystals grown at pH 7.25 and pH 8.5 shared the same crystallographic symmetry ($P6_122$) (crystallographic monomer) and same cryo-protectant (2-methyl-2,4-pentanediol), the structures were different. Specifically, W50 in position A was only observed at pH 7.25, with histidine (as part of a hexahistidine tag) bound, whereas W50 is only observed in position B at pH 8.5, with 2-methyl-2,4-pentanediol (MPD) bound (Figure 4.3**a, b**). The 105° rotation of the chi1 angle (CA:CB) of W50 at the higher pH occurred with side chain movements of the proximal residues, including S48 and T78 and water rearrangement (Figure 4.4). For example, a water molecule (B-factor 33 Å^2), which was not present at pH 7.25, was observed at pH 8.5 (Figure 4.4). The rearrangement of the S48 and T78 side chains at pH 8.5, relative to pH 7.25, resulted in the formation of a new hydrogen bond network behind the active site (see Figure 4.4).

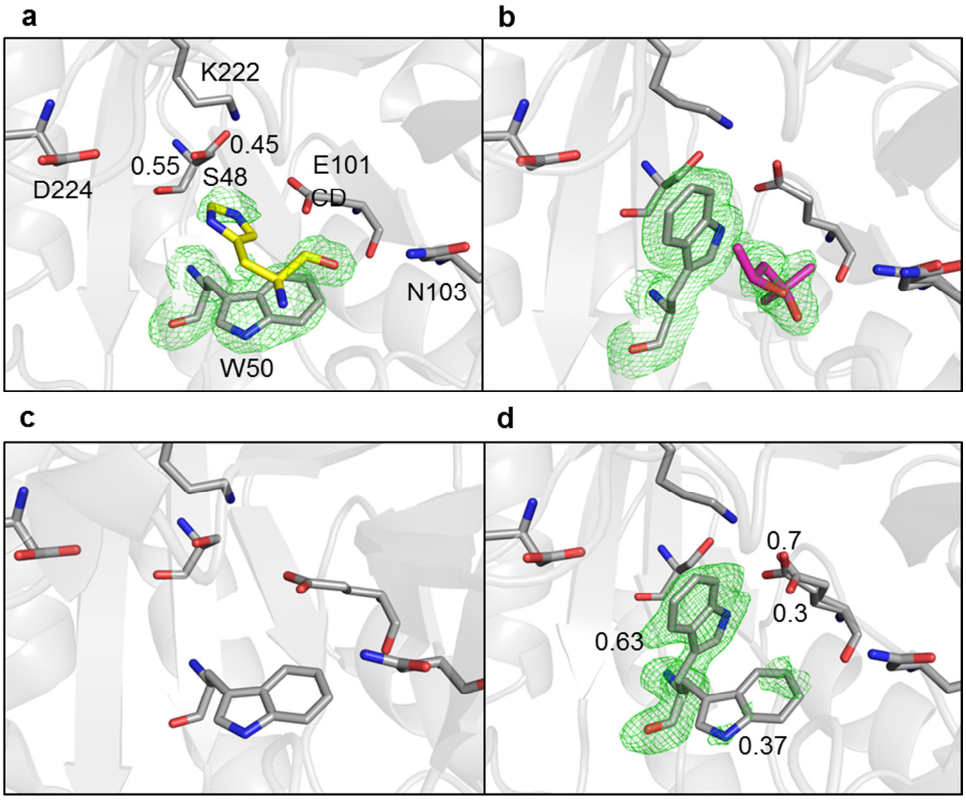


Figure 4. 3 | Effect of pH on the preorganization of active site residues in KE07 variants. The active site side chains of R5 (**a, b**), R7 (**c, d**) are shown as grey sticks. Crystals of **b, d** were grown at pH 8.5. Crystal of **a** was grown at pH 7.25. The structure **c** was published previously and was crystallized at pH 8.0 (5). 2-Methyl-2,4-pentanediol (MPD) is shown as pink sticks and H263 from another subunit (chain A, symmetry number 2) is shown as yellow stick. Occupancies of alternative conformations are shown next to the residues. Electron densities centered on W50, the histidine, and the MPD are shown (**a, b, d**), except **c** which is published previously (5). The F_o-F_c omit maps are contoured at 3.0σ (**a, b**) and 2.5σ (**d**). PDB ID of structures: **a** (5D2Y), **b** (5D2X), **c** (3IIV), **d** (5D33).

Although the crystal symmetry and assembly of R7 ($P2_12_12_1$, crystallographic dimer) were different to R5, the structural changes in response to increased pH were similar. The R7 crystals grown at different pH shared the same crystallographic symmetry ($P2_12_12_1$). Structures of chain A of R7 in both pH 8 and 8.5 were almost identical (RMSD < 0.5 Å), and W50 was in position B. This is most likely due to the presence of histidine within the active site (as part of the hexahistidine tag of chain B). In contrast, the structure of chain B of R7 at pH 8.5 was different to the structure at pH 8.0: position B of W50 was observed in equilibrium with position A at pH 8.5, whereas only position A was observed at pH 8.0 (Figure 4.3**c**, **d**). At pH 8.5 the side chains of S48 and E101 adopt alternate conformations and form new hydrogen bonds with K222. The alternate conformation of E101 probably contributes to the enrichment of W50 in position B by removing steric hindrance (Figure 4.3**c, d**).

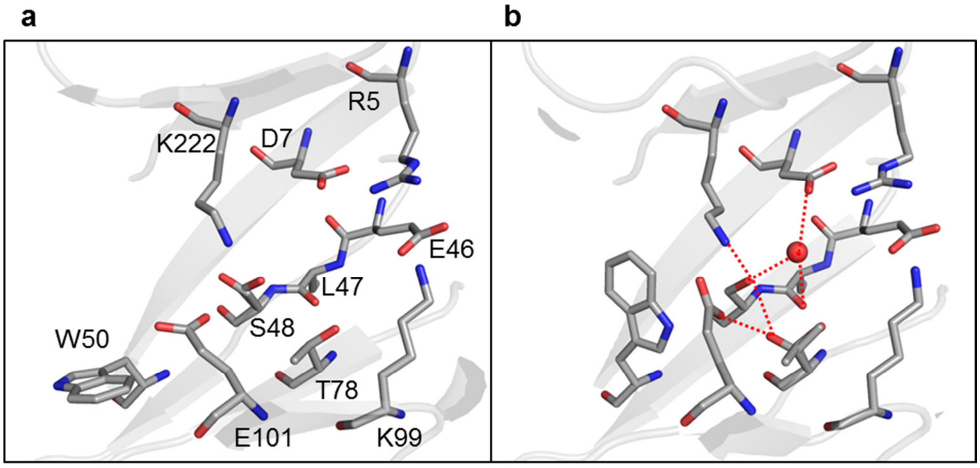


Figure 4. 4 | Effect of pH on active site side chains and water placement in R5 KE07. Residues involved in a hydrogen-bonding network located near the active site of R5 are shown as grey sticks. The additional water molecule at pH8.5 (**b**) comparing to pH 7.25 (**a**) is shown as red sphere. The new hydrogen bonds due to the additional water and the side chain movement at pH 8.5 are shown as red dotted lines (**b**). PDB ID of structures: **a** (5D2Y), **b** (5D2X).

Crystal packing of KE07 R6 crystals

We obtained a crystal structure of R6 from crystals that were soaked in substrate containing solution. Previously, a structure of R6 was solved from crystals that were obtained in the absence of substrate (5). Our R6 protein (PDB ID: 5D32) was crystalized under different conditions (pH 8.5 vs. pH 7.5), the crystals had different crystallographic symmetry ($P2_12_12_1$ vs. $P3_1$), and had a different number of chains in the asymmetric unit (2 vs. 6 chains) compared to the previous structure (PDB ID: 3IIP). Although the two structures have different crystal contacts, the differences between the structures are consistent with those induced by pH and soaking in later rounds of KE07 evolution. In the pH 8.5 structure the S48 and K222 side chains were pointing away from the active site. In addition W50 was rotated 105° (position B) and E101 adopted conformation B (Figure 4.5). Although the structure of R6 that we obtained was apparently altered by the combination of soaking and higher pH, there was no unambiguous density for substrate or product in the active site.

The W50 rotation in R6 may be made more favourable due to the K146T mutation, in addition to three mutations from earlier rounds (N224D, G202R, I7D). The basic residue K146 on the binding loop L5 was mutated to polar T146 which formed a new hydrogen bond between side chain oxygen of T146 and the backbone oxygen of Y143 within loop L5. This hydrogen bond likely stabilizes and repositions loop L5 at R6. The remodelling of loop L5 resulted in the repositioning of the side chain of S144 ~1 Å closer to the side chain of W50 (~7 Å at R4 vs ~6 Å at R6). In addition the loop L5 is proximal to α4 which is connected to E101 on β4. β4 is also connected to β3 and β2 through backbone hydrogen bonds. Furthermore, α4 is proximal to L3, which is proximal to L2. L2 is also connected to β2 where W50 is located through peptide bonds.

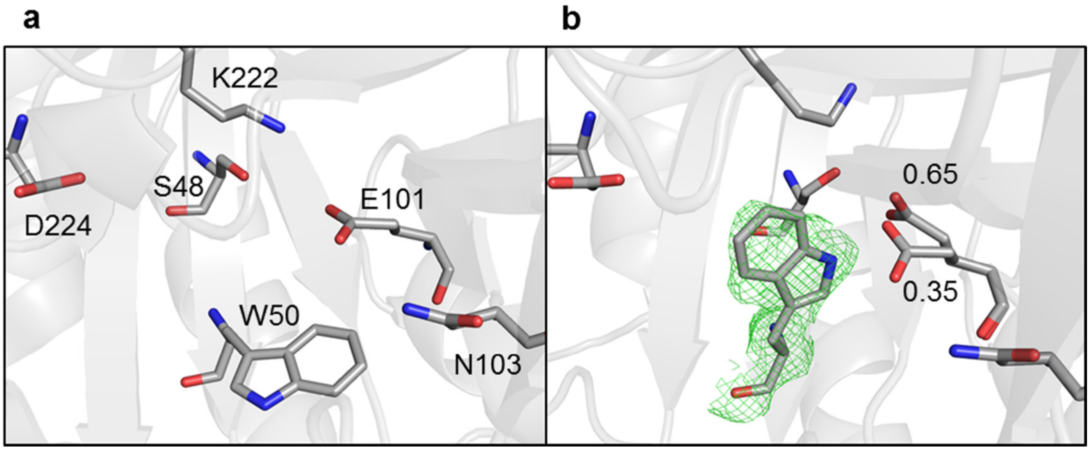


Figure 4. 5 | Comparison of R6 structures. A previously published structure of R6 (**a**, chain B) (5) was obtained from crystals in the $P3_1$ space group. The R6 structure solved here was soaked with substrate (**b**, chain B) and was in the $P2_12_12_1$ space group. The PDB IDs of the structures are 3IIP (**a**), 5D32 (**b**). Occupancies of alternative conformations are shown next to the residues. Electron densities centered on W50 for **b** (no electron density for **a**, which is previously published). The F_o-F_c omit map is contoured at 3.0σ (**b**).

Active site conformation: R7 vs. R7-2

We obtained a crystal structure of most active KE07 variant, R7-2. In previous papers R7-2 was called R7 10/11G (4, 5). The R7-2 enzyme has 2-fold higher activity than R7 (in previous papers R7 was called R7 1/3H) (4, 5). The R7-2 protein was crystallized in the same condition as R7 (pH 8.5) and had the same crystallographic symmetry ($P2_12_12_1$), assembly (crystallographic dimer) and the same cryo-protectant (MPD). Although chain A of R7 and R7-2 were highly similar (RMSD <0.2 Å), chain B was different. In particular, W50 and E101 were observed in a single conformation and the side chain of K222 pointed away from the active site (Figure 4.6). Put simply, in variant R7-2, position B was fully occupied.

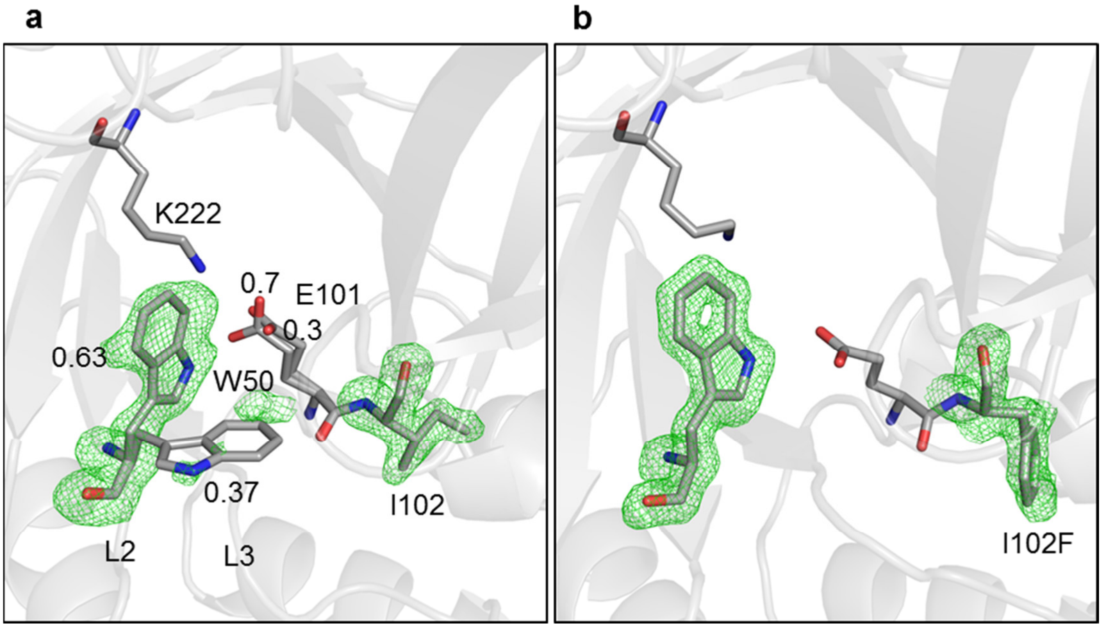


Figure 4. 6 | Crystal structures of R7 and R7-2. Crystal structures of R7 (**a**, chain B, R7 1/3H in previous publications) and R7-2 (**b**, chain B, R7 10/11G in previous publications (4, 5)). Catalytic residues and mutated residue I102F are shown as sticks. Histidine (H263) from another subunit bound in R7-2 was omitted for clarity of structure comparison. PDB ID for each structure: 5D33 (**a**), 5D38 (**b**). Occupancies of alternative conformations are shown next to the residues. Electron densities centered on W50 and I102F. The F_o-F_c omit maps are contoured at 2.5σ (**a**) and 3.0σ (**b**).

The I102F mutation in R7-2 appears to be responsible for the solidification of the E101 and W50 conformations, in combination with the L12M mutation (which is also present in R5), which likely affects W50 indirectly by through steric hindrance in the buried area connected to W50 backbone. R7 and R7-2 differ by four mutations (L12M, I102F, Y84H, T207M) (see Table 3.3 in Chapter 3). Residues 84 and 207 are on the surface of the protein and appear to have no effect on the active site. I102 is adjacent to the catalytic E101 and surrounded by hydrophobic residues on β4, α4, β5, L3, α3, and α4 (the following residues are within 5 Å of I102: I83, V100, A105, A106, L112, I113, I116, V125, V127, Figure 4.7). The I102F mutation results in phenylalanine filling a hydrophobic cavity and a small adjustment of the main chain, which appears to rigidify the neighbouring residue, E101.

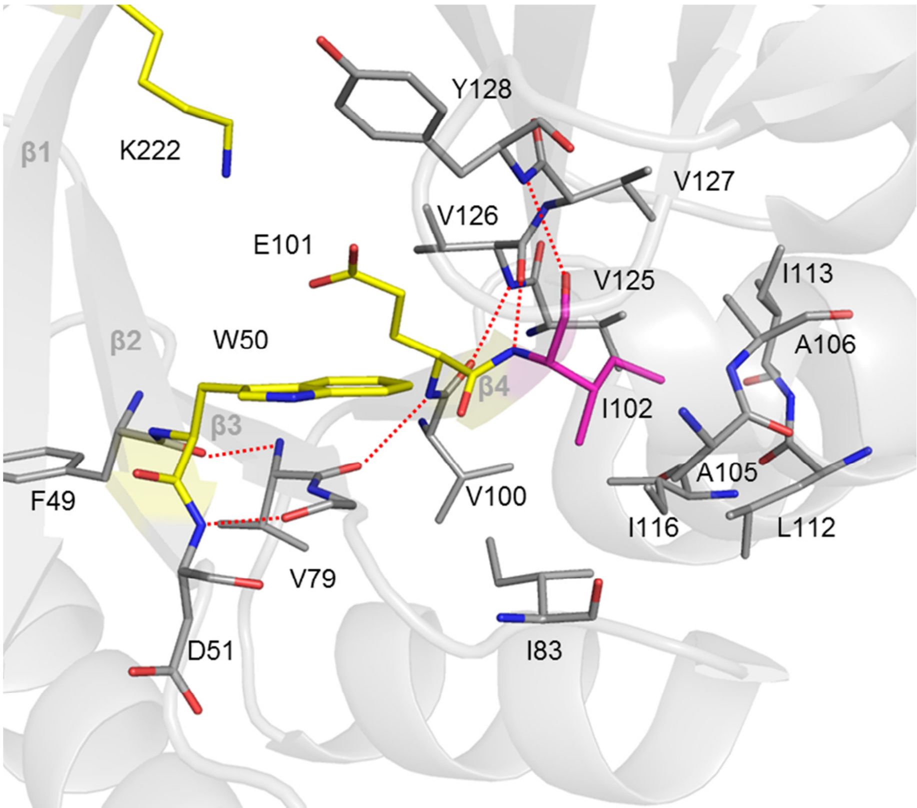


Figure 4. 7 | Residues interacting with I102 in the crystal structure of KE07 R7. I102, which is mutated to a larger phenylalanine in R7-2, is shown as magenta stick, residues within 5 Å distance (I83, V100, A105, A106, L112, I113, I116, V125, V127) are shown as grey sticks, and active site residues are shown as yellow sticks. The red dotted lines (<3 Å) show the connection from I102 to E101 on β4 and W50 on β2. PDB ID: 3IIV.

To sum up, the I102F mutation appears to be responsible for the rigidification and repositioning of the catalytic residue E101 and, indirectly, W50 in R7-2 compared to R7. The I102F mutation accounts for the increased structural homogeneity (preorganization) of R7-2 that led to a 2-fold higher k_{cat} in R7-2 compared with R7. A previous study suggested that a rigid catalytic glutamate was essential for thermodynamic stability and to tolerate nearby internal charges (20). Mutations that result in repositioning of the catalytic base and tryptophan to improve alignment of the substrate have been identified in the evolutionary trajectories of other designed Kemp eliminases (21–23).

Active site conformation: R1 vs. R7-2

The most evolved variant (R7-2) was compared with the designed variant (R1) (Figure 4.8) (See Figure 3.7 in Chapter 3 for the mutations that occurred during evolution). The biggest difference between the active sites was the rotation of W50 (Figure 4.8). In R7-2 the side chains of both K222 and S48 were pointed away from the active site, relative to R1. In R7-2, E101 moves into the active site, increasing the distance between E101 and K222 (R1: 2.9 Å, R7-2: 5.7 Å).

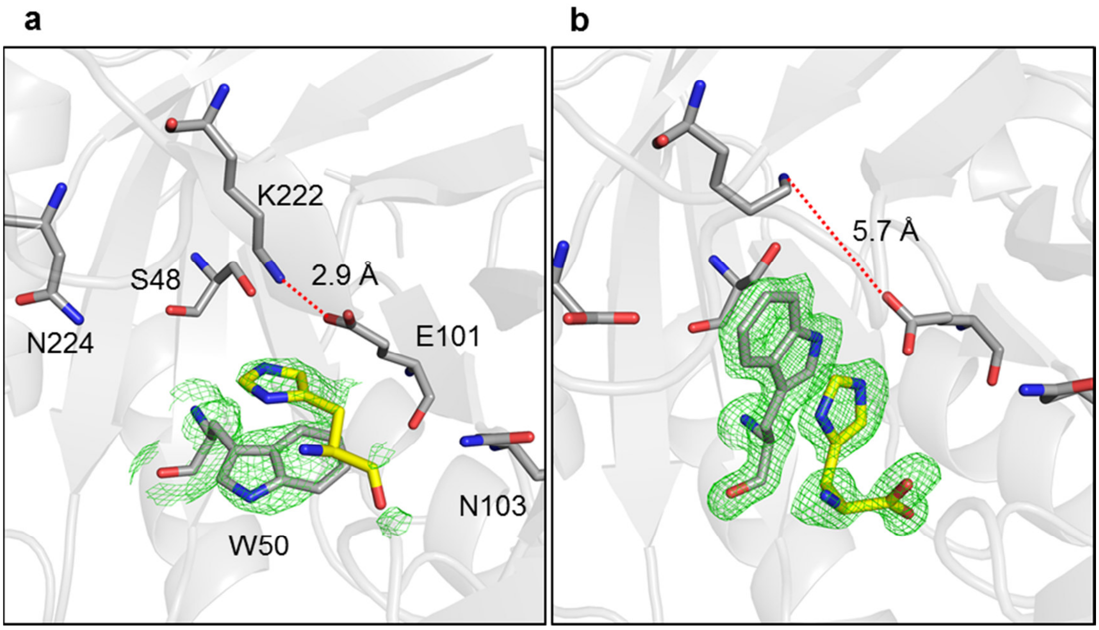


Figure 4. 8 | Comparison of KE07 R1 (designed) and the most evolved KE07 variant (R7-2). Active site residues of R1 (**a**) and R7-2 (**b**) are shown as grey sticks. The distance between E101 and K222 are shown with red dotted lines. H263 from the other subunit is shown as yellow stick. PDB ID for each structure: 4ZO8 (**a**), 5D38 (**b**). Electron densities centered on W50 and H263. The F_o-F_c omit maps are contoured at 3.0σ.

Electrostatics of R7-2

Local and long-range electrostatic networks were remodelled in R7-2 due to the alternate W50 conformation (Figure 4.9). The hydrogen bond between E101 and K222 was broken in R7-2. In R7-2, E101 forms a hydrogen bond with W50, histidine

(H263) from another subunit, and a water molecule. The water, that previously formed an H-bond with E101, is part of a long-range H-bond network extending to the surface residues (R5, E46, K99, E167). The orientation of the histidine (H263) was almost perpendicular to the original binding mode and positively charged residue to stabilize the developing negative charge on the transition state was changed from K222 at R1 to R202 at R7-2 (Figure 4.9). The orientation of H263 is probably analagous to the substrate binding orientation as the imidazole and oxazole rinds share many similarities, thus it is most likely a reasonable substrate analogue.

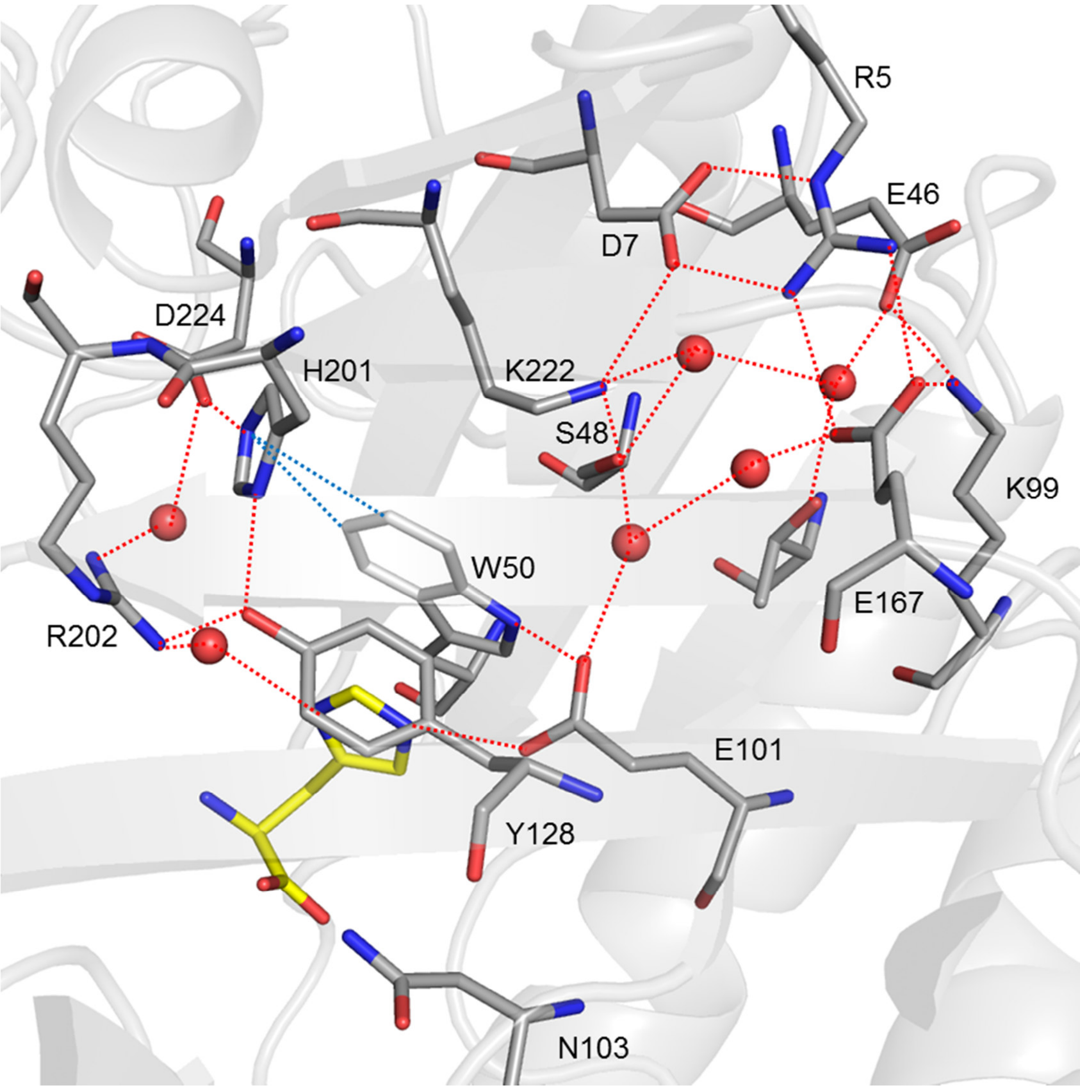


Figure 4. 9 | Electrostatic interactions in R7-2. Residues in the active site and electrostatic network behind active site are shown as grey sticks. Waters in the electrostatic network are shown as red spheres. Hydrogen bonds are shown as red dotted lines and the electrostatic interaction between W50 and H201 (3.5 Å) is shown as blue dotted lines. Histidine (H263) from the other subunit is shown as yellow stick. PDB ID: 5D38.

4.3.2 Site-directed mutagenesis

W50A: R1 vs. R7-2

To quantify the contribution of W50 to KE07 catalysis in the designed (R1) and the most evolved variant (R7-2), we performed site directed mutagenesis, mutating W50 to alanine. The proteins were purified by nickel affinity column. Figure 4.9 indicates the variants were of similar purity. The catalytic parameters of the variants are presented in Table 4.4. The W50A mutants showed significant (>95% activity loss) loss of activity of in both the designed (R1) and most evolved (R7-2) backgrounds. Interestingly, the K_M of R7-2/W50A increased 7-fold whilst the K_M of R1/W50A was unchanged. This is somewhat consistent with the reduction in K_M of R7-2 *vs.* R1 (0.76 mM *vs.* 0.45 mM, respectively; Table 4.4). The W50A mutation in R1 led to a greater loss of activity (45-fold reduction in k_{cat}) than the E101A mutation (20-fold reduction in k_{cat}) (4). However, both mutants (W50A or E101A) led to the reaction rate being very close to the background rate (4, 24), indicating that both W50 and E101 are essential for the proton transfer.

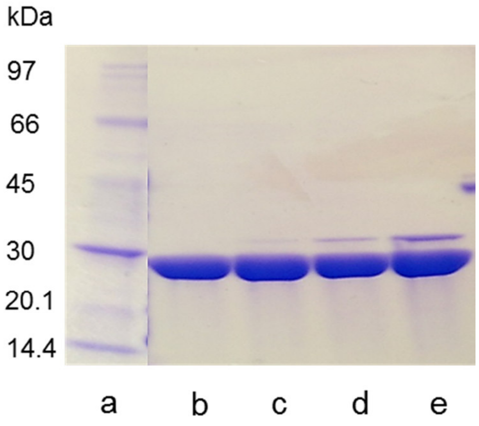


Figure 4. 10 | SDS-PAGE of KE07 variants. Low range protein ladder (**a**, Invitrogen) and KE07 variants (10µM): KE07 R1 (**b**), KE07 R1 W50A mutant (**c**), KE07 R7-2 (**d**), KE07 R7-2 W50A mutant (**e**).

Table 4. 4 | Mutagenesis of W50 to Alanine in KE07 variants.

	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M ($s^{-1}\cdot M^{-1}$)
R1	0.016 ± 0.001	0.76 ± 0.11	20.60
R1/ W50A	0.0003 ± 0.00004 (2.2 %)	0.82 ± 0.16	0.42 (2 %)
Fold change	**45** fold (-)	**1**	**49** fold (-)
R7-2	2.57 ± 0.12	0.45 ± 0.04	5769.74
R7-2/ W50A	0.04 ± 0.01 (1.7 %)	3.32 ± 1.32	13.41 (0.2 %)
Fold change	**58** fold (-)	**7** fold (+)	**430** fold (-)

The mean and standard deviations of the kinetic parameters for the mutants were calculated from three independent measurements at 298 K (pH 7.25) with 5-nitrobenzisoxazole. The relative magnitudes of the kinetic parameter for the W50A mutant compared to the non-mutated variant are given as percentages.

4.4 Summary

In this chapter, I have demonstrated that the π-stacking residue W50 adopts an alternate conformation in the later rounds of KE07 evolution (R5-R7), which correlates with the KIEs observed in Chapter 3, i.e. from a conformationally homogeneous R1, W50 becomes conformationally heterogeneous in R5, and the new conformation of W50 is stabilized by R7-2. In addition, the essential role of W50 (as much as E101) in catalysis was demonstrated in this chapter by site directed mutagenesis. The importance of the conformation of the active site tryptophan has been shown in other enzyme systems including most active antibody Kemp eliminase Ab34E4 (19, 25). In the Ab34E4 system, the substrate was bound to the productive conformation of W91. To be specific, W91 in Ab34E4 (substrate π-stacking residue) shows alternative conformations in the apo-structure and selects one conformation in complex with 5-nitrobenzisoxazole, while maintaining E50 as a general base, which is similar to the structural observation in intermediate KE07 variants here. Furthermore, alternative conformations of active site tryptophan residues have been observed in TIM-barrel xylanase (26) and other antibody catalysts (25). This indicates that conformational selection of alternative conformations of active site tryptophans to increase enzymatic activity is not limited to KE07, and may be a general evolutionary strategy to optimize active sites for catalysis.

4.5 References

1. Frushicheva MP, Cao J, Warshel A (2011) Challenges and advances in validating enzyme design proposals: The case of kemp eliminase catalysis. *Biochemistry* 50(18):3849–3858.
2. Świderek K, Tuñón I, Moliner V, Bertran J (2015) Protein flexibility and preorganization in the design of enzymes. The Kemp elimination catalyzed by HG3.17. *ACS Catal* 5(4):2587–2595.
3. Osuna S, Jiménez-Osés G, Noey EL, Houk KN (2015) Molecular dynamics explorations of active site structure in designed and evolved enzymes. *Acc Chem Res* 48(4):1080–1089.
4. Röthlisberger D, et al. (2008) Kemp elimination catalysts by computational enzyme design. *Nature* 453(7192):190–195.
5. Khersonsky O, et al. (2010) Evolutionary optimization of computationally designed enzymes: Kemp eliminases of the KE07 series. *J Mol Biol* 396:1025–1042.
6. Kabsch W (2010) Xds. *Acta Crystallogr Sect D Biol Crystallogr* 66(2):125–132.
7. Karplus PA, Diederichs K (2012) Linking crystallographic model and data quality. *Science* 336(6084):1030–1033.
8. Winn MD, et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr Sect D Biol Crystallogr* 67(4):235–242.
9. Afonine P V., et al. (2012) Towards automated crystallographic structure refinement with phenix. refine. *Acta Crystallogr Sect D Biol Crystallogr* 68(4):352–367.
10. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr Sect D Biol Crystallogr* 60(12):2126–2132.
11. Chen VB, et al. (2009) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr Sect D Biol Crystallogr* 66(1):12–21.
12. Afonine P V, et al. (2010) phenix. model_vs_data: a high-level tool for the calculation of crystallographic model and data statistics. *J Appl Crystallogr* 43(4):669–676.
13. DeLano WL (2002) The PyMOL molecular graphics system.
14. Gibson DG, Smith HO, Hutchison III CA, Venter JC, Merryman C (2010) Chemical synthesis of the mouse mitochondrial genome. *Nat Methods* 7(11):901–903.
15. Weiss MS, Hilgenfeld R (1997) On the use of the merging R factor as a quality indicator for X-ray data. *J Appl Crystallogr* 30(2):203–205.
16. Diederichs K, Karplus PA (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat Struct Biol* 4(4):269–275.
17. Diederichs K, Karplus PA (2013) Better models by discarding data? *Acta Crystallogr Sect D Biol Crystallogr* 69(7):1215–1222.
18. Meyer MP, Tomchick DR, Klinman JP (2008) Enzyme structure and dynamics affect hydrogen tunneling: The impact of a remote side chain (I553) in soybean lipoxygenase-1. *Proc Natl Acad Sci* 105(4):1146–1151.
19. Debler EW, Müller R, Hilvert D, Wilson IA (2008) Conformational isomerism can limit antibody catalysis. *J Biol Chem* 283(24):16554–16560.
20. Isom DG, Castañeda CA, Cannon BR, Velu PD, García-Moreno E B (2010) Charges in the hydrophobic interior of proteins. *Proc Natl Acad Sci* 107(37):16096–16100.
21. Khersonsky O, et al. (2012) Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed Kemp eliminase KE59. *Proc Natl Acad Sci* 109(26):10358–10363.
22. Blomberg R, et al. (2013) Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature* 503(7476):418–421.
23. Khersonsky O, et al. (2011) Optimization of the in-silico-designed Kemp eliminase KE70 by computational design and directed evolution. *J Mol Biol* 407(3):391–412.
24. Frushicheva MP, Cao J, Chu ZT, Warshel A (2010) Exploring challenges in rational enzyme design by simulating the catalysis in artificial kemp eliminase. *Proc Natl Acad Sci U S A* 107(39):16869–16874.
25. Foote J, Milstein C (1994) Conformational isomerism and the diversity of antibodies. *Proc Natl Acad Sci U S A* 91(22):10370–4.
26. Lo Leggio L, et al. (2001) Substrate specificity and subsite mobility in T. aurantiacus xylanase 10A. *FEBS Lett* 509(2):303–308.

Chapter 5: Solution and Computational Analysis of the Conformational Dynamics of KE07 and Variants

5.1 Introduction

Preamble

The alteration of the conformation of W50 in the later rounds of KE07 evolution was identified from X-ray crystal structures, as presented in Chapter 4 and is consistent with KIE analysis as presented in Chapter 3. In this Chapter, the conformational change of W50 was investigated in three ways. Firstly, the tryptophan fluorescence intensities of the KE07 variants were measured to test the prediction, based on the crystal structures, that the alternative conformation of W50 causes fluorescence quenching. Secondly, molecular dynamics (MD) simulations of KE07 variants were performed to see if we can observe the conformational change of W50 during 25 ns MD simulations. Finally, quantum mechanical calculations of the transition state were performed using active site fragments with the two different W50 positions, to determine whether the conformational change affects the ability of KE07 to stabilize the transition state.

Pre-organized active site electrostatics and fluorescence

The ability of enzymes to accelerate reactions up to the diffusion limit (1, 2) mainly originates from TS stabilization by the pre-organized electrostatics of their active sites (3–5). In addition to charged groups, many active sites also contain aromatic amino acids, including tryptophan, tyrosine, and phenylalanine. Each of these aromatic amino acids have different emission spectra (6): phenylalanine has a fluorescence emission maximum at 282 nm; tyrosine has fluorescence emission maximum at between 300 and 310 nm; tryptophan has fluorescence emission maximum between 308 and 355 nm, depending on the local environment.

Tryptophan fluorescence emission is highly sensitive to the local electric field, and tryptophan has a much greater quantum yield than tyrosine or phenylalanine (6–9). Among these intrinsic amino acid fluorophores, tryptophan is the only one that displays significant changes in its fluorescence emission spectrum depending on the local environment (6, 10). Moreover, most enzymes have only a few tryptophan residues, but many tyrosine and phenylalanine residues. Thus, the emission maxima and intensities of tryptophan fluorescence spectra have been used as probes for rotamer prediction (11), electrostatic interactions (12) or to analyse the thermostability (13) of proteins.

Tryptophan fluorescence quenching in proteins

There are two lowest singlet excited states of tryptophan ($\pi\pi^*$), ^{1}La and ^{1}Lb, and it has been shown that the ^{1}La excited state is involved in tryptophan fluorescence in all proteins (14). The electron density transfers from the pyrrole ring to the benzene ring upon photon energy absorption in the ^{1}La excited state. Ground-state interactions of tryptophan with other side chains inside pre-organized active sites of proteins significantly affect the energy and lifetime of the excited state, resulting in different quantum yields. The proximity or orientation of effective amino acid quenchers, such as glutamic and aspartic acid, cysteine, and histidine side chains to the tryptophan residue affects the excited state electron transfer rate, and thus, the fluorescence lifetime (10, 15). There are two kinds of quenching mechanisms: static (11), and dynamic quenching (16, 17), in addition to protein unfolding (13). First, static quenching occurs due to the formation of a non-fluorescent ground state complex between tryptophan and proximal side chains, whereby the excited state returns to the ground state too quickly to detect emission. Second, dynamic quenching occurs due to the different diffusion rate of quenchers, which changes the collision frequency. For example, higher temperature leads larger collision frequency between tryptophan and quenchers (e.g. diffusive oxygen or solvents), thereby increasing solvent relaxation, lowering the quantum yield (6).

5.2 Methods

Tryptophan fluorescence was measured using a Cary Eclipse Fluorescence Spectrophotometer with a Single Cell Peltier Accessory for temperature control (Agilent Technology). Tryptophan was excited at 280 nm (slit width 5 nm) and emission was monitored between 300-400 nm (1.5 nm slit width). Purified KE07 protein was diluted (1.8-5 μM) in buffer (25 mM HEPES pH 7.25, 100 mM NaCl) and the spectra were measured at 283-323 K.

Molecular dynamics simulations were performed using GROMACS 4.5.5 (18). The force field applied for simulation was GROMOS96 53a6 (19). Protonation states of histidine were selected based on the pH of the experimental assay (7.25) and their p*K*a values calculated by PROPKA3.0 (20). The coordinates of KE07 variants (R1, 4, 5, 6, 7, 7-2) were immersed into a dodecahedron-shaped box of water with a buffering distance of 1.0 nm between the protein and the edge of the box. Sodium ions were added to neutralize charge. Energies were subsequently minimized with the Steepest Descent method for 2,500 steps. The last frame of each energy-minimized structure was used as the initial frame for the MD simulation. Electrostatic interactions were calculated with the Particle-Mesh Ewald (PME) method (21). The cut-off for PME was 1.0 nm. The time step was set at 2 fs and the temperature was set at 305 K. The number of steps set to 50,000 (0.1 ns for NVT) for generating randomized velocities drawn from a distribution before position restrained, 500,000 (1 ns for NPT) for equilibrium at 1 atm, and 12,500,000 (25 ns) for the production MD simulation. MD trajectories were visualized using Pymol (22).

Quantum cluster optimization was performed using Gaussian09 (23). Initial structures were determined based on the crystal structures (PDB ID: 5D2T for R1 and 5D38 for R7-2). Transition state geometry optimizations were carried out *in vacuo* using the M06-2X (24) level of theory with a 6-31+G (d) basis set. All atoms were frozen except the substrate and the carboxyl anion. After optimizing the position of the substrate and carboxyl group of E101, further optimization was performed *in vacuo* with all atoms relaxed.

5.3 Results

5.3.1 Tryptophan fluorescence

Remodelled electrostatics of active site by tryptophan conformational change in KE07 evolution

In the early rounds of KE07 evolution (R1-4), the accumulated mutations occurred in the immediate area of the active site (N224D, G202R, I7D), whereas in the later rounds of evolution (R5-7), the mutations (V12M, K146T, F77I, F229S, I102F) primarily occurred outside the active site, as described in Chapters 3 and 4.

In the crystal structures, presented in Chapter 4, we observe an alternative conformation of W50 (>90° rotation, position B) from the designed conformation (position A) emerge in R5 and become stabilized in later rounds. The rotation of W50 changes the electrostatics of the active site, and particularly the electrostatics surrounding the fluorophore (Figure 5.1). In position A, the NE1 of the pyrrole ring of W50 is exposed to solvent (the closest charged residue was the carboxyl oxygen of D51, >4.3 Å away, far more than hydrogen bonding distance). In addition, the CZ3 of the benzene ring of W50, which is negatively charged in the ^1L_a ($\pi\pi^*$) state was <3 Å away from the negatively charged carboxylic oxygen OE2 of E101. On the other hand, in position B, a hydrogen bond was formed between the NE1 of the pyrrole ring of W50 and the OE1 of E101 (Figure 5.1). In addition, the CZ3 of W50, which is negatively charged in the ^1L_a state, was within 3.5 Å of the positively charged NE2 of H201.

The crystal structures and KIEs suggest that W50 undergoes a conformational change via a conformationally disordered intermediate in R5 & 6. However, crystal structures could be confounded by crystal packing artefacts, and the KIEs are indirect, with alternative explanations possible. Probing the local electrostatic environment of the active site through analysis of tryptophan fluorescence offers a means by which the “flipping” of W50 can be directly analysed in solution.

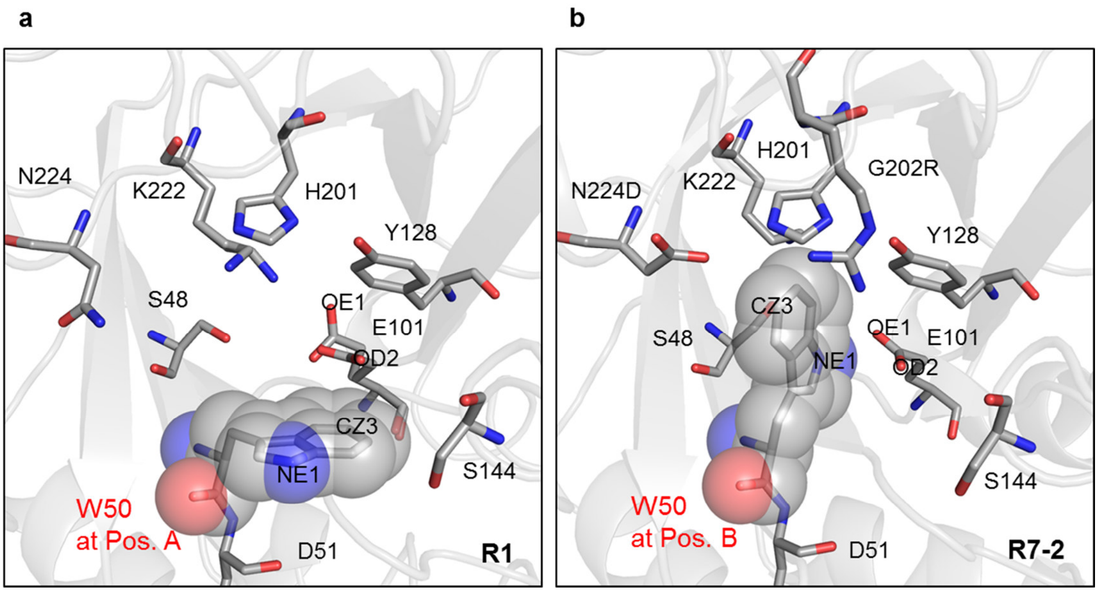


Figure 5. 1 | Pre-organized electrostatics of the KE07 active site in R1 and R7-2. The crystal structures of R1 (**a**, PDB ID: 5D2T) and R7-2 (**b**, PDB ID: 5D38) are shown with grey cartoon, with sticks for charged local residues and spheres for W50. The names of residues and specific atoms are labeled near the residues.

Fluorophores in KE07 variants and reference system (W50A)

Although there were some substitutions of tyrosine and phenylalanine residues during the KE07 evolution, tryptophan residues (W50, W156) were preserved throughout (Figure 5.2). Moreover, X-ray crystal structures of KE07 variants show no conformational change of any aromatic residues except W50. To be specific, KE07 has only one tryptophan in the highly polar active site, which is the only fluorophore residue showing a conformational change in the KE07 crystal structures, as shown in Chapter 4. The other tryptophan at position 156 is located in helix $\alpha 5$ and is surrounded by hydrophobic residues (I129, V127, V140, V160, V107). There is no change in tyrosine residues among KE07 variants, except in R7, which has one additional tyrosine at position 84 on the surface, far from active site. There were a few phenylalanine mutations outside the active site in R1 (F227L), R7 (F229S, F77I) and R7-2 (F229S, F77I, I102F) (for details see Fig 5.2).

a

	R1	R3	R4	R5	R6	R7	R7-2
W	50, 156	50, 156	50, 156	50, 156	50, 156	50, 156	50, 156
Y	128, 143, 240 (39, 182)	128, 143, 240 (39, 182)	128, 143, 240 (39, 182)	128, 143, 240 (39, 182)	128, 143, 240 (39, 182)	**+84**, 128, 143, 240 (39, 182)	128, 143, 240 (39, 182)
F	23, 38, 141, 229, 189, 138, 77, 120, 86 (49, 210, 214, ~~227~~)	23, 38, 141, 229, 189, 138, 77, 120, 86 (49, 210, 214, 227)	23, 38, 141, 229, 189, 138, 77, 120, 86 (49, 210, 214, 227)	23, 38, 141, 229, 189, 138, 77, 120, 86 (49, 210, 214, 227)	23, 38, 141, 229, 189, 138, 77, 120, 86 (49, 210, 214, 227)	23, 38, 141, ~~229~~, 189, 138, ~~77~~, 120, 86 (49, 210, 214, 227)	**+102**, 23, 38, 141, ~~229~~, 189, 138, ~~77~~, 120, 86 (49, 210, 214, 227)

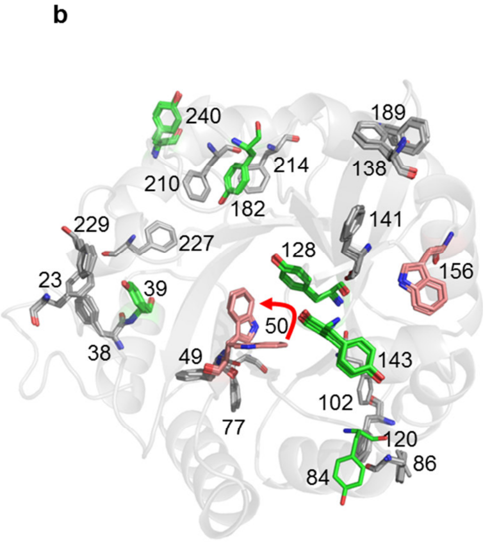


Figure 5. 2 | | Aromatic residues of KE07 variants. Residue numbers of aromatic residues of KE07 variants are listed (**a**). Additional aromatic residues are shown in bold in red. Residues that were mutated to non-aromatic residues are strikethrough in red. Non-solvent accessible residues are in brackets. Superimposed aromatic residues, including tryptophan (pink), tyrosine (green), and phenylalanine (grey), are shown as sticks on the grey cartoon of R1 (5D2W) (**b**). The conformational change of W50 observed during evolution is indicated by red arrow. PDB ID of the superimposed structures: R1 (5D2W), R5 (5D2Y), R6 (5D32), R7 (5D37), and R7-2 (5D38).

Fluorescence intensity with and without W50

Tryptophan fluorescence was measured for all KE07 variants over the same temperature range used for kinetics (283–323 K) (Figure 5.3). Site directed mutagenesis was performed to construct W50A mutations in the R1 and R7-2 backgrounds, which were used as a reference system. Although the fluorescence intensity of R7-2/W50A was about 20% lower than R1/W50A, as expected due to the addition of two phenylalanine residues in R7-2, the maximum wavelength and shape were identical between W50A mutants of R1 and R7-2 ($\lambda_{\max}$ = 320 nm, width ~50 nm at ½ I_o). In addition the two W50A mutants had temperature independent fluorescence emission intensities (I_o), which fluctuated (<10%) over the temperature range (283–323 K), suggesting that the protein remains folded over the temperature range. Therefore, the tryptophan fluorescence changes observed in the KE07 variants (R1 to R7-2) likely reflect local environmental changes of W50. From round R5 to

R7-2 there were no mutations within the active site and changes in fluorescence between these rounds likely reflects changes in the tryptophan rotamer population.

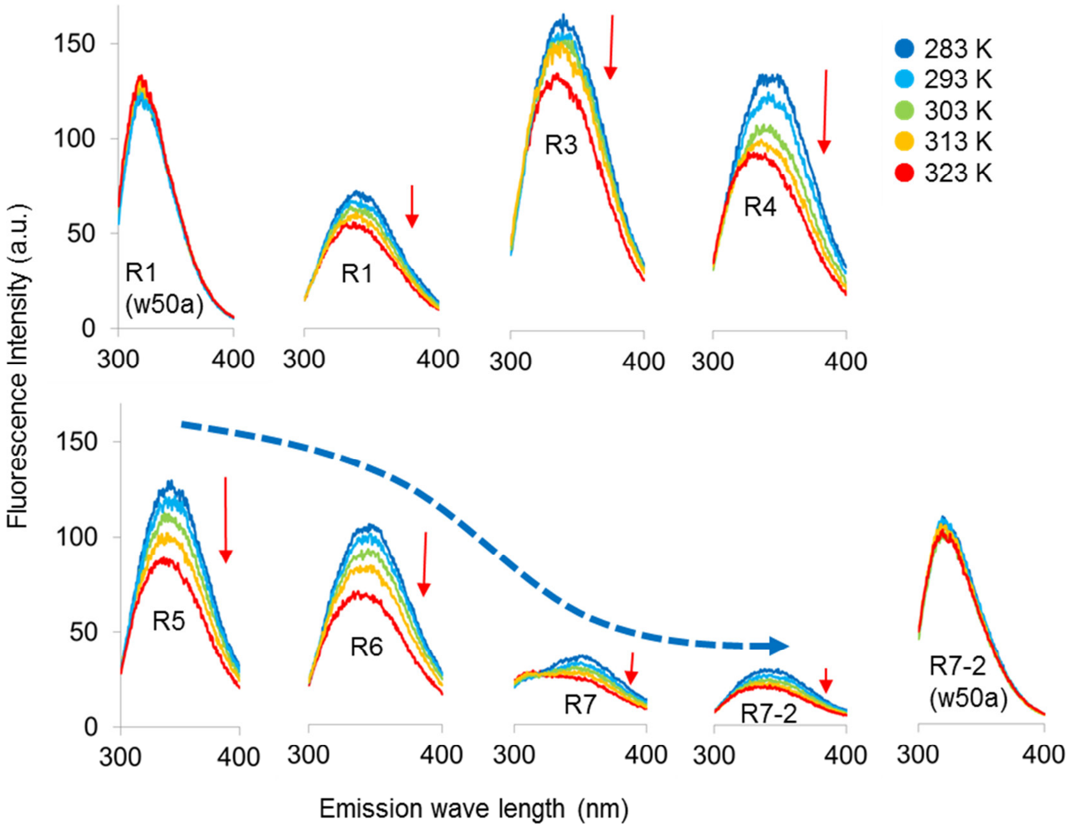


Figure 5. 3 | Tryptophan fluorescence of KE07 variants. Fluorescence emission spectra (excitation at 280 nm) of KE07 variants (1.8-5.0 μM) and W50A mutants (R1/W50A and R7-2/W50A) were measured at 283-323 K (temperatures for each spectra are indicated with different colors). The detailed F_{max} and λ_{max} data are presented in Table 5.1. The static fluorescence quenching by tryptophan conformational change is indicated by the blue dotted line. The thermodynamic quenching is shown in red arrows. The fluorescence emission intensities were concentration-normalized (R1/W50A, R1, R6, R7, R7-2, and R7-2/W50A were at 5 μM, R3 and R4 were at 3.5 μM, and R5 was at 1.8 μM).

The fluorescence spectra without W50 were narrow (width ~50 nm at ½ I_o , λ_{max} = 320 nm) and had temperature independent intensities (I_o), which fluctuated (<10%) over the temperature range (283–323 K). On the other hand, spectra of the KE07 variants with W50 showed three overall differences. First, spectra were red-shifted about 20 nm (λ_{max} = 340 nm at ½ I_o). Second, they had broader spectra (width ~70 nm at ½ I_o). Last, the intensity (I_o) became temperature dependent whereby the intensity decreased about 22-33% at 323 K relative to at 283 K of KE07 variants with presence of W50 (%ΔF_{max}, Table 5.1, Figure 5.3). Unlike W156 (buried and surrounded by hydrophobic residues), W50 which is partially solvent accessible and surrounded by ionic polar side chains (e.g. S48, E101, K222 and so on) would be a favourable pathway for fluorescence transfer in KE07. Thus, the significant spectral difference with and without W50 suggested that the fluorescence of other residues including W156 transfers to W50 and W50 becomes the major contributor of fluorescence relaxation. This may account for the lower intensity of the KE07 variants than the W50A mutants of R1 and R7-2.

Table 5. 1 | Tryptophan fluorescence parameters

KE07 variants		Temperature (K)					ΔF_{max}[c] (283 – 323 K)	% ΔF_{max}[d]
		283	293	303	313	323		
R1 (w50a)	F_{max}[a]	124	124	126	131	133	-9	-8%
	λ_{max}[b]	321	320	320	318	320		
R7-2 (w50a)	F_{max}	107	106	99	103	101	7	6%
	λ_{max}	320	322	321	322	319		
R1	F_{max}	71	66	63	60	55	16	23%
	λ_{max}	338	334	334	338	338		
R3	F_{max}	163	154	150	148	133	30	19%
	λ_{max}	340	338	341	338	335		
R4	F_{max}	132	123	106	99	92	40	30%
	λ_{max}	337	342	338	336	339		
R5	F_{max}	129	121	113	102	89	40	31%
	λ_{max}	340	340	336	338	336		
R6	F_{max}	103	98	91	82	69	35	33%
	λ_{max}	345	347	345	347	337		
R7	F_{max}	35	32	29	27	24	11	32%
	λ_{max}	352	350	348	348	348		
R7-2	F_{max}	29	26	24	22	20	8	29%
	λ_{max}	338	338	336	342	342		

[a] Fluorescence maximum intensity and [b]corresponding wave length

[b] $\Delta F_{max} = F_{max}$ (283 K) – F_{max} (323 K): Intensity difference between 283 K and 323 K

[c] $\% \ \Delta F_{max} = \frac{\Delta F_{max}\ (283 - 323\ \mathrm{K})}{F_{max}(283\ \mathrm{K})} \times 100$

Temperature dependent fluorescence intensity in KE07 evolution

The degree of thermodynamic quenching was the lowest in R1 and R3 (23% and 19% respectively) and from R4 the degree of quenching varied little (29 -33%) (Table 5.1, Figure 5.4). The active site residues are conserved from R4 until R7-2. The observed thermodynamic quenching is seemingly due to temperature promoted collision of W50 with solvent or side chains nearby. For example, the backbone amide of G81 is 3.3 Å away from the benzene ring of the W50 (Figure 5.6). The intensity was temperature independent in the absence of W50, and the enzyme remained active over the temperature range (Chapter 3), suggesting that temperature induced unfolding was not responsible for the observed changes in fluorescence emission. Increasing the population of the alternate W50 rotamer due to increased temperature is also unlikely as an underlying mechanism because the temperature induced quenching was observed in all KE07 variants. In addition, R7-2 showed only the non-fluorescence conformation of W50 in the crystal structure, yet it showed the same degree of quenching with R4-R7. Thus, the temperature induced quenching is seemingly thermodynamic quenching and the amount of the quenching (%$\Delta F_{\max}$, Table 5.1, Figure 5.6) was correlated with the active site composition rather than configuration.

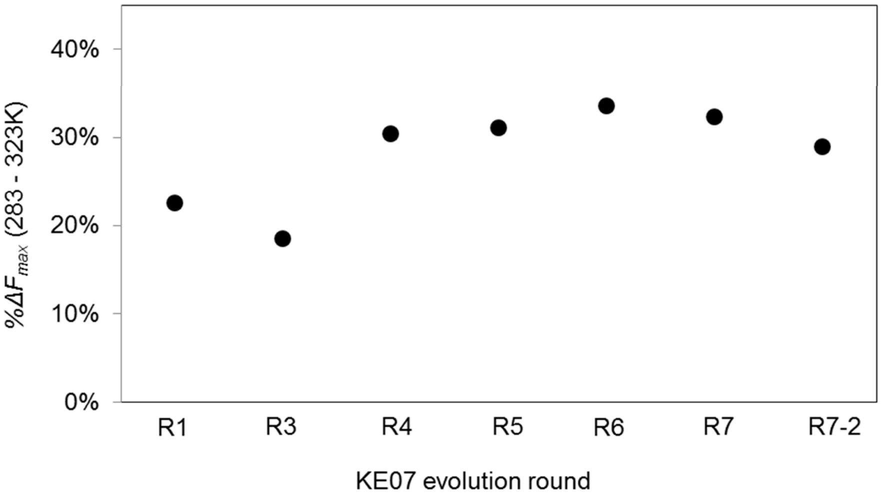


Figure 5. 4 | The relative temperature dependent fluorescence intensity change during KE07 evolution. $\%\Delta F_{\max} = \frac{\Delta F_{max}\,(283-323K)}{F_{max}(283K)} \times 100$. Detailed $F_{\max}$ and $\lambda_{\max}$ at each temperature are presented in Table 5.1.

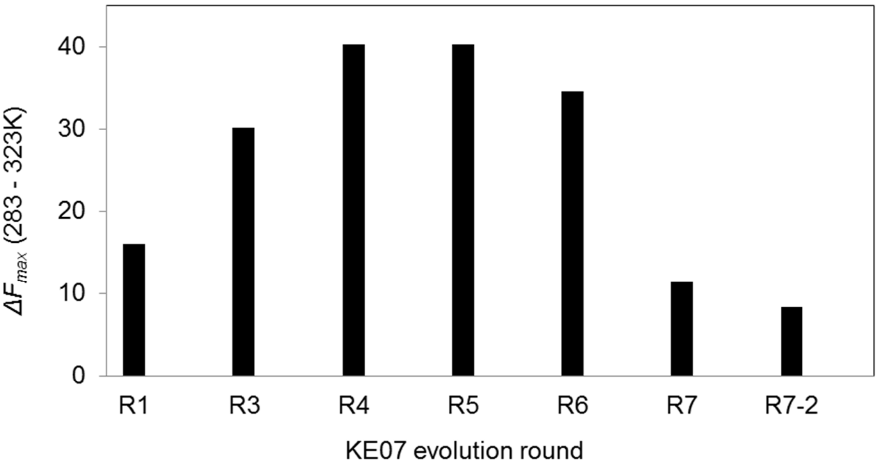


Figure 5. 5 | The temperature dependent fluorescence intensity change during KE07 evolution. The wavelength at the maximum intensity (λ_{max}) for each KE07 variant is presented in Table 5.1. $\Delta F_{max} = F_{max}\,(283K) - F_{max}\,(323K)$.

Fluorescence intensity in R1-R4

In contrast to the temperature dependence of the fluorescence intensity, we observe large changes in the overall fluorescence intensity of the variants. Over the first four rounds of evolution, the majority of the mutations are located within, or proximal to, the active site. These mutations largely account for the changes in fluorescence intensity in the early rounds of directed evolution. For example in R3 there are three mutations (N224D, G202R, I7Q) relative to R1, which may affect the local environment of W50. The N224D and G202R mutations formed a new electrostatic network and decreased the solvent accessibility of the active site (Chapter 3). The I7Q mutation may interact with S48 and K222 which form hydrogen bonds with catalytic E101 in the R1 protein (Chapter 4). Together the three mutations likely increase the nucleophilicity of E101 resulting in the increased catalytic activity of KE07 R3 relative to R1 (Chapter 3). Glutamic acid 101 is the closest charged residue to W50; the benzene ring of W50 is 2.8 Å away from the carboxylic oxygen of E101 (Figure 5.1).

Upon excitation of tryptophan, the electron density is shifted from the pyrrole ring to the benzene ring (12, 14, 15). Therefore, the increased charge of the E101 at R3 would result in depolarization of W50 and reduced relaxation by charge repulsion leading to increased fluorescence emission (Figure 5.3). The residue at position 7 changes twice; I7 at R1 was mutated to polar glutamine at R3 and subsequently mutated to the aspartic acid at R4. D7 formed an electrostatic network with R5 and K222 (Figure 5.6). The benzene ring of W50 is 3.3 Å away from backbone amide of G81, which may be an electron transfer pathway from W50 upon tryptophan excitation. The backbone amide of G81 is linked to T78 on β3 by peptide bonds. The side chain oxygen of T78 is part of an electrostatic network behind the active site (T78-water-E46-R5-D7-K222-S48, Figure 5.6). Therefore, the newly introduced aspartic acid (I7D mutation) may be indirectly connected to the W50 through a long-range electrostatic network, which may contribute to the decreased intensity at R4 compared to R3 (Figure 5.3) due to facilitated electron transfer-tryptophan fluorescence quenching through backbone amide.

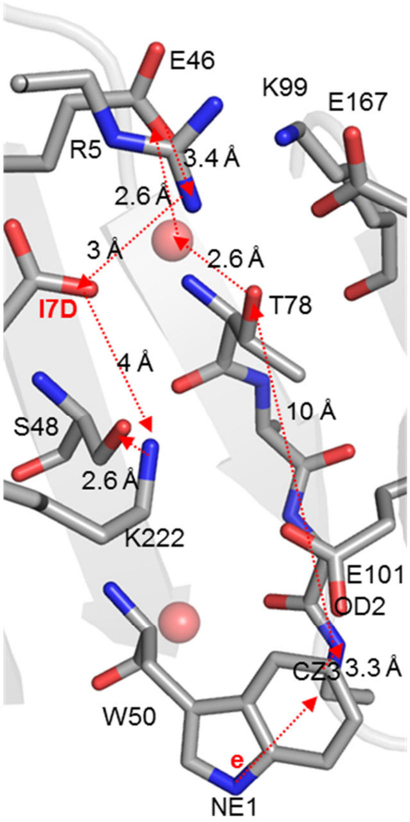


Figure 5. 6 | Possible additional electron pathways upon tryptophan excitation at R4 by I7D mutation. Residues within the putative electron transfer pathways are shown in grey sticks. Possible additional electron pathway upon tryptophan excitation by I7D mutation at R4 is shown in red dotted arrows. The distances are shown near each arrow. The PDB ID for the structure is 3IIO (chain A). Waters are shown in red spheres.

Fluorescence intensity in later round (R5 – R7-2)

The increase in fluorescence intensity due to W50 can be rationalized through the changes in the active site through R1-4, but there are no further changes to the local environment of W50 over R5-7-2. Despite this, we observe a consistent decrease in the fluorescence over these final rounds of evolution. This suggests that the environment around W50 is changed due to another effect, such as W50 changing its conformation. Crystal structures in the later rounds showed different rotamer population of W50 and E101, regulated by remote mutations. In particular, the NE1 atom of the W50 rotamer (position B) forms a hydrogen bond with E101, forming a non-fluorescent ground-state complex. Thus, any changes in the tryptophan fluorescence in later rounds of the KE07 evolution are expected to be entirely due to local conformational changes. Our result showed that the tryptophan fluorescence intensity decreased gradually in the later rounds of KE07 evolution, suggesting gradual enrichment of the non-fluorescent complex by increasing population of W50 position B. This is exactly what is observed in the crystal structures (Chapter 4), providing consistent data in solution and in the crystal lattice, and suggesting that this conformational variability is unlikely to be a crystallization artefact.

To summarize, we demonstrated a change of the tryptophan fluorescence spectra of the variants by proximal and distant mutations by using KE07 evolutionary variants. In particular, the increasing “B” rotamer population of W50 was demonstrated by observing decreasing fluorescence intensity in later round of KE07 evolution in the absence of any direct active site effects.

5.3.2 Molecular dynamics simulation of KE07 variants

To test whether MD simulation can predict the effect of mutations on the rotamer population of W50, we performed 25 ns molecular dynamics simulations of a series of KE07 variants. Crystal structures of KE07 variants with W50 in position A were used as initial structures, except for R7-2, which exhibited only one conformer of W50, position B. One chain was used for all simulations because all KE07 variants were shown to be monomeric in solution (25). Substrate was not used for the simulations as crystal structures showed that both conformers of W50 are present in the apo structure (Chapter 4). Apo structures were used to rule out any "induced-fit" effects and to confirm that the sampling of W50 is a pre-existing, equilibrium, state.

The hydrogen of the 3'C of the substrate transfers to the carboxylic oxygen of the E101 in the reaction catalysed by KE07. In the reaction, W50 stabilizes the substrate through π-stacking. Thus, altered pre-organization of W50 and E101 would affect the hydrogen transfer mechanism by changing the C-H-O reaction coordinate and thus kinetics (e.g. k_{cat} and KIE). Furthermore, it has been shown that the stability of catalytic residues is important for efficient catalysis in highly ionic active site environments (26).

The 25 ns MD simulations were performed at 305 K. We were able to observe sampling of the alternative conformation of W50 (position B) in later rounds (R5-R7), while we never observed the alternative conformation of W50 in early round variants of KE07 (R1, R4) (Figure 5.7). R7-2, which started with W50 in position B, did not show W50 in position A, consistent with the crystal structures. Position B of W50 is perpendicular to that of position A (Figure 5.7**g**). The substrate is stacked against W50 (Chapter 3, 4). Therefore, the emergence of position B and reduced sampling of position A suggest the emergence of a new substrate binding position and reduced sampling of the designed substrate binding position in later rounds (R5-7).

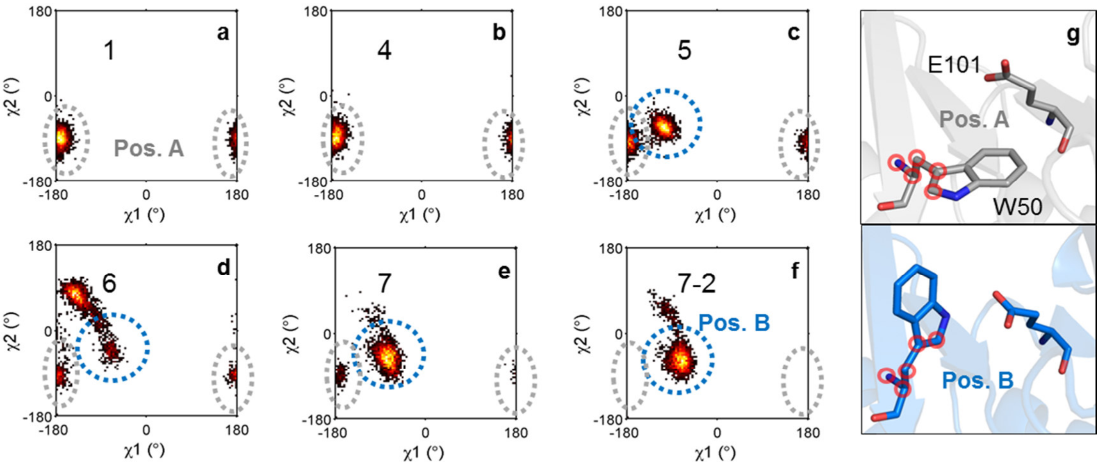


Figure 5. 7 | Population of chi1 versus chi2 angles of W50 in 25 ns molecular dynamics simulations. KE07 R1 (**a**), R4 (**b**), R5 (**c**), R6 (**d**), R7 (**e**), and R7-2 (**f**) are shown. All W50 started from at position A (grey stick, **g**) except R7-2 which started at position B (blue stick, **g**). The atoms used for dihedral angle analysis are highlighted by red circles in the crystal structure of R1 (grey stick, PDB ID: 4ZO8) *vs*. R7-2 (blue stick, PDB ID: 5D38) (**g**) $\chi1$: N:CA:CB:CG, $\chi2$: CA:CB:CG:CD1 of W50.

The carboxylates of the catalytic residue E101, which is proximal to W50 (<3 Å) in all KE07 variants, underwent subtle changes in position and flexibility. The 25 ns MD simulation showed the motion (number of discrete conformations sampled) of E101 increases in the early round of KE07 evolution (R1-4) (Figure 5.8**ab**). The increased motion of E101 at R4 compared to R1 presumably is due to the loss of a salt bridge between E101 and K222 by I7D mutation at R4 in the increased local charges by N224D and G202R mutations (Chapter 3, Figure 3.9). This happened at the same time as a decrease in E_a ($\propto \Delta H^{\ddagger}$) and ln$A$ ($\propto -T\Delta S^{\ddagger}$) at R4 (Chapter 3, Figure 3.10). Therefore, those results suggest that the increased charge of E101 and the local environment reduced the activation enthalpy ($E_a \propto \Delta H^{\ddagger}$) and at the same time destabilized E101 and thus destabilized transition state resulting in increased entropic cost at R4 (= decreased ln$A$).

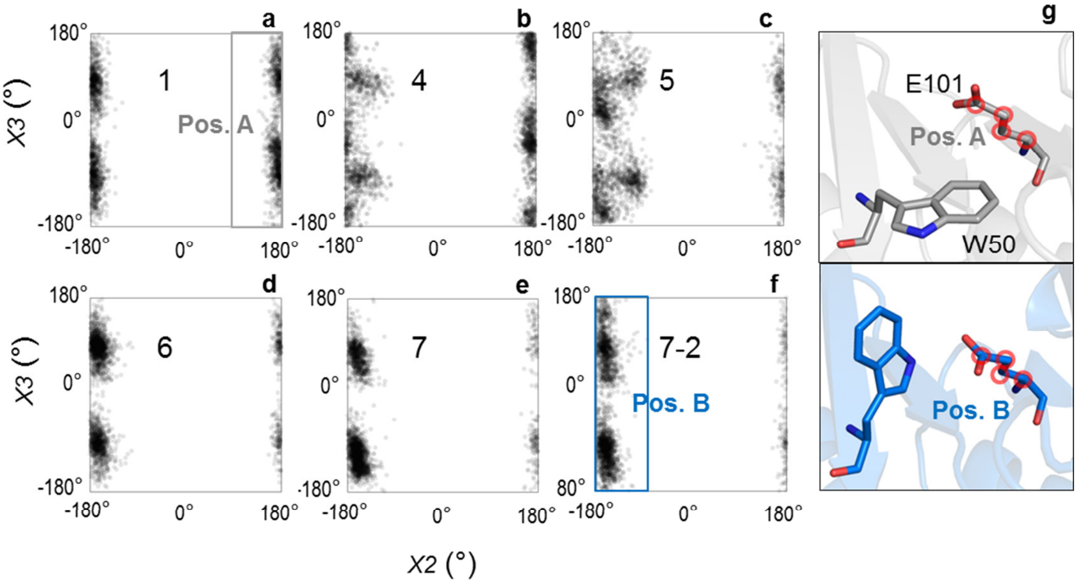


Figure 5. 8 | Population of chi2 versus chi3 angles of E101 in 25 ns molecular dynamics simulations. KE07 R1 (**a**), R4 (**b**), R5 (**c**), R6 (**d**), R7 (**e**), and R7-2 (**f**) are shown. The atoms used for chi2 dihedral angle analysis are highlighted by red circles in the crystal structure of R1 (grey, PDB ID: 4ZO8) *vs.* R7-2 (blue, PDB ID: 5D38) (**g**) $\chi2$: CA:CB:CG:CD, $\chi\mathbf{3}$: CB:CG:CD:OE2 of E101.

The 25 ns MD simulation showed the population of chi2 ($\chi2$) = 170° position of E101 was decreased and the motion around chi2 ($\chi2$) = 170° position of E101 was further increased at R5 (Figure 5.8**c**). In addition to the change of E101 sampling, position B of W50 emerged at R5 (Figure 5.7**c**), implying the new substrate-binding mode emerges at R5. Together, the changes in E101 and W50 orientation presumably change the C-H-O coordinate and thus kinetics and KIEs at R5. When W50 is in position B, the pyrrole nitrogen likely forms a hydrogen bond with the carboxylic oxygen of E101 (Chapter 4), which probably decreases the charge of E101 and stabilizes W50 and thus stabilizes the substrate. Those changes coincide with the subtle increases in E_a and ln$A$ and significant decrease in KIE $E_{a(D)} - E_{a(H)}$ and significant increase in KIE A_H/A_D at R5 compared to R4.

The MD simulation showed that the motion of the catalytic E101 residue decreased as the population of position A of W50 decreased (Figure 5.7, 5.8) in the later rounds (R5-7). This coincides with the increase in KIE $E_{a(D)} - E_{a(H)}$ and the decrease in KIE A_H/A_D resulting in restoration of the values at R7 to levels seen in R1-4 (Chapter 3). The stabilization of catalytic E101 (or the equivalent catalytic D or

D-His) during KE07 evolution observed here has been shown in other *de novo* Kemp eliminases as well as in MD simulations (27–29).

To summarize, MD simulations could recapitulate the crystallographic, spectroscopic (fluorescence) and kinetic results. For example, in the MD simulations, E101 showed increased motion at R4 which correlates with decreased E_a and $\ln A$ in R4 compared to R1 (Chapter 3). MD simulations of R5 KE07 showed the emergence of position B of W50. The sampling of E101 becomes more heterogeneous from R1-R4, which corresponds with a decrease in the activation entropy (included in $\ln A$). In the later rounds (5-7), the E101 motion was stabilized and the original positions of E101 and W50 were diminished in the MD simulations.

5.3.3 Substrate modelling: Transition state optimization

The conformational fluctuations observed *via* crystallography, fluorescence spectroscopy and MD establish that there is a consistent trend throughout the evolutionary trajectory for W50 to change its primary conformation. This correlates with the increase in activity across the trajectory. However, it does to directly establish that the new flipped active site can catalyse the Kemp elimination reaction. Accordingly, the dependence of the substrate position and the relative energy of the transition state (TS) in KE07 on the conformation of W50 was evaluated using quantum chemical (QC) calculations at the M06-2X/6-31+G(d) level of theory, in collaboration with Prof. Michelle Coote and co-workers. The product and the histidine binding modes in the crystal structures (PDB ID: 5D2T, 3D38, see Chapter 4) were used for the initial positioning of the substrate into the active site with W50 in positions A (Figure 5.9**a**) and B (Figure 5.9**b**). Substrate binding residues (W50, Y128, H201, D224, N103, S48) and catalytic residues (E101, R202, K222) were included in the quantum cluster model for the TS optimization (Figure 5.9**ab**). To start with, the substrate and the side chain of E101 were allowed to move while the other residues were frozen during TS optimization. The converged transition state geometries show that the conformational change of W50 remodels the substrate binding site (Figure 5.9**cd**). The energy difference between the two TS geometries is ~9 kJ/mol in favour of the new "flipped" conformation, showing that the transition state is indeed stabilized during KE07 evolution. Subsequently, the system was reduced to a less computationally-intensive size (53 heavy atoms) and optimized further with no constraints (Figure 5.9**ef**). During the full relaxation, the R1-like system required significant reorganization to stabilize the TS (and appeared to be adopting a "flipped"-like state), while the evolved system changed very little, implying that the evolved conformation is much better pre-organized to stabilize the TS geometry.

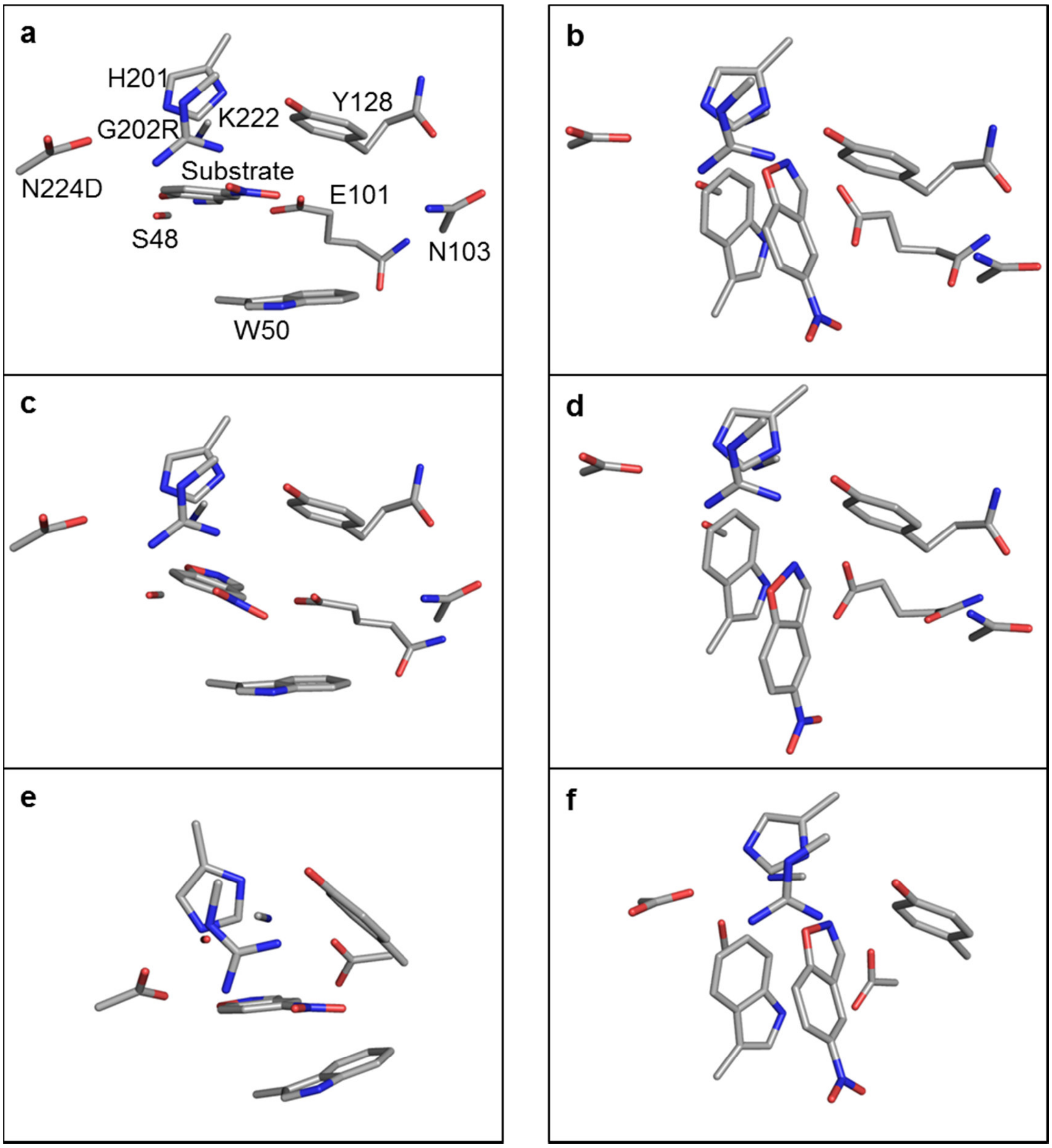


Figure 5. 9 | Transition state optimizations with W50 in position A and B using the KE07 active site fragments and the substrate. The designed model system is presented on the left side panel and evolved model is presented on the right side panel. The initial KE07 active site fragments based on the crystal structures with W50 in position A (**a**) and in position B (**b**) were presented on the top. Optimized transition states with all atoms frozen except the reactive core are shown under the each initial structure (**c, d**). Further optimized transition states with no constraint using reduced size are shown at the bottom (**e, f**). M06-2X/6-31+G(d) level of theory was used for all calculations. *** Figure 5.9 was made based on the results provided by Prof. Michelle Coote and co-workers, Research School of Chemistry, Australian National University.**

5.4 Summary

In this chapter, the alteration of the conformation of W50 during the evolution of KE07 was evaluated in three ways: tryptophan fluorescence intensity, molecular dynamics simulations, and quantum chemical calculations. Firstly, the increased quenching of tryptophan fluorescence in later rounds 5-7-2, despite the absence of any mutations in the active site, suggests that the alternative conformation of W50 is gradually enriched *via* remote mutations over the evolutionary trajectory. Secondly, the alternative conformation of W50 was only observed in the later rounds of KE07 evolution (from R5). In addition, the catalytic base (E101) showed the increased degree of motion in R4 comparing to R1, consistent with the increasing entropic cost of the reaction in R1-4. Finally, transition state geometries and thermodynamic energies with the designed and evolved conformations of W50 demonstrate that the alteration of the conformation of W50 changes the substrate binding mode and decreases the transition state energy. In conclusion, the remote mutations accumulated in later rounds of KE07 evolution gradually increase sampling of the alternative conformation of W50 to better stabilize the transition state, resulting in an enhancement of k_{cat}.

5.5 References

1. Putman SJ, Coulson AFW, Farley IRT, Riddleston B, Knowles JR (1972) Specificity and kinetics of triose phosphate isomerase from chicken muscle. *Biochem J* 129:301–310.
2. Christensen H, Martin MT, Waley SG (1990) Beta-lactamases as fully efficient enzymes. Determination of all the rate constants in the acyl-enzyme mechanism. *Biochem J* 266(3):853–861.
3. Villa J, Warshel A (2001) Energetics and dynamics of enzymatic reactions. *J Phys Chem B* 105(33):7887–7907.
4. Ram Prasad B, Warshel A (2011) Prechemistry versus preorganization in DNA replication fidelity. *Proteins: Struct, Funct, Bioinf* 79(10):2900–2919.
5. Warshel A (1998) Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites. *J Biol Chem* 273(42):27035–27038.
6. Lakowicz JR (2006) *Principles of fluorescence spectroscopy* doi:10.1007/978-0-387-46312-4.
7. Callis PR, Vivian JT (2003) Understanding the variable fluorescence quantum yield of tryptophan in proteins using QM-MM simulations. Quenching by charge transfer to the peptide backbone. *Chem Phys Lett* 369(3–4):409–414.
8. Callis PR, Liu T (2004) Quantitative prediction of fluorescence quantum yields for tryptophan in proteins. *J Phys Chem B* 108:4248–4259.
9. Burstein EA (1976) Luminescence of protein chromophores (model studies). *Adv Sci Technol (Itogi Nauk i Tekhniki), ser Biophys* 6:60–83.
10. Chen Y, Barkley MD (1998) Toward understanding tryptophan fluorescence in proteins. *Biochemistry* 37(28):9976–9982.

11. Gasymov OK, Abduragimov AR, Glasgow BJ (2012) Tryptophan rotamer distribution revealed for the α-helix in tear lipocalin by site-directed tryptophan fluorescence. *J Phys Chem B* 116(45):13381–13388.
12. Vivian JT, Callis PR (2001) Mechanisms of tryptophan fluorescence shifts in proteins. *Biophys J* 80(5):2093–2109.
13. Weichel M, Bassarab S, Garidel P (2008) Probing thermal stability of MAbs by intrinsic tryptophan fluorescence. *Bioprocess Int* (5):42–52.
14. Callis PR (1997) [7] 1 L a and 1 L b transitions of tryptophan: Applications of theory and experimental observations to fluorescence of proteins. *Methods Enzymol* 278:113–150.
15. Osysko AP, Muíño PL (2011) Fluorescence quenching of tryptophan and tryptophanyl dipeptides in solution. *J Biophys Chem* 2(3):316–321.
16. He Z, Chen J, Moser SE, Jones OG, Ferruzzi MG (2016) Interaction of β-casein with (−)-epigallocatechin-3-gallate assayed by fluorescence quenching: effect of thermal processing temperature. *Int J Food Sci Technol* 51(2):342–348.
17. Li H, et al. (2016) Caffeic acid phenethyl ester exhibiting distinctive binding interaction with human serum albumin implies the pharmacokinetic basis of propolis bioactive components. *J Pharm Biomed Anal* 122:21–28.
18. Pronk S, et al. (2013) GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29(7):845–854.
19. Oostenbrink C, Villa A, Mark AE, Van Gunsteren WF (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force field parameter sets 53A5 and 53A6. *J Comput Chem* 25(13):1656–1676.
20. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH (2011) PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *J Chem Theory Comput* 7(2):525–537.
21. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An N.log (N) method for Ewald sums in large systems. *J Chem Phys* 98(12):10089–10092.
22. DeLano WL (2002) The PyMOL molecular graphics system.
23. Frisch M, et al. (2009) Gaussian 09, Revision A. 02, Gaussian. *Inc, Wallingford, CT* 200.
24. Zhao Y, Truhlar DG (2008) The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other function. *Theor Chem Acc* 120(1–3):215–241.
25. Röthlisberger D, et al. (2008) Kemp elimination catalysts by computational enzyme design. *Nature* 453(7192):190–195.
26. Isom DG, Castañeda CA, Cannon BR, Velu PD, García-Moreno E B (2010) Charges in the hydrophobic interior of proteins. *Proc Natl Acad Sci* 107(37):16096–16100.
27. Khersonsky O, et al. (2012) Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed Kemp eliminase KE59. *Proc Natl Acad Sci* 109(26):10358–10363.
28. Khersonsky O, et al. (2011) Optimization of the in-silico-designed Kemp eliminase KE70 by computational design and directed evolution. *J Mol Biol* 407(3):391–412.
29. Blomberg R, et al. (2013) Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature* 503(7476):418–421.

Chapter 6

Conformational tinkering drives evolution of a promiscuous activity through indirect mutational effects

Yang, G.*, **Hong, NS** *, Baier, F., Jackson, C.J., Tokuriki, N.

Submitted to Biochemistry 2016, May

This manuscript was submitted to Biochemistry in May 2016. Nobuhiko Tokuriki and co-workers, Gloria Yang and Florian Baier, designed and performed directed evolution, kinetics and mutagenic characterisations. I purified and crystallized the proteins, solved the structures, performed docking and molecular dynamic simulations, provided corresponding figure designs or figures, and wrote the experimental methods with guidance and under supervision of Colin Jackson. Paul Carr collected crystallographic data and helped refinement process. Galen Correy helped in the docking using AutoDock Vina. The paper was written by N. Tokuriki and C. Jackson, with input from all authors, including me.

Conformational tinkering drives evolution of a promiscuous activity through indirect mutational effects.

Gloria Yang[1,3], Nansook Hong[2,3], Florian Baier[1], Colin J Jackson[2,*] and Nobuhiko Tokuriki[1,*]

[1] Michael Smith Laboratories, University of British Columbia, Vancouver, BC, Canada.
[2] Research School of Chemistry, Australian National University, Canberra, ACT, 0200, Australia.
[3]These authors contributed equally to this work

*To whom correspondence should be addressed: Nobuhiko Tokuriki, Ph. D., Michael Smith Laboratories, University of British Columbia, Vancouver, V6T 1Z4, BC, Canada, Telephone: +1-604-822-8156, Fax: +1-604-822-2114, Email: tokuriki@msl.ubc.ca, Colin Jackson, Ph. D., Research School of Chemistry, Australian National University, Canberra, ACT, 0200, Australia, Telephone: +61-2-61258325, Fax: +61-2-6125-0750, Email: colin.jackson@anu.edu.au

ABSTRACT

How remote mutations can lead to changes in enzyme function at a molecular level is a central question in evolutionary biochemistry and biophysics. Here, we combine laboratory evolution with biochemical, structural, genetic and computational analysis to dissect the molecular basis for the functional optimization of phosphotriesterase activity in a bacterial lactonase (AiiA) from the metallo-β-lactamase (MBL) superfamily. We show that a thousand-fold increase in phosphotriesterase activity is caused by a more favourable catalytic binding position of the paraoxon substrate in the evolved enzyme that resulted from conformational tinkering of the active site through remote mutations. A non-mutated active site residue, Phe68, was displaced by ~3 Å through the indirect effects of two peripheral mutations, enabling molecular interactions between the residue and paraoxon. Comparative mutational scanning, *i.e.* examining the effects of alanine-mutagenesis on different genetic backgrounds, revealed significant changes in the functional roles of Phe68 and other non-mutated active site residues caused by the indirect effects of trajectory mutations. Our work reveals the impact of remote mutations on the catalytic contributions of non-mutated resides, which is the result of an intramolecular network of strong epistatic mutational relationships between active site and remote residues. Defining these long-range conformational and functional epistatic

relationships has allowed us to better understand the subtle, but cumulatively significant, role of remote mutations in evolution.

INTRODUCTION

The evolution of new enzymatic functions requires remodelling of the active site in order to accommodate and properly position the new substrate. Direct contributions of mutations, *i.e.*, mutations that lead to gain or loss of interaction(s) with the substrate, can cause a significant change in enzyme function[1,2]. Direct contributions can be often rationally elucidated and many tools have been developed to predict changes in enzyme function[3,4]. However, it has been observed in enzyme evolution studies that function-altering mutations can occur at residues that are distant from the active site. These residues do not directly interact with the substrate, and thus affect catalytic activity through indirect means[1,2,5-9]. Such indirect contributions of mutations often improve catalytic activity through "molecular tinkering"[10,11], including fine-tuning the shape of the active site[12] and the position of catalytic residue(s)[6], catalytic metal ion(s)[7,13], and/or altering the dynamics of the enzyme[14]. Although the importance of indirect contributions during enzyme evolution is widely recognized, it is still challenging to elucidate and predict indirect mutational effects[15]. Moreover, no study has provided direct evidence and quantitative experimental measurement of the extent by which remote mutations affect catalytic function through fine-tuning of active site residue(s). Elucidating the structural and functional role of remote mutations and quantifying their contributions to catalytic activity is crucial for our understanding of sequence-structure-function relationships.

Laboratory (or directed) evolution has been successfully employed to gain fundamental insights in the molecular mechanisms of enzyme evolution[16]. Strong selection pressure can be utilized to prevent the fixation of neutral mutations, thereby reducing the number of mutational steps required for the adaptation to a selective pressure. In this study, we used laboratory evolution to generate an evolutionary trajectory toward phosphotriesterase (PTE) activity from an N-acyl homoserine lactonase in the metallo-β-lactamase (MBL) superfamily. Bacterial phosphotriesterases evolved within the last half-century to hydrolyze the P-O bond of xenobiotic organophosphate (OP) pesticides, such as paraoxon and parathion[17-19]. To date, two major classes of bacterial phosphotriesterases have been identified, one belonging to the amidohydrolase (AH) superfamily and one to the MBL superfamily[20]. While the evolutionary origin and mechanisms of phosphotriesterases in the AH superfamily have been extensively studied[20-24], comparatively little is known about phosphotriesterases from the MBL superfamily. Methyl-parathion hydrolase (MPH) was the first phosphotriesterase identified in the MBL superfamily[25,26]. Phylogenetic and enzymatic characterizations of MPH homologous enzymes have indicated that, as with phosphotriesterases from the AH superfamily, MPH most likely evolved from lactonases, enzymes that hydrolyze the C-O ester bond of lactone ring and are involved in interfering bacterial quorum sensing[27]. Amongst other lactonases from the MBL superfamily, the N-acyl homoserine lactonase (AHLase) of *Bacillus thuringiensis*, AiiA, exhibits promiscuous phosphotriesterase activity[28]. However, the evolutionary potential and plausible evolutionary trajectories of these MBL lactonases toward phosphotriester hydrolysis have not been examined.

In this study, we evolved the promiscuous phosphotriesterase activity of the AiiA lactonase in the laboratory. Six rounds of random mutagenesis and screening of the library of mutants for higher phosphotriesterase activity yielded a variant with eight amino acid substitutions that exhibited a ~1000-fold improvement in its activity towards the organophosphate paraoxon. We performed structural characterizations of the wild-type enzyme and an evolved variant to reveal the molecular mechanism underlying the functional change. Additionally, we conducted extensive mutational analyses to elucidate the functional role of the remote mutations on repositioning and changes in the catalytic contributions of non-mutated, active-site residues. Our work provides a detailed molecular description of the evolution of a new enzymatic function and demonstrates how mutations indirectly affect catalytic activity through conformational tinkering.

RESULTS

Directed evolution of AiiA towards increased phosphotriesterase activity.

B. thuringiensis AiiA, as well as most enzymes in the MBL superfamily, utilize two active site metal ions to catalyze chemical reactions, by activating a water molecule for the nucleophilic attack and stabilizing the developing charge on the transition state (**Fig. S1A**). Many enzymes of the MBL superfamily are considered to be zinc-dependent enzymes[29,30]; however, our previous study showed that diverse metal ions can be accommodated and that the activity profile of the MBL enzymes can vary significantly depending on which metal ion is incorporated into the active site[31]. Therefore, we examined the functional effect of various metal ions (Cd^{2+}, Co^{2+}, Mn^{2+}, Ni^{2+} and Zn^{2}) for the phosphotriesterase activity of AiiA (**Fig. S2**). Out of the five tested AiiA metal-ion isoforms, Mn^{2+}-AiiA exhibited the highest phosphotriesterase activity (v_0= 15.9 nMs^{-1} for 5 μM of enzyme and 500 μM of paraoxon), which is ~100-fold higher than Zn^{2+}-AiiA (v_0 = 0.23 nMs^{-1} for 5 μM of enzyme and 500 μM of paraoxon). Moreover, the phosphotriesterase activity of AiiA was easily detectable in crude cell lysate when Mn^{2+} was supplied to the cell culture media and enzyme activity assay buffer. However, the activity of AiiA in cell lysate was barely detectable in the presence of Zn^{2+}. Thus, throughout the directed evolution experiment and subsequent biochemical characterizations we supplied Mn^{2+} to media and buffers.

AiiA was subjected to six rounds of directed evolution to improve its promiscuous phosphotriesterase activity. Mutagenized libraries of AiiA were generated using error-prone PCR and DNA shuffling, subcloned into the expression vector pET27(b), and transformed into *E. coli* BL21 (DE3) cells. The library (~10^4 transformants) was plated on agar plates and colonies were transferred onto nitrocellulose membranes for protein expression. Subsequently, colonies were lysed and assayed on membranes for improved phosphotriesterase activity, by measuring the release of *p*-nitrophenol, which produces an observable yellow color. Then, ~200 of the most active colonies were re-grown in 96-well plates, and phosphotriesterase activity of the clarified cell-lysate was assayed spectrophotometrically. The variants with the highest activity improvements in cell-lysate were then selected, sequenced, and used as a template for the next round of directed evolution (**Fig. S3**). During the fourth round of evolution our initial screen failed to yield any improved variants. Thus, we lowered the substrate concentration from 500 μM to 200 μM for the membrane pre-screen and from 250 μM to 150 μM for the cell-lysate screening to enable selection for variants with lower K_M (**Table S1**). In total, six iterative rounds of directed evolution yielded an AiiA variant with a ~270-fold increase in phosphotriesterase activity in cell lysate (**Fig. 2A**).

The functional adaptation of AiiA yielded a generalist enzyme.

We purified the most improved variant from each round and determined its kinetic parameters for native lactonase and evolved phosphotriesterase activities, using the substrates N-hexanoyl-L-homoserine lactone and paraoxon-ethyl, respectively (**Fig. 2B** and **C**). The evolutionary trajectory of AiiA exhibits similar characteristics observed in other directed evolution experiments. First, the improvements in phosphotriesterase activity exhibit diminishing returns[1,12,32], *i.e.*, drastic improvements during initial rounds (~100-fold improvement in k_{cat}/K_M in the first 3 rounds), followed by smaller increments in the later rounds (~10-fold increase in k_{cat}/K_M in the final 3 rounds). The most evolved variant, AiiA-R6, exhibits a >1000-fold higher k_{cat}/K_M for paraoxon hydrolysis compared to the starting point, AiiA-wt ($k_{cat}/K_M = 5.4 \times 10^5$ $M^{-1}s^{-1}$ *vs.* 5.1×10^2 $M^{-1}s^{-1}$). Over the first four rounds changes in k_{cat} and K_M contributed to the improvement in catalytic efficiency, whereas in the last two rounds only K_M changed (decreased) significantly. Overall, the improvement in k_{cat}/K_M is largely due to a 130-fold decrease in K_M (from 3400 to 26 μM), while k_{cat} increased by only 8-fold (from 1.8 to 14 s^{-1}). Second, the trade-off between the native lactonase and evolved phosphotriesterase activities was weakly negative[33,34]. The lactonase activity decreased by ~3-fold during the evolution, while the phosphotriesterase activity increased ~1000-fold (**Fig. 2C**), yielding a generalist enzyme with high catalytic activities for lactones and phosphotriesters (10^5 $M^{-1}s^{-1} \geq k_{cat}/K_M$). In addition, catalytic activities towards two related organophosphate compounds, methyl-paraoxon and methyl-parathion, co-evolved along with the target substrate, ethyl paraoxon. Specifically, the k_{cat}/K_M for methyl-paraoxon and methyl-parathion increased by ~100-fold from AiiA-wt to AiiA-R6 (1.1×10^5 $M^{-1}s^{-1}$ and 7.1×10^2 $M^{-1}s^{-1}$, respectively; **Table 1**).

Genotypic changes that led to increased phosphotriesterase activity.

In total, eight amino acid substitutions accumulated during six rounds of directed evolution (**Fig. 3A**). Four mutations, L33M, V69G, K139T and I230M, occurred in the first round, but only two, L33M and V69G, are proximally located with respect to the active site (**Fig. 3A**). We generated single point mutants of all four mutations in the background of AiiA-wt to examine their individual contributions, and found that V69G alone accounted for the improvement observed in round 1. The other three mutations were functionally neutral (**Table S4**). Only single mutations occurred at each subsequent round: F64C in round 2, M33V in round 3, S20F in round 4, K218R in round 5 and H18Q in round 6. Although most mutations occurred proximal to the active site, only S20F appears to be directly part of active site cavity (**Fig. 3A** and **3B**). V69G and F64C are located in the loop 3 above the active site (**Fig. 3B**). Residue 33, which is located on the ß-strand connected to loop 1, adjacent to a metal-binding residue, His235, was mutated twice during the trajectory: L33M at round 1 was neutral, but the subsequent mutation at round 3, M33V, increased phosphotriesterase activity by 2-fold (**Fig. 2B**). S20F (round 4) and H18Q (round 6) occurred in an active site loop 8, although H18Q does not appear to be part of the active site cavity (**Fig. 3A**). K218R (round 5) is not located near the active site, yet it caused a 1.5-fold decrease in K_M during round 5 (**Fig. 2B)**.

The combination of indirect and direct mutations underlies the conformational active site tinkering.

To understand the molecular basis underlying the increase in phosphotriesterase activity, we solved the crystal structures of AiiA-wt and AiiA-R4 to a resolution of 1.29 Å under identical crystallization conditions (**Table S3**). Both structures exhibit the same crystal packing, which allows a detailed structural comparison, with minimal crystallographic artifacts (**Table S3**). We selected AiiA-R4 for structural analysis because it exhibits the highest k_{cat} value of all evolved variants, and no significant improvements in cell lysate activity were obtained in round 5 and 6 (**Fig. 2A**). Overall, the backbone of the structures of AiiA-wt and AiiA-R4 align with a maximum r.m.s.d. of 0.108 Å over all C-α atoms (**Fig. 3A**) and no significant changes in the positions of the catalytic metal ions and the coordinating residues were observed (**Fig. S4A** and **S4B**). However, the active site cavity of AiiA-R4 was remodelled through two major structural modifications (**Fig. 3B** and **3C**). First, the position of loop 3 was altered and Phe68, which is located on the tip of loop 3, was shifted downwards by 2.9 Å, which resulted in a narrowing of the active site entrance (**Fig. 3B**). The conformational tinkering of loop 3 and Phe68 is most likely caused by two mutations on the same loop: V69G and F64C. The substitution of the larger hydrophobic residues with smaller ones created a space underneath loop 3, enabling the loop to collapse slightly downwards, thus causing the reposition of Phe68. Second, the active site mutation S20F in round 4 resulted in the introduction of a bulky aromatic residue, which directly caused a narrowing of the active site (**Fig. 3B**).

To understand how the reshaping of the active site affected substrate positioning and phosphotriesterase activity, we attempted to soak paraoxon into the protein crystals. However, we were unable to obtain unambiguous substrate or product density to infer substrate binding. Therefore, we computationally docked paraoxon into the structures of AiiA-wt and AiiA-R4 and conducted MD simulations (**Fig. 4**). For each variant we conducted two independent MD simulations runs (25 ns), which provide similar results, but suggest a distinct substrate binding position in the active sites of AiiA-wt and AiiA-R4 (**Fig. S6**). In AiiA-R4, the paraoxon substrate is reasonably well positioned in the active site where the scissile P-O bond of paraoxon can be in-line with the nucleophilic water molecule, which is activated and bridged between the two metal ions, indicating a catalytically favorable substrate position (**Fig. 4B**). This orientation is supported by the repositioned Phe68, which forms hydrophobic and π-π-stacking interactions with the *p*-nitrophenol leaving group of paraoxon. The C_4 of Phe68 and N of paraoxon maintain an average distance of 3.9 Å during the simulation, which is within the optimal distance range for π-π-stacking interactions[35]. Additionally, the mutation S20F might also contribute to position the substrate by hydrophobic interactions with one ethyl group of paraoxon (**Fig. 4B**). On the contrary, in AiiA-wt, the paraoxon is not placed in the position where an activated water molecule can perform nucleophilic attack, indicating catalytically inactive, non-productive binding (**Fig. 4A**). Phe68 in AiiA-wt remained relatively distant from the substrate during the MD simulations with average distance of 5.7 Å between the C_4 of Phe68 and N of paraoxon, indicating no interaction between Phe68 and the substrate (**Fig. 4C**). Taken together, these results suggest that the remodelling of the active site with the displacement of Phe68 and the active site mutation, S20F, collectively cause a catalytically more favourable paraoxon binding position in AiiA-R4, which may underlie the 200-fold increase in phosphotriesterase activity. Furthermore, we speculate that AiiA uses the same catalytic machinery for phosphotriesterase as well as for its native lactonase activity, in which an activated bridging water molecule between the two metal ions serve as a nucleophile (**Fig. S1B**). This is supported by the fact that mutating Asp108, which stabilizes the water molecule that participates in nucleophilic attack on the carbonyl carbon of the AHL substrate[36], to asparagine is highly deleterious (>100-fold decrease) to lactonase activity and phosphotriesterase activity in AiiA-wt and AiiA-R4 (**Table S4**).

Epistatic interactions altered the contribution of Phe68.

In order to quantify how mutations of the trajectory alter the catalytic contribution of Phe68, we performed "comparative" alanine scanning mutagenesis, *i.e.* the activity change caused by F68A was compared in different genetic backgrounds of AiiA. As suggested by the structural analysis, we hypothesized that two adjacent mutations, V69G and F64C (from round 1 and round 2, respectively), could be responsible for the displacement of Phe68, and would potentially alter its functional contribution (**Fig. 3B**). Thus, we generated AiiA variants with the individual mutations V69G and F64C and their combination, V69G-F64C. Subsequently we introduced F68A in all four variants, including AiiA-wt, and compared their effect on phosphotriesterase activity (**Fig. 5**). If the effect of F68A on phosphotriesterase activity differ between in the background of the wild-type and in that of the mutants (V69G and/or F64C), it would indicate that V69G and F64C epistasically interact with Phe68 and alter its contribution to catalysis. The individual mutations V69G and F64C increased phosphotriesterase activity in the background of the wild-type (Phe68) by 11.8- and 1.8-fold, respectively (**Fig. 5B**). The double mutant F64C-V69G revealed significant positive epistasis between the two mutations; the combination of the two mutations synergistically increased phosphotriesterase activity by 70-fold compared to an expected increase of 21-fold based on the additive null-model prediction, *i.e.,* the contributions of the two mutations are independent (**Fig. 5B**). Introducing F68A in AiiA-wt was advantageous, and resulted in a 2-fold activity increase (**Fig. 5B**). However, F68A became deleterious in the background of either V69G or F64C, causing a 2.9-fold and 1.9-fold decrease in activity, respectively. F68A became even more deleterious in the background of the F64C-V69G double mutant, with a 11.6-fold decrease in activity (**Fig. 5B**). Thus, the overall effect of F68A changed by 23-fold (2-fold positive *vs.* 11.6-fold negative) between AiiA-wt and the double mutant V69G-F64C. These results indicate that the role of Phe68 for catalysis was substantially altered during the evolution, which is consistent with the structural and MD simulation analyses. Interestingly, the deleterious effect of F68A was buffered through the accumulation of other mutations in the trajectory. F68A caused a 5-fold decrease in the background of the triple mutant V69G, F64C and S20F (acquired in round 4, **Table 4**), and only a 2-fold decrease in AiiA-R4, *i.e.,* the triple mutant plus L33V, K139T, and I230M (**Fig. 5B**). Thus, the contribution of Phe68 to catalysis was not only affected by adjacent mutations, *e.g*., F64C and V69G, but also by more remote ones, indicating that epistatic interactions can occur between very distant mutations.

Extended comparative mutational scanning reveals changes in the functional contribution of other non-mutated active residues.

Next, we investigated whether epistatic interactions during the evolutionary process drove any change in the catalytic contribution of other non-mutated active site residues, in addition to Phe68. We extended our comparative mutational scanning analysis to several other active site residues, Gln72, Arg134, Glu135, Glu136, and Tyr194 (**Fig. 6**). These residues were selected because their side chain position was shifted in the crystal structures and/or MD simulations, suggesting that their interactions with the substrate could be different in AiiA-wt and AiiA-R4 (**Fig. 4** and **Fig. S4**). Each of these residues was mutated to alanine with the exception of Tyr194, which was mutated to phenylalanine to avoid destabilization of the scaffold. Three out of five residues, Gln72, Glu136, and Tyr194, changed their functional contribution during the evolutionary trajectory between AiiA-wt and AiiA-R4 (**Fig. 6A**). Similar to Phe68, the functional contribution of Tyr194 increased during the evolution: it caused a 10-fold improvement in phosphotriesterase activity in AiiA-wt, but its effect was reduced to a 5.5-fold increase in AiiA-R4 (**Fig. 6A)**. In contrast, an opposite tendency in functional contribution was observed for Q72A and E136A. Q72A was deleterious (1.6-fold decrease) in the wild-type background, but became beneficial in AiiA-R4 (2.3-fold increase, **Fig. 6A**). Introducing E136A resulted in a 2.6-fold increase in AiiA-wt, and an even greater 11-fold increase in AiiA-R4 (**Fig. 6A**). It should be noted that while the overall catalytic efficiency (k_{cat}/K_M) increased with the introduction of the Q72A, R134A, E135A, E136A, and Y194F into AiiA-R4, none of the mutations have been selected for during the evolution. This could be explained by the fact that the improvements conferred by these mutations are largely a result of reduction in K_M and the mutations therefore could provide no or only marginal activity increase in the cell-lysate screening employed during the directed evolution experiment (**Fig. 6C** and **Table S4**). The results suggest that the catalytic contributions of the active site residues were indeed altered by remote mutations accumulated during the evolution.

Epistatic interactions between mutations of the evolutionary trajectory

Finally, we investigated the extent by which epistatic interactions affected the mutational trajectory in the directed evolution. We selected the six mutations that had effects on the catalytic parameters, V69G, F64C, S20F, L33V, K139T and I230M, and compared their catalytic effect in the genetic background of when they appeared in the trajectory to their effect in the genetic background of AiiA-wt (**Fig. 7**). Two out of the six mutations, F64C of round 2 and S20F of round 4, exhibited positive epistasis, *i.e.*, a more positive effect on the phosphotriesterase activity in the trajectory than individually in the wild-type background. In particular, S20F was almost neutral in the background of AiiA-wt, but became positive (2.5-fold) in the trajectory, which indicates that the positive effect of S20F was permitted by earlier mutations of the trajectory (**Fig. 7A**). On the contrary, H18Q was more positive in the background of AiiA-wt compared to its effect in round 6 (**Fig. 7A**). Taken together, the results indicate that intertwined epistatic interaction networks prevail the evolution of AiiA towards improved phosphotriesterase activity, and altered not only the catalytic contributions of non-mutated active site residues but also the effects of mutations that appeared during the later rounds of the evolution.

DISCUSSION

Many directed evolution studies have observed that the accumulation of remote mutations can cause significant improvements in protein function[1,2]. However, only a handful of studies have elucidated the molecular mechanisms underlying the effects of remote mutations on enzyme function[6-8,14,32,37,38]. Here, we provide a detailed molecular explanation for the evolution of a new enzyme function: the ~1000-fold increase in phosphotriesterase activity of AiiA was largely driven by remote mutations that caused the reconfiguration of an active site residue Phe68, resulting in more favorable positioning of the substrate in the active site for catalysis. Moreover, we provide a quantitative measure of remote mutational effects; the comparative mutational scanning revealed that the functional contribution of Phe68 to the phosphotriesterase activity markedly increased (>20-fold) in the background of the two adjacent mutations V69G-F64C.

Remote and active site mutations each possess their advantages and disadvantages, owing to the level of chemical and structural consequences they cause. For example, mutating an active site residue can result in significant changes in the active site, such as the introduction of new electrostatic interactions between the enzyme and substrate, or significantly changing the size and shape of the active site cavity. However, the level of structural disruption that they cause means that active site mutations can also be detrimental to both the activity and the structural integrity of proteins[39,40]. On the other hand, remote mutations might not change use the existing chemistry in the active site, but can fine-tune the shape of the cavity and the position of key catalytic residues and cofactors in subtle ways that direct active site mutations cannot[41]. For example, directly mutating Phe68 might not provide the same structural modifications that are required to increase the activity as its indirect repositioning through V69G and F64C can achieve. In this instance, it appears that subtle conformational tinkering, *i.e.*, collapsing loop 3 downward into the active site, and shifting the position of Phe68, was a better adaptation strategy for the evolution of AiiA towards phosphotriesterase activity.

Our work provided experimental evidence to link epistatic interactions and active site conformational tinkering. Over the last decade, growing experimental and theoretical evidence has indicated that epistasis, or non-additive interactions among mutations, is prevalent during adaptive evolution and plays a central role in shaping the accessibility of mutational trajectories[42-44]. Mutations that occur at an early stage in an adaptive evolution could permit or restrict the potential of other mutations and affect the appearance of mutations in later stages[1,43-46]. Here we expand the view of epistatic interactions to residues that are not mutated during evolution, and demonstrate that their functional contribution can change during a functional transition. Such epistatic interactions might be prevalent and play important roles in many other enzyme evolution and engineering examples, in which functional adaptation is achieved through the molecular tinkering of active site residues by remote mutations. It is essential that we further develop our knowledge of the molecular mechanisms underlying such indirect effects for the design and engineering of proteins[47]. Thus, along with a detailed structural characterization, comparative mutational scanning, as described in this study, would serve as a robust approach to investigate and connect the role of remote mutations to functional changes in many other protein evolution studies.

Phosphotriesters and lactones are different chemical compounds in respect to their chemical structure and the bonds that are broken during hydrolysis (P-O bond *via* a pentacoordinate transition state *vs.* C-O bond *via* a tetrahedral transition state; **Fig. 1**). However, enzymes in three different lactonase families, each with distinct structural folds and active sites (the AH superfamily (TIM barrel fold), the MBL superfamily (αββα-fold) and the paraoxonase (PON) superfamily (β-propeller fold)), all exhibit promiscuous phosphotriesterase activity[48]. Elias and Tawfik proposed that the two activities, phosphotriesterase and AHL lactonase are commonly shared because two substrates could bind in a similar geometry in the active site and thus can utilize the same catalytic machinery[27]. The plausible evolutionary trajectories underlying the functional transition of enzymes in the AH and PON superfamily have been previously explored using directed evolution[23,33,49]. In this work, we have made similar observations to previous studies, *e.g.,* the two reactions can be traversed within only several mutational steps, supporting the notion that phosphotriesterases promptly evolved in parallel from ancestral lactonases in the time frame of a few decades[27]. Interestingly, the molecular changes associated with the increase in phosphotriesterase activity appear to be different in each case. In mammalian PON1, a displacement of the catalytic Ca^{2+} was largely responsible for the transition between lactonase and phosphotriesterase activities[7]. Phosphotriesterase activity in DrPLL, a lactonase in the AH superfamily, increased through enlargement of the active site cavity[23]. OphC2, a phosphotriesterase in the MBL superfamily, has been shown to have evolved from a dihydrocoumarin lactonase by only two active site mutations[50]. The evolution of AiiA, presented in this study, demonstrated that improvements in the phosphotriesterase activity were achieved by a reduction in the active site cavity. Therefore, enzymes with similar native functions might be equally good springboards for the evolution of a particular new function. The caveat to this conclusion is that parallel evolution from different starting enzymes might involve different molecular and mechanistic adaptations to achieve the same goal.

MATERIALS AND METHODS

Molecular cloning of AiiA variants and mutant libraries. AiiA gene variants and mutant libraries were sub-cloned into a modified pET27(b) vector (Novagen), which replaced a N-terminal pelB leader sequence to Strep-tag II sequence (WSHPQFEK). PCR products and vectors were digested with *Nco* I and *Hind* III (Thermo Scientific) for 3 h at 37°C. The digested vector was further treated with FastAP (alkaline phosphatase, Thermo Scientific) for an additional hour. Digested DNA was purified from a 1% agarose gel using a gel extraction kit (Qiagen). Ligations were performed in 20 µL reactions at a vector:insert molar ratio of 1:3 using T4 DNA ligase (Thermo Scientific) with approx. 30 ng vector DNA, and incubated at room temperature for 2 h. The ligation mixtures were transformed into *E. cloni* 10G cells (Lucigen), yielding $>10^5$ colonies for mutant libraries. Colonies containing mutant library variants were pooled, and the plasmids were purified using a plasmid purification kit (Qiagen) and retransformed into *E. coli* BL21 cells (DE3) for enzyme expression and activity screening.

Site-directed mutagenesis. Single-point mutant variants were constructed by site-directed mutagenesis as described in the QuikChange Site-Directed Mutagenesis manual (Agilent) using specific primers, which are listed in **Table S2**. Genotype of all variants were confirmed by DNA sequencing.

Generation of mutagenized libraries. Random mutant libraries were generated with error-prone PCR using nucleotide analogues (8-oxo-2'-deoxyguanosine-5'-Triphosphate (8-oxo-dGTP) and 2'-deoxy-P-nucleoside-5'-Triphosphate (dPTP); TriLink). Two independent PCRs were prepared, one with 8-oxo-dGTP and one with dPTP. Each 50 µL reaction contained 1 × GoTaq Buffer (Promega), 3 µM $MgCl_2$, 1 ng template DNA, 1 µM of primers (forward (T7 promoter): taatacgactcactataggg; reverse (T7 terminator):

gctagttattgctcagcgg), 0.25 mM dNTPs, 1.25 U GoTaq DNA polymerase (Promega) and either 100 µM 8-oxo-dGTP or 1 µM dPTP. PCR cycling conditions: initial denaturation at 95°C for 2 min followed by 20 cycles of denaturation (30 sec, 95°C), annealing (60 sec, 58°C) and extension (70 sec, 72°C) and a final extension step at 72°C for 5 min. Subsequently, each PCR was treated with *Dpn* I (Thermo Scientific) for 1 h at 37°C to digest the template DNA. PCR products were purified using the Cycle Pure PCR purification kit (Omega Bio-tek) and further amplified with a 2 × Master mix of Econo TAQ DNA polymerase (Lucigen) using 10 ng template from each initial PCR and the same primer set at 1 µM in a 50 µL reaction volume. PCR cycling conditions: Initial denaturation at 95°C for 2 min followed by 30 cycles of denaturation (30 sec, 95 °C), annealing (20 sec, 58°C) and extension (70 sec, 72°C) and a final extension step (72°C, 2 min). The PCR products were purified and cloned as described above. The protocol yielded 1-2 amino acid substitutions per gene (1-2 bp) per round.

Generation of DNA shuffling libraries. The staggered extension process (StEP) protocol was used to recombine DNA sequence of improved variants [51]. Plasmids of variants were mixed in equimolar amounts to 500 ng of total DNA and used as a template for the StEP reaction. PCR cycling conditions: Initial denaturation at 95°C for 5 min, followed by 100 cycles 95°C for 30 sec, followed by 58°C for 5 sec. PCR products were purified using the Cycle Pure PCR purification kit and further amplified with a 2 × Master mix of Econo TAQ DNA polymerase. StEP libraries were cloned as described above.

Activity prescreen on agar plates. Transformed cells were plated on LB agar plates with kanamycin (50 µg/ml) and incubated at 37°C overnight yielding about 3,000–5,000 colonies in total on 4–6 plates. Colonies were replicated onto nitrocellulose membranes, which were placed onto LB agar containing kanamycin and 1 mM IPTG for protein expression and incubated at room temperature overnight. After expression, the membranes were placed into an empty petri dish and the cells were lysed by alternating incubations at -20°C and 37°C three times for 10 min each. To assay activity, 25 mL of 0.5% agarose buffer (50 mM Tris–HCl pH 7.5, 100 mM NaCl and 200 µM $MnCl_2$)

containing 500 µM paraoxon (Sigma), or 250 µM after round 4, was poured onto the membrane. Colonies with active enzymes developed a yellow color due to the release of a product, *p*-nitrophenol. The most active colonies (~200 variants) were directly picked from plates for subsequent screening in 96-well plates.

Cell lysate activity screen in 96-well plates. Colonies picked into in 96-well plates were grown in 200 uL of LB with 50 µg/ml kanamycin at 30°C overnight. 20 µL of each culture were used to inoculate 400 µL of LB with 50 µg/ml kanamycin, and incubated at 37°C for 3 h, then protein expression was induced by adding IPTG at the final concentration of 1mM, and incubated at 30°C for another 3 h. Cells were harvested by centrifugation at 4000 × g for 10 min at 4°C and pellets were frozen at -80°C for at least 1 h. Cells were lysed by adding 200 µL of 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 0.1% (w/v) Triton-X100, 200 µM $MnCl_2$, 100 µg/mL lysozyme, and 0.5 U benzonase (Novagen). After 30 min of incubation at room temperature the lysate was clarified at 4000 × g for 20 min at 4°C. To assay enzymatic activity, 100 µL of the clarified lysate was mixed with 100 µL paraoxon solution at a final concentration of 250 µM (or 150 µM after round 4) in 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 0.02% Triton-X100 and the reaction was monitored at 405 nm. The activity of the best variants was subsequently confirmed in triplicate cultures and activity assays. The variant with the highest activity was sequenced and used as template for the next round. When multiple variants were identified StEP shuffling was used to recombine beneficial mutations.

Enzyme purification for kinetic analysis. All variants were cloned as described above, transformed and overexpressed in *E. coli* BL21 (DE3) cells and purified using Strep-tactin resin (IBA lifesciences) as described previously[28]. During expression, purification and storage 200 µM $MnCl_2$ was supplied in media and buffers.

Enzyme kinetics. The kinetic parameters and activity levels of purified of enzyme variants were obtained as described previously[28]. Briefly, the activity for paraoxon-methyl, paraoxon-ethyl, and parathion-ethyl (Sigma) was monitored following the release of *p*-nitrophenol at 405 nm with an extinction coefficient of 18,300 $M^{-1}cm^{-1}$ [28]. The activity for N-hexanoyl-L-homoserine lactone was monitored at 560 nm using a phenol red based pH indicator assay. A standard curve was prepared using HCl to calculate an extinction coefficient of 1334 $M^{-1}cm^{-1}$. The kinetic parameters K_M and k_{cat} were determined by fitting the initial rates to the Michaelis–Menten model ($v_0=k_{cat}[E]_0[S]_0/(K_M+[S]_0)$) using KaleidaGraph (Synergy Software).

Crystallization of AiiA-wt and AiiA-R4. Proteins were expressed with a N-terminal His_{10}-tag in *E. coli* BL21 (DE3) cells in TB media containing 1% glycerol, 50 µg/mL kanamycin and 200 µM $ZnCl_2$. Cells were grown at 30°C for 6 h and another 16 h at 22°C and harvested at 8,000 × g for 15 min at 4°C. Pellets were resuspended in 50 mM HEPES pH 7.5, 500 mM NaCl, 25 mM Imidazole and 200 µM $ZnCl_2$ (Buffer A). Lysis was performed by sonication (OMNI sonic ruptor 400) and the lysate was clarified at 20,000 × g for 60 min at 4°C. The proteins were purified using a Ni-NTA column (Qiagen) and eluted in 50 mM HEPES pH 7.5, 500 mM NaCl, 500 mM Imidazole, 200 µM $ZnCl_2$. To remove the N-terminal His_{10}-tag TEV protease cleavage was performed at 4°C for 5 days in 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.5 mM EDTA and 1 mM DTT containing 10% TEV protease relative to the purified protein[52]. Subsequently, TEV protease and His_{10}-tags were removed using a Ni-NTA column. The proteins were further purified using size exclusion (HiLoad 16/600 Superdex 75 column, GE-Healthcare), eluted in buffer containing 10 mM HEPES pH 7.0, 2 mM DTT, 100 µM $ZnCl_2$ and concentrated to 20 mg/mL. Crystallization was performed using the hanging drop method by mixing a protein solution (1 µL) and a well solution (2 µL) containing 25% (w/v) PEG 4 K, 20% (v/v) glycerol, 80 mM Tris-HCl pH 8.5 and 160 mM $MgCl_2$ as described elsewhere[53]. Crystals appeared in two weeks at 18°C and continued to grow for several months. Crystals were briefly soaked into 35% PEG 4 K before nitrogen gas flash frozen. Crystallographic data were collected at 100 K at the Australian Synchrotron using the MX1 beam line with a wavelength of 0.9537 Å. The diffraction data obtained were indexed, processed and scaled with the programs XDS[54] and AIMLESS[55]. All structures were solved by molecular replacement using MOLREP[56] as implemented in the CCP4 suite of programs[57]. The models were subsequently optimized by iterative model building with the program COOT[58] and refinement with phenix.refine[59]. The structures were then evaluated using MolProbity[60]. Resolution estimation and data truncation were performed by using overall half-dataset correlation CC (1/2) > 0.5[61]. Details of the refinement statistics are produced by Phenix and summarized in **Table S3**. Structures of wild-type AiiA and AiiA-R4 were deposited in the PDB with accession codes 5EH9 and 5EHT, respectively.

Molecular docking and molecular dynamics simulations. AutoDock Vina[62] was used to dock paraoxon into apo AiiA-wt and AiiA-R4 structures. AutoDockTools4[63] was used to generate polar hydrogens and add partial charges to the proteins using KollmanUnited charges. The search space was included in a box of 40 × 40 × 40 Å and centred on the binding site of the ligand. Both rigid and flexible-docking were performed. Flexible torsions of side chains and ligand were assigned based on B-factors of the crystal structures with Autotors. The ligand paraoxon-ethyl was generated using PRODRG[64]. For each calculation eight poses were obtained and ranked according to the scoring-function of Autodock Vina. Selected protein-ligand structures were used for the molecular dynamics simulation using GROMACS 4.6.5[65]. The force field applied for the simulation was GROMOS96-53a6[66]. Ligand topology was generated by the Automated Topology Builder 2.0. Proteins were immersed into a dodecahedron shaped box of water at a buffering distance of 1.0 nm between the protein and the edge of the box. Sodium ions were added to neutralize charge. Energies were subsequently minimized with the Steepest Decent method for 2,500 steps. The last frame of each energy-minimized structure was used as the initial frame for MD simulation. Electrostatic interactions were calculated with the Particle-Mesh Ewald (PME) method[67]. The cut-off for PME was 1.0 nm. The time step was set at 2 fs at 300 K. The number of steps set to 25,000 (0.1 ns for NVT) for generating random velocities with position restrained, 250,000 (1 ns for NPT) for equilibrium at 1 atm, and 12,500,000 (25 ns) for the production MD simulation. V-rescale for temperature coupling and Parrinello-Rahman for pressure coupling were used.

SUPPORTING INFORMATION

Supplementary Figure S1: Proposed catalytic mechanism of AiiA in N-homoserine lactone hydrolysis.

Supplementary Figure S2: Metal dependency of AiiA for paraoxon activity.

Supplementary Figure S3: Overview of directed evolution scheme.

Supplementary Figure S4: Structural comparison between AiiA-wt and AiiA-R4.

Supplementary Figure S5: B-factor comparison between AiiA-wt and AiiA-R4

Supplementary Figure S6: Molecular dynamics simulations of AiiA-wt and AiiA-R4

Supplementary Table S1: Library information and amino acid changes.

Supplementary Table S2: Site-directed mutagenesis primers.

Supplementary Table S3: Crystallography data collection and refinement statistics.

Supplementary Table S4: Cell lysate activities.

Supplementary Table S5: Kinetic parameters for paraoxonase activity.

ACKNOWLEDGMENTS

We thank members in the Tokuriki lab for comments on the manuscript. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC, RGPIN: 418262-12). N.T. is a Canadian Institute of Health Research (CIHR) new investigator and a Michael Smith Foundation for Health Research (MSFHR) career investigator.

REFERENCES

(1) Miton, C. M.; Tokuriki, N. *Protein Sci.* **2016**.
(2) Morley, K. L.; Kazlauskas, R. J. *Trends in Biotechnology* **2005**, *23* (5), 231–237.
(3) Chica, R. A.; Doucet, N.; Pelletier, J. N. *Current Opinion in Biotechnology* **2005**, *16* (4), 378–384.
(4) Kries, H.; Blomberg, R.; Hilvert, D. *Current Opinion in Chemical Biology* **2013**, *17* (2), 221–228.
(5) Oelschlaeger, P. *Protein Science* **2005**, *14* (3), 765–774.
(6) Jiménez-Osés, G.; Osuna, S.; Gao, X.; Sawaya, M. R.; Gilson, L.; Collier, S. J.; Huisman, G. W.; Yeates, T. O.; Tang, Y.; Houk, K. N. *Nature Chemical Biology* **2014**, *10* (6), 431–436.
(7) Ben-David, M.; Wieczorek, G.; Elias, M.; Silman, I.; Sussman, J. L.; Tawfik, D. S. *Journal of Molecular Biology* **2013**, *425* (6), 1028–1038.
(8) Sykora, J.; Brezovsky, J.; Koudelakova, T.; Lahoda, M.; Fortova, A.; Chernovets, T.; Chaloupkova, R.; Stepankova, V.; Prokop, Z.; Smatanova, I. K.; Hof, M.; Damborsky, J. *Nature Chemical Biology* **2014**, *10* (6), 428–430.
(9) Oue, S.; Okamoto, A.; Yano, T.; Kagamiyama, H. *J. Biol. Chem.* **1999**, *274* (4), 2344–2349.
(10) Bridgham, J. T.; Eick, G. N.; Larroux, C.; Deshpande, K.; Harms, M. J.;

Gauthier, M. E. A.; Ortlund, E. A.; Degnan, B. M.; Thornton, J. W. *PLoS Biol* **2010**, *8* (10), e1000497.

(11) Jacob, F. *Science* **1977**, *196* (4295), 1161–1166.

(12) Tokuriki, N.; Jackson, C. J.; Afriat-Jurnou, L.; Wyganowski, K. T.; Tang, R.; Tawfik, D. S. *Nature Communications* **2012**, *3*, 1257.

(13) Tomatis, P. E.; Rasia, R. M.; Segovia, L.; Vila, A. J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (39), 13761–13766.

(14) Dellus-Gur, E.; Elias, M.; Caselli, E.; Prati, F.; Salverda, M. L. M.; de Visser, J. A. G. M.; Fraser, J. S.; Tawfik, D. S. *Journal of Molecular Biology* **2015**, *427* (14), 2396–2409.

(15) Brodkin, H. R.; DeLateur, N. A.; Somarowthu, S.; Mills, C. L.; Novak, W. R.; Beuning, P. J.; Ringe, D.; Ondrechen, M. J. *Protein Science* **2015**, *24* (5), 762–778.

(16) Fischer, M.; Kang, M.; Brindle, N. P. *Protein Sci.* **2016**, *25* (2), 352–359.

(17) Singh, B. K.; Walker, A. *FEMS Microbiol Rev* **2006**, *30* (3), 428–471.

(18) Sethunathan, N.; Yoshida, T. *Can. J. Microbiol.* **1973**, *19* (7), 873–875.

(19) Singh, B. K. *Nat. Rev. Microbiol.* **2009**, *7* (2), 156–164.

(20) Bigley, A. N.; Raushel, F. M. *Biochim. Biophys. Acta* **2013**, *1834* (1), 443–453.

(21) Afriat, L.; Roodveldt, C.; Manco, G.; Tawfik, D. S. *Biochemistry* **2006**, *45* (46), 13677–13686.

(22) Jackson, C. J.; Foo, J.-L.; Kim, H.-K.; Carr, P. D.; Liu, J.-W.; Salem, G.; Ollis, D. L. *Journal of Molecular Biology* **2008**, *375* (5), 1189–1196.

(23) Meier, M. M.; Rajendran, C.; Malisi, C.; Fox, N. G.; Xu, C.; Schlee, S.; Barondeau, D. P.; Höcker, B.; Sterner, R.; Raushel, F. M. *J. Am. Chem. Soc.* **2013**, *135* (31), 11670–11677.

(24) Afriat-Jurnou, L.; Jackson, C. J.; Tawfik, D. S. *Biochemistry* **2012**, *51* (31), 6047–6055.

(25) Zhongli, C.; Shunpeng, L.; Guoping, F. *Applied and Environmental Microbiology* **2001**, *67* (10), 4922–4925.

(26) Dong, Y.-J.; Bartlam, M.; Sun, L.; Zhou, Y.-F.; Zhang, Z.-P.; Zhang, C.-G.; Rao, Z.; Zhang, X.-E. *Journal of Molecular Biology* **2005**, *353* (3), 655–663.

(27) Elias, M.; Tawfik, D. S. *Journal of Biological Chemistry* **2011**, *287* (1), 11–20.

(28) Baier, F.; Tokuriki, N. *Journal of Molecular Biology* **2014**, *426* (13), 2442–2456.

(29) Crowder, M. W.; Spencer, J.; Vila, A. J. *Acc. Chem. Res.* **2006**, *39* (10), 721–728.

(30) González, J. M.; Medrano Martín, F. J.; Costello, A. L.; Tierney, D. L.; Vila, A. J. *Journal of Molecular Biology* **2007**, *373* (5), 1141–1156.

(31) Baier, F.; Chen, J.; Solomonson, M.; Strynadka, N. C. J.; Tokuriki, N. *ACS Chem. Biol.* **2015**, 150416121657000.

(32) Kaltenbach, M.; Jackson, C. J.; Campbell, E. C.; Hollfelder, F.; Tokuriki, N. *eLife* **2015**, *4*.

(33) Aharoni, A.; Gaidukov, L.; Khersonsky, O.; McQ Gould, S.; Roodveldt, C.; Tawfik, D. S. *Nat Genet* **2005**, *37* (1), 73–76.

(34) Khersonsky, O.; Tawfik, D. S. *Biochemistry* **2010**, *79*, 471–505.

(35) Meyer, E. A.; Castellano, R. K.; Diederich, F. *Angew. Chem. Int. Ed. Engl.* **2003**, *42* (11), 1210–1250.

(36) Momb, J.; Wang, C.; Liu, D.; Thomas, P. W.; Petsko, G. A.; Guo, H.; Ringe, D.; Fast, W. *Biochemistry* **2008**, *47* (29), 7715–7725.

(37) Tomatis, P. E.; Fabiane, S. M.; Simona, F.; Carloni, P.; Sutton, B. J.; Vila, A. J. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105* (52), 20605–20610.

(38) Perica, T.; Kondo, Y.; Tiwari, S. P.; McLaughlin, S. H.; Kemplen, K. R.; Zhang, X.; Steward, A.; Reuter, N.; Clarke, J.; Teichmann, S. A. *Science* **2014**, *346* (6216), 1254346–1254346.

(39) Shoichet, B. K.; Baase, W. A.; Kuroki, R.; Matthews, B. W. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92* (2), 452–456.

(40) Tokuriki, N.; Stricher, F.; Serrano, L.; Tawfik, D. S. *PLoS Comput Biol* **2008**, *4* (2), e1000002.

(41) Mesecar, A. D.; Stoddard, B. L.; Koshland, D. E. *Science* **1997**, *277* (5323), 202–206.

(42) de Visser, J. A. G. M.; Cooper, T. F.; Elena, S. F. *Proceedings of the Royal Society B: Biological Sciences* **2011**, *278* (1725), 3617–3624.

(43) Kaltenbach, M.; Tokuriki, N. *J. Exp. Zool. (Mol. Dev. Evol.)* **2014**, *322* (7), 468–487.

(44) Harms, M. J.; Thornton, J. W. *Nat Rev Genet* **2013**, *14* (8), 559–571.

(45) Harms, M. J.; Thornton, J. W. *Nature* **2014**, *512* (7513), 203–207.

(46) Noor, S.; Taylor, M. C.; Russell, R. J.; Jermiin, L. S.; Jackson, C. J.; Oakeshott, J. G.; Scott, C. *PLoS ONE* **2012**, *7* (6), e39822.

(47) Hilvert, D. *Annu. Rev. Biochem.* **2013**, *82*, 447–470.

(48) Draganov, D. I. *Chem. Biol. Interact.* **2010**, *187* (1-3), 370–372.

(49) Hawwa, R.; Larsen, S. D.; Ratia, K.; Mesecar, A. D. *Journal of Molecular Biology* **2009**, *393* (1), 36–57.

(50) Luo, X. J.; Kong, X. D.; Zhao, J.; Chen, Q.; Zhou, J.; Xu, J. H. *Biotechnology and Bioengineering* **2014**, *111* (10), 1920–1930.

(51) Zhao, H.; Giver, L.; Shao, Z.; Affholter, J. A.; Arnold, F. H. *Nat. Biotechnol.* **1998**, *16* (3), 258–261.

(52) Cabrita, L. D.; Gilis, D.; Robertson, A. L.; Dehouck, Y.; Rooman, M.; Bottomley, S. P. *Protein Sci.* **2007**, *16* (11), 2360–2367.

(53) Liu, D.; Momb, J.; Thomas, P. W.; Moulin, A.; Petsko, G. A.; Fast, W.; Ringe, D. *Biochemistry* **2008**, *47* (29), 7706–7714.

(54) Kabsch, W. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66* (Pt 2), 125–132.

(55) Evans, P. R.; Murshudov, G. N. *Acta Crystallogr. D Biol. Crystallogr.* **2013**, *69* (Pt 7), 1204–1214.

(56) Vagin, A.; Teplyakov, A. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66* (Pt 1), 22–25.

(57) Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G. W.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67* (Pt 4), 235–242.

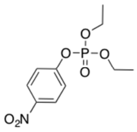
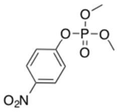
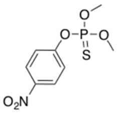
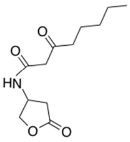
(58) Emsley, P.; Cowtan, K. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60* (Pt 12 Pt 1), 2126–2132.

(59) Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N.

W.; Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D. *Acta Crystallogr. D Biol. Crystallogr.* **2012**, *68* (Pt 4), 352–367.
(60) Chen, V. B.; Arendall, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66* (Pt 1), 12–21.
(61) Pateman, R. M.; Hill, J. K.; Roy, D. B.; Fox, R.; Thomas, C. D. *Science* **2012**, *336* (6084), 1028–1030.
(62) Trott, O.; Olson, A. J. *J. Comput. Chem.* **2010**, *31* (2), 455–461.
(63) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. *J. Comput. Chem.* **2009**, *30* (16), 2785–2791.
(64) Schüttelkopf, A. W.; van Aalten, D. M. F. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60* (Pt 8), 1355–1363.
(65) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J Chem Theory Comput* **2008**, *4* (3), 435–447.
(66) van Gunsteren, W. F., et al., *Biomolecular simulation: The GROMOS96 manual and user guide*. **1996**.
(67) Darden, T.; York, D.; Pedersen, L. *The Journal of chemical physics*. **1993**.

Table 1

Table 1 | Kinetic parameters of AiiA variants and MPH for selected substrates.

												
	Paraoxon-ethyl			**Paraoxon-methyl**			**Parathion-methyl**			**C6-HSL**		
Enzyme	k_{cat} [s^{-1}]	K_M [μM]	k_{cat}/K_M [$s^{-1}M^{-1}$]	k_{cat} [s^{-1}]	K_M [μM]	k_{cat}/K_M [$s^{-1}M^{-1}$]	k_{cat} [s^{-1}]	K_M [μM]	k_{cat}/K_M [$s^{-1}M^{-1}$]	k_{cat} [s^{-1}]	K_M [μM]	k_{cat}/K_M [$s^{-1}M^{-1}$]
AiiA-wt	**1.8** ± 0.1	**3400** ± 180	**5.1 × 10^2**	**5.5** ± 0.4	**2400** ± 400	**2.3 × 10^3**	*n.d.*	*n.d.*	**3.4**	**340** ± 33	**89** ± 27	**3.8 × 10^6**
AiiA-R1	**8.2** ± 0.6	**1000** ± 170	**7.9 × 10^3**	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	**340** ± 58	**350** ± 150	**9.6 × 10^5**
AiiA-R2	**10** ± 0.3	**400** ± 40	**2.6 × 10^4**	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	**200** ± 5.5	**290** ± 21	**6.9 × 10^5**
AiiA-R3	**25** ± 1.8	**470** ± 70	**5.3 × 10^4**	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	**500** ± 9.4	**460** ± 31	**1.1 × 10^6**
AiiA-R4	**29** ± 6.6	**220** ± 65	**1.3 × 10^5**	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	**470** ± 19	**120** ± 22	**6.9 × 10^6**
AiiA-R5	**15** ± 1.3	**40** ± 7	**3.8 × 10^5**	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	**900** ± 69	**750** ± 180	**1.2 × 10^6**
AiiA-R6	**14** ± 1.2	**26** ± 5	**5.4 × 10^5**	**13** ± 0.4	**120** ± 8	**1.1 × 10^5**	**7.2** ± 0.5	**11000** ± 1200	**7.1 × 10^2**	**940** ± 72	**820** ± 190	**1.2 × 10^6**
MPH	**14** ± 1.1	**900** ± 200	**1.6 × 10^4**	**55** ± 6	**690** ± 200	**8.0 × 10^4**	**62** ± 3	**880** ± 140	**7.0 × 10^4**	**0.7** ± 0.1	**675** ± 151	**1.1 × 10^3**

n.d. indicates not determined.

Figure 1

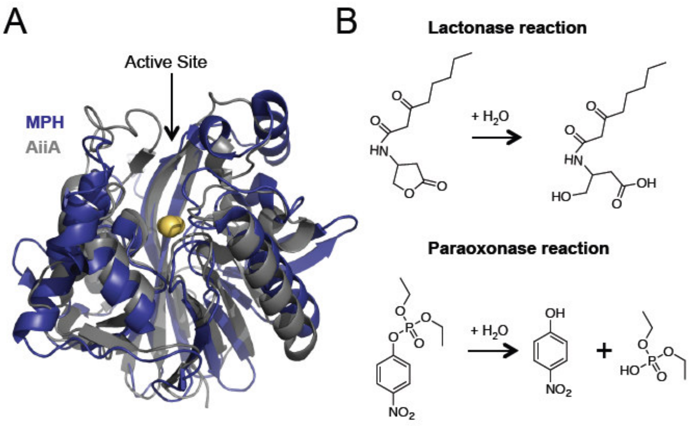


Fig. 1. Protein structures and enzymatic reactions of AiiA and MPH. (**A**) Structural overlay of *B. thuringiensis* N-Acyl homoserine lactonase, AiiA (grey, PDB ID: 3dhb), and the *Pseudomonas sp.* methyl-parathion hydrolase, MPH (deep blue, PDB ID: 1p9e). The active site metals are shown as beige spheres. (**B**) Reaction scheme of the native homoserine lactonase (top) and promiscuous phosphotriesterase (bottom) of AiiA. A proposed mechanism of AiiA for lactonase activity is described in **Fig. S1**[36].

Figure 2

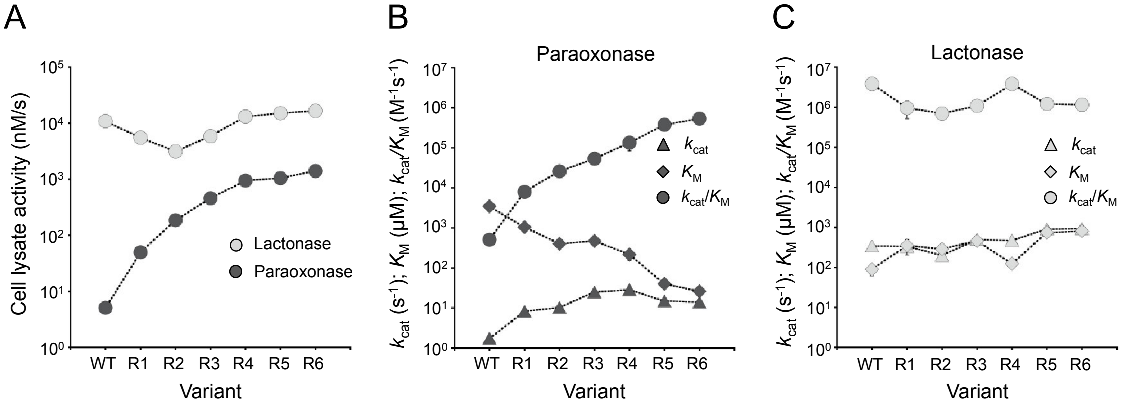


Fig. 2. Activity changes of AiiA over the evolutionary trajectory towards improved phosphotriesterase activity. (**A**) Lactonase (light grey) and phosphotriesterase (dark grey) activity in cell lysate of wild-type (wt) and evolved AiiA variants. Experiments were performed in triplicate and values were averaged with errors (standard deviation) less than 10% **(Table S4).** (**B**) Catalytic parameters, k_{cat}, K_M, and k_{cat}/K_M, of wt and evolved variants for phosphotriesterase activity. Each value represents the average of three parallel measurements with the standard deviation (bars). Individual values are listed in **Table S5.** (**C**) Catalytic parameters, k_{cat}, K_M, and k_{cat}/K_M of WT and evolved variants for the native lactonase activity. Each value represents the average of three measurements with the standard deviation (bars).

Figure 3

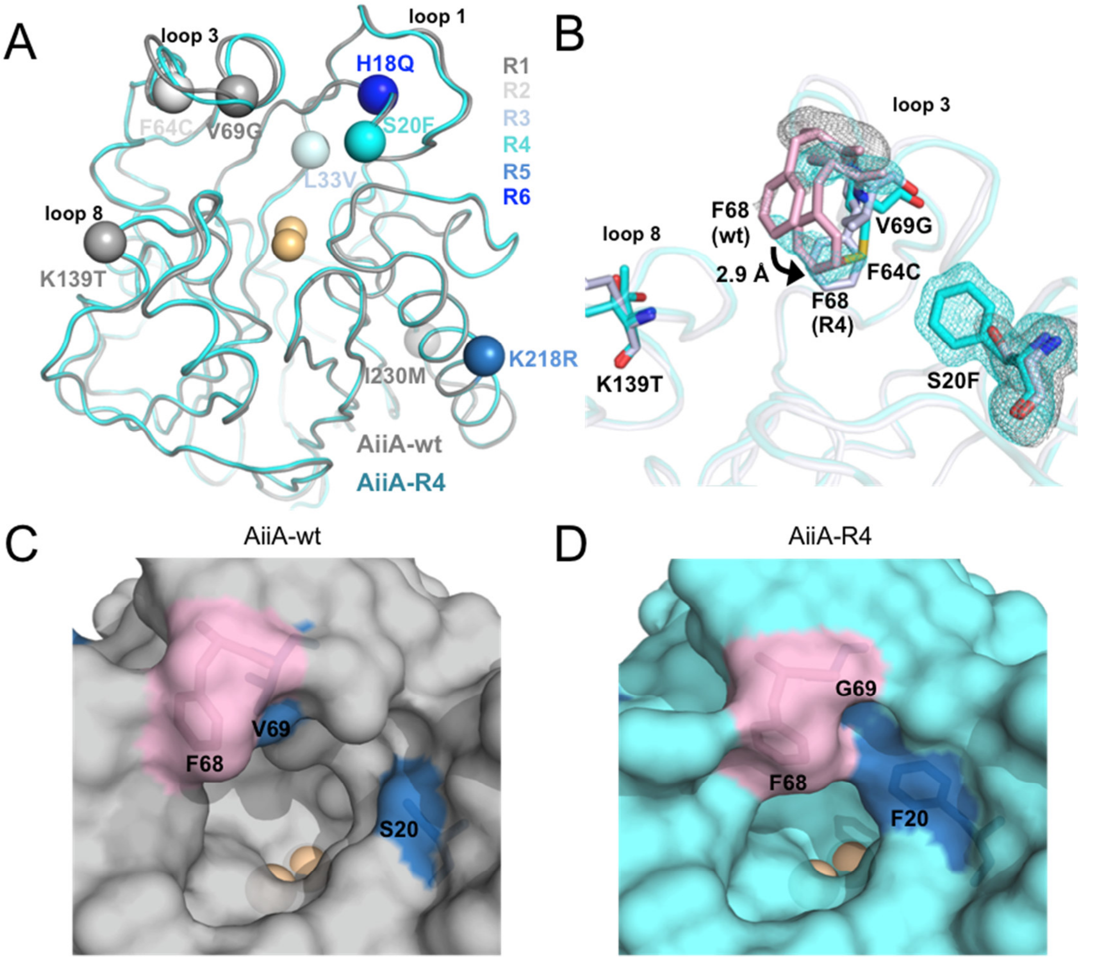


Fig. 3. Structural changes of AiiA during the evolution. (**A**) Overlay of crystal structures of AiiA-wt (grey, PDB ID: 5eh9) and AiiA-R4 (cyan, PDB ID: 5eht). The C-α atoms of residues mutated during the trajectory are shown as spheres and colored according to their occurrence in the trajectory from grey (R1) to dark blue (R4). (**B**) Highlights of direct and indirect structural changes occurring in the active site. Residues that were mutated during the evolution are shown as sticks and colored in grey (AiiA-wt) and cyan (AiiA-R4) respectively. The non-mutated residue Phe68 is highlighted in pink. Electron densities (*F*o-*F*c omit maps) of Phe68 and Ser20 in AiiA-wt, and Phe68 and Phe20 in AiiA-R4 are shown in grey and cyan meshes, respectively, and contoured at 3σ. (**C** and **D**) Surface representation of the active site of (**C**) AiiA-wt and (**D**) AiiA-R4. Phe68 is highlighted in pink and the two mutations at the active site entrance, S20F and V69G, are in shown in blue. Metal ions are shown as beige spheres.

Figure 4

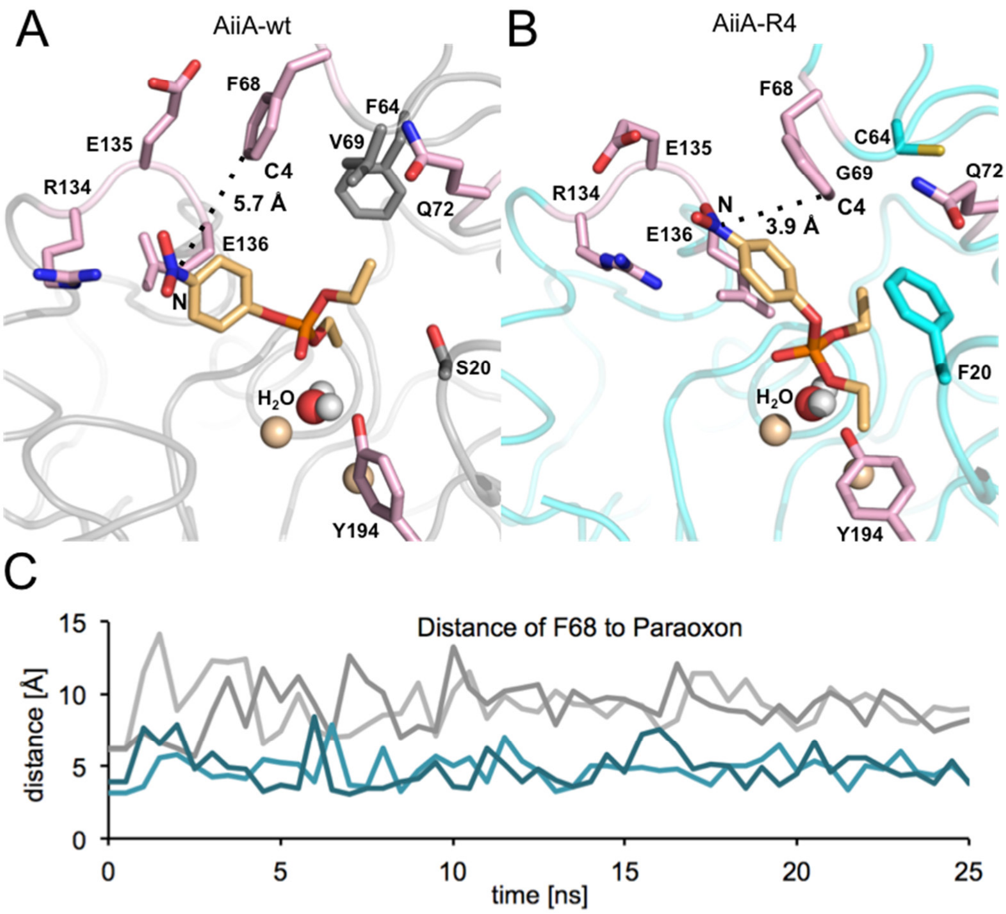


Fig. 4. A snapshot of active site configuration and substrate position in the molecular dynamics (MD) simulations at 25ns of (**A**) AiiA-wt and (**B**) AiiA-R4. Residues that have been repositioned and have not been mutated are shown as sticks and highlighted in pink. Residues mutated during the evolution are highlighted grey and cyan respectively. A plausible nucleophilic water molecule bridging the two metal ions is shown as spheres. The distance between Phe68 and paraoxon was calculated between the C4 atom of the phenyl and the N atom of the nitro group, respectively. (**C**) Distance between Phe68 and paraoxon during two independent MD simulations of 25 ns for AiiA-wt (light and dark grey) and AiiA-R4 (cyan and blue). The distance between Phe68 and paraoxon was calculated between the C4 atom of the phenyl and the N atom of the nitro group, respectively.

Figure 5

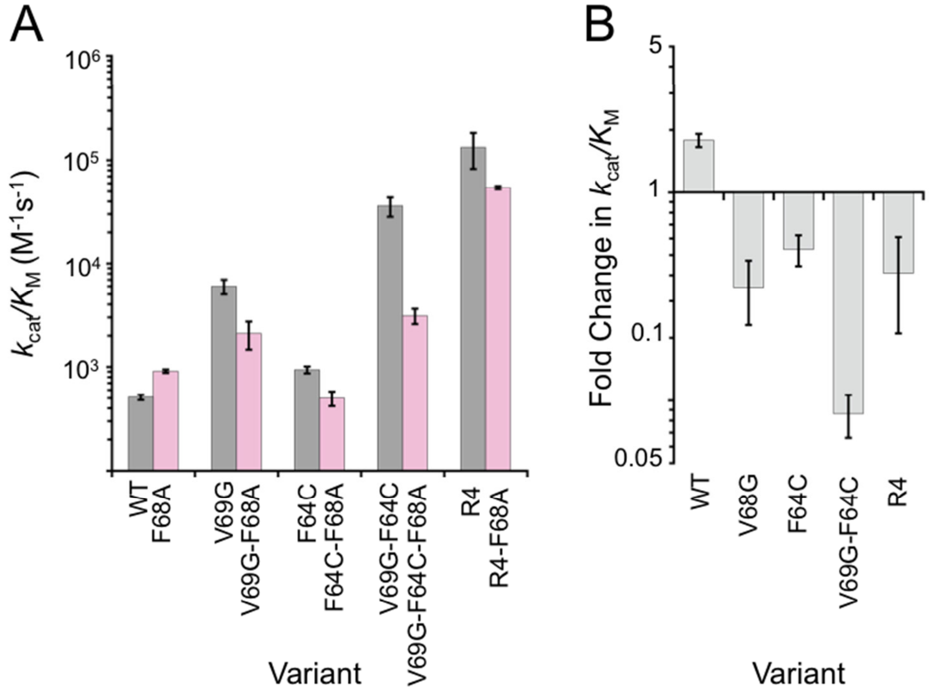


Fig. 5. Comparative mutational scanning of Phe68 in the background of different AiiA variants. (A) The functional contribution of Phe68 was quantified in various genetic background by measuring the effect of F68A on phophotriesterase activity. Catalytic efficiencies (k_{cat}/K_M) of variants. Variants possess Phe68 is highlighted in grey, and variants with Ala68 in pink. Each value represents the average of three measurements with the standard deviation. **(B)** The fold-change in catalytic efficiency that resulted from the introduction of F68A (*e.g.*, AiiA-F68A/AiiA-wt).

Figure 6

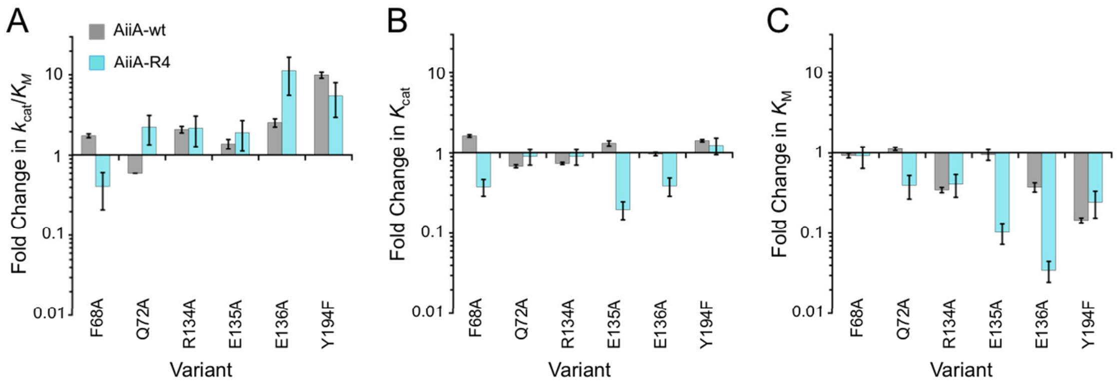


Fig. 6. Comparative mutational scanning of selected non-mutated active site residues between AiiA-wt and AiiA-R4. The functional contribution of various active site residues, Phe68, Gln72, Arg134, Glu135 and Glu136, was measured in the genetic background of AiiA-wt and AiiA-R4. The bars represent fold change in phosphotriesterase activity between AiiA-wt and mutant (grey), or AiiA-R4 and Aii-R4 mutant (cyan). Each value represents the average of three measurements with the standard deviation (bars). (**A**) The fold-change in k_{cat}/K_M. (**B**) The fold-change in k_{cat}. (**C**) The fold-change in K_M.

Figure 7

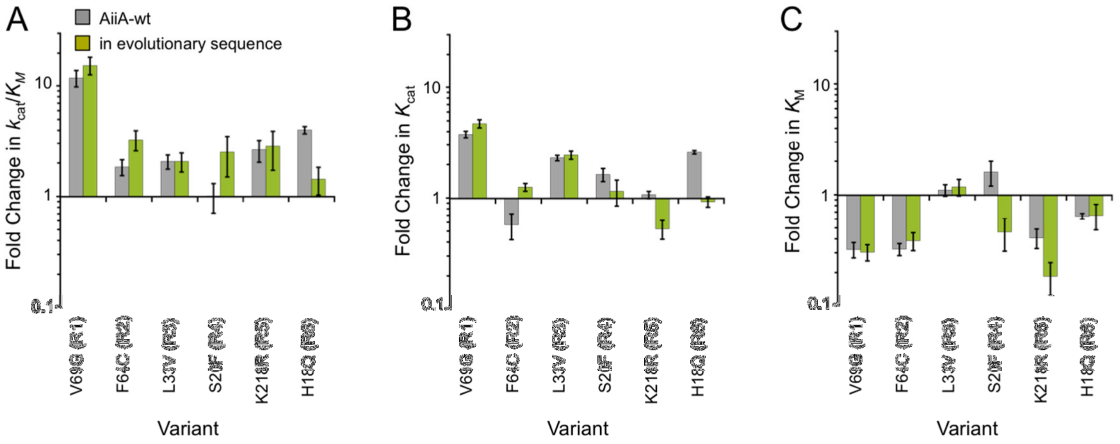
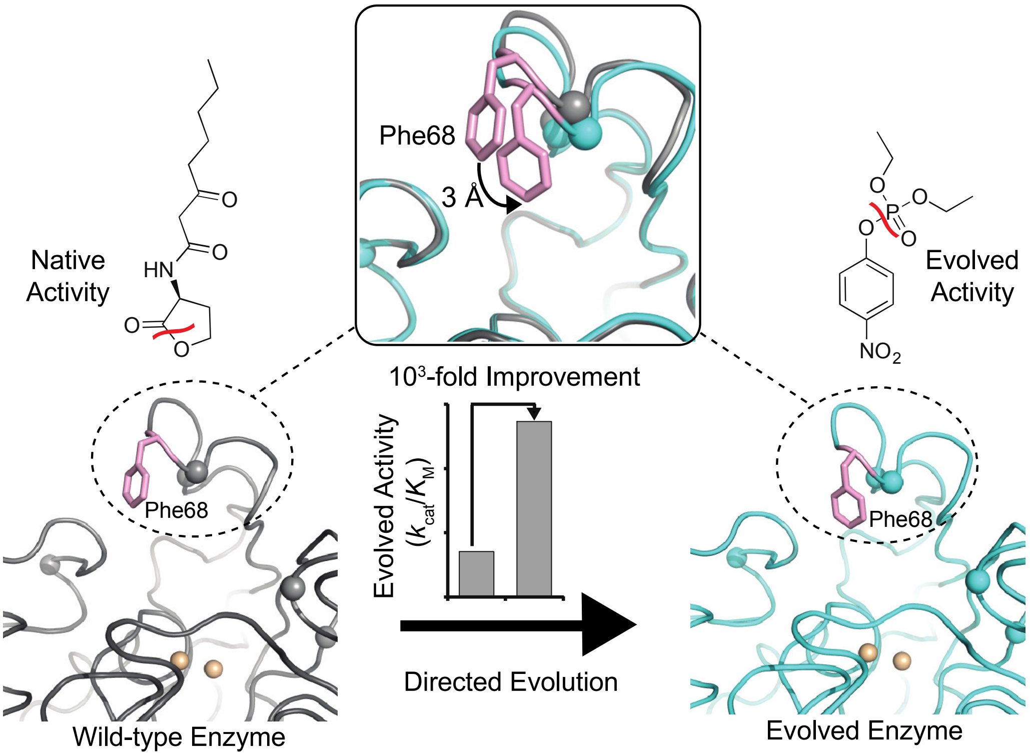


Fig. 7. Change in mutational effect in the background of AiiA-wt and in the evolutionary trajectory. Mutations that occurred in the trajectory were introduced into AiiA-wt and their catalytic efficiencies (grey) were compared to their respective improvement in catalytic efficiency during the evolutionary trajectory (green). Each value represents the average of three measurements with the standard deviation (bars). **B**) The fold-change in k_{cat}/K_M. (**B**) The fold-change in k_{cat}. (**C**) The fold-change in K_M.

TABLE OF CONTENTS



Supplementary Figure S1

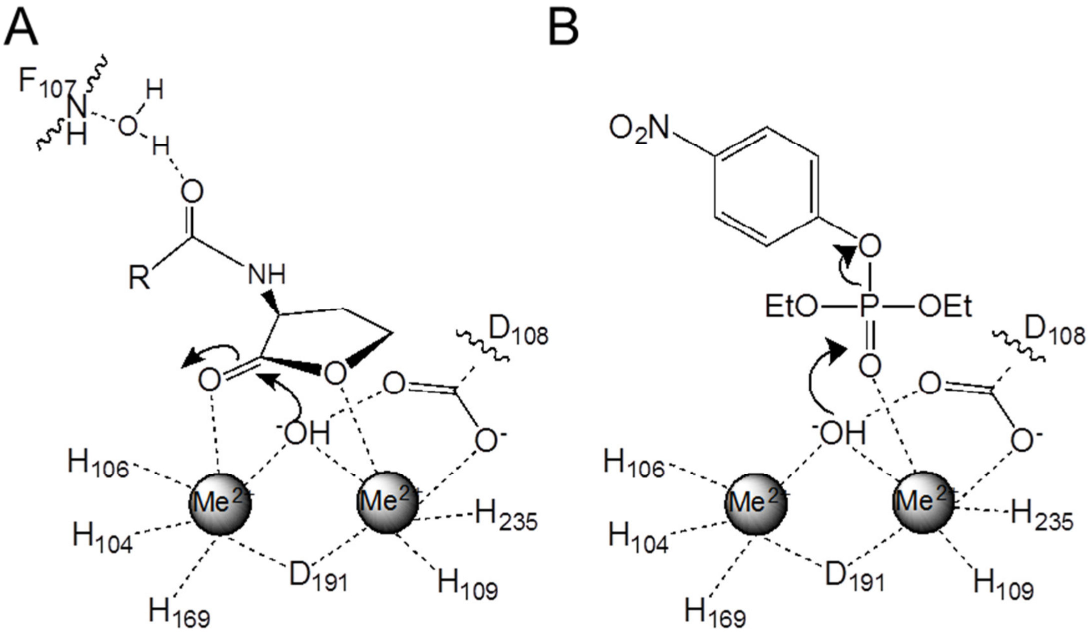


Fig. S1: Proposed catalytic mechanism of AiiA for native and promiscuous substrates. (**A**) Acyl homoserine lactone hydrolysis of AiiA as described by Momb *et al.* (35)**.** Metal cations are coordinated by adjacent histidines, Asp191, and the bridging hydroxide ion, which serves as a nucleophile that attacks the carbonyl carbon of the AHL substrate (35). Binding of the substrate is stabilized by hydrogen bond interactions between the two carbonyl groups of the substrate with the metal ions, as well as interactions between the O_3 of the substrate with Phe107 via a water molecule[36]. (**B**) Proposed binding and mechanism of paraoxon-ethyl hydrolysis of AiiA. The hydroxide ion that serves as the nucleophile in lactone hydrolysis (35) likely plays the same role in paraoxon hydrolysis.

Supplementary Figure S2

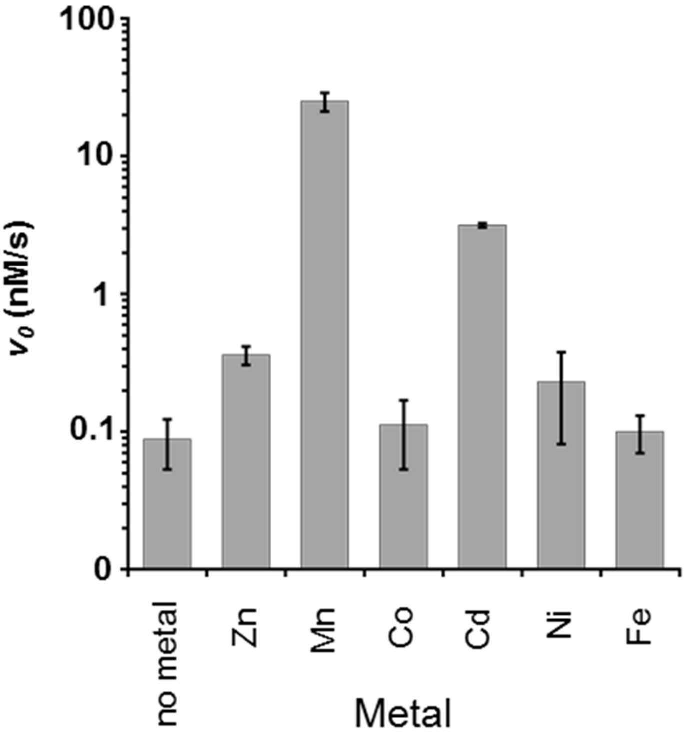


Fig. S2: Metal dependency of AiiA for paraoxon activity. AiiA was expressed and purified in the presence of 100 µM of the corresponding metal or without any metal supplied (no metal). Metals were added to the LB media during expression and to all buffers for purification and activity assays. Enzymatic activities were measured using a final concentration of 5 µM of enzyme and 500 µM of paraoxon substrate. The error bars represent the standard deviation of triplicate measurements.

Supplementary Figure S3

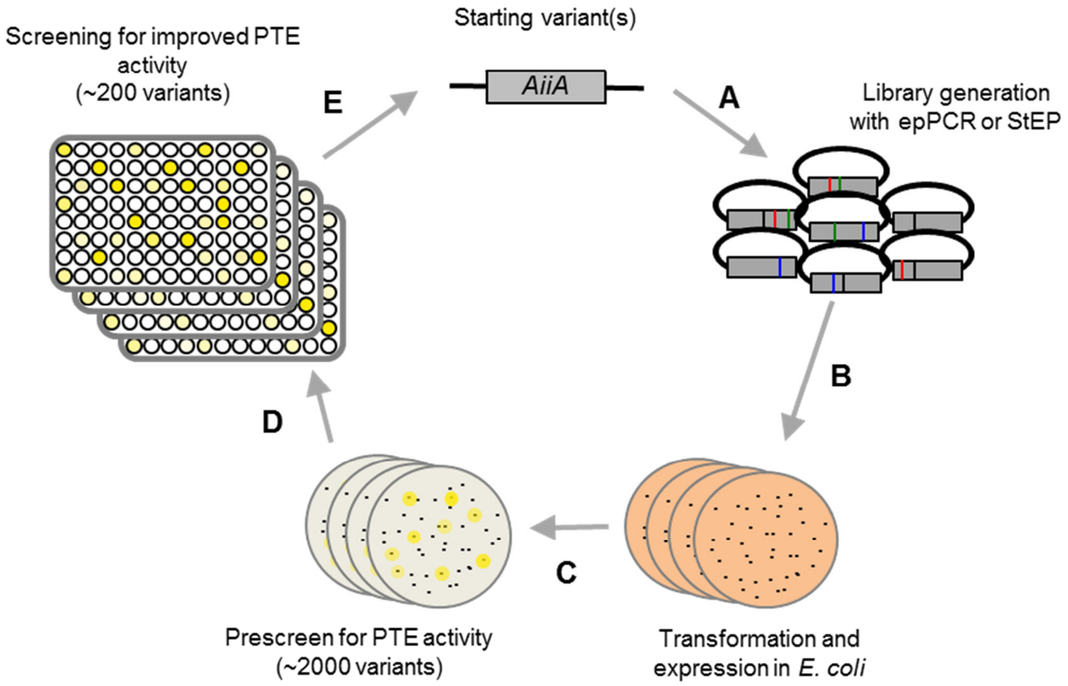


Fig. S3: Overview of the directed evolution scheme. (**A**) Starting variant(s) were mutagenized using error-prone PCR or StEP (staggered extension process) recombination and subcloned into a vector containing a N-terminal Strep-tag. (**B**) The resulting library was transformed into *E. coli* BL21 (DE3) cells for protein expression. The plates were replicated onto nitrocellulose membranes. (**C**) Colonies on the nitrocellulose membrane were lysed and pre-screened for paraoxonase activity, which can be identified through the development of yellow colour due to the release of *p*-nitrophenol. (**D**) Approximately 200 of the most active colonies were picked from nitrocellulose membrane for screening in ~2 × 96-well plates liquid culture. (**E**) The most improved variant(s) served as a starting point for the next round of directed evolution

Supplementary Figure S4

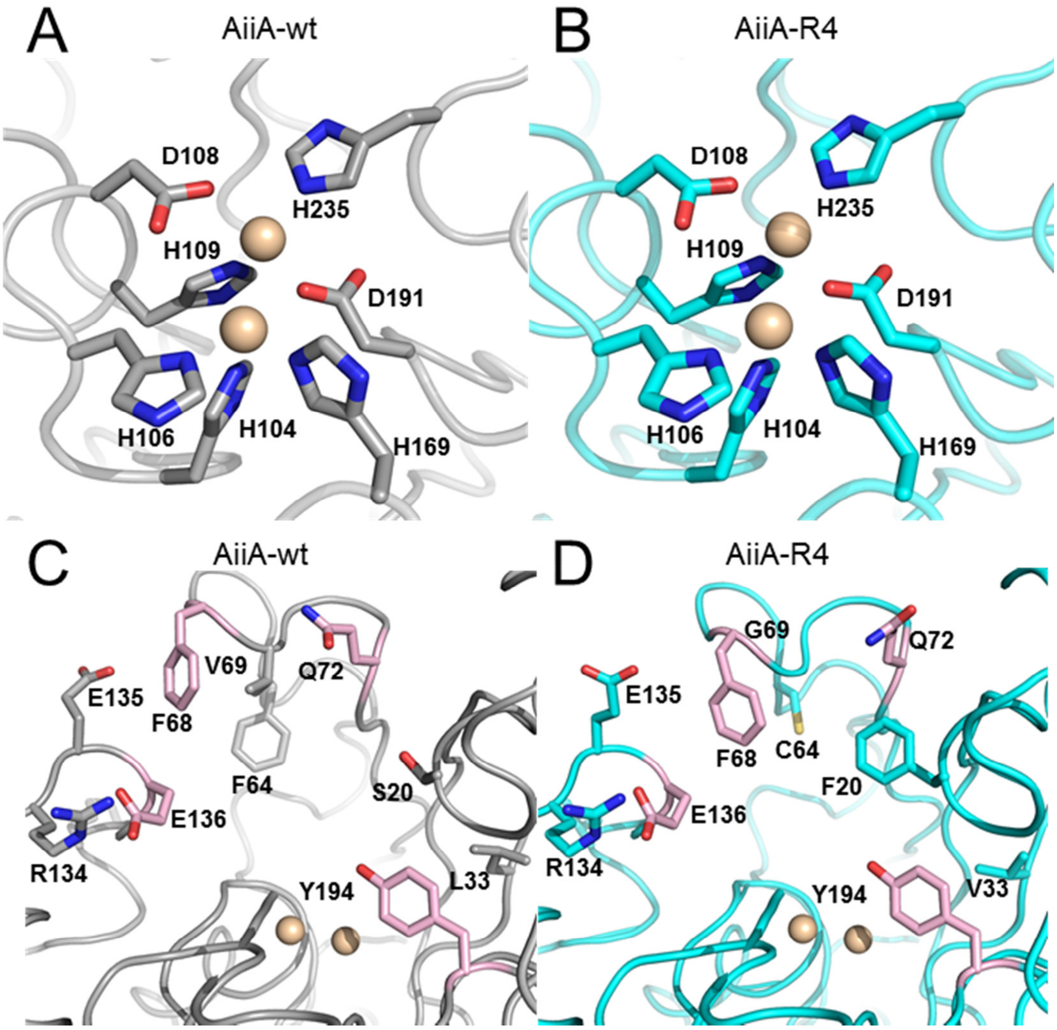


Fig. S4: Structural comparison between AiiA-wt and AiiA-R4. Configuration of active site metals (spheres) and metal binding residues (sticks) of (**A**) AiiA-wt and (**B**) AiiA-R4 are shown. Close up view of the active site of (**C**) AiiA-wt and (**D**) AiiA-R4. Residues that are repositioned in the crystallized structure are highlighted in pink. Residues mutated during the directed evolution are highlighted in cyan.

Supplementary Figure S5

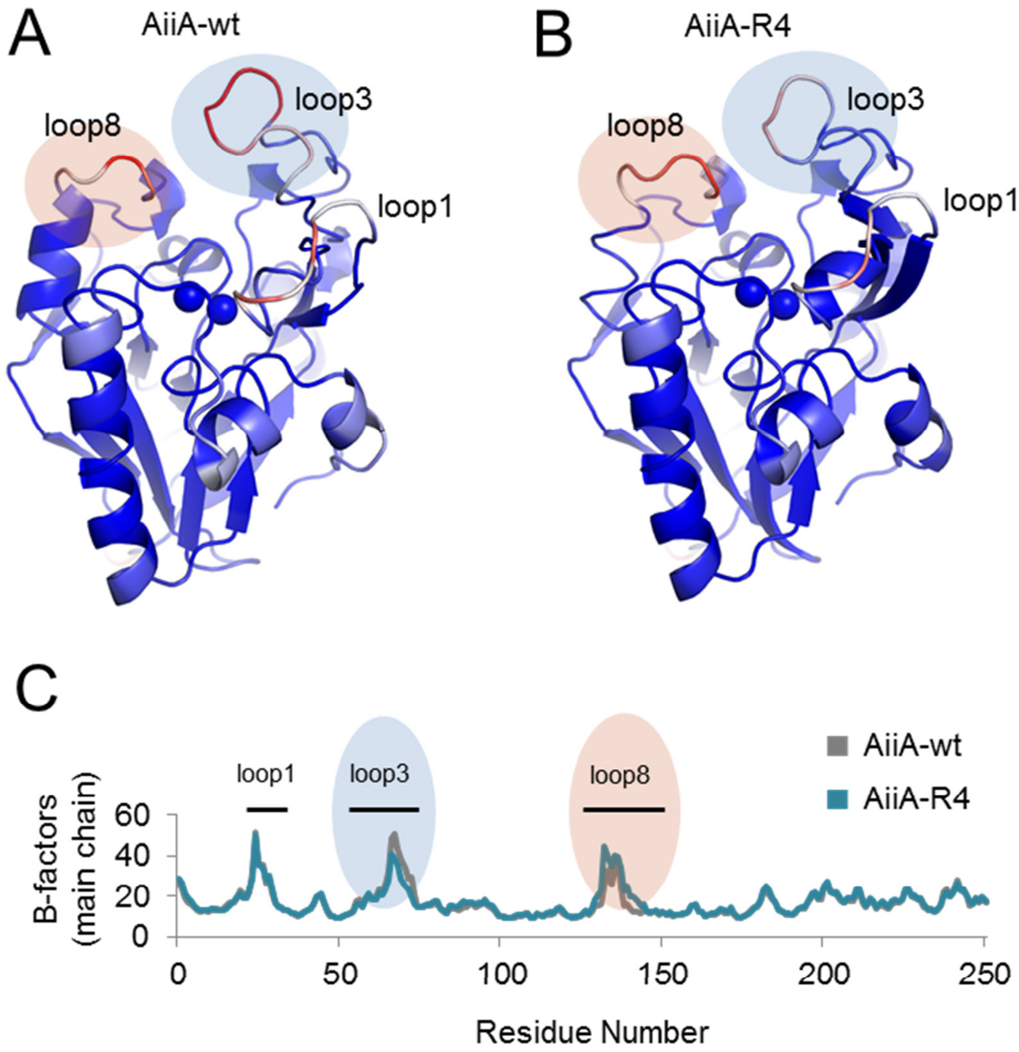


Fig. S5: B-factor in the crystal structures of (**A**) AiiA-wt and (**B**) AiiA-R4. The B-factor value increases from blue to red and loop regions where the B-factor changed between AiiA-wt and AiiA-R4 are highlighted with circles. (**C**) Comparison of the B-factors of the main chains of AiiA-wt and AiiA-R4 and loop regions where the B-factor changed between AiiA-wt and AiiA-R4 are highlighted with circles.

Supplementary Figure S6

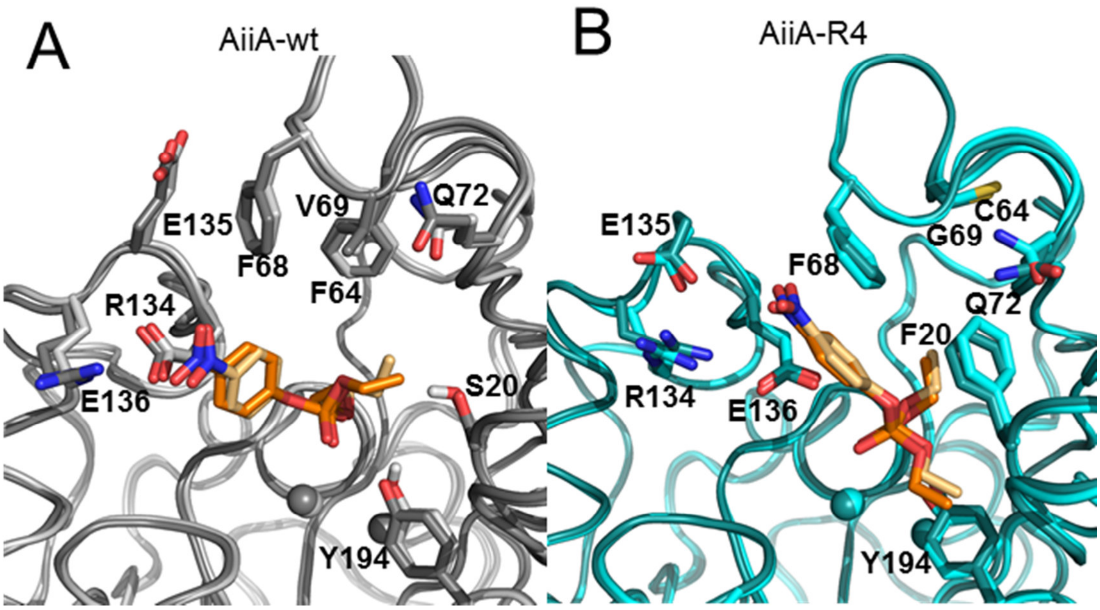
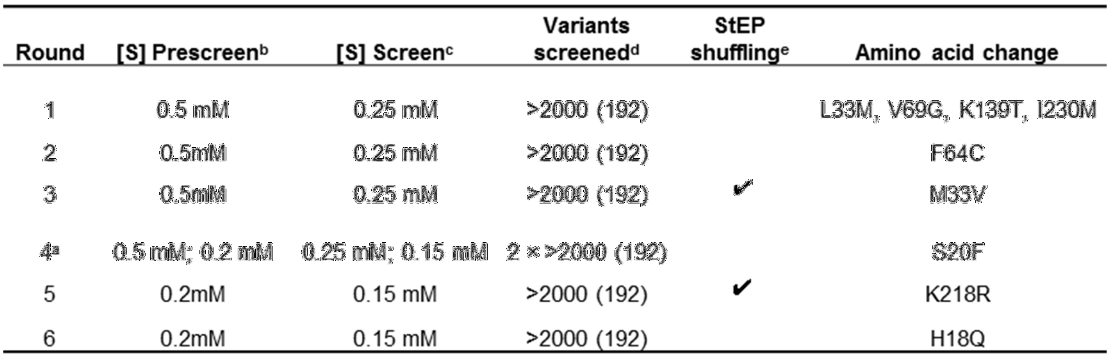


Fig. S6: Overlay of two independent molecular dynamics (MD) simulations of (A) AiiA-wt (first and second trials colored light and dark grey, respectively) (**B**) AiiA-R4 (first and second trials colored cyan and blue, respectively). Residues that were mutated during the trajectory or appear to have been repositioned are shown as sticks. Paraoxon substrate for the first and second trials are colored yellow and orange, respectively.

Supplementary Table S1

Table S1. Library information and amino acid changes.



Round	[S] Prescreen[b]	[S] Screen[c]	Variants screened[d]	StEP shuffling[e]	Amino acid change
1	0.5 mM	0.25 mM	>2000 (192)		L33M, V69G, K139T, I230M
2	0.5mM	0.25 mM	>2000 (192)		F64C
3	0.5mM	0.25 mM	>2000 (192)	✓	M33V
4[a]	0.5 mM; 0.2 mM	0.25 mM; 0.15 mM	2 × >2000 (192)		S20F
5	0.2mM	0.15 mM	>2000 (192)	✓	K218R
6	0.2mM	0.15 mM	>2000 (192)		H18Q

[a] In round 4 two independent rounds of screening were performed with different substrate concentrations, because no improved variant could be identified with the higher substrate concentrations.

[b] Paraoxon concentration used for agar plate prescreen as described in material & methods.

[c] Paraoxon concentration used in 96-well plate screening as described in material & methods.

[d] The number of variants screened in agar plate prescreen and in 96-well plates (brackets) per round.

[e] DNA shuffling was performed in rounds where several improved variants with distinct mutations were identified.

Supplementary Table S2

Table S2. Site-directed mutagenesis primers.

Primer name	Template	Mutation	Sequence (5' to 3')
V69G-Fwd	AiiA-wt	V69G	ctttttaacggtacatttggtgaaggacagatcttaccg
V69G-Rev			cggtaagatctgtccttcaccaaatgtaccgttaaaaag
L33M-Fwd	AiiA-wt	L33M	gggaaactattaaacatgccggtgtggtgttatcttttg
L33M-Rev			caaaagataacaccacaccggcatgtttaatagtttccc
K139T-Fwd	AiiA-wt	K139T	gcacttcatagagaagaatatatgacagaatgtatattaccgcatttg
K139T-Rev			caaatgcggtaatatacattctgtcatatattcttctctatgaagtgc
I230M-Fwd	AiiA-wt	I230M	gttgtgaaaaaagagaaaccaatgattttctttggtcatgatatagagc
I230M-Rev			gctctatatcatgaccaaagaaaatcattggtttctcttttttcacaac
F64C-Fwd	AiiA-wt	F64C	gttaataatgaagggctttgtaacggtacatttgttgaagg
F64C-Rev			ccttcaacaaatgtaccgttacaaagcccttcattattaac
L33V-Fwd	AiiA-wt	L33V	gggaaactattaaacgtgccggtgtggtgttatcttttg
L33V-Rev			caaaagataacaccacaccggcacgtttaatagtttccc
S20F-Fwd	AiiA-wt	S20F	gttgcatgttggatcattcgtttgttaacagtgcgttaac
S20F-Rev			gttaacgcactgttaacaaacgaatgatccaacatgcaac
K218R-Fwd	AiiA-wt	K218R	gatccagaattagctttatcttcaattagacgtttaaaagaagttgtg
K218R-Rev			cacaacttcttttaaacgtctaattgaagataaagctaattctggatc
H18Q-Fwd	AiiA-wt	H18Q	gtcgttgcatgttggatcaatcgtctgttaacagtgcgttaac
H18Q-Rev			gttaacgcactgttaacagacgattgatccaacatgcaacgac
F68A-Fwd	AiiA-wt	F68A	gggctttttaacggtacagctgttgaaggacagatcttac
F68A-Rev			gtaagatctgtccttcaacagctgtaccgttaaaaagccc
Q72A-Fwd	AiiA-wt	Q72A	gcttgctcttctaggagcgatcttaccgaaaatgactgaggaag
Q72A-Rev			gcttgctcttctccttcaacaaatgtaccgttaaaaagcccttcattattaac
R134A-Fwd	AiiA-wt	R134A	gaggcagcacttcatgcagaagaatatatgaaag
R134A-Rev			ctttcatatattcttctgcatgaagtgctgcctc
E135A-Fwd	AiiA-wt	E135A	gaggcagcacttcatagagcagaatatatgaaagaatgtatattacc
E135A-Rev			ggtaatatacattctttcatatattctgctctatgaagtgctgcctc
E136A-Fwd	AiiA-wt	E136A	gcagcacttcatagagaagcatatatgaaagaatg
E136A-Rev			cattctttcatatatgcttctctatgaagtgctgc
D108N-Fwd	AiiA-wt	D108N	gttctcacttacattttaatcatgcaggaggaaacggtgc
D108N-Rev			gcaccgtttcctcctgcatgattaaaatgtaagtgagaac
Y194F-Fwd	AiiA-wt, AiiA-R4	Y194F	cgattgatgcatcgttcacgaaagagaattttgaagatgaagtg
Y194F-Rev			cacttcatcttcaaaattctctttcgtgaacgatgcatcaatcg
F68A-F64C-Fwd	AiiA-F68	F64C	gttaataatgaagggctttgtaacggtacagctgttgaagg
F68A-F64C-Rev			ccttcaacagctgtaccgttacaaagcccttcattattaac
V69G-F64C-Fwd	AiiA-V69G	F64C	gttaataatgaagggctttgtaacggtacatttggtgaagg
V69G-F64C-Rev			ccttcaccaaatgtaccgttacaaagcccttcattattaac
F68A-V69G-F64C-Fwd	AiiA-F68A-V69G	F64C	gttaataatgaagggctttgtaacggtacagctggtgaagg
F68A-V69G-F64C-Rev			ccttcaccagctgtaccgttacaaagcccttcattattaac
R1-F68A-Fwd	AiiA-R1	F68A	gggctttttaacggtacagctggtgaaggacagatcttac
R1-F68A-Rev			gtaagatctgtccttcaccagctgtaccgttaaaaagccc
R4-F68A-Fwd	AiiA-R4	F68A	gggctttgtaacggtacagctggtgaaggacagatcttac
R4-F68A-Rev			gtaagatctgtccttcaccagctgtaccgttacaaagccc
R4-Q72A-Fwd	AiiA-R4	Q72A	gcttgctcttctaggagcgatcttaccgaaaatgactgaggaag
R4-Q72A-Rev			gcttgctcttctccttcaccaaatgtaccgttacaaagcccttc
R4-R134A-Fwd	AiiA-R4	R134A	gaggcagcacttcatgcagaagaatatatgacag
R4-R134A-Rev			ctgtcatatattcttctgcatgaagtgctgcctc
R4-E135A-Fwd	AiiA-R4	E135A	gaggcagcacttcatagagcagaatatatgacagaatgtatattacc
R4-E135A-Rev			ggtaatatacattctgtcatatattctgctctatgaagtgctgcctc
R4-E136A-Fwd	AiiA-R4	E136A	gcagcacttcatagagaagcatatatgacagaatg
R4-E136A-Rev			cattctgtcatatatgcttctctatgaagtgctgc

Supplementary Table S3

Table S3. Crystallographic data collection and refinement statistics.

	AiiA-wt	**AiiA-R4**
PDB ID	5EH9	5EHT
Wavelength (Å)	0.9537	0.9537
Resolution range (Å)	45.47-1.29 (1.34-1.29)	32.49-1.29 (1.34-1.29)
Space group	P 2_1 2_1 2_1	P 2_1 2_1 2_1
Unit cell (Å)	54.51 55.47 79.358	54.64 55.74 79.88
Unit cell (°)	90.00 90.00 90.00	90.00 90.00 90.00
Total reflections	430254 (42089)	826812 (69935)
Unique reflections	61128 (6032)	61760 (6093)
Multiplicity	7.0 (7.0)	13.4 (11.5)
Completeness (%)	99.80 (99.83)	99.47 (99.33)
Mean I/sigma (I)	16.33 (1.32)	20.78 (1.44)
Wilson B-factor (Å^2)	15.87	13.34
R-merge	0.065 (1.56)	0.082 (1.76)
R-means	0.070	0.085
CC1/2	0.999 (0.566)	0.999 (0.636)
CC*	1 (0.85)	1 (0.88)
R-work	0.12 (0.24)	0.14 (0.28)
R-free	0.17 (0.27)	0.17 (0.35)
Number of non-hydrogen atoms	2259	2204
Macromolecules	2036	2013
Ligands	56	32
Water	167	159
Protein residues	253	253
RMS (bonds, Å)	0.022	0.022
RMS (angles, °)	2.27	2.31
Ramachandran favoured (%)	96	96
Ramachandran outliers (%)	0.39	0.79
Clash score	4.32	3.46
Average B-factor (Å^2)	23.9	20.8
macromolecules	22.8	20.1
ligands	36.6	30.5
solvent	32.7	27.3

* Highest-resolution shell is shown in parentheses.

Supplementary Table S4

Table S4. Cell lysate activities.

Variant	Lactonase[a] Rate (nM/s)	Lactonase[a] Relative to Previous Round	Paraoxonase[b] Rate (nM/s)	Paraoxonase[b] Relative to Previous Round
WT	**1.1** (± 0.05) **x** 10^{4}	-	**5.1** ± 0.8	-
R1	**5.5** (± 0.2) **x** 10^{3}	0.5	**50** ± 4.9	9.8
R2	**3.1** (± 0.04) **x** 10^{3}	0.6	**190** ± 11	3.7
R3	**5.9** (± 0.2) **x** 10^{3}	1.9	**460** ± 16	2.5
R4	**1.3** (± 0.1) **x** 10^{4}	2.2	**950** ± 7.0	2.1
R5	**1.5** (± 0.1) **x** 10^{4}	1.2	**1000** ± 48	1.1
R6	**1.7** (± 0.03) **x** 10^{4}	1.1	**1400** ± 88	1.3
		Relative to WT		Relative to WT
V69G	**4.9** (± 0.1) **x** 10^{3}	0.4	**52** ± 2.5	10.1
L33M	*n.d.*	*n.d.*	**4.7** ± 0.1	1.0
I230M	*n.d.*	*n.d.*	**5.3** ± 0.05	1.1
K139T	*n.d.*	*n.d.*	**5.6** ± 0.1	1.1
F64C	**2.5** (± 0.1) **x** 10^{3}	0.2	**17** ± 2.2	3.4
L33V	**2.7** (± 0.1) **x** 10^{4}	2.5	**18** ± 0.5	3.5
S20F	**6.0** (± 0.05) **x** 10^{3}	0.5	**8.8** ± 1.0	1.7
K218R	**1.3** (± 0.1) **x** 10^{4}	1.2	**5.9** ± 1.3	1.2
H18Q	**1.1** (± 0.03) **x** 10^{4}	1.0	**13** ± 1.8	2.5
F68A	**1.7** (± 0.02) **x** 10^{4}	1.6	**10** ± 0.9	2.0
Q72A	**8.4** (± 0.4) **x** 10^{3}	0.8	**2.5** ± 0.2	0.5
R134A	**8.8** (± 0.3) **x** 10^{3}	0.8	**5.0** ± 2.0	1.0
E135A	**1.1** (± 0.05) **x** 10^{4}	1.0	**18** ± 0.2	3.4
E136A	**4.6** (± 0.1) **x** 10^{3}	0.4	**4.3** ± 1.2	0.8
Y194F	**6.8** (± 0.1) **x** 10^{3}	0.6	**39** ± 1.1	7.6
V69G-F64C	**3.8** (± 0.2) **x** 10^{3}	0.3	**200** ± 13	38.8
F64C-S20F	**3.3** (± 0.2) **x** 10^{3}	0.3	**22** ± 0.3	4.3
V69G-F64C-S20F	**5.9** (± 0.2) **x** 10^{3}	0.5	**350** ± 16	68.0
V69G-F68A	**1.6** (± 0.1) **x** 10^{4}	1.5	**21** ± 0.2	4.0
F64C-F68A	**4.3** (± 0.2) **x** 10^{3}	0.4	**18** ± 0.4	3.6
V69G-F64C-F68A	**4.4** (± 0.02) **x** 10^{3}	0.4	**21** ± 1.2	4.0
		Relative to R4		Relative to R4
R4-F68A	**1.5** (± 0.02) **x** 10^{4}	1.1	**310** ± 9.4	0.3
R4-Q72A	**1.2** (± 0.03) **x** 10^{4}	0.9	**880** ± 25	0.9
R4-R134A	**9.1** (± 0.3) **x** 10^{3}	0.7	**810** ± 13	0.9
R4-E135A	**1.2** (± 0.01) **x** 10^{4}	0.9	**680** ± 17	0.7
R4-E136A	**1.8** (± 0.7) **x** 10^{3}	0.1	**180** ± 9.4	0.2
R4-D108N	**1.5** (± 1.1) **x** 10^{2}	0.0	**5.3** ± 1.3	0.0
R4-Y194F	**3.9** (± 0.2) **x** 10^{3}	0.3	**950** ± 35	1.0

All variants were measured using the same amount of lysate (100 µL in 200 µL reaction) and the same substrate concentration. Measurements were performed in triplicate and values were averaged. Errors represent standard deviation.

n.d. means not determined.

[a] To assay lactonase activity a final concentration of 250 µM N-acyl homoserine lactone was used.

[b] To assay paraoxonase activity a final concentration of 250 µM paraoxon was used.

Chapter 7

Enzyme evolvability is contingent on the initial sequence background

Baier, F., **Hong, NS**, Yang, G., Carr, P.D., Jackson, C.J.,

and Tokuriki, N.

Submittion ready manuscript 2016, July

This manuscript has been prepared in collaboration with the laboratory of Nobuhiko Tokuriki, University of British Columbia, Canada, in June 2016. Nobuhiko Tokuriki and co-workers, Florian Baier and Gloria Yang, designed and performed directed evolution experiments, kinetic analysis and mutagenic characterisations. I purified the proteins, including TEV proteases and performed SDS-PAGE analysis, size exclusion chromatography, crystallized the enzymes and solved the collection of structures under the guidance and supervision of Colin Jackson. Paul Carr helped in the crystal data and in data processing and refinement. I prepared the corresponding method and results sections of this manuscript, including figures, with Colin Jackson.

Enzyme evolvability is contingent on the initial sequence background

Florian Baier[a], Nansook Hong[b], Gloria Yang[a], Paul D Carr[b], Colin J Jackson[b]
and Nobuhiko Tokuriki[a]

[a]Michael Smith Laboratories, University of British Columbia, Vancouver, BC, Canada.

[b]Research School of Chemistry, Australian National University, Canberra, ACT, Australia.

Corresponding author:
Nobuhiko Tokuriki
Michael Smith Laboratories
University of British Columbia, Vancouver, V6T 1Z4, BC, Canada
Tel: +1-604-822-8156
Fax: +1-604-822-2114
Email: tokuriki@msl.ubc.ca

Abstract

Can an enzyme evolve high catalytic activity from any sequence background? We address this question by evolving two related β-lactamases, in parallel, towards a shared promiscuous phosphonate monoester hydrolase (PMH) activity. We observe striking differences in the evolvability and phenotypic solutions for each enzyme: the soluble expression of NDM1 (the amount of functional enzyme in the cell) is compromised to improve catalytic efficiency by over 20,000-fold. In contrast, the catalytic efficiency of VIM2 improved only 60-fold, but its soluble expression increased significantly. Furthermore, both enzymes adapt *via* different but deterministic mutational solutions. For example, a single mutation, W93G, improves the PMH activity of NDM1 by >300-fold, but decreases PMH activity by 10-fold in the background of VIM2, and is only partially compatible with other B1 β-lactamases. Our results suggest that not all enzymes can evolve to the same catalytic efficiency, and thus a diverse set of initial genotypes might be important for the evolution of an efficient enzyme function.

Introduction

The evolvability of a new enzyme function has to satisfy two major prerequisites. First, an enzyme already needs to be endowed with the new function at a sufficient level to provide a fitness advantage when a selection pressure is applied for the function (1, 2). Second, the new function must be adaptable enough to reach high efficiency through the accumulation of mutations (evolutionary plasticity) (3). Numerous studies have reported evidence for the first prerequisite; many enzymes are shown to exhibit latent (weak) promiscuous activities towards different chemical reactions than that which the enzyme may have evolved for (4). Such promiscuous activities can provide a "springboard" for the evolution of new functions (5, 6). The second prerequisite is less clear. Would two different enzymes with the same promiscuous activity serve equally well as evolutionary starting points to generate a new and efficient enzyme? Or is enzyme evolution contingent on the starting sequence? The theoretical framework of molecular evolution indicates that only certain genotypes can access specific adaptive pathways for the acquisition of a new function (7). However, experimental evidence supporting the idea that there are significant differences between different genotypes in evolvability and evolutionary contingency is still scarce.

A more profound question relates to the molecular basis for differences in evolvability. It has been formulated that protein stability is associated with evolvability, that an excess of protein stability can buffer the destabilizing effects of function-changing mutations, and thus enhance evolvability (8). Another molecular determinant of evolvability is mutational epistasis, i.e. the functional effects of mutations vary depending on the genetic background (9). Thus, evolution towards new functions relies on the accessibility of beneficial functional mutations; if a certain sequence cannot access a function-changing mutation with a single mutation step, and instead requires the fixation of an epistatic permissive mutation(s) to allow incorporation of a function-changing mutation, the sequence may encounter an evolutionary dead-end. Such a scenario could result in differences in evolvability even between orthologous enzymes with similar physicochemical properties. There is growing evidence that indicates intramolecular mutational epistasis is prevalent (10–12). Furthermore, adaptive trajectories for new function can be highly restricted in terms of the order in which functional mutations can accumulate (13). However, only a few studies have reported experimental evidence for epistasis causing differences in evolvability between orthologous enzymes (14, 15). In particular, previous work has demonstrated that the effects of the initial functional mutation(s) in an evolutionary trajectory on a promiscuous activity can vary significantly depending on genetic backgrounds among orthologous enzymes (16). How these initial sequence differences specifically shape complete evolutionary trajectories, due to epistasis, and lead to different evolutionary outcomes is still a matter of debate (17–23). Indeed, studying molecular basis of evolvability in a long-term evolutionary scenario, i.e. whether orthologous enzymes will plateau at different fitness levels, has not been addressed.

In this work, we have conducted an empirical test for evolutionary contingency by performing parallel experimental evolution, starting from two orthologous enzymes, towards a single promiscuous activity, examined differences in evolvability between the enzymes, and elucidated its molecular basis. Our model is the shared promiscuous PMH activity of B1 β-lactamases, which comprise a functional family in the metallo-β-lactamase (MBL) superfamily. The physiological function of β-lactamase is to confer β-lactam antibiotic resistance to bacteria by hydrolyzing the β-lactam ring (k_{cat}/K_M is typically approximately 10^6 $M^{-1}s^{-1}$, **Figure 1 A** and **Table 1**). B1 β-lactamases share a highly conserved structure, function and catalytic mechanism, despite their extensive sequence divergence (amino acid sequence identity between enzymes can be as low as 20%; **Table 2**). In addition to the conserved native function, these enzymes exhibit a variety of promiscuous activities towards catalytically distinct chemical reactions, such as phosphodiester, phosphotriester and phosphonate monoester hydrolysis (PMH) activity (4).

Here, we have performed a comparative laboratory evolution experiment starting from four β-lactamases, NDM1, EBL1, VIM2 and VIM2, towards promiscuous PMH activity. We have evolved each enzyme using directed evolution. For two enzymes, NDM1 and VIM2, we continued the evolution until the enzyme fitness (enzymatic activity in cell lysate) plateaus. We then quantified to what extent the evolvability of the two enzymes differed, i.e. the rate of improvement and the level of the fitness plateau. We characterized phenotypic (catalytic activity, solubility and stability) and genotypic (sequence) changes that lead to differences in evolvability. High-resolution structural analysis of the most evolved variants with and without bound substrate/product *via* X-ray crystallography allowed us to elucidate the molecular basis of the different evolutionary contingency between the two enzymes/trajectories. Finally, we discuss the implications of our observations for the evolution of proteins in nature and the laboratory.

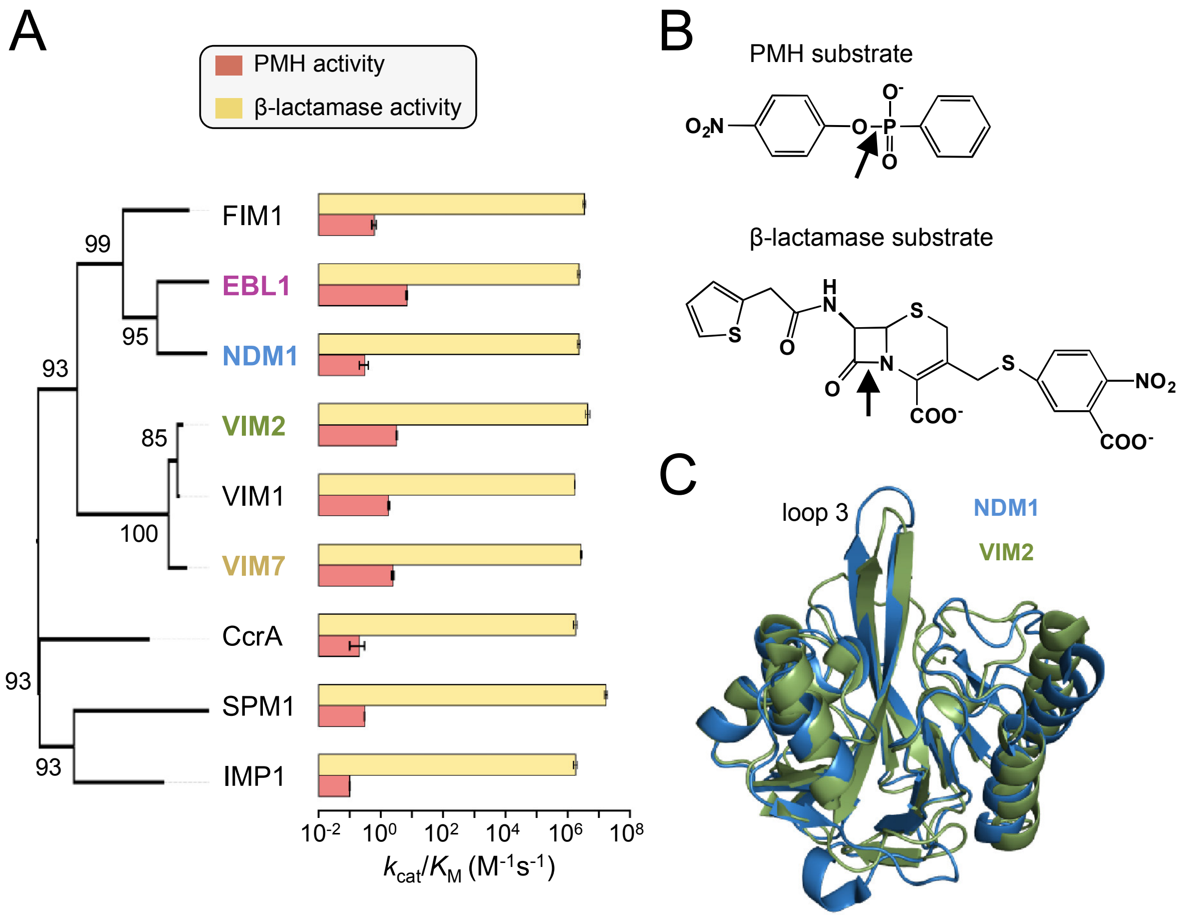


Figure 1. The functional and structural similarity of B1 β-lactamases. (**A**) Catalytic efficiencies (k_{cat}/K_M) of nine B1 β-lactamases for native β-lactamase and promiscuous PMH activities. The phylogenetic relationship is shown on the left with bootstrap values at each node. Error bars of catalytic efficiencies represent standard deviation of triplicate measurements. (**B**) Chemical structures of the chromogenic β-lactamase substrate (CENTA) and PMH substrate (4-nitrophenyl phenylphosphonate) used in this study to assay enzymatic activity. The arrows indicate the bond broken during catalysis (**C**). Structural overlay of the two B1 β-lactamases NDM1 (blue, PDB ID: 3SPU) and VIM2 (green, PDB ID: 1KO3).

Table 1. Information on enzymes used in this study.

Enzyme	Uniprot ID	Organism	PDB ID	Structural resolution (Å)
FIM1	K7SA42	*Pseudomonas aeruginosa*	*n.a.*	*n.a.*
EBL1	Q2N9N3	*Erythrobacter litoralis*	*n.a.*	*n.a.*
NDM1	C7C442	*Klebsiella pneumonia*	3spu	2.1
VIM2	Q9K2N0	*Pseudomonas aeruginosa*	1ko3	1.9
VIM1	Q9XAY4	*Pseudomonas aeruginosa*	*n.a.*	*n.a.*
VIM7	Q840P9	*Pseudomonas aeruginosa*	2y87	1.9
CcrA	P25910	*Bacteroides fragilis*	1znb	1.9
SPM1	Q8G9Q0	*Pseudomonas aeruginosa*	2fhx	1.9
IMP1	Q79MP6	*Pseudomonas aeruginosa*	1ddk	3.1

n.a. No structural information has been deposited in the protein data bank (www.rcsb.org).

Table 2. Pairwise sequence and structure comparison.

	FIM1*	**EBL1***	**NDM1**	**VIM2**	**VIM1***	**VIM7**	**CcrA**	**SPM1**	**IMP1**
FIM1*		-	-	-	-	-	-	-	-
EBL1*	46%		-	-	-	-	-	-	-
NDM1	45%	56%		1.03	-	0.97	1.13	1.05	1.14
VIM2	36%	32%	35%		-	0.36	1.07	1.03	1.11
VIM1*	36%	31%	34%	94%		-	-	-	-
VIM7	32%	27%	32%	80%	81%		0.85	0.94	1.15
CcrA	32%	31%	30%	31%	31%	29%		0.97	1.25
SPM1	25%	24%	22%	25%	25%	27%	24%		1.13
IMP1	29%	29%	31%	30%	29%	30%	34%	32%	

RMSD [Å]

Sequence identity

Pairwise structural similarity (root mean standard deviation (RMSD)) was computed using the sequence independent command "super" in pymol (24).

*indicates that no structural information is available.

Pairwise amino acid sequence identities were calculated using the web-based program SIAS (http://imed.med.ucm.es/Tools/sias.html; standard parameters with gaps not taken into account) from a structure-based sequence alignment using Espresso (25) (standard parameters).

Material and Methods

Calculation of sequence identity, structural similarity and phylogeny. Pairwise sequence identities were calculated with the web-based program SIAS (http://imed.med.ucm.es/Tools/sias.html) using a multiple sequence alignment generated with ClustalW2 (26) using default parameters. Pairwise structural similarity (root mean square deviation (RMSD)) of available structures were computed using the sequence-independent structure alignment command "super" in PyMOL (The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC). A multiple sequence alignment was performed with the 3D-Expresso algorithm (25), which incorporates available structural information. Phylogenetic trees were constructed based on the alignment using the maximum likelihood method (500 Bootstraps) in Mega5 (27) with default parameters.

Bacterial expression. The B1 β-lactamase genes were sub-cloned into overexpression vectors without their native signal peptide. For protein purification, the genes were sub-cloned into a pET27(b) vector (Novagen) containing a N-terminal Strep-tag II sequence (MASWSHPQFEKGAG). For directed evolution, the genes were sub-cloned into a pET29(b) vector (Novagen) containing a N-terminal MBP-tag with a periplasmic expression tag. In detail, PCR products and vectors were digested in 50 and 100 μL reactions, respectively, with *Nco* I (*BamH* I for FIM1) and *Hind* III (FastDigest, Fermentas) restriction enzymes for 1 h at 37 °C. The vectors were further treated with alkaline phosphatase (FastAP, Fermentas) for an additional hour at 37 °C. The digested DNA fragments were purified from 1% agarose gels using a gel extraction kit (Qiagen). Ligations were performed in 20 μL reactions at a vector:insert mass ratio of 1:3 using T4 DNA ligase (Fermentas) with approx. 30 ng vector DNA and incubated at least for 1 hour at RT. Heat shock transformation into chemically competent *Escherichia coli* 10G cells (Lucigen) yielded $>10^5$ colonies for mutant libraries. Colonies of mutant libraries were pooled, purified using plasmid mini prep kit (Qiagen) and retransformed into *E. coli* BL21 (DE3) for activity screening as described below.

Generation of mutagenized library. Random mutant libraries were generated with error-prone PCR using nucleotide analogues: 8-Oxo-2'-deoxyguanosine-5'-Triphosphate (8-oxo-dGTP) and 2'-Deoxy-P-nucleoside-5'-Triphosphate (dPTP); (TriLink). Two independent PCR reactions were prepared, one with 8-oxo-dGTP and one with dPTP. Each 50 μL reaction contained 1 × GoTaq Buffer (Promega), 3 μM $MgCl_2$, 1 ng template DNA, 1 μM of primers (forward (T7 promoter): taatacgactcactataggg; reverse (T7 terminator): gctagttattgctcagcgg), 0.25 mM dNTPs, 1.25 U GoTaq DNA polymerase (Promega) and either 200 μM 8-oxo-dGTP or 200 μM dPTP. Cycling conditions: initial denaturation at 95 °C for 2 minutes followed by 20 cycles of denaturation (30 seconds, 95 °C), annealing (60 seconds, 58 °C) and extension (70 seconds, 72 °C) and a final extension step at 72 °C for 5 minutes. Subsequently, each PCR product was treated with *Dpn* I (Fermentas) for 1 h at 37 °C to digest the template DNA. PCR products were purified using the Cycle Pure PCR purification kit (E.N.Z.A) and further amplified with a 2 x Master mix of Econo TAQ DNA polymerase (Lucigen) using 10 ng template from each initial PCR and the same primers at 1 μM in a 50 μL reaction volume. Cycling conditions: initial denaturation at 95 °C for 2 minutes followed by 30 cycles of denaturation (30 seconds, 95 °C), annealing (20 seconds, 58 °C) and extension (70 seconds, 72 °C) and a final extension step at 72 °C for 2 minutes. The PCR products were purified and cloned as described above. The protocol yielded 1-2 amino acid substitutions per gene.

Generation of DNA shuffling libraries. The staggered extension process (StEP) protocol (28) was used to recombine equally improved mutants. Plasmids of variants were mixed in equimolar amounts to 500 ng of total DNA and used as a template for the StEP reaction. Cycling conditions: Initial denaturation at 95 °C for 5 minutes followed by 100 cycles 95 °C for 30 s followed by 58 °C for 5 s. PCR products were purified using the Cycle Pure PCR purification kit and further amplified with a 2 x Master mix of Econo TAQ DNA polymerase. Libraries were cloned into the pET29(b) vector as described above.

Site-directed mutagenesis. Single-point mutant variants were constructed by site-directed mutagenesis as described in the QuikChange Site-Directed Mutagenesis manual (Agilent) using specific primers. All variants contained only the desired mutation, which was confirmed by DNA sequencing.

Pre-screen on agar plates. Libraries in pET29-pMBP were electroporated into *E. coli* BL21 (DE3) cells and incubated for 1 h at 37 °C prior to plating. For a low concentration antibiotic/substrate pre-screen, the transformants were plated on agar plates (150 mm diameter) containing 4 μg/mL ampicillin, 0.1 mM IPTG, 200 μM $ZnCl_2$ and 40 μg/mL kanamycin, yielding >500 colonies. The minimum inhibitory concentration (MIC) of ampicillin for *E. coli* cells expressing NDM1 and VIM2 is 256 μg/mL, whereas for the *E. coli* cells alone it is <2 μg/mL. Subsequently, surviving colonies were directly picked from plates for rescreening in 96-well plates. For a phosphonate hydrolase activity pre-screen, transformation reactions were plated onto six agar plates (150 mm diameter) containing 40 μg/mL kanamycin, such that each plate contained between 400-2000 colonies. Colonies were replica plated onto nitrocellulose membrane (BioTrace NT Pure Nitrocellulose Transfer Membrane 0.2 μm, PALL Life Sciences), which was then placed onto LB agar plates containing 1 mM IPTG, 200 µM $ZnCl_2$ and 40 µg/mL kanamycin for overnight protein expression at room temperature. After expression, the membrane was placed into an empty petri dish and the cells were lysed by alternating incubations at -20 °C and 37 °C three times for 10 minutes each. To assay activity, 25 mL of 0.5% agarose in 50 mM Tris-HCl buffer pH 7.5 containing 200 µM $ZnCl_2$ and 250 µM *p*-nitrophenyl phenylphosphonate (Sigma) was poured onto the membrane. Colonies with active enzymes developed a yellow color due to the hydrolyzed substrate. The most active colonies (~200 variants) were directly picked from plates for screening in 96-well plates.

Cell lysate activity screen in 96-well plates. To test the fitness and solubility of variants, individual wells of a 96-well plate containing 400 µl of LB media supplemented with 40 µg/ml kanamycin were inoculated with 20 µl of overnight culture and incubated at 30 °C for 3 hours. Protein expression was induced by adding IPTG to a final concentration of 1 mM and further incubation at 30 °C (20 °C and 37 °C for testing temperature effect on expression) for 3 hours. Cells were harvested by centrifugation at 4,000 × *g* for 10 minutes and pellets were frozen -80 °C for at least 30 min. For lysis, cell pellets were resuspended in lysis buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 200 µM $ZnCl_2$, containing 0.1% Triton X-100, 100 µg/ml lysozyme and 1 U/ml of benzonase) and incubated at 25 °C with shaking at 1200 rpm for 1 hour. The cell lysates were clarified by centrifugation at 4,000 × *g* for 20 minutes at 4 °C. Clarified lysates were diluted (1000-fold for β-lactamase activity 2-fold for phosphonate hydrolase activity) in order to obtain linear initial rates and measured against a single substrate concentration (90 µM for β-lactamase activity and 500 µM for phosphonate hydrolase activity).

Solubility assay. The solubility of variants in their respective cell lysates was assayed from a 96-well plate expression trial at 30 °C, as described above. The soluble fraction was assayed by loading a fraction (10 µL + 2 µL of 5 × SDS-PAGE sample buffer) of the clarified lysate (200 µl) on a SDS-PAGE gel. The insoluble fraction was assayed by resuspending the pellet in lysis buffer (200 µL) and loading a fraction (10 µL + 2 µL 5 × of SDS-PAGE sample buffer) on a SDS-PAGE gel. The intensity of the protein band was measured using the Gel Doc^{TM} XR System (BioRad) and analyzed using the Image Lab^{TM} Software (Bio-Rad). The solubility was calculated as Is/ (Is + Ip) × 100, with Is being the intensity of the supernatant band and Ip being the intensity of the pellet band.

Purification of Strep-tagged proteins. All variants were cloned as described above, transformed, overexpressed in *E. coli* BL21 (DE3) cells and purified as described previously (29).

Enzyme kinetics. The kinetic parameters and activity levels of purified enzyme variants were obtained as described previously. Briefly, phosphonatase activity was monitored following the release of *p*-nitrophenol at 405 nm with an extinction coefficient of 18,300 M^{-1} cm^{-1}. The β-lactamase activity was monitored at 405 nm for the CENTA substrate (30), and molar product formation was calculated with the extinction coefficient of 6,300 M^{-1} cm^{-1}. Extinction coefficients were calculated using the ProtParam tool of the ExPASy server (31).

Thermostability measurements. The thermostabilities of the variants were measured with a thermal shift assay as described previously (32). Briefly, enzyme variants (2 μM) were mixed with 5 × SYPRO Orange dye (Invitrogen) in a 20 μl reaction and heated from 25 °C to 95 °C in a 7500 Fast Real-Time PCR system (Applied Biosystems). Measurements were conducted in triplicate and unfolding was followed by measuring the change in fluorescence caused by binding of the dye (excitation, 488 nm; emission, 500–750 nm). The melting temperature (T_m) is calculated from midpoint of the denaturation curve and values were averaged.

Protein purification for crystallization. The NDM1 and VIM2 protein variants were expressed in *E. coli* BL21 (DE3) cells in TB medium (400 ml) supplemented with 1% glycerol, 50 μg/mL kanamycin and 200 μM ZnCl2. Cells were grown at 30 °C for 6 hours. The temperature was lowered to 22 °C and the cells were incubated for a further 16 hours and harvested by centrifugation for 15 minutes at 8,500 × *g* (R9A rotor, Hitachi), then resuspended in buffer A (50 mM HEPES pH 7.5, 500 mM NaCl, 20 mM imidazole, and 200 μM $ZnCl_2$) and lysed by sonication (OMNI sonic ruptor 400). Cellular debris was removed by centrifugation at 29,070 × *g* for 60 minutes (R15A rotor, Hitachi). The supernatant was loaded onto a 5 ml Ni-NTA superflow cartridge (Qiagen) followed by extensive washing with buffer A prior to elution of proteins in buffer B (50 mM HEPES pH 7.5, 500 mM NaCl, 500 mM imidazole, and 200 μM ZnCl2). Protein containing fractions were analyzed by SDS-PAGE (Bolt Mini Gels, Novex). The buffer B containing the proteins was exchanged to TEV reaction buffer (50 mM Tris-HCl pH 8.0, 0.5 mM EDTA, 1 mM DTT, and 150 mM NaCl) using HiPrep 26/10 desalting column (GE healthcare). 20% TEV (33) (w/w) was added and incubated at 4 °C for 4 days. The TEV reaction buffer was exchanged to buffer A before TEV protease and His-tag containing debris were removed by Ni-NTA superflow column (5 mL, Quagen). His-tag cleaved protein was then concentrated using a10 kDa molecular weight cut-off MWCO ultrafiltration membrane (Amicon, Millipore) and loaded on HiLoad 16/600 Superdex 75 pg column (GE Healthcare). Proteins were eluted into crystallization buffers (described below).

Crystallization of NDM1.R10. NDM1.R10 protein in crystallization buffer (20 mM HEPES pH 7.5, 2 mM β-mercaptoethanol, 150 mM NaCl, and 100 μM $ZnCl_2$) was concentrated to 15 mg/mL using a 10 kDa molecular weight cut-off ultrafiltration membrane (Amicon, Millipore) and crystallized by the hanging drop method. The hanging drops were prepared by mixing protein solution (1 μL) and well solution (2 μL). NDM1.R10 crystals appeared after two weeks in 0.1 M MES (pH 6.75) and 1.3 M $MgSO_4$ at 18 °C and continued to grow. Crystals were soaked in cryoprotectant solution for 30 seconds (precipitant, and 25% glycerol), and flash cooled in liquid nitrogen. The MES bound crystal diffracted to 1.67 Å at beam line MX1 at the Australian Synchrotron. The product bound structure was obtained by soaking the crystal in precipitant solution, containing 15 mM substrate for 3 minutes to 30 minutes before soaking in cryoprotectant solution and flash cooling in liquid nitrogen. The crystals diffracted to 1.68-2 Å at a beam line MX2 (0.9537 Å) at the Australian Synchrotron.

Crystallization of VIM2.R10. The first size exclusion peak of VIM2.R10 protein (dimeric fractions), in buffer containing 50 mM HEPES pH 7.5, 150 mM NaCl, and 200 μM $ZnCl_2$, was concentrated to 2.6 mg/mL using a 10 kDa molecular weight cut-off ultrafiltration membrane (Amicon, Millipore) and crystallized by the hanging drop method. The drops were prepared by mixing a protein solution (2 μL), well solution (4 μL), 1 mM TCEP, and 2.5 mM PNPP. Crystals appeared after two weeks in 0.1 M HEPES (pH 7.5) and 1.2 M sodium citrate at 18 °C. Crystals were soaked in cryo-protectant solution for several minutes (precipitant, and 10% glycerol), and flash cooled in liquid nitrogen. The crystal diffracted to 2 Å at beam line MX1 (0.9537 Å) at the Australian Synchrotron.

Data collection and structure determination. The crystallographic data were collected at 100 K at the Australian Synchrotron. Data were processed using XDS (34). Scaling was performed using Aimless in the CCP4 program suite (35–37). Resolution estimation and data truncation were performed by using overall half-dataset correlation CC(1/2) > 0.5 (38, 39). Molecular replacement was used to solve all structures with MOLREP (40) using the structures deposited under PDB accession codes 3SPU and 1KO3 as starting models for NDM1 and VIM2, respectively. The model was refined using phenix.refine (41) and Refmac v5.7 (42) in CCP4 v6.3 program (37), and the model was subsequently optimized by iterative model building with the program COOT v0.7 (43).

Results

The selection of evolutionary starting points

Our previous study revealed that several B1 β-lactamases possess low levels of promiscuous PMH activity (4). Here, we expand this characterization to nine B1 β-lactamases (22-94% sequence identity between nine enzymes) and have confirmed this tendency. All nine enzymes showed similar catalytic efficiencies for their native β-lactamase activity (k_{cat}/K_M = ~10^6 $M^{-1}s^{-1}$, **Figure 1 A**), but the catalytic efficiency for PMH activity varies by two-orders of magnitude (k_{cat}/K_M = 10^{-1} ~ 10^1 $M^{-1}s^{-1}$, **Figure 1 A**). To test the initial evolvability, we selected four enzymes of the NDM1 (NDM1 and EBL1) and VIM (VIM2 and VIM7) clusters as starting points for the directed evolution experiment. These proteins share amino acid sequence identity ranging from 35% to 80% and high structural similarity (**Figure 1** and **Table 2**). The initial fitness (the level of PMH activity in *E. coli* cell lysate) of the enzymes varied considerably, with VIM2 exhibiting the highest fitness - 15-fold, 7-fold and 4-fold higher than EBL1, VIM7 and NDM1, respectively (**Figure 2A**). Clearly, the "fitness" is determined by both catalytic efficiency (k_{cat}/K_M) for the PMH activity and the amount of soluble and functional enzyme in the cell lysate ([E]). For example, EBL1 exhibits the highest k_{cat}/K_M but also the lowest fitness, which can be attributed to the relationship between catalytic efficiency and amount of functional enzyme in the cell ([E]).

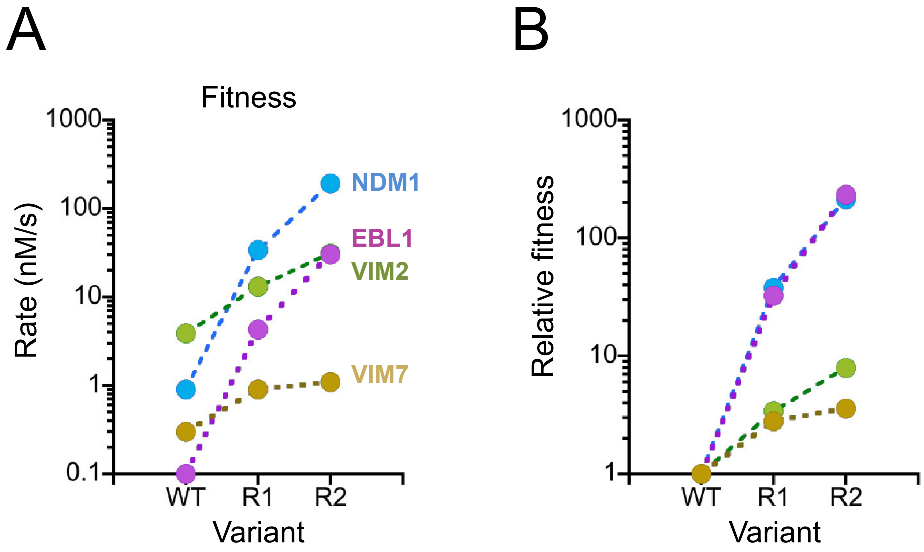


Figure 2. Initial fitness improvements of NDM1, EBL1, VIM2 and VIM7 during directed evolution towards higher PMH activity. (**A**) Fitness improvement (PMH activity in cell lysate) of NDM1 (blue), EBL1 (magenta), VIM2 (green) and VIM7 (beige), in response to the selection criteria used in the directed evolution experiment. WT indicates the wild-type enzyme, and R1 and R2 represent the isolated variant of each round. Values were averaged from three replicates. (**B**) Relative fitness improvement normalized to the initial wild-type levels of each enzyme.

The comparative directed evolution experiment

The four enzymes, NDM1, EBL1, VIM2 and VIM7, were subjected to the same directed evolution procedure **(Figure 3**). NDM1 and VIM2 were subjected to ten rounds of directed evolution, whereas only two rounds were performed for the homologs EBL1 and VIM7 (**Figure 1**). Briefly, the enzymes were fused to a maltose binding protein (MBP)-tag for periplasmic expression in *E. coli*. Randomly mutagenized libraries of the enzymes were generated by error-prone PCR, resulting in 1~2 amino acid substitutions per gene. Due to the low PMH activity of the initial variants, a direct high-throughput screening of PMH activity on agar plates was not feasible. Therefore, we employed a β-lactam (ampicillin) antibiotic selection (purifying selection for the enzymes' native activity) for the first 8 rounds of directed evolution **(Table 3)**. This was done to "purge" the library of enzymes with no activity and provide a smaller library of variants that retained catalytic activity. The ampicillin concentration was ~64-fold below the enzymes' minimum inhibitory concentration (MIC), thus the process purges out non-functional variants but retains variants that are neutral or only mildly compromised in their native activity. Colonies grown on the screening plates were picked into 96-well plates (in total 396 variants for each enzyme per round), regrown, lysed and screened for PMH activity. The most improved variants were isolated, sequenced and used as templates for the next round of evolution. For the last two rounds (R9 and R10) the antibiotic screening was replaced with a direct colorimetric phosphonatase activity screening on agar plates, allowing us to screen >2,000 variants per round. Additionally, DNA shuffling was performed at each round to recombine the beneficial mutations when several improved variants were identified. Overall, ten rounds of directed evolution were performed, with 2-3 additional DNA shuffling library screening for each trajectory (**Table 3**).

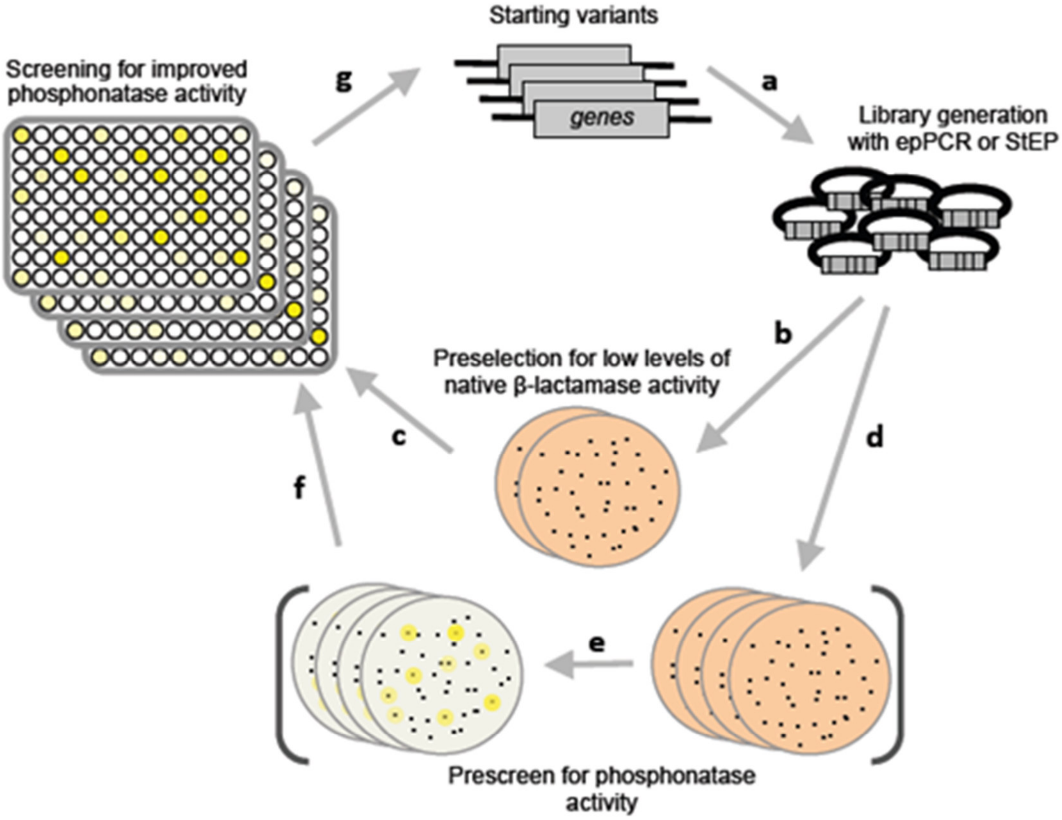
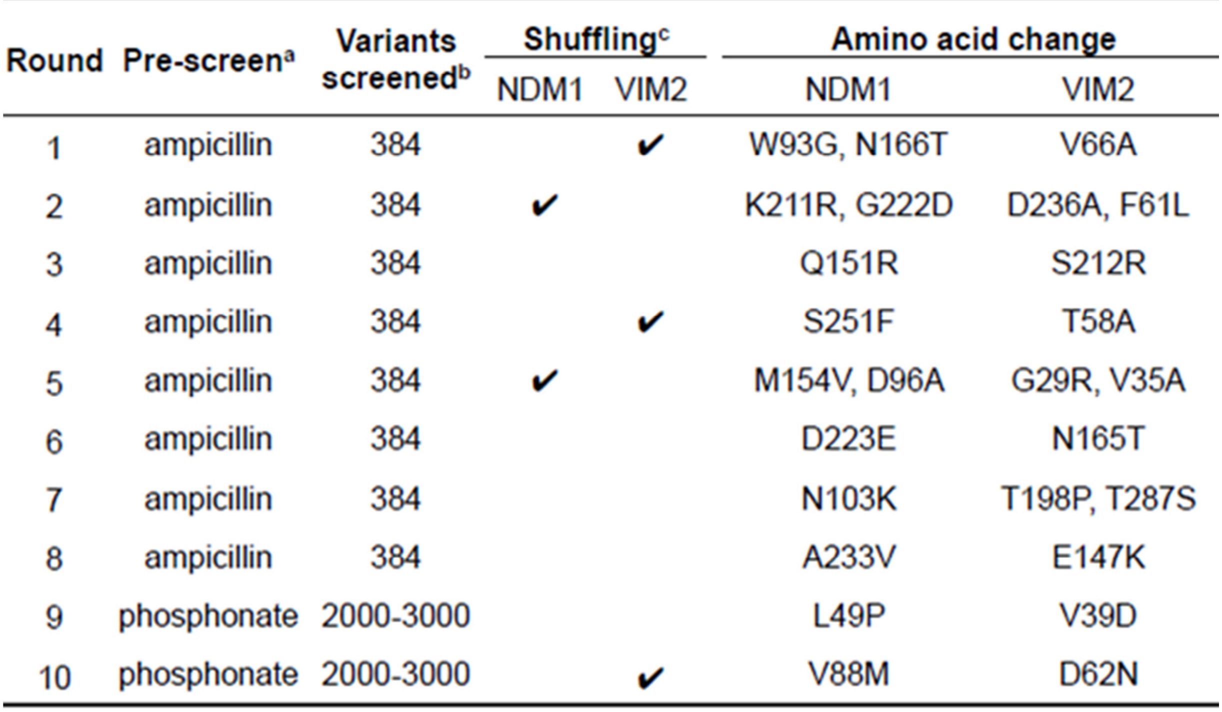


Figure 3. Overview of the laboratory evolution strategy. (**a**) Starting variants were mutagenized using error-prone PCR or StEP (staggered extension process) recombination and ligated into a vector containing an N-terminal MBP-tag and periplasmic expression signal peptide. The resulting library was then transformed into *E. coli* BL21 DE3 cells for selection. (**b**) In the first eight rounds the library was plated on agar plates containing low levels of ampicillin in order to select for functional variants. (**c**) Surviving colonies were picked from plates for rescreening in 96-well plates. (**d**) In the last two rounds a direct screen for phosphonatase activity was used and colonies were replicated onto a nitrocellulose membrane. (**e**) After expression and cell lysis the phosphonatase activity of the variants was assayed and active variants could be identified through the development of a yellow colour. (**f**) The most active colonies (~200 variants) were directly picked from plates for screening in 96-well plates. (**g**) The most improved variant(s) served as the starting point for the next round of directed evolution.

Table 3. Library information and amino acid changes.



Round	Pre-screen[a]	Variants screened[b]	Shuffling[c]		Amino acid change	
			NDM1	VIM2	NDM1	VIM2
1	ampicillin	384		✔	W93G, N166T	V66A
2	ampicillin	384	✔		K211R, G222D	D236A, F61L
3	ampicillin	384			Q151R	S212R
4	ampicillin	384		✔	S251F	T58A
5	ampicillin	384	✔		M154V, D96A	G29R, V35A
6	ampicillin	384			D223E	N165T
7	ampicillin	384			N103K	T198P, T287S
8	ampicillin	384			A233V	E147K
9	phosphonate	2000-3000			L49P	V39D
10	phosphonate	2000-3000		✔	V88M	D62N

[a] Selection with ampicillin (8 µg/ml) in order to enrich the library for functional variants. Screening with the phosphonate substrate (4-Nitrophenyl phenylphosphonate 250 µM) selects directly for improved phosphonatase activity. Details are described in Material & Methods.
[b] The number of variants screened for improved phosphonatase activity. In the case of ampicillin screen 384 colonies (functional variants) were picked and screened in 96-well plates. In R9 and R10, colonies were directly screened on agar plates for improved phosphonatase activity (2000-3000 variants). Colonies with high activity (~200) were picked and rescreened in 96-well plates. Details are described in Material & Methods.
[c] DNA shuffling was performed in rounds were multiple improved variants were identified with unique mutations.

EBL1 and VIM7 differ in their evolvability despite similar genotypic solutions

The question arises as to whether other closely related sequences would differ in their evolvability and find alternative genotypic solutions to the same problem. To address this question, we performed two rounds of evolution towards PMH activity using EBL1 and VIM7 as starting points. EBL1 is closely related to NDM1 (56% sequence identity), whereas VIM7 is closely related to VIM2 (80% sequence identity). Interestingly, the improvements in fitness of EBL1 and VIM7 are very similar to that of NDM1 and VIM2, respectively (**Figure 1**). NDM1 and EBL1 improve fitness by 200-fold, whereas VIM2 and VIM7 improve by only <10-fold over the two rounds of evolution. The improvements in k_{cat}/K_M correlate similarly, thus the discrepancy between fitness and k_{cat}/K_M is most likely due to solubility. Interestingly, the initial genotypic solutions for NDM1, EBL1 and VIM7 are almost identical; all three enzymes acquire a mutation at position W93 to either G or L. However, the effect on fitness of the W93 mutation varies (**Figure 1**).

The NDM1 and VIM exhibit substantial differences in evolvability

The fitness of both NDM1 and VIM2 improved over the ten rounds of directed evolution (**Figure 2 A**). Both trajectories exhibit a typical pattern for evolutionary optimization – diminishing returns (44). This involves a large improvement in the initial phase of the evolution was followed by gradual increments in the later ones, and finally plateauing of the improvement. Note that, even though a direct PMH screen was employed in the last two rounds (with no selection for β-lactamase activity) no variant with significantly improved PMH activity was isolated. Thus, the fitness plateaus in both trajectories are not due to the initial β-lactamase selection to maintain the native activity, but rather due to exhaustion of available mutations that could improve PMH activity. Despite the same trends, the two trajectories exhibited substantial differences in the rate of improvement, and in the height of their fitness peaks. Initially, VIM2 showed 4-fold higher fitness compared to NDM1. Yet, NDM1 surpassed VIM2 in the first round, and its fitness increased 3,600-fold over ten rounds of directed evolution. On the other hand VIM2 improved overall only 50-fold. Thus, NDM1 fitness improved >70-fold more than VIM2, and the evolved NDM1-R10 exhibits 16-fold higher fitness compared to the evolved VIM2-R10.

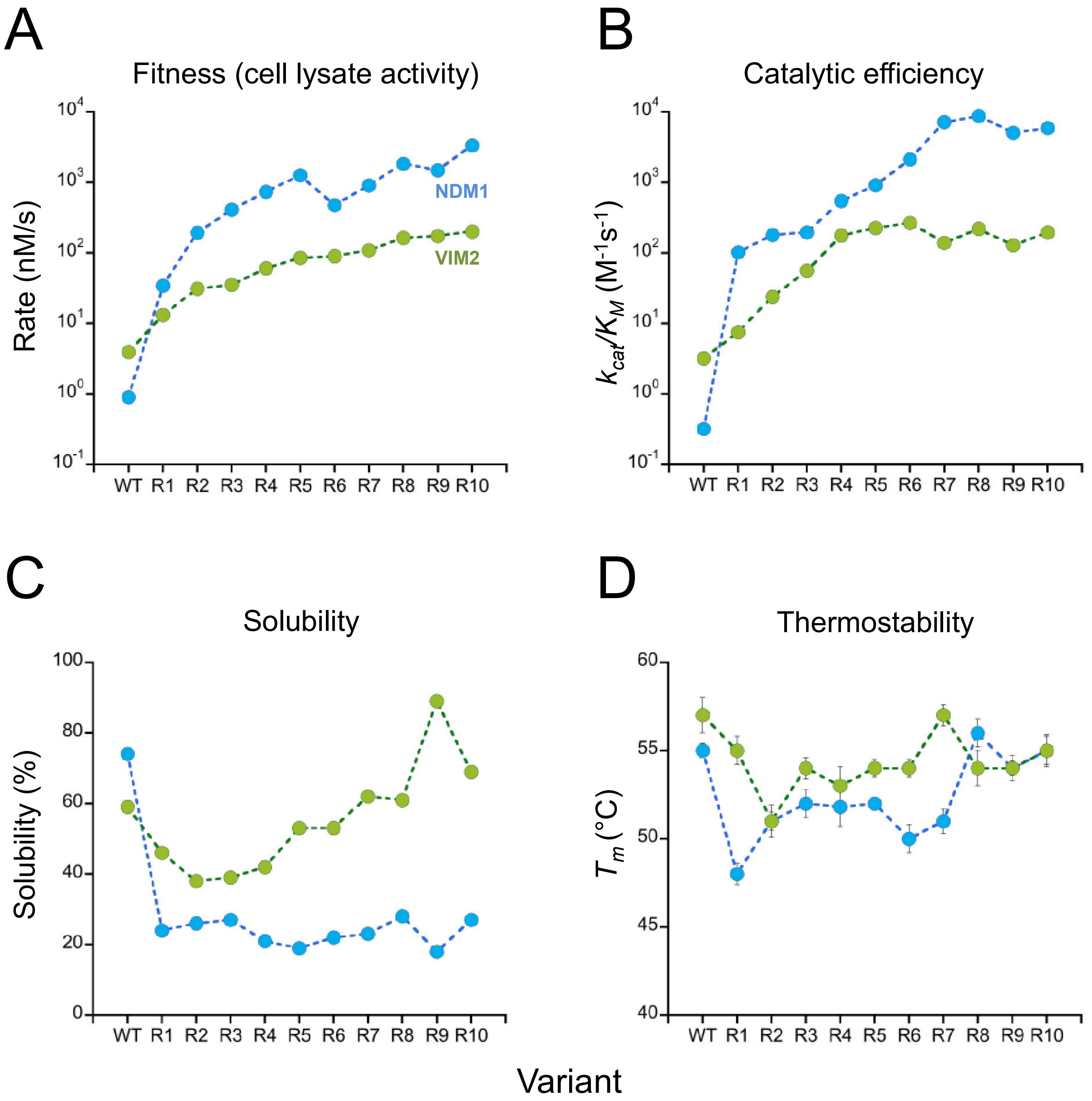


Figure 4. Phenotypic adaptation of NDM1 and VIM2 during the directed evolution towards higher phosphonatase activity. (**A**) Fitness improvement of NDM1 (blue) and VIM2 (green) (phosphonatase activity in cell lysate), which represents the selection criteria in the directed evolution experiment. WT indicates the wild-type enzyme, and R1 to R10 represents the isolated variant of each round. Values represent the average of three independent experiments. (**B**) Catalytic efficiencies (k_{cat}/K_M) of purified variants for the phosphonatase activity (individual parameters are listed in **Tables 4 and 5** (**C**). The soluble expression of variants as determined by SDS–PAGE analysis (**Figure 5**). (**D**) Thermostability of purified variants, which are calculated from the midpoint of the thermal denaturation curve in a thermal shift assay (**Table 6**). Error bars represent standard deviation from three independent assays.

Table 4. Kinetic parameters of NDM1 variants

	phosphonatase			β-lactamase		
Variant	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M ($s^{-1}M^{-1}$)	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M ($s^{-1}M^{-1}$)
WT	0.004 ± 0.0006	~10,000 ± 2200	3.2×10^{-1}	39 ± 1	17 ± 2	2.3×10^{6}
R1	0.5 ± 0.02	5100 ± 340	1.0×10^{2}	63 ± 1	30 ± 2	2.1×10^{6}
R2	1.2 ± 0.1	6500 ± 700	1.8×10^{2}	140 ± 5	60 ± 7	2.3×10^{6}
R3	0.9 ± 0.02	4500 ± 250	2.0×10^{2}	87 ± 2	70 ± 6	1.3×10^{6}
R4	1.8 ± 0.02	3300 ± 100	5.5×10^{2}	57 ± 1	32 ± 3	1.8×10^{6}
R5	2.1 ± 0.002	2400 ± 100	9.1×10^{2}	54 ± 2	20 ± 4	2.8×10^{6}
R6	18 ± 0.6	8500 ± 400	2.1×10^{3}	86 ± 2	42 ± 3	2.0×10^{6}
R7	11 ± 0.1	1600 ± 170	7.2×10^{3}	120 ± 3	65 ± 6	1.9×10^{6}
R8	15 ± 0.1	1700 ± 160	8.7×10^{3}	96 ± 2	61 ± 4	2.4×10^{6}
R9	7 ± 0.04	1400 ± 300	5.1×10^{3}	115 ± 3	63 ± 5	1.8×10^{6}
R10	8 ± 0.4	1957 ± 130	5.9×10^{3}	103 ± 6	48 ± 6	2.1×10^{6}

Measurements were performed in triplicate and averaged. ± indicates error of the fitting to the Michaelis-Menten equation.

Table 5. Kinetic parameters of VIM2 variants.

	phosphonatase			β-lactamase		
Variant	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M ($s^{-1}M^{-1}$)	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M ($s^{-1}M^{-1}$)
WT	0.003 ± 0.0003	970 ± 30	3.2×10^{0}	46 ± 2	14 ± 2	3.3×10^{6}
R1	0.016 ± 0.002	2200± 50	7.5×10^{0}	40 ± 2	28 ± 6	1.4×10^{6}
R2	0.025 ± 0.0003	1100 ± 30	2.3×10^{1}	26 ± 2	20 ± 5	1.3×10^{6}
R3	0.06 ± 0.0004	1020 ± 16	5.6×10^{1}	63 ± 3	33 ± 7	1.9×10^{6}
R4	0.23 ± 0.004	1300 ± 115	1.8×10^{2}	110 ± 9	91 ± 21	1.2×10^{6}
R5	0.25 ± 0.002	1100 ± 23	2.2×10^{2}	195 ± 10	130 ± 23	1.5×10^{6}
R6	0.36 ± 0.02	1400 ± 200	2.7×10^{2}	39 ± 2	26 ± 7	1.5×10^{6}
R7	0.22 ± 0.01	1600 ± 80	1.4×10^{2}	70 ± 3	33 ± 4	2.1×10^{6}
R8	0.34 ± 0.02	1500 ± 200	2.2×10^{2}	19 ± 1	8 ± 1	2.4×10^{6}
R9	0.2 ± 0.002	1600 ± 300	1.3×10^{2}	24 ± 1	16 ± 3	1.5×10^{6}
R10	0.4 ± 0.01	1957 ± 130	2.0×10^{2}	34 ± 2	15 ± 4	2.3×10^{6}

Measurements were performed in triplicate and averaged. ± indicates error of the fitting to the Michaelis-Menten equation.

Different phenotypic solutions underlie the increase in fitness of each enzyme

In order to examine which phenotypic changes, e.g. catalytic efficiencies (k_{cat}/K_M) and/or amount of functional enzyme in the cell ([E]), change the fitness of the two trajectories, the phenotypic parameters of all intermediate variants, k_{cat}/K_M, protein solubility and thermostability (T_m) were determined (**Figure 4)**. Protein solubility of a variant was determined as a fraction of the MBP-tagged fusion protein expressed in the supernatant fraction of the cell lysate (**Figure 5**). For catalytic efficiency and T_m measurements, all variants were sub-cloned, expressed and purified as Strep-tag fusions.

The difference between the two trajectories appears more striking when comparing the changes in the k_{cat}/K_M for PMH activity. Initially, VIM2 exhibits 10-fold higher k_{cat}/K_M compared to NDM1 (**Figure 4 B**). However, NDM1 improved by 20,000-fold during the evolution (from 0.3 to 5900 $M^{-1}s^{-1}$), and the catalytic efficiency of VIM2 increased only by 60-fold (from 3.2 to 200 $M^{-1}s^{-1}$). Overall, the two enzymes differ by >300-fold in improvement of the k_{cat}/K_M, and NDM1 evolved to a 30-fold higher final catalytic efficiency than VIM2. The native β-lactamase activity was only marginally impaired in both trajectories (**Figure 4**).

The discrepancy between the improvement in the fitness and k_{cat}/K_M of the evolved NDM1-R10 (3,600-fold *vs.* 20,000-fold, respectively) is due to tradeoffs between function and protein solubility; the drastic increase in the catalytic efficiency is accompanied by a loss in the amount of functional, soluble, protein in the cell. The first round mutations (W93G and N166T) of the NDM1 trajectory increased fitness by 35-fold. This increase consists of a 300-fold increase in k_{cat}/K_M and reduction in protein solubility from 74% to 24% (**Figure 4 C**). The solubility of the NDM1 variants then remained low (~20%) throughout the trajectory while the PMH activity gradually increased until round 8.

Tradeoffs between function and solubility are observed in the VIM2 trajectory, but to a lesser extent than in NDM1. The catalytic activity increased by 7-fold in the first 2 rounds, whereas the solubility decreased slightly, from 59% to 38%. However, in contrast to NDM1, the increases in fitness in the latter rounds of the VIM2 evolution were largely due to increases in protein solubility (**Figure 4**). The catalytic efficiency for PMH activity stagnated after round 4, but the solubility increased to >80% in R9 and >60% in R10.

Thermostability seems to only partially correlate with the solubility of the variants

(**Figure 4 C** and **D**). The decrease in protein solubility in Round 1 for both NDM1 and VIM2 is correlated with the decrease in thermostability. For NDM1 the thermostability decreased from 55 to 48 °C and the solubility decreased from 74% to 28%. For VIM2, the thermostability decreased from 56 to 51 °C in the first two rounds accompanied with the reduction of protein solubility from 60% to 40%. However, from R3-10 there was little correlation between thermostability and solubility, e.g. for VIM2, thermostability changed little, whereas solubility increased >20%. Hence, it is likely that protein folding, rather than the thermostability of the native folded state, mainly influences the soluble expression of the proteins.

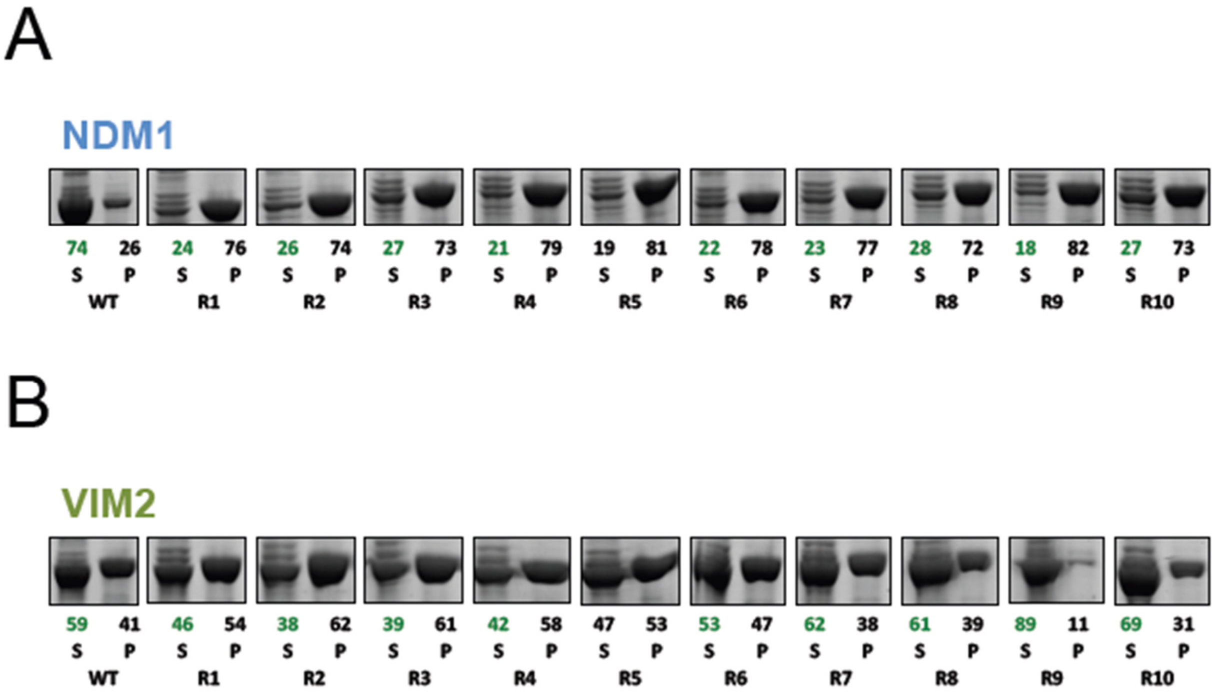


Figure 5. SDS-PAGE analysis of solubility of NDM1 and VIM2 variants. The soluble and insoluble pellet fractions of cell lysates (S and P, respectively) were analysed by SDS–PAGE. The percentage of protein in the soluble and insoluble fraction was determined by measurement of the relative intensities of the supernatant and pellet bands of the protein variant.

Table 6. Activity levels and thermostability of purified variants.

Enzyme	Variant	Activitiy (nM/s)		T_m (°C)
		phosphonatase	β-lactamase	
FIM1	WT	**0.4** ± 0.01	**5.1** ± 0.2	**53.5** ± 0.7
	H64V	**0.3** ± 0.02	**18.3** ± 0.6	**54.0** ± 0.9
	H64A	*n.d.*	*n.d.*	*n.d.*
	W85G	**2.6** ± 0.1	**4.6** ± 1.0	**46.2** ± 1.3
EBL1	WT	**11.6** ± 0.4	**12.3** ± 0.4	**66.2** ± 0.5
	W66G	**57.5** ± 1.2	**28.6** ± 0.2	**64.0** ± 0.8
	P46V	**4.7** ± 0.1	**20.2** ± 0.7	**64.8** ± 0.6
	P46A	**5.7** ± 0.1	**16.3** ± 0.6	**65.9** ± 0.6
NDM1	WT	**0.4** ± 0.02	**61.2** ± 1.0	**55.0** ± 0.4
	W93G	**38.1** ± 1.1	**126.2** ± 2.5	**47.4** ± 0.4
	A72V	**0.3** ± 0.04	**81.8** ± 2.6	**54.8** ± 0.5
VIM2	WT	**8.1** ± 0.1	**51.4** ± 0.3	**56.3** ± 0.7
	V66A	**13.9** ± 1.0	**54.4** ± 0.6	**54.9** ± 0.3
	W87G	**0.6** ± 0.03	**6.9** ± 0.1	**56.2** ± 0.6
VIM1	WT	**3.0** ± 0.1	**123.3** ± 1.0	**59.0** ± 0.3
	V66A	**8.9** ± 0.4	**130.8** ± 2.3	**55.1** ± 0.5
	W87G	**1.0** ± 0.02	**13.6** ± 1.2	**55.4** ± 1.8
VIM7	WT	**3.7** ± 0.1	**50.1** ± 1.1	**56.8** ± 1.0
	V66A	**4.1** ± 0.1	**40.2** ± 2.2	**54.1** ± 0.5
	W87G	**12.0** ± 0.4	**9.9** ± 0.3	**53.2** ± 0.8

± indicates s.d. from three measurements.
n.d. means not determined because the variant could not purified in sufficient quantities.
Activity levels of purified enzymes were measured at one enzyme (1 μM for phosphonatase and 1 nM for β-lactamase) and substrate (500 μM for phosphonatase and 100 μM for β-lactamase) concentration.

Different genotypic solutions underlie the distinct phenotypic adaptations

In order to examine the genetic changes causing the distinct phenotypic adaptation of the two enzymes, all intermediate variants were sequenced. Each enzyme accumulated in total 13-15 mutations over the ten rounds of directed evolution (**Figure 6 A** and **Table 2**). The positions of the mutations in the two trajectories differ significantly, but are largely located around the enzymes' active sites (**Figure 6 B and C**). Almost half (6 out of 13) of the mutations of

VIM2 are located in or next to loop 3, which lies above the enzymes' active site (**Figure 6 C**). NDM1 did not incorporate a single mutation in Loop 3, but acquired mutations spread around the active site (**Figure 6 B**). The first round mutations, and the ones with the highest fitness effects, are V72A (located in loop 3) for VIM2 and W93G (located below loop 3) and N166T (remote periphery) for NDM1 (see **Figure 7** for amino acid numbering based on sequence alignment). Out of the 26 total mutations to both enzymes, only 3 out of 24 positions (G35, W93, D223) that were mutated in each enzyme are conserved between NDM1 and VIM2. Thus, the differences in fitness and phenotypic adaptation of the two trajectories towards PMH are the result of distinct genotypic solutions to the problem of PMH selective pressure.

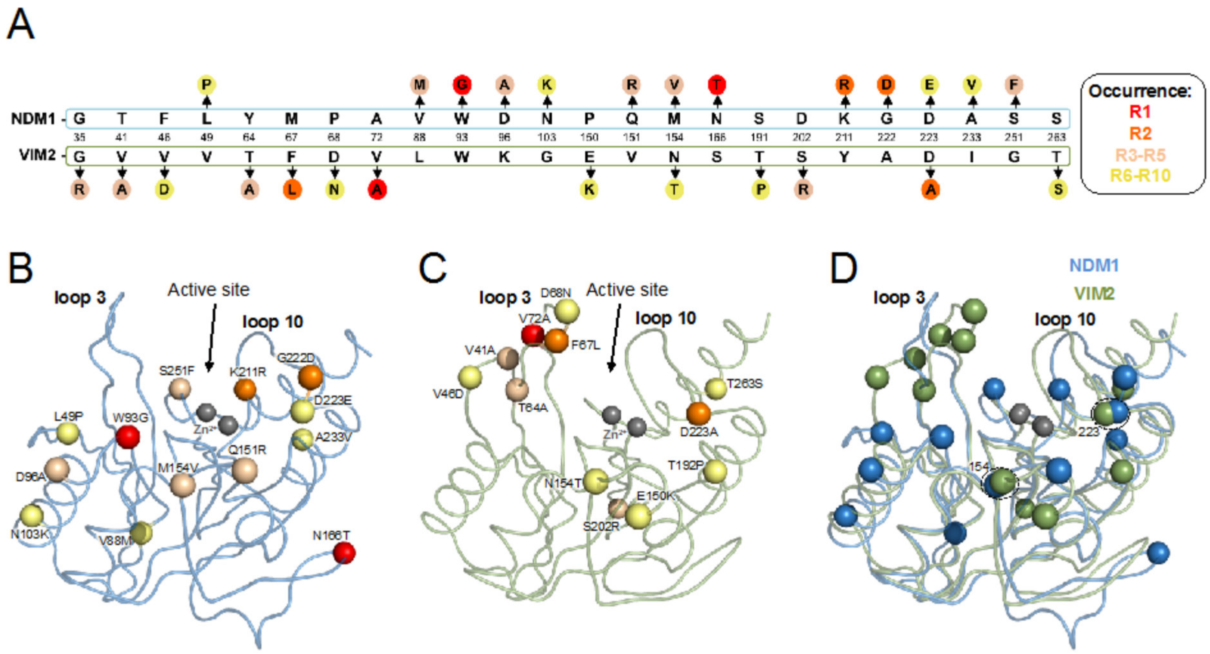


Figure 6. The mutations accumulated in the evolutionary trajectories of NDM1 and VIM2. (**A**) The mutated residues of both trajectories are aligned (a full sequence alignment is shown in **Figure 7**) and arrows indicate mutations. Colored circles describe the occurrence with red for R1 (first round), orange for R2, beige for R3-R5 and yellow for R6-R10. The structural location of the mutations is mapped on the wild-type structures of (**B**) NDM1 (blue; PDB ID: 3SPU) and (**C**) VIM2 (green; PDB ID: 1KO3) with the C-α of mutated residues shown as spheres. Active site metal ions are shown as grey spheres. (**D**) A close up view of the aligned active sites of NDM1 (blue) and VIM2 (green) with mutated residues shown as spheres with different color and active site metals as grey spheres.

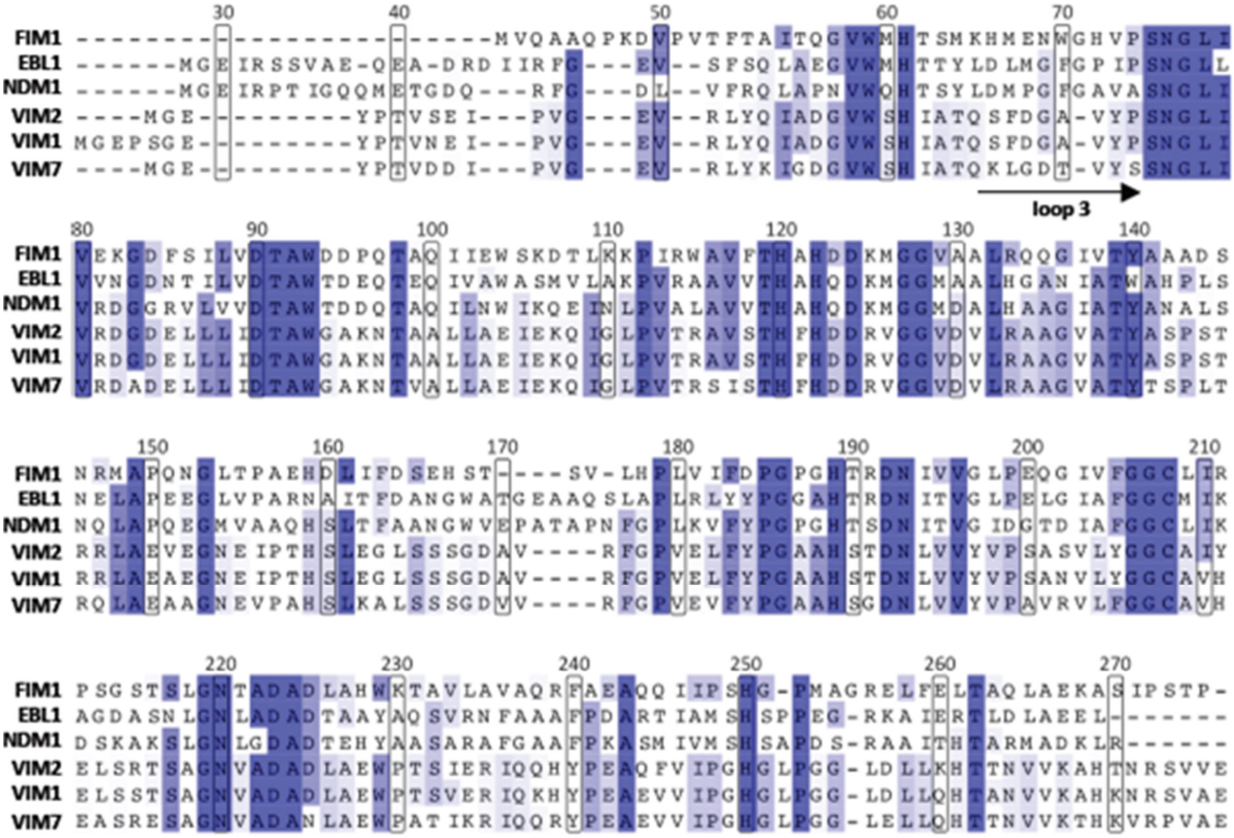


Figure 7. Sequence alignment of selected B1 β-lactamases. Conserved residues are colored in shades of blue. Numbering is based on the NDM1 (PDB ID: 3SPU).

The structural basis for increased expression in VIM2-R10

To obtain molecular insight into how the activities of NDM1 and VIM2 were improved and why the activities were not improved equally through directed evolution, despite the same initial structure and function, we solved the crystal structures of most active variants NDM1-R10 and VIM2-R10 (wild-type structures of NDM1 and VIM2 are published (45, 46), PDB IDs 3SPU and 1KO3, respectively. Interestingly, although size exclusion peaks for WT VIM2, WT NDM1 and NDM1-R10 all showed a single monomeric peak, the size exclusion chromatogram of VIM2-R10 showed an addition peak, corresponding to dimeric VIM2-R10 by molecular weight estimates (**Figure 8 B**). Thus, both peak-1 and peak-2 of VIM2-R10 were screened to obtain crystals. Although peak-2 (monomer) was never crystallized, peak-1 (dimer) produced crystals in conditions of 0.1 M HEPES (pH 7.5) and 1.2 M sodium citrate at 18 °C. Crystal data collection and refinement statistics are presented in **Table 7** for NDM1-R10 and VIM2-R10 crystals.

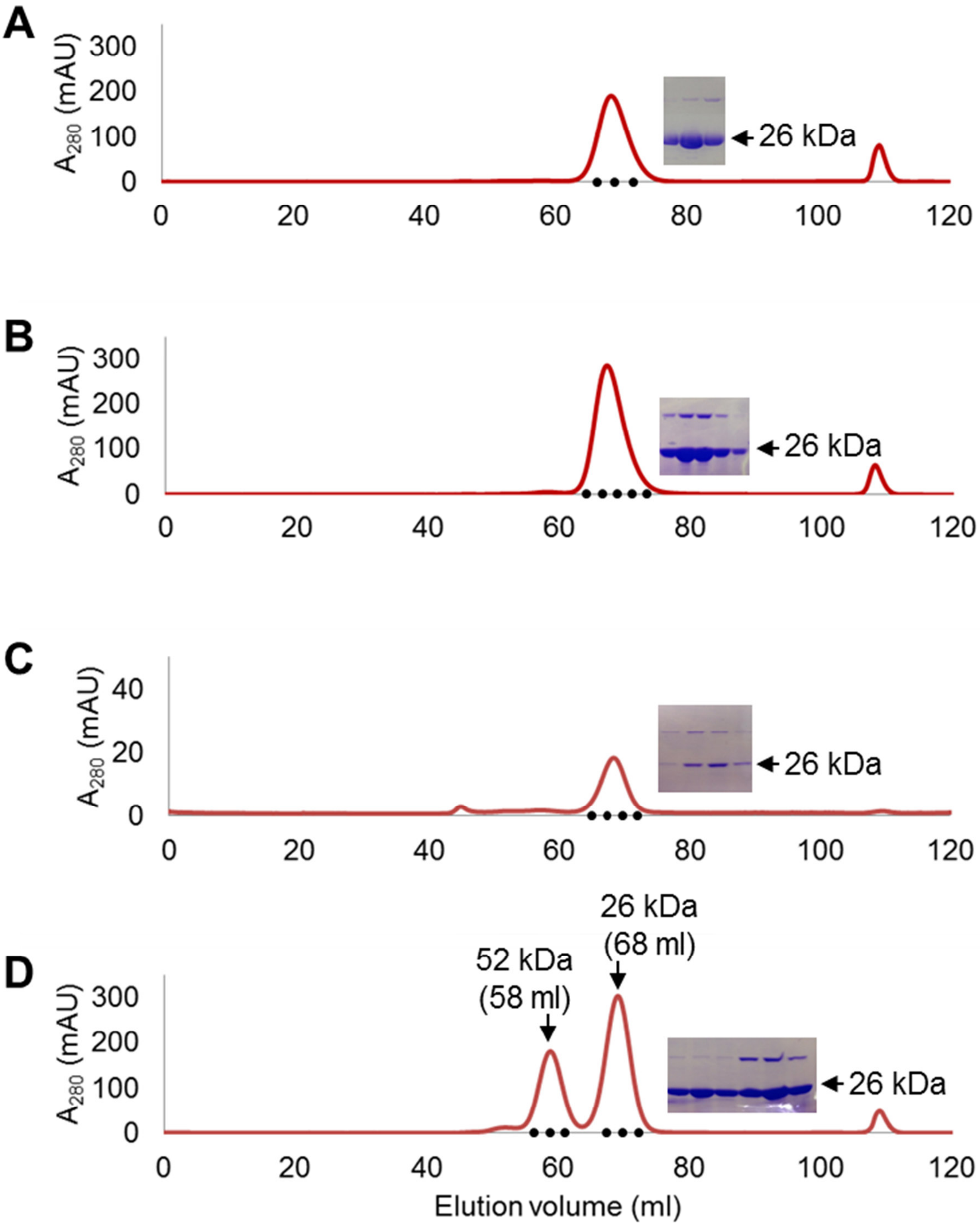


Figure 8. Size exclusion chromatography of NDM1, NDM1-R10, VIM2 and VIM2-R10. Size exclusion chromatograms of NDM1 (**A**), NDM1-R10 (**B**), VIM2 (**C**), VIM2-R10 (**D**) are shown (Superdex 75, GE Healthcare). SDS-PAGE (Novex) results of each protein corresponding to the peak fractions (black dots under each peak) are shown near the peaks. Protein sizes were identified based on the calibration curve on the manufacturer's instruction (GE Healthcare) for Gel filtration calibration kits LMW (low molecular weight) (https://www.gelifesciences.com/gehcls_images/GELS/Related%20Content/Files/1326706518989/litdoc28951560_20120420131146.pdf).

Table 7. Crystallographic data collection and refinement statistics.

	NDM1-R10	NDM1-R10	NDM1-R10	VIM2-R10
PDB ID	**5JQJ**	**5K4N**	**5K4M**	**To be submitted**
Wavelength (Å)	0.9537	0.9537	0.9537	0.9537
Resolution range (Å)	34.45-1.67 (1.73-1.67)	39.02-1.68 (1.74-1.68)	39.13-1.98 (2.05-1.98)	37.89-2.19 (2.27-2.19)
Space group	C 2 2 2_1	C 2 2 2_1	C 2 2 2_1	C 1 2 1
Unit cell (Å, °)	37.80 137.72 77.46 90.00 90.00 90.00	37.94 138.09 77.97 90.00 90.00 90.00	37.82 138.16 78.20 90.00 90.00 90.00	128.60 41.67 156.76 90.00 99.41 90.00
Total reflections	46608 (3668)	15313 (810)	29257 (2838)	150766 (13607)
Unique reflections	23318 (1837)	23845 (412)	14613 (1421)	41963 (4024)
Multiplicity	2.0 (2.0)	1.9 (2.0)	2.0 (2.0)	3.6 (3.4)
Completeness (%)	97.22 (77.72)	99.88 (98.89)	99.72 (99.37)	97.82 (96.08)
Mean I/sigma(I)	17.93 (1.89)	15.99 (12.19)	11.13 (5.78)	11.28 (1.86)
Wilson B-factor (Å^2)	19.27	14.20	14.37	33.13
[a]R-merge	0.029 (0.46)	0.053 (0.07)	0.029 (0.08)	0.094 (0.76)
[b]R-meas	0.04	0.075	0.04	0.11
[c]CC1/2	0.999 (0.588)	0.994 (0.991)	0.999 (0.978)	0.996 (0.613)
[d]CC*	1.00 (0.861)	0.999 (0.998)	1.00 (0.994)	0.999 (0.872)
R-work	0.144 (0.240)	0.163 (0.265)	0.142 (0.150)	0.229 (0.341)
R-free	0.187 (0.287)	0.187 (0.323)	0.208 (0.198)	0.288 (0.363)
Number of non-hydrogen atoms	2052	1944	1982	7217
macromolecules	1758	1747	1749	6964
ligands	28	45	46	16
water	266	152	187	237
Protein residues	230	228	230	922
RMS(bonds) (Å)	0.020	0.021	0.017	0.005
RMS(angles) (°)	2.08	2.16	1.95	0.99
Ramachandran preferred (%)	98.05	97.12	96.63	90.5
Ramachandran allowed (%)	0.98	1.92	2.40	5.6
Ramachandran outliers (%)	0.98	0.96	0.96	3.9
Clashscore	3.95	4.26	5.67	6.82
Average B-factor (Å^2)	24.0	16.2	16.8	41.6
macromolecules	22.1	15.2	15.4	41.7
ligands	36.8	28.8	29.7	38.4
solvent	35.0	23.7	26.6	40.2

* Highest-resolution shell is shown in parentheses.

[a] R-merge = $\sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$ (47).

[b] R-meas = $\sum_{hkl}\{N(hkl)/[N(hkl) - 1]\}^{1/2} \times \sum_i |I_i(hkl) - \langle I(hkl)\rangle| \sum_i I_i(hkl)$ (48).

[c] CC1/2 = $\sum(x - \langle x\rangle)(y - \langle y\rangle)/[\sum(x - \langle x\rangle)^2 \sum(y - \langle y\rangle)^2]^{1/2}$ (38)(39).

[d] CC* = $\sqrt{\frac{2CC_{1/2}}{1 + CC_{1/2}}}$ (38)(39).

NDM1-R10 crystal soaking experiments

The crystal structure of NDM1-R10 included a 2-(*N*-morpholino)-ethanesulfonic acid molecule (MES, B-facor: ~50 Å^2) from the crystallization buffer (0.1 M) in the active site (**Figure 9**). The MES molecule was replaced by the product of PMH hydrolysis, (phenyl phosphonate) after soaking for three minutes into a solution containing 15 mM PMH, followed by soaking into cryo-protectant solution (precipitant, and 25% glycerol). Non-hydrolysed PMH (**1**) was additionally found near Y64, on the surface of the protein, which rotated to form a hydrogen bond (2.6 Å) with the phosphoryl oxygen of PMH. Extended soaking into PMH (for 30 minutes) produced a crystal with very similar overall structure and product positioning to that observed in the crystal soaked for 3 minutes. The only differences were that the non-hydrolysed PMH on the surface was replaced by product, and the PMH leaving group (4-nitrophenol**)** was observed near A96, which was mutated from aspartic acid in NDM1. The structural changes, specifically the positions and identity of the various ligands, due to crystal soaking, are shown in **Figure 9**. Although VIM2-R10 crystals were also soaked with substrate containing solution, no substrate or product was identified in the structures.

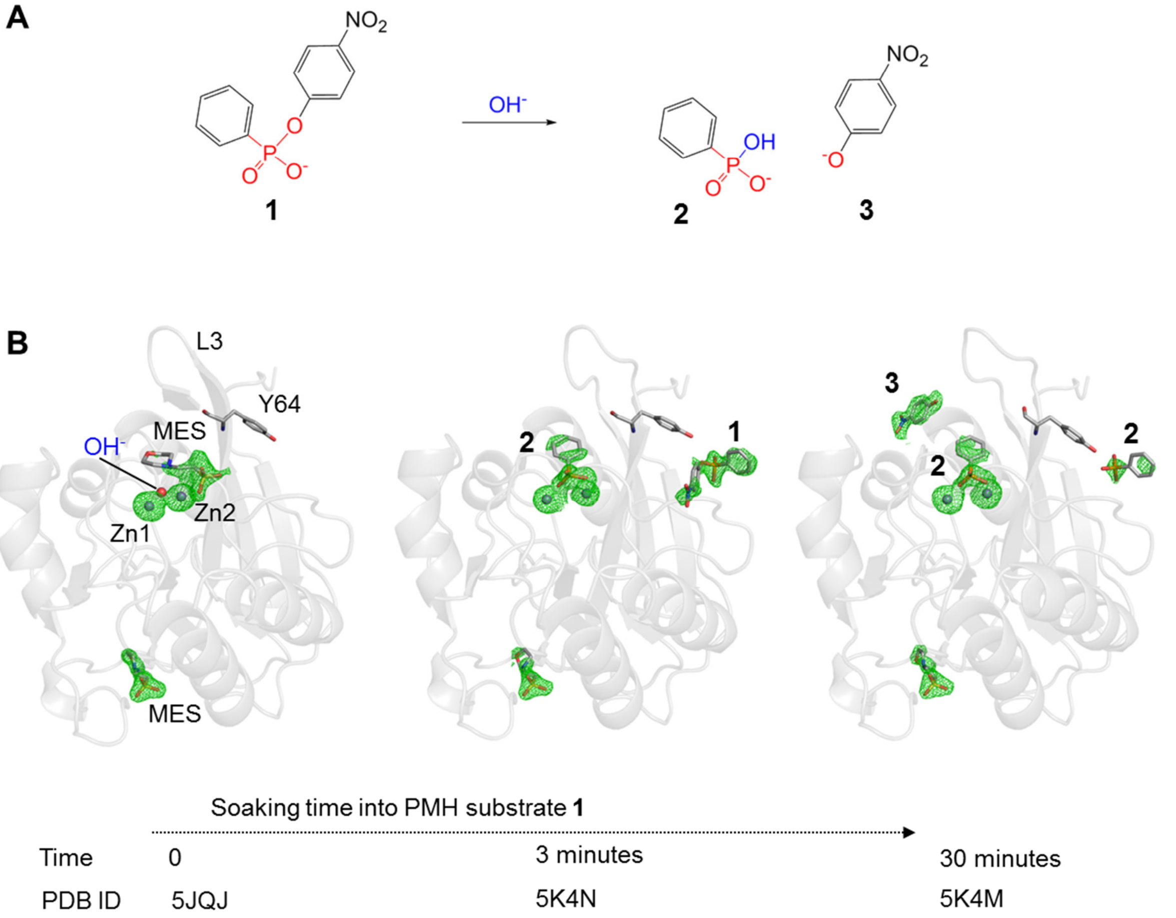


Figure 9. The hydrolysis of PMH and crystal strcutures of NDM1-R10 after crystal soaking. (**A**) The PMH substrate (1, CAS: 57072-35-2) is hydrolyzed by a hydroxide anion into phenylphosphonic acid (2, CAS: 1571-33-1) and 4-nitrophenol (3, CAS: 100-02-7). (**B**) The crystal structure without soaking included one MES on the active site (metal binding site) and another MES on the surface (PDB ID: 5JQJ). The crystals were soaked into PMH containing solution for 3 minutes (PDB ID: 5K4N) and 30 minutes (PDB ID: 5K4M) contained hydrolyzed and non-hydrolyzed PMH. The $Fo - Fc$ electron density for the ligands (MES, PMH substrate, product, and leaving group, Zn1, Zn2) is contoured at 3σ and shown as a green mesh.

Active site loop remodeling drives the evolution of the NDM1

To investigate how the structure of NDM1-R10 was optimized to increase the weak phosphonatase activity (k_{cat}/K_M) of NDM1 ~20,000-fold through 10 rounds of directed evolution (**Figure 4**), we analyzed the structural changes. A total of 13 mutations accumulated all over the protein, including five mutations in the active site (W93G, D223E, K211R, G222D, S251F), four mutations in the second and third shells from the active site (A233V, L49P, M154V, V88M), and four mutations on the surface (Q151R, D96A, N103K, N166T) (**Figure 10**).

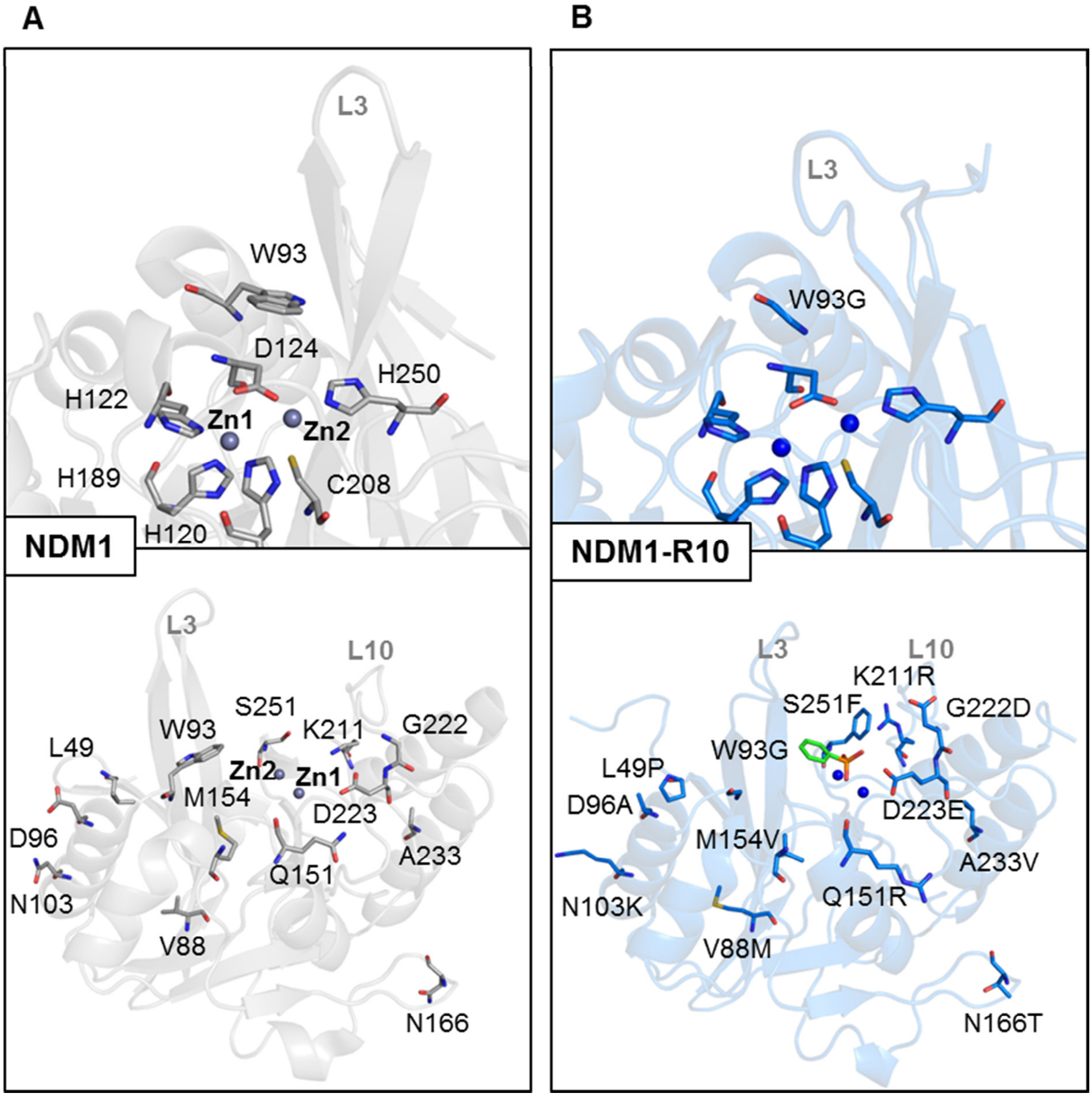


Figure 10. Crystal structures of NDM1 and NDM1-R10. Metal coordinating residues and active site tryptophan residue W93 are shown with sticks in addition to the cartoon style representation of NDM1 (**A**), NDM1-R10 (**B**). NDM1 is shown in grey and NDM1-R10 is in blue. Mutated residues are shown as sticks in the lower panel. Zinc atoms are shown as grey spheres for NDM1 and blue spheres for NDM1-R10. Product presented in the NDM1-R10 crystal structure are shown in green stick (**B**, lower panel), excluded in top figure of **B** for clarity. PDB ID: 3SPU (**A**), 5K4N (**B**).

The overall structure and metal ion positions of NDM1-R10 were highly conserved with NDM1-WT, despite the large number of mutations (RMS < 1Å, **Figure 10-11**). The Zn-Zn distance was 3.6 Å for the MES bound structure, which was slightly shorter than the Zn-Zn distance in the published apo-NDM1-WT structure (3.8 Å, PDB ID: 3SPU).

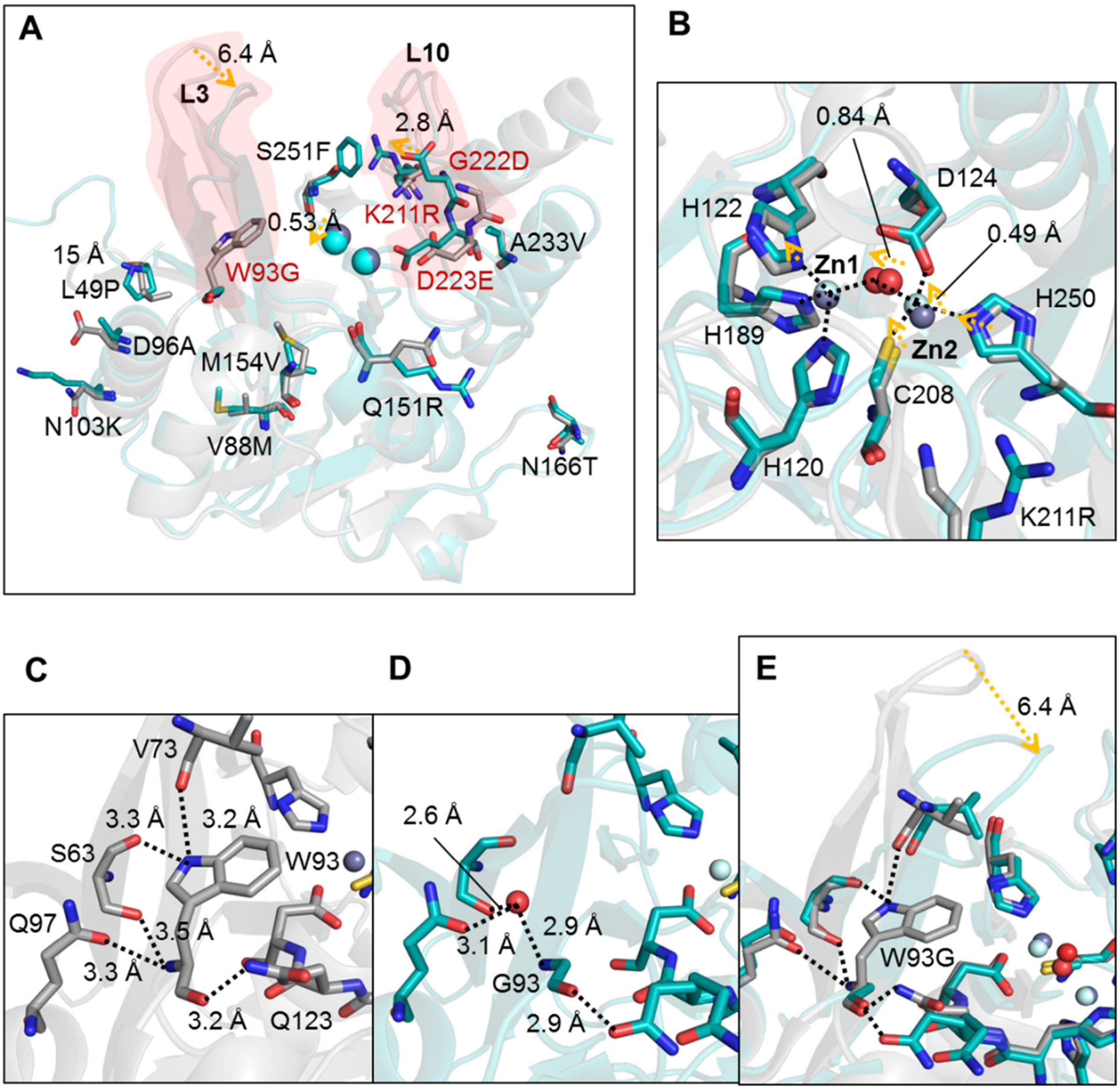


Figure 11. Crystal structures of NDM1 and NDM1-R10. NDM1 (grey, PDB ID: 3SPU) and NDM1-R10 (cyan, PDB ID: 5JQJ) are shown as cartoon. Mutated residues (**A**) and metal coordinating residues (**B**) and residues form electrostatic networks with W93 (**C-E**) are shown as sticks. Two zinc ions (grey spheres for NDM1 and light cyan spheres for NDM1-R10) and the bridging water molecule (red sphere) are shown. Displacements are shown by yellow arrows with distances. Mutations proximal to the displaced main chains (red highlighted) are in red character of panel A.

NDM1-R10 showed significant movement of the active site loop L3 (**Figure 10-11**). To be specific, loop L3 moved toward the active site more than 6 Å and contained fewer hydrogen bonds to the indole nitrogen of W93, due to the loss of the tryptophan side chain by the W93G mutation (**Figure 12 C-E**). The side chain of W93 on loop L5 is involved in a hydrogen bond network (the indole nitrogen of W93 forms electrostatic network with the backbone atoms of S63 on β3 and V73 on β4) and in a hydrophobic interaction (the benzene ring of W93 ~4 Å distance from the side chains of L65 on β3 and V73 on β4) underneath loop L3, which provides steric support for the loop (**Figure 12**). The elimination of the tryptophan side chain underneath loop L3 by the W93G mutation weakens the interaction between β3 and β4 and removes steric support for the positioning of those beta strands. Interestingly, the conformation of the repositioned loop L3 in NDM1-R10 is similar to the conformation of loop L3 in VIM2 (PDB ID: 1KO3). The product bound NDM1-R10 structure showed that the W93G mutation repositions L65 on loop L3, where the hydrophobic side chain of L65 interacts with the benzene ring of the PMH substrate (i.e. van der Waals interactions, distance between CD1 of L65 and CAE of PMH: 3.6 Å) (**Figure 12**). The position of L65 in NDM1-R10 is approximately in the same place as the W93 sidechain in NDM1. Thus, the steric hindrance between the benzene ring of W93 and the benzene ring of PMH substrate most likely causes steric hindrance for the proper substrate positioning in the WT enzyme, which is alleviated through evolution (**Figure 12**). In addition to the W93G mutation in the active site, remote mutations, including V88M and L49P, also potentially affect the active site loop L3. The side chain of L88 on β5 is ~4 Å to the hydrophobic side chains on α1 (I105), α2 (L132), and L6 (V113) and the backbone atoms form hydrogen bonds with β4, which supports L3. The larger methionine sidechain at this position in NDM1-R10 may indirectly affect to the conformation and the flexibility of this region by better filling the cavity and increasing hydrophobic interactions (I101, L102 on α1: >5 Å to V88 → <4 Å to M88). α1 forms hydrophobic interaction with β2-4, which support L3. Therefore, the additional hydrophobic interactions between α1 and β5 likely affect L3.

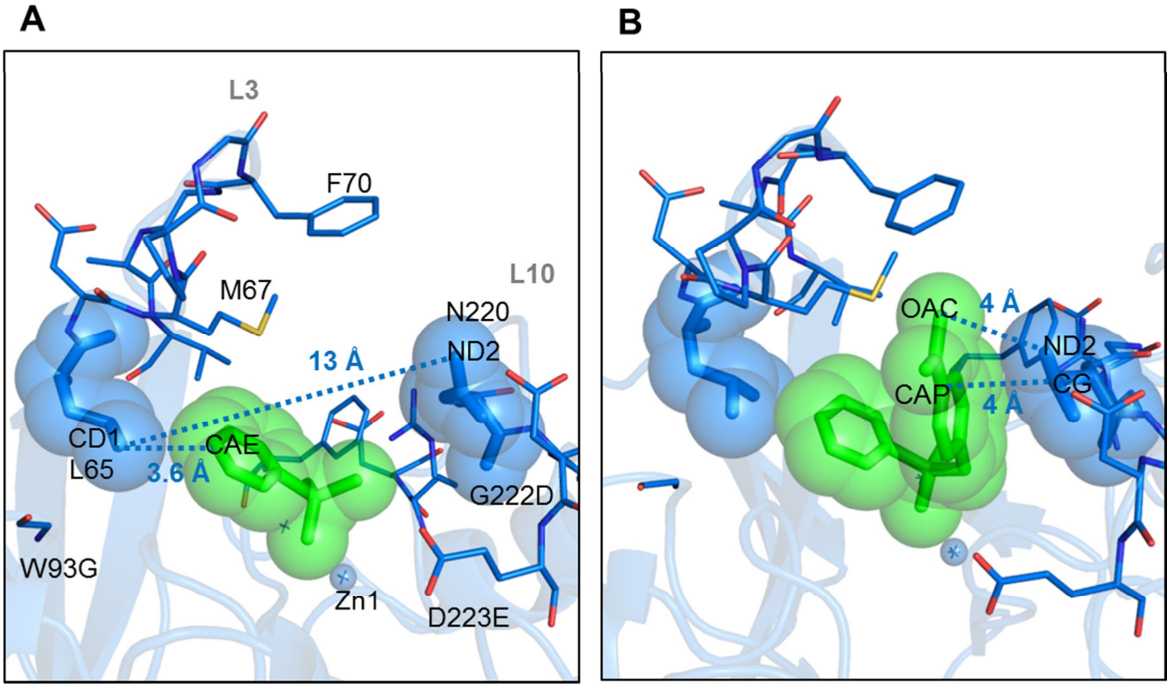


Figure 12. Crystal structures of PMH product complex and possible PMH substrate positioning in NDM1-R10. The PMH product (**A**), (green sticks and green spheres) and potential PMH substrate (**B**), (green sticks and green spheres) in NDM1-R10. The active site loop L3 and L10 residues displaced during evolution are shown in sticks. The distances between side chains of NDM1-R10 and PMH product and potential substrate are shown in blue dotted lines with distances and atom names. Zn1 is shown as sphere. PDB ID for the structure is 5K4N.

Many other mutations that occurred during the evolution of NDM1 were seemingly involved in remodelling and stabilizing the active site loop L10, potentially to improve binding of the PMH substrate. Four mutations occurred on or near L10 (G222D, K211R, D223E, S251F), forming new hydrogen bonds within and with loop L10 (**Figure 13**). The side chain NH of N220 formed a new hydrogen bond with the carboxyl group of E223 in L10 in NDM1-R10. Thus, the additional hydrogen bond between N220 and E223 within loop L10 potentially stabilizes the loop and the position of N220. This could be important because N220 is located in the substrate binding site, and has been shown to form a hydrogen bond with the lactam carbonyl group, stabilizing the ligand in antibiotic-bound structures, such as ampicillin (PDB ID: 3Q6X) (45) and cefuroxime (PDB ID: 4RLO) (49) in NDM1. In NDM1-R10, the side chain of N220 is pointed away from the active site due to its rotation to interact with D222 (**Figure 13**). This sidechain rotation removes a possible steric clash with the 4-nitrophenyl group of PMH and results in a more complementary pocket for the new PMH substrate (the predicted distance between nitro phenyl ring and the N220 side chain is ~4 Å) (**Figure 12 B**).

The side chain nitrogen (Nε) of R211 (K211R) forms a new hydrogen bond with the backbone carbonyl of S217 on L10 (**Figure 13**). The benzene ring of the F251 (S251F) on L12 is ~4 Å from R211, providing stabilization of the R211 side chain configuration (**Figure 10-11**). Again, this stabilizes a new conformation, which (unlike K211) points away from the active site, changing the shape of the substrate binding pocket.

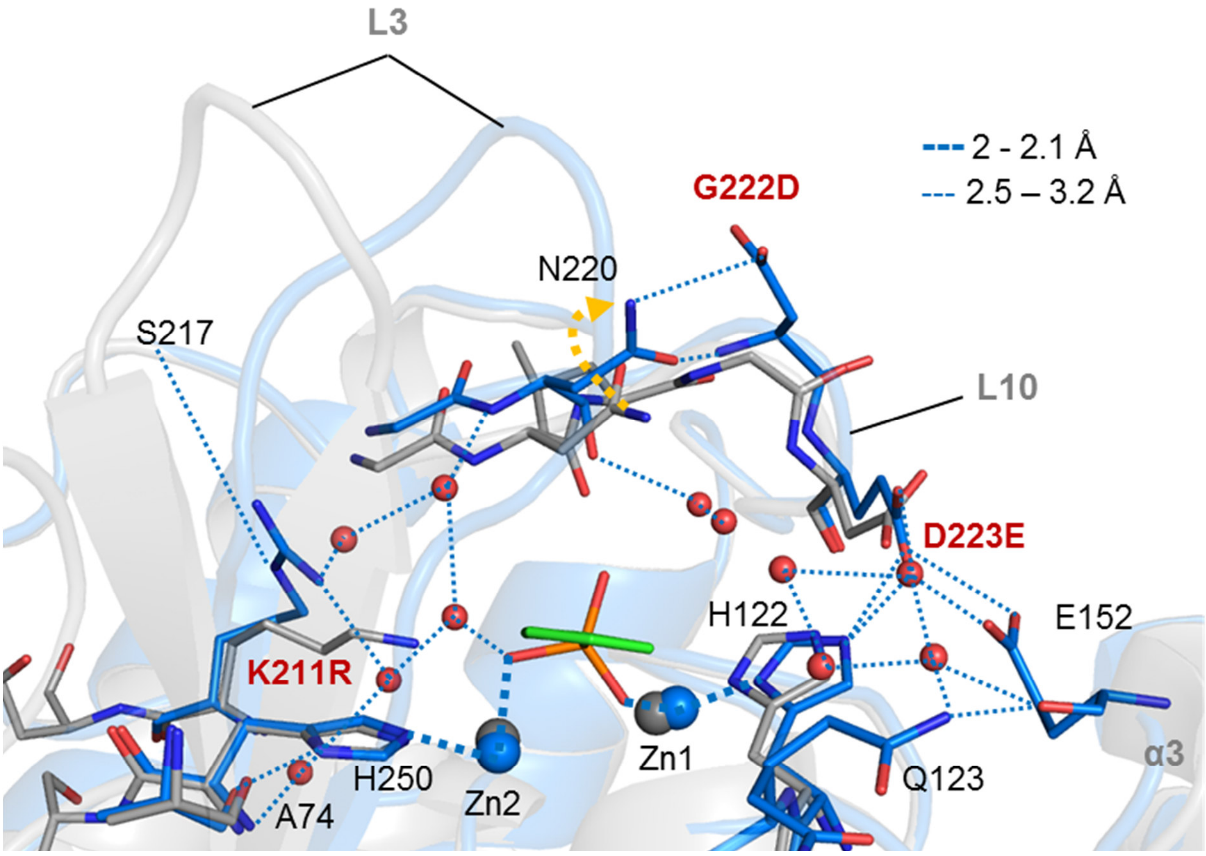


Figure 13. Super imposed structures of NDM1 and NDM1-R10. Electrostatic interactions in NDM1-R10. The structure of NDM1 (grey cartoon and stick, PDB ID: 4RL0) was superimposed with PMH product (green stick) complex of NDM1-R10 (blue cartoon and stick, PDB ID: 5K4N). Mutations on loop L10 are red. Thick dotted lines indicate distances of 2–2.1 Å (strong interaction) and thin dotted lines indicate distances of 2.5-3.2 Å. N220 rotational conformational change due to the new interaction with the G222D mutation is presented as a yellow dotted line.

The longer side chain of E223 (D223E) results in the formation of new hydrogen bonds with E152 (2.9 Å) on α3, in addition to maintaining the previous hydrogen bond with H122 (2.6 Å) on α2. Three additional waters in this region form new hydrogen bonds with E223, H122, E152 and Q123. The new hydrogen bond between the backbone oxygen of E152 and side chain nitrogen of Q123 on loop L6 appears to results from the loss of steric hindrance due to the M154V mutation proximal to the Q123 (**Figure 14**). All of the changes in this region appear to stabilize a new conformation in this part of the active site, in which side chains of E223 and Q123 point away from the active site, changing the shape of the substrate binding pocket.

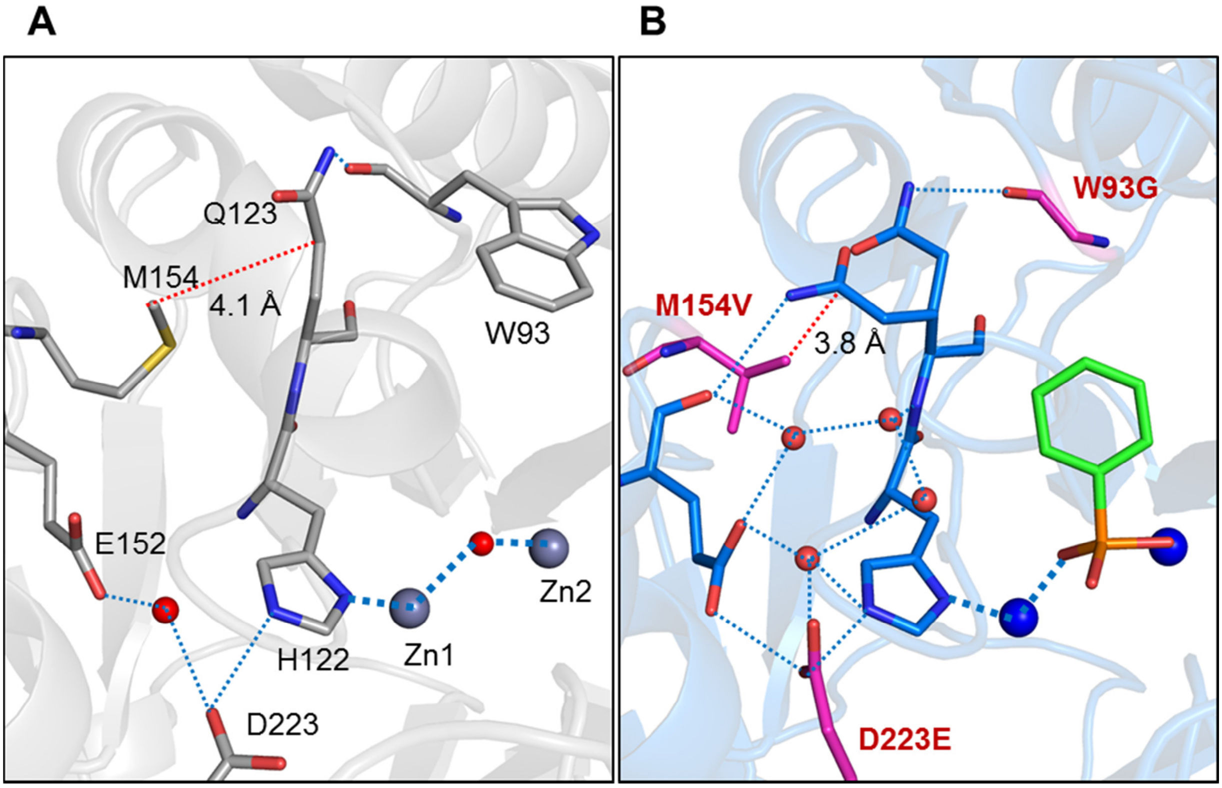


Figure 14. Structures of NDM1 and NDM1-R10. NDM1 (**A**, grey, PDB ID: 3SPU) and NDM1-R10 (**B**, blue, PDB ID: 5K4N). Interactions are shown as blue dotted thin lines (2.5-3.2 Å), blue dotted thick lines (2-2.1 Å), and red dotted lines (~4 Å). Mutations M154V, W93G, and D223E in R10 are shown as magenta sticks. PMH product is shown as green stick. Zinc ions are shown as grey (WT) and blue (R10) spheres and waters are shown as red spheres.

Finally, the A233V mutation α4 appears to stabilize the catalytic residues by improving the hydrophobic interactions inside of the buried area, from ~5 Å to <4 Å to L10 (L209) and α5 (I259, Y229) (**Figure 15**). In particular, A233V stabilize the metal ion coordinating residues H189 and C208 by stabilizing the proximal residues Y229 and L209. The stabilization of the catalytic residues may enhance the activity by suppressing unnecessary dynamics, which may produce unproductive configurations (50–52).

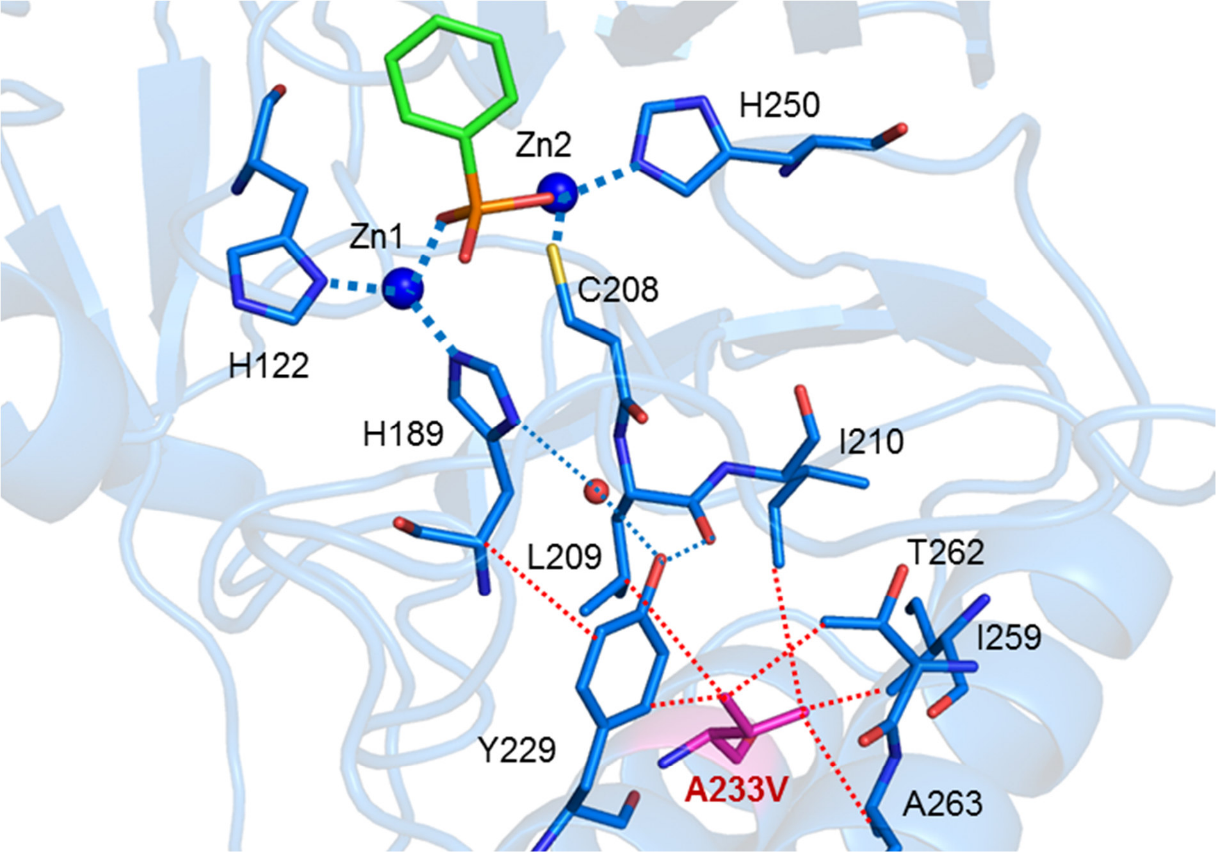


Figure 15. Stabilization of hydrophobic interaction due to A233V mutation in NDM1-R10. V233 is shown as magenta stick. Hydrophobic interactions (~4 Å) are shown as red dotted lines. Hydrogen bonds are shown as thin blue dotted lines. Strong interactions (~2 Å) are shown as thick blue dotted lines. PMH product is shown as green stick. PDB ID: 5K4N.

To sum up, the mutations that occurred during 10 rounds of directed evolution primarily adjusted the shape of the substrate binding pocket *via* alterations to the positioning of the active site loops L3 and L10 to improve the binding of PMH, while maintaining the same overall protein fold of NDM1. Interestingly, the electrostatic network observed for the cefuroxime product in NDM1 (PDB ID: 4RL0) was very similar to the network for the PMH product in NDM1-R10 (**Figure 16**). This suggests that the catalytic machinery of NDM1-WT and NDM1-R10 remains very similar, which is consistent with the similar hydrolysis mechanisms, as shown in **Figure 17**.

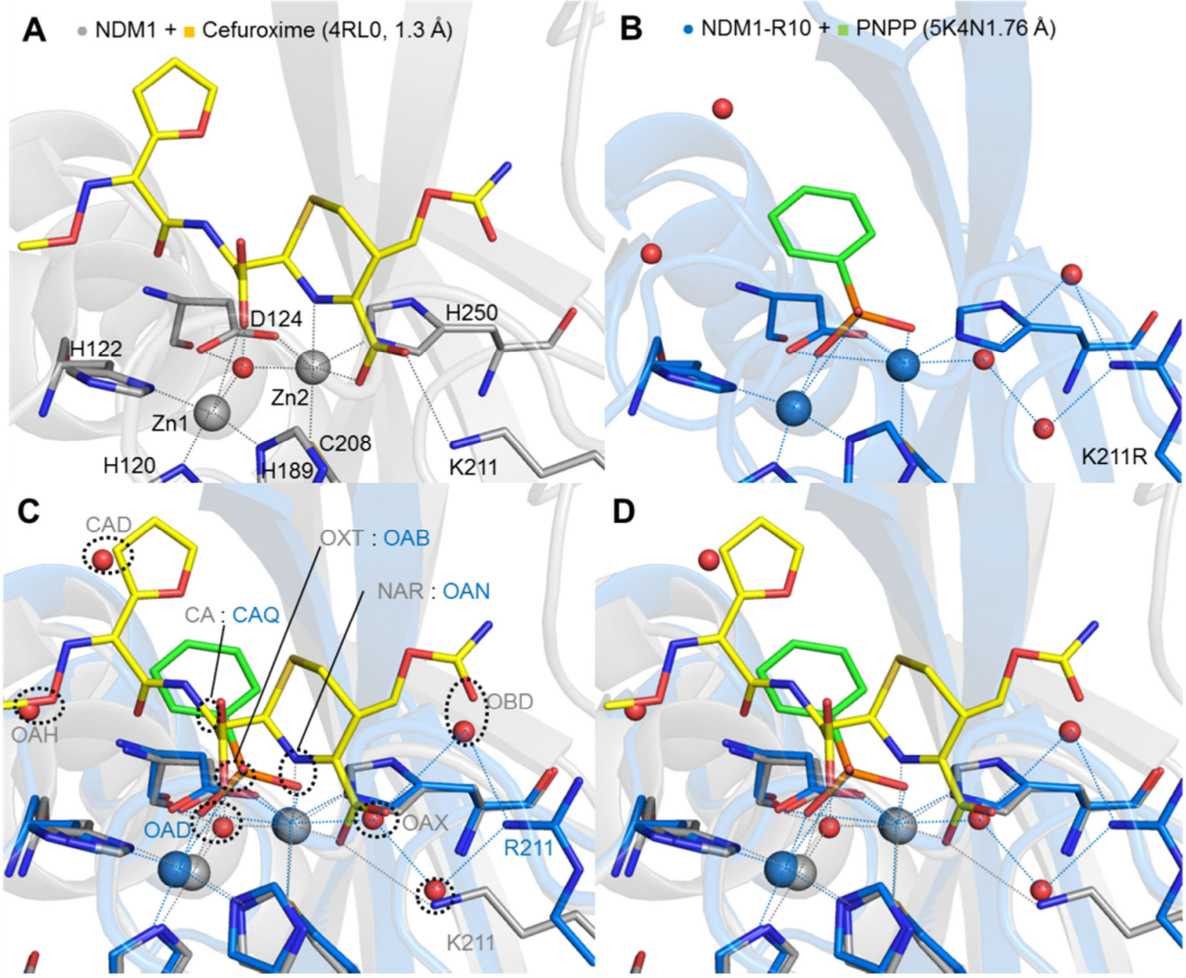


Figure 16. Crystal structures of the cefuroxime-NDM1 complex and PMH product-NDM1-R10 complex. (**A**) The structure of cefuroxime product (yellow stick) complex of NDM1 (grey cartoon and stick) (PDB ID: 4RL0). (**B**) PMH product (green stick) complex of NDM1-R10 (PDB ID: 5K4N, blue cartoon and stick). (C, D) Super imposed of two structures and electrostatic interactions (dotted lines). Zinc ions are shown in grey spheres for NDM1 and blue for NDM1-R10 and waters are shown in red spheres).

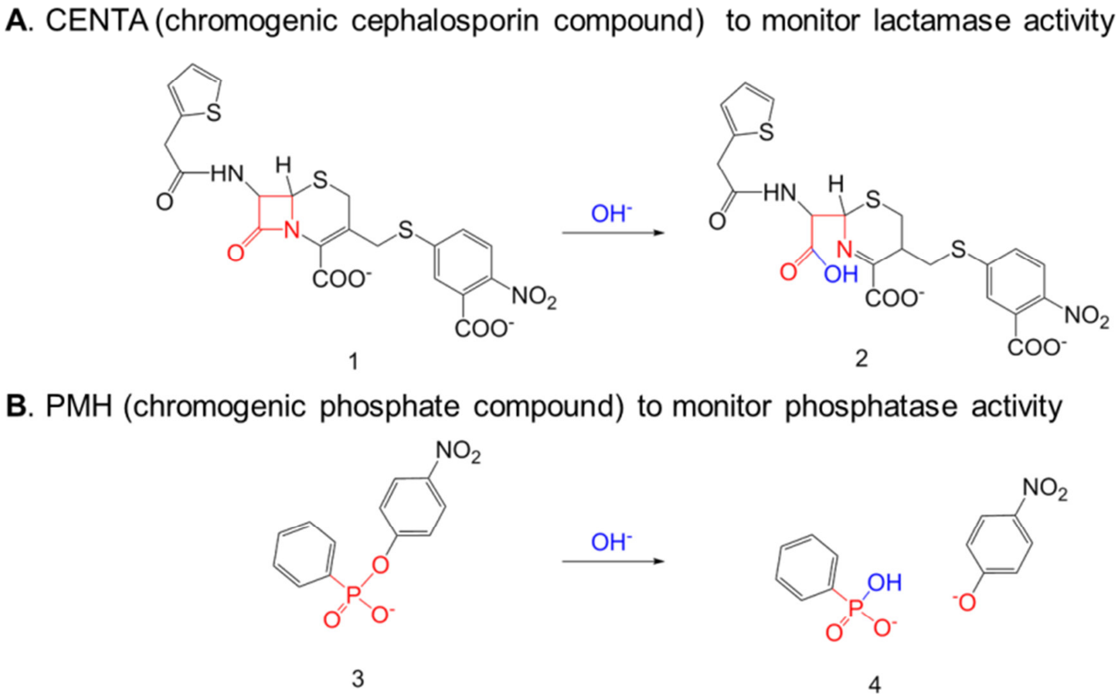


Figure 17. The hydrolysis of CENTA and PMH. (**A**) Lactam hydrolysis of CENTA and (**B**) phosphate hydrolysis of PMH are shown.

A domain-swapping event in the evolution of VIM2-R10.

To investigate how the structure of the most evolved VIM2-R10 was optimized to increase the weak phosphatase activity of VIM2 60-fold through 10 rounds of directed evolution, and to improve the solubility of the protein (**Figure 4-6**), we solved the crystal structure of VIM2-R10 (**Table 7**). A total of 15 mutations accumulated through 10 rounds of directed evolution, including five mutations in the active site (F67L, N154T, S66P, T191P, D223A), with the other ten mutations located on the surface (G27R, V41A, V46D, T64A, V72A, E150K, V195I, S202R, T263S, N264D). The mutations on the surface are far from the active site and do not involve in significant changes of the intramolecular hydrogen bond network.

Our size exclusion results showed that VIM2 primarily exists as a monomer, whereas VIM2-R10 exists as both monomer and dimer in solution (**Figure 8**). The overall scaffold of VIM2 is composed of two half β/α barrels where the core beta sheets of each half barrel are sandwiched to form one monomer subunit $\beta_{12}\alpha_5$ fold. Interestingly, our crystal structure of VIM2-R10 was a heterodimer, where the two half β/α barrels of chain A were disassembled and domain swapped its half β/α barrel with another half β/α barrel of chain B resulting in heterodimer composed of two sub units (**Figure 18**). Specifically, β1–β7 and α1 of chain B combined with β8–β12 and α2–α5 of chain A to form subunit 1. To form the other part of the dimer (subunit 2) the β1–β7 and α1 of chain A combined with β8–β12 and α2–α5 of chain B. In this dimer conformation, α2 was located above the active site of the other subunit, acting as an active site lid for each other's subunit (**Figure 18-20**).

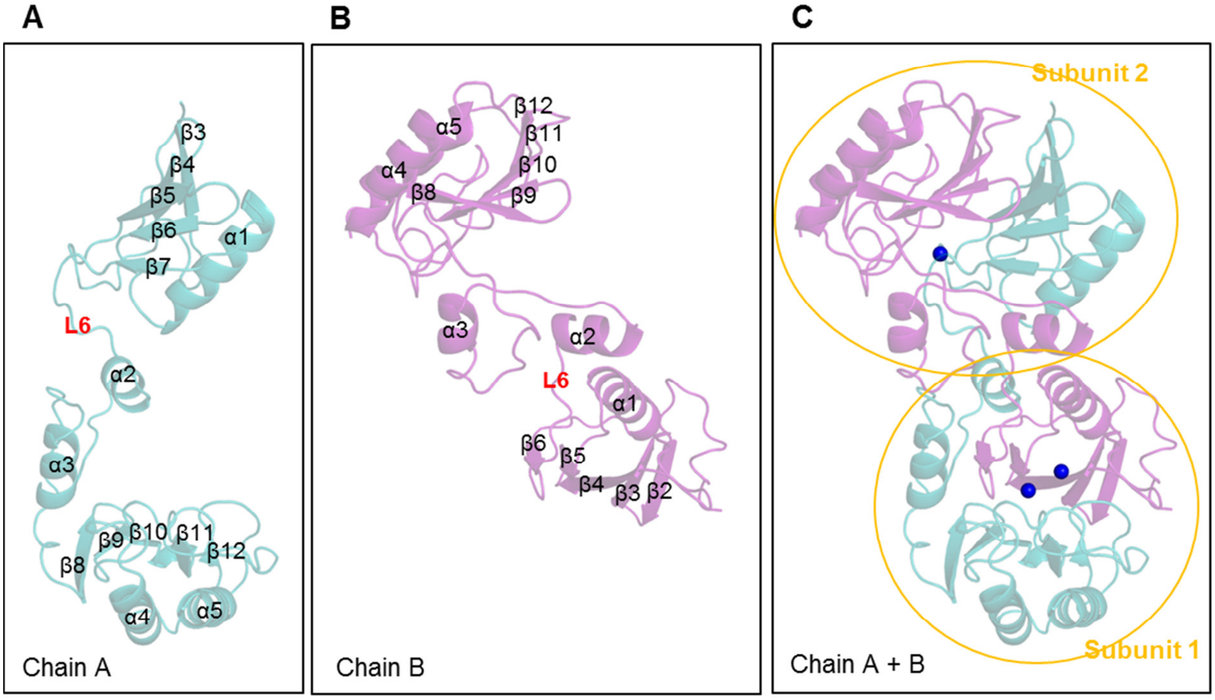


Figure 18. VIM2-R10 crystal structure showing domain swapping. Chain A (cyan, **A**) and chain B (purple, **B**) and Chain A plus chain B (**C**) of VIM2.R10 crystal structure are shown in cartoon style. Zinc ions are shown in blue spheres. Subunits of a VIM2-R10 dimer are shown in yellow.

The structural change from monomer to dimer suggests that the evolutionary nature of VIM2 is a gene duplication and fusion which has also been suggested for the HisA and HisF proteins involved in histidine synthesis pathway isolated from the hyper-thermophile *Thermotoga maritima* (53–55). They suggested that the ancestral β/α barrel of those proteins (HisA and HisF) is created by gene duplication and fusion of the half β/α barrels (55, 56). In our structure, the reason for the disassembly of VIM2 into half β/α barrels is seemingly due to loss of a salt-bridge between L6 (located in the middle of the two half βα barrels) and α3 and L10, which belongs to the other half β/α barrel (**Figure 18-19**). To be specific, the side chain nitrogen of N154 on α3 formed a salt-bridge with backbone oxygen of F121 on loop L6. The carboxylic oxygen of D223 on loop L10 formed a salt-bridge with imidazole nitrogen of H120 on loop L6. Thus, the N154T and D223A mutations in VIM2-R10 removed the salt-bridges and therefore allowed the disassembly of the $\beta_{12}\alpha_5$ fold into two half β/α barrels, which were then reassembled into a heterodimer with the two chains swapping each other's half β/α barrels. D124 on loop L6 (in between the two half β/α barrels) was originally a Zn^{2} coordinating residue. Due to half barrel swapping, D124 departed from the metal center and formed new salt bridge with backbone nitrogen of G128 on α2 of the other chain, which placed the α2 above the active site of another subunit (**Figure 19**).

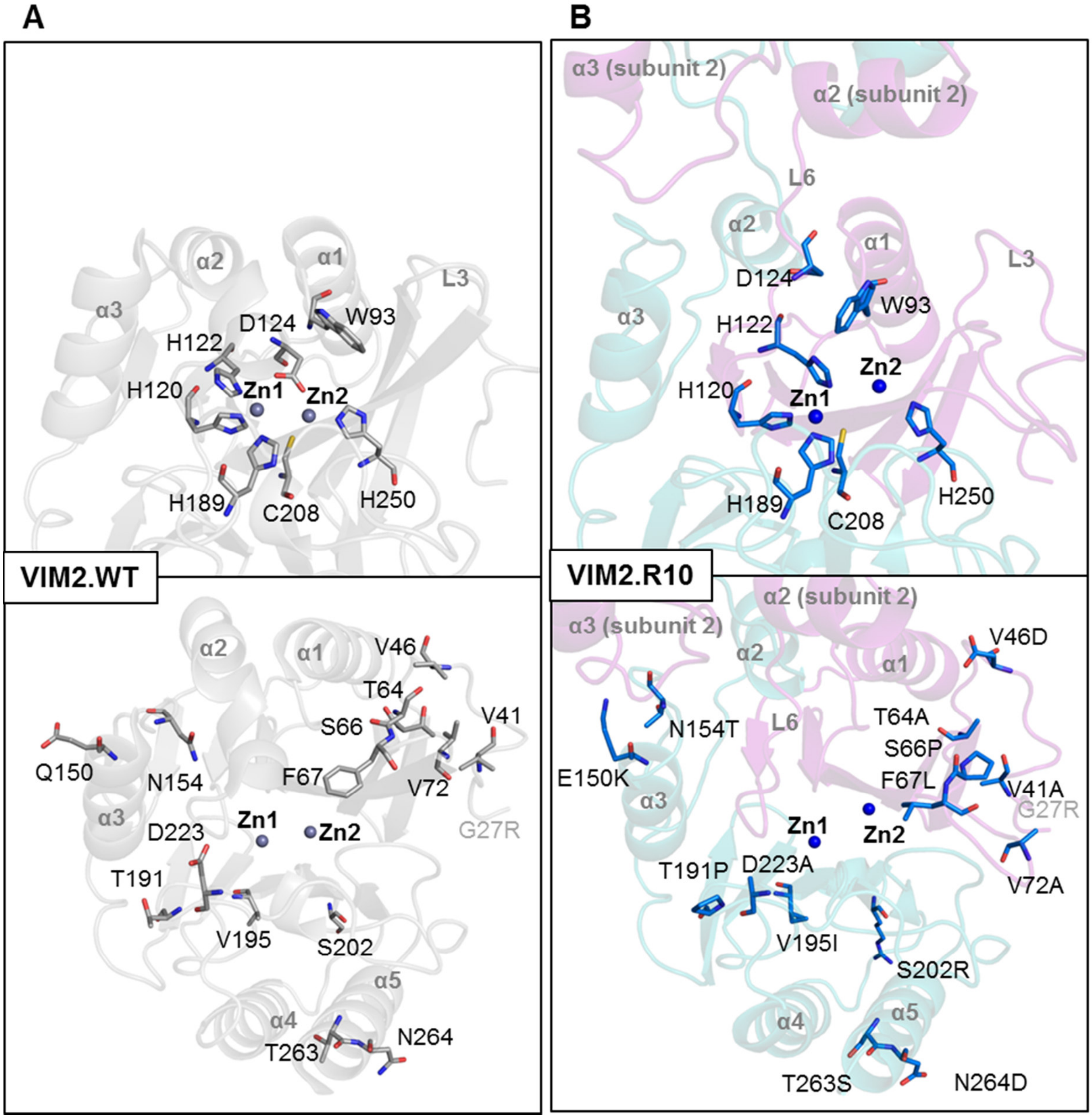


Figure 19. Crystal structure of VIM2 and VIM2-R10. Metal coordinating residues and the active site tryptophan residue W93 are shown with sticks in addition to the cartoon style representation of VIM2 (**A**), VIM2.R10 (subunit 1, **B**). VIM2 structure is shown in grey. VIM2-R10 structure is shown in blue sticks for mutations, purple cartoon for chain B for subunit 1 and cyan cartoon for chain A for subunit 1. The alpha helix α2 of chain A of subunit 2 located above the subunit 1 active site for VIM2-R10 are presented. Alpha helix numbers are presented with grey character on the cartoon. Mutated residues are shown as sticks in the lower panel. Zinc atoms are shown as grey for VIM2 and blue spheres for VIM2-R10.

Although VIM2-R10 maintained the $\beta_{12}\alpha_5$ fold for each sub unit, despite the disassembly and reassembly to transform into dimer by swapping half part of each other's chain (**Figure 18**), the active site and the loops L3 and L10 undergo significant changes (**Figure 19-20)**. First, the metal coordination changed dramatically, mainly due to the transformation of loop L6 which bridges the two half β/α barrels. To be specific, the loop L6 movement caused the departure of the D124 on the loop L6 from the metal center. The departure of the Zn^{2+} coordinating residue D124 caused reduced occupancy of the Zn^{2+} metal ion from the β-site. Only subunit 1 unequivocally showed binding of Zn^{2+}. In addition, the removal of the bond between N154 and the backbone oxygen of the F115 and the movement of the loop L6 caused conformational change in H122. The conformational change of the Zn^{2+} coordinating residue in the α-site H122, caused a 1.5 Å movement of the Zn^{2+} into a more buried area. In the absence of D124 in the active site of VIM2-R10, the weakly bound Zn^{2+} in the β-site moved further apart from Zn^{2+} in the α-site. The Zn-Zn distance changed from 4.2 Å in VIM2 to 5.2 Å in VIM2-R10. In addition, the active site shape was changed by the mutations (S66P, F67L, N154T, T191P, D223A), and conformational changes of the side chains (W93, F121, H122, D123, D124, N220) (**Figure 19**). Specifically, the active site tryptophan W93 adopted an alternative conformation by rotating ~90° from it's original position. The distance between the active site loop L3 and L10 was increased. The distance between F67 on loop L3 and N220 on loop L10 was 5.9 Å in VIM2. The distance was 7.1 Å in VIM2-R10 (**Figure 20**). This was mainly due to the conformational change with the F67L and S66P mutations on the loop L3, and the conformational change of N220 on loop L10. Moreover, the loop L6 and the α2 of another subunit created a wall and a lid for the active site (**Figure 19-20**).

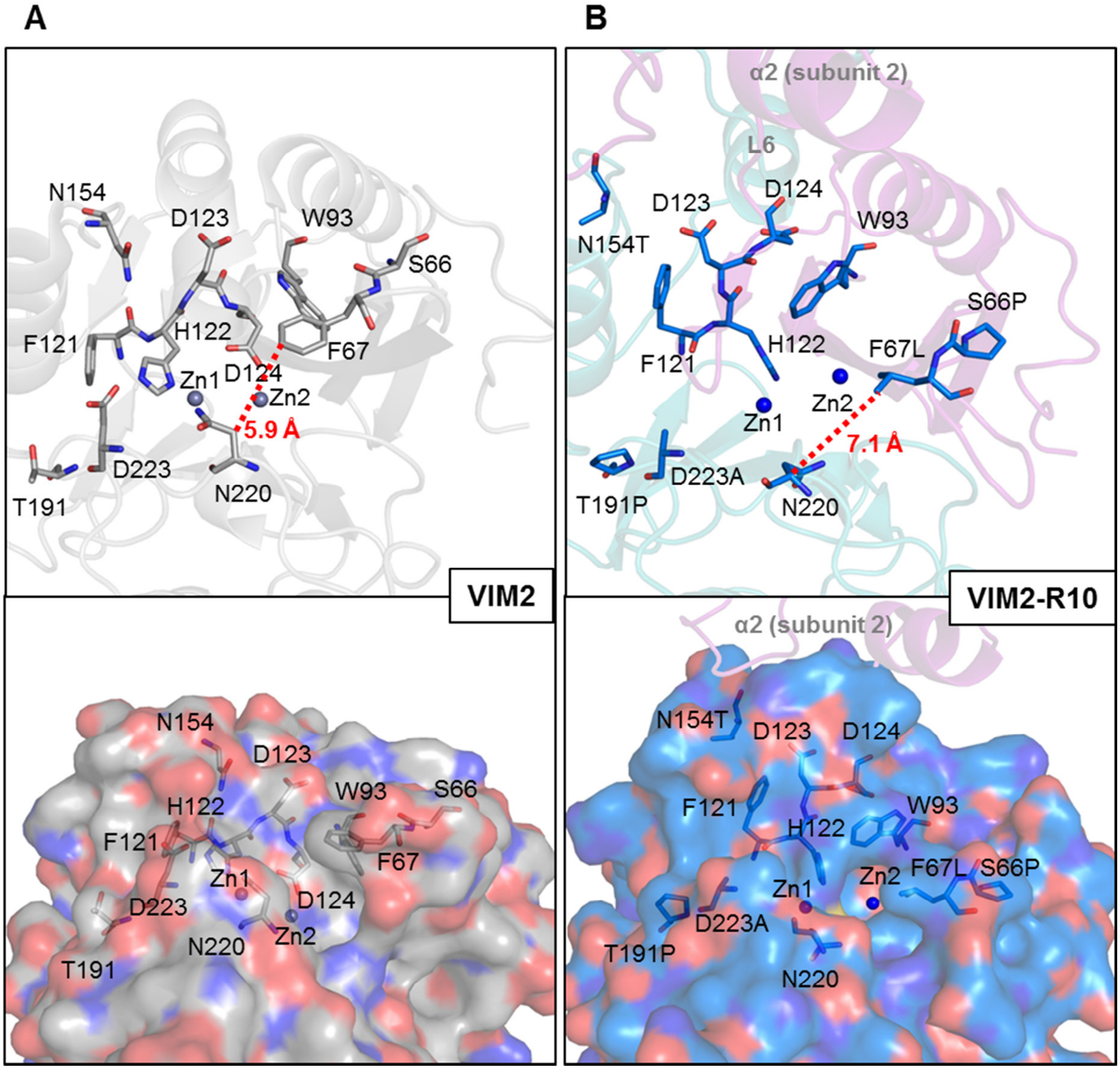


Figure 20. Active sites of the crystal structures VIM2 and VIM2-R10. (**A**) VIM2 is shown in grey cartoon (top) and surface (lower panel). (**B**) VIM2.R10 chain A is shown in cyan cartoon and chain B in purple cartoon (top panel) and the surface is blue in lower panel. Active site residues that were mutated (S66P, F67L, N154T, T191P, D223A) or changed their conformations significantly (W93, F121, H122, D123, D124, N220) are shown in sticks (grey sticks for VIM2 and blue sticks for VIM2-R10). PDB IDs for the structures are1KO3 for VIM2 and XXX for VIM2-R10.

Structural evolutionary changes of NDM1-R10 versus VIM2-R10

The increased activity of NDM1-R10 was due to the remodeling of the active site loops L3 and L10 to properly orient the PMH substrate binding mode while maintaining remaining structural features including metal coordination. The increased solubility of VIM2-R10 (**Figure 5**) may result from the domain swapping observed in VIM2-R10, which appears to stabilize a low occupancy metal ion state.

Discussion

Our observations suggest that the evolution followed a limited but accessible adaptive pathway from each starting point. Despite the limitations of our directed evolution scheme, i.e., we employed a pre-selection using ampicillin antibiotic resistance (native function) and subsequently screened only ~400 variants for PMH activity per round and enzyme, our results indicate that this screening system was sufficient to identify function-changing mutations. The direct prescreening with PMH activity, which expanded our screening capacity up to ~2,000 variants per round, did not yield any further improvement in the last two rounds (**Figure 4**). The adaptive evolution from each enzyme also appears to be deterministic. Such determinism in evolutionary trajectories has been observed in other protein evolution (20, 57, 58). Although a few studies revealed multiple trajectories are sometimes accessible (15), the multiple pathways often shared the same mutations, and the alternative pathways deviated with few mutations (7, 59). Therefore, it may be common that only a limited change of adaptive mutations are accessible, and that protein evolution is largely deterministic from a given starting point.

Starting from orthologous enzyme sequences, the evolution may follow alternative trajectories and yield substantially different outcomes, and thus evolution of new enzyme function can be contingent on genetic backgrounds. Sequence changes among orthologous enzymes are, by definition, neutral in terms of their native functions, i.e., mutations have been accumulated under purifying neutral selection pressure to maintain the native function (60). However, due to the prevalence of epistasis, accumulation of these presumably neutral mutations could cause changes on the effects of mutated as well as non-mutated residues. Sequence incompatibilities observed, (e.g. if the exchange of amino acids between orthologs is functionally deleterious, also dubbed Dobhanzhky-Muller incompatibilities), have been observed in various systems (61–63). It has also been shown that the effect of a single substitution, A156T, on 56 different hepatitis C virus NS3 proteases varied from nearly lethal to beneficial (2-fold increase) for its native protease activity (64). One may expect that orthologous mutations alter the level and evolvability of promiscuous functions even more because there is no selection pressure to maintain such physiologically irrelevant functions. A few recent studies demonstrated the effect of a functional mutation could vary significantly between orthologous enzymes (65). The effect of E383A on nine gamma-glutamyl phosphate reductase (ProA) orthologs towards a promiscuous activity, NAGSA dehydrogenase activity, varied from 50- to 770-fold (16). Likewise, the double mutations, H205I/I263W on dihydrocoumarin lactonases increase methyl-parathion hydrolase activity by 8- to 6000-fold depending on genetic backgrounds (66). Our work demonstrates that a highly adaptive mutation in one background can be deleterious in another background: the effect of W93G on PMH activity varies from a 2-fold decrease in VIM2 to a 100-fold increase in NDM1. The prevalence of epistasis caused a striking difference in accessibility to function-changing mutations, and thus, adaptability in terms of both the rate of adaptation as well as the level of fitness plateau. The two comparative trajectories adopted different genotypic solutions and phenotypic solutions; NDM1 improved catalytic efficiency at the cost of solubility, whereas VIM2 mostly improved solubility. Therefore, the evolution of new enzyme functions can be highly contingent on starting genetic backgrounds. Contingency can play an important role during natural adaptation, where the evolution of a new enzyme function depends on the presence of evolvable protein sequences. However their occurrence as well as their spatial and temporal presence can be stochastic (58).

Our observations also have important implications for protein design, engineering and evolution in the laboratory. Protein engineers tend to choose a single starting sequence based on the availability of biochemical and structural information. However, it might be important to explore various starting sequences and examine which sequence is more evolvable. Importantly, in our work, an initially less fit variant evolved at a higher rate and ultimately reached a higher fitness plateau, thus the initial activity may not be a good indicator for evolvability. The molecular causes for contingency can be specific for each protein sequence, and it may be impossible predict evolvability from sequences at this point. Understanding of molecular basis of mutational epistasis and evolutionary contingency remain a challenge that must be met for us to develop our ability decipher evolutionary dynamics of proteins as well as generate novel proteins in the laboratory.

References

1. Tawfik OK and DS, Khersonsky O, Tawfik DS (2010) Enzyme Promiscuity : A Mechanistic and Evolutionary Perspective. *Annu Rev Biochem* 79:471–505.
2. Brien PJO, Herschlag D (1999) Catalytic promiscuity and the evolution of new enzymatic activities. *Chem Biol* 6(April):91–105.
3. Ng T-K, Gahan LR, Schenk G, Ollis DL (2015) Altering the substrate specificity of methyl parathion hydrolase with directed evolution. *Arch Biochem Biophys* 573:59–68.
4. Baier F, Tokuriki N (2014) Connectivity between catalytic landscapes of the metallo-β-lactamase superfamily. *J Mol Biol* 426(13):2442–2456.
5. Luo J, van Loo B, Kamerlin SCL (2012) Examining the promiscuous phosphatase activity of Pseudomonas aeruginosa arylsulfatase: A comparison to analogous phosphatases. *Proteins: Struct, Funct, Bioinf* 80(4):1211–1226.
6. Afriat L, Roodveldt C, Manco G, Tawfik DS (2006) The latent promiscuity of newly identified microbial lactonases is linked to a recently diverged phosphotriesterase. *Biochemistry* 45(46):13677–13686.
7. Harms MJ, Thornton JW (2013) Evolutionary biochemistry: revealing the historical and physical causes of protein properties. *Nat Publ Gr* 14(8):559–571.
8. Wyganowski KT, Kaltenbach M, Tokuriki N (2013) GroEL / ES Buffering and Compensatory Mutations Promote Protein Evolution by Stabilizing Folding Intermediates. *J Mol Biol* 425(18):3403–3414.
9. Harms MJ, Thornton JW (2014) Historical contingency and its biophysical basis in glucocorticoid receptor evolution. *Nature* 512(7513):203–207.
10. Lehner B (2011) Molecular mechanisms of epistasis within and between genes. *Trends Genet* 27(8):323–331.
11. Tufts DM, et al. (2014) Epistasis constrains mutational pathways of hemoglobin adaptation in high-altitude pikas. *Mol Biol Evol*:msu311.
12. Weinreich DM, Delaney NF, DePristo MA, Hartl DL (2006) Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 312(5770):111–114.
13. Bridgham JT, et al. (2008) Evolution of a new function by degenerative mutation in cephalochordate steroid receptors. *PLoS Genet* 4(9):e1000191.
14. Bridgham JT, Ortlund EA, Thornton JW (2009) An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature* 461(7263):515–519.
15. Phillips PC (2008) Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet* 9(11):855–867.
16. Khanal A, McLoughlin SY, Kershner JP, Copley SD (2015) Differential effects of a mutation on the normal and promiscuous activities of orthologs: implications for natural and directed evolution. *Mol Biol Evol* 32(1):100–108.
17. Tokuriki N, Tawfik DS (2009) Stability effects of mutations and protein evolvability. *Curr Opin Struct Biol* 19(5):596–604.
18. Tokuriki N, Tawfik DS (2009) Protein dynamism and evolvability. *Science* 324(5924):203–207.
19. Szendro IG, et al. (2013) Predictability of evolution depends nonmonotonically on population size. *Proc Natl Acad Sci* 110(2):571–576.
20. Roy SW (2009) Probing evolutionary repeatability: neutral and double changes and the predictability of evolutionary adaptation. *PLoS One* 4(2):e4500.
21. Lobkovsky AE, Wolf YI, Koonin E V (2011) Predictability of evolutionary trajectories in fitness landscapes. *PLoS Comput Biol* 7(12):e1002302.
22. Bloom JD, Labthavikul ST, Otey CR, Arnold FH (2006) Protein stability promotes evolvability. *Proc Natl Acad Sci* 103(15):5869–5874.
23. Bloom JD, Raval A, Wilke CO (2007) Thermodynamics of neutral protein evolution. *Genetics* 175(1):255–266.
24. DeLano WL (2002) The PyMOL molecular graphics system.
25. Armougom F, et al. (2006) Expresso : automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Res* 34(suppl 2):604–608.
26. Larkin MA, et al. (2007) Clustal W and Clustal X version 2 . 0. *Bioinformatics* 23(21):2947–2948.
27. Tamura K, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28(10):2731–2739.
28. Zhao H, Giver L, Shao Z, Affholter JA, Arnold FH (1998) Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat Biotechnol* 16(3):258–261.

29. Studier FW (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Expr Purif* 41:207–234.
30. Bebrone C, et al. (2001) CENTA as a chromogenic substrate for studying β-lactamases. *Antimicrob Agents Chemother* 45(6):1868–1871.
31. Gasteiger E, et al. (2005) *Protein identification and analysis tools on the ExPASy server* (Springer).
32. Niesen FH, Berglund H, Vedadi M (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* 2(9):2212–2221.
33. Cabrita LD, et al. (2007) Enhancing the stability and solubility of TEV protease using in silico design. *Protein Sci* 16(11):2360–2367.
34. Kabsch W (2010) Xds. *Acta Crystallogr Sect D Biol Crystallogr* 66(2):125–132.
35. Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr Sect D Biol Crystallogr* 62(1):72–82.
36. Evans PR (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr Sect D Biol Crystallogr* 67(4):282–292.
37. Winn MD, et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr Sect D Biol Crystallogr* 67(4):235–242.
38. Karplus PA, Diederichs K (2012) Linking crystallographic model and data quality. *Science* 336(6084):1030–1033.
39. Diederichs K, Karplus PA (2013) Better models by discarding data? *Acta Crystallogr Sect D Biol Crystallogr* 69(7):1215–1222.
40. Vagin A, Teplyakov A (1997) MOLREP: an automated program for molecular replacement. *J Appl Crystallogr* 30(6):1022–1025.
41. Afonine P V., et al. (2012) Towards automated crystallographic structure refinement with phenix. refine. *Acta Crystallogr Sect D Biol Crystallogr* 68(4):352–367.
42. Murshudov GN, et al. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr Sect D Biol Crystallogr* 67(4):355–367.
43. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr Sect D Biol Crystallogr* 60(12):2126–2132.
44. Tokuriki N, et al. (2012) Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme. *Nat Commun* 3:1257.
45. King D, Strynadka N (2011) Crystal structure of New Delhi metallo-β-lactamase reveals molecular basis for antibiotic resistance. *Protein Sci* 20(9):1484–1491.
46. Garcia-Saez I, Docquier J-D, Rossolini GM, Dideberg O (2008) The three-dimensional structure of VIM-2, a Zn-beta-lactamase from Pseudomonas aeruginosa in its reduced and oxidised form. *J Mol Biol* 375(3):604–11.
47. Weiss MS, Hilgenfeld R (1997) On the use of the merging R factor as a quality indicator for X-ray data. *J Appl Crystallogr* 30(2):203–205.
48. Diederichs K, Karplus PA (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat Struct Biol* 4(4):269–275.
49. Feng H, et al. (2014) Structural and mechanistic insights into NDM-1 catalyzed hydrolysis of cephalosporins. *J Am Chem Soc* 136(42):14694–7.
50. Isom DG, Castañeda CA, Cannon BR, Velu PD, García-Moreno E B (2010) Charges in the hydrophobic interior of proteins. *Proc Natl Acad Sci* 107(37):16096–16100.
51. Jiménez-Osés G, et al. (2014) The role of distant mutations and allosteric regulation on LovD active site dynamics. *Nat Chem Biol* 10(6):431–436.
52. Khersonsky O, et al. (2012) Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed Kemp eliminase KE59. *Proc Natl Acad Sci* 109(26):10358–10363.
53. Huber R, et al. (1986) Thermotoga maritima sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C. *Arch Microbiol* 144(1976):324–333.
54. Nelson KE, et al. (1999) Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of Thermotoga maritima. *Nature* 399(6734):323–329.
55. Sterner R, Höcker B (2005) Catalytic versatility, stability, and evolution of the (βα) 8-barrel enzyme fold. *Chem Rev* 105(11):4038–4055.
56. Lang D, Thoma R, Henn-Sax M, Sterner R, Wilmanns M (2000) Structural evidence for evolution of the β/α barrel scaffold by gene duplication and fusion. *Science* 289(5484):1546–1550.
57. Bauer B, Gokhale CS (2015) Repeatability of evolution on epistatic landscapes. *Sci Rep* 5:20–25.
58. Lenormand T, Roze D, Rousset F (2009) Stochasticity in evolution. *Cell* 24(January):157–165.
59. Steinberg B, Ostermeier M (2016) Environmental changes bridge evolutionary valleys. *Sci Adv* 2(1):e1500921.

60. Wang X, Minasov G, Shoichet BK (2002) Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. *J Mol Biol* 320(1):85–95.
61. Orr HA, Turelli M (2001) The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities. *Evolution (N Y)* 55(6):1085–1094.
62. Kondrashov AS, Sunyaev S, Kondrashov FA (2002) Dobzhansky–Muller incompatibilities in protein evolution. *Proc Natl Acad Sci* 99(23):14878–14883.
63. Haerty W, Singh RS (2006) Gene regulation divergence is a major contributor to the evolution of Dobzhansky–Muller incompatibilities between species of Drosophila. *Mol Biol Evol* 23(9):1707–1714.
64. Sarrazin C, et al. (2007) Dynamic hepatitis C virus genotypic and phenotypic changes in patients treated with the protease inhibitor telaprevir. *Gastroenterology* 132(5):1767–1777.
65. Christin P-A, Weinreich DM, Besnard G (2010) Causes and evolutionary significance of genetic convergence. *Trends Genet* 26(9):400–405.
66. Luo X, et al. (2014) Switching a newly discovered lactonase into an efficient and thermostable phosphotriesterase by simple double mutations His250Ile/Ile263Trp. *Biotechnol Bioeng* 111(10):1920–1930.

Chapter 8. Discussion

In this thesis, I have studied three model enzyme systems: the designed Kemp Eliminase, KE07, the N-acyl homoserine lactonase AiiA, and the bacterial beta-lactamases NDM1 and VIM2. In each case, my work focused on the evolution of these enzymes for increased, or novel activity. The aim in each study was to obtain a detailed molecular understanding of how changes in protein structure and dynamics change the function of the protein. To accomplish this, I have solved a large number of high-resolution crystal structures, as well as performing molecular dynamics simulations, and in the case of KE07 carrying out detailed kinetic analysis and solution fluorescence spectroscopy. Although the insights into protein evolution that I have gained from each system are slightly different, there are a number of clear molecular processes that appear to be broadly important. By following the molecular improvements in KE07, AiiA, NDM1 and VIM2 as they occur I have been able to characterize the importance of active site preorganization during evolution and catalysis, how the preorganization of active sites can be modulated by the effects of remote second and third shell mutations, and the sometimes extreme effects of intramolecular epistasis and how this manifests in evolutionary contingency – differential evolvability of some protein scaffolds. These concepts are discussed below.

8.1 Preorganization and kinetic isotope effects

The correlation that I observed in the KE07 evolutionary trajectory between the optimization of active site preorganization and the primary kinetic isotope effects (KIEs) were revealed through Arrhenius plots ($Ea_{(D)} - Ea_{(H)}$, A_H/A_D) and protein crystallography (Chapter 3-6). The KIEs significantly changed when the active site was heterogeneous at R5, in which catalytic residues W50 and E101 adopted new rotamer conformations (Figure 8.1). This is consistent with analysis of soybean lipoxygenase-1, which shows similar correlation between conformational preorganization state and KIEs (1) (discussed in Chapter 3). This correlation has also been demonstrated previously in other enzymes (AADH, MADH, DHFR), using heterogeneous conformation model-QM/MM studies (2–5).

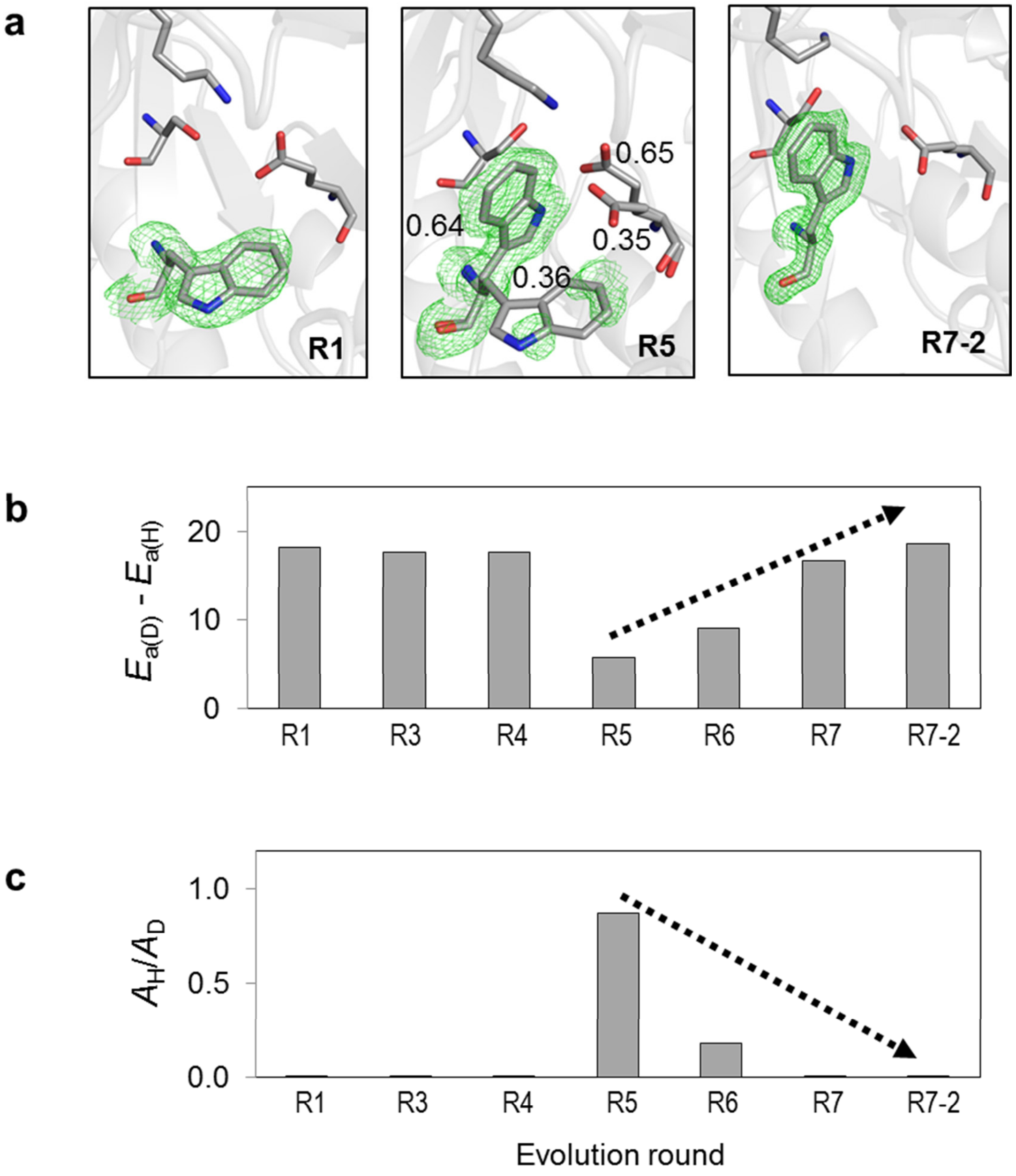


Figure 8. 1 | Changed Kinetic Isotope Effects (KIEs) due to preorganization optimization during *de novo* KE07 evolution. Active site residues, E101, W50, K222, S48, are shown as grey sticks. *F*o – *F*c omit maps (green mesh) are centered on W50 and contoured at 3.0σ. Occupancies (<1) are shown near the each side chain. PDB ID: 5D2T (R1), 5D3O (R5), 5D38 (R7-2).

8.2 Preorganization by remote mutations – conformational tinkering

In every directed evolution experiment the activity changed by orders of magnitude, yet the mutations rarely occurred in the active sites. Rather, they were located in the second and third shells (Figure 8.2). However, the actives sites changed dramatically in some instances and it appears that the primary role of the remote mutations was to manipulate the conformations of residues within the active site/substrate binding site through differential stabilization of certain rotamers (Figure 8.2), for e.g. by forming or removing local interactions (hydrogen bonds and van der Waals interactions). This observation that remote mutations appear to optimize active site preorganization, is consistent to the other studies of enzyme evolutions (6–11), but has rarely been shown so clearly. Moreover, while allosteric regulation by remote mutations has been rationalized (10, 12, 13), enzyme design is still limited to active site remodelling.

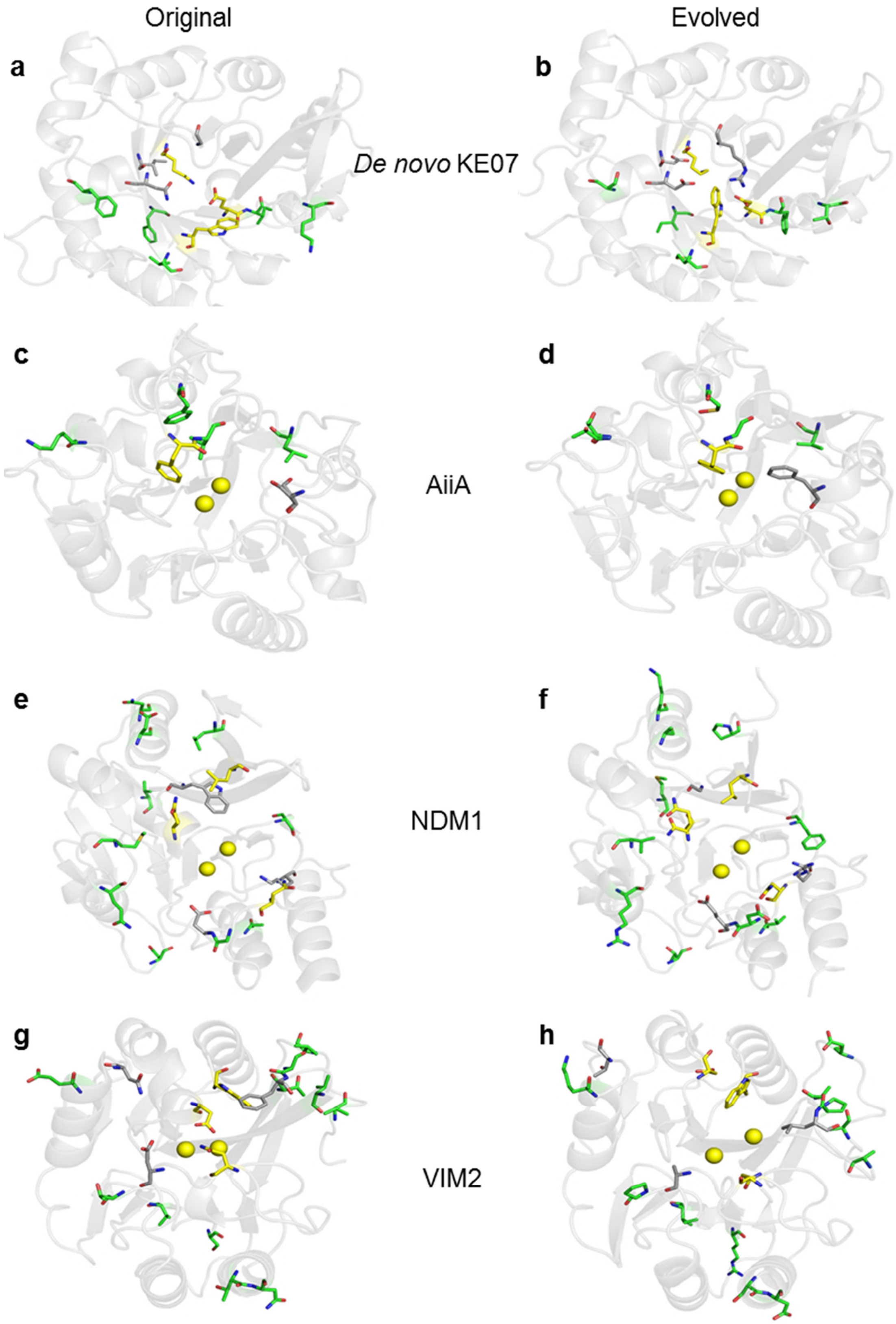


Figure 8. 2 | Preorganization by remote mutations. Crystal structures of KE07 designed (**a**) and evolved R7-2 (**b**), AiiA wild type (**c**) and evolved R4 (**d**), NDM1 wild type (**e**) and evolved R10 (**f**), VIM2 wild type (**g**) and evolved R10 (**h**). Remote mutations are shown as green sticks. Key active site residues are shown as yellow sticks (zinc ions are shown as yellow spheres, zinc positioning residues are excluded for clarity). Active site mutations are shown as grey sticks. PDB ID: 5D2T (**a**), 5D38 (**b**), 5EH9 (**c**), 5EHT (**d**), 3SPU (**e**), 5K4N (**f**), 1KO3 (**g**).

8.3 Epistasis and contingency in evolution

Intramolecular epistasis, the context-dependence, or interaction, between mutations, and contingency, where early mutations constrain later mutations (14), was observed in the evolution of the four enzymes towards small synthetic molecules, presented in this thesis: KE07, AiiA, NDM1, and VIM2. As discussed above, mutations that accumulated during enzyme evolution mainly increased activity (e.g. 200-20,000-fold k_{cat}/K_M) by optimizing the preorganization of active site for the new substrates (Figure 8.2-8.3). VIM2 was something of an outlier in this analysis, as the mutations mainly increased solubility by stabilizing an unusual domain-swapped dimer state (Figure 8.3). The direct comparison between NDM1 and VIM2, which belong to same phenotype and functional group genotype, revealed that these enzymes accumulated different mutations to achieve different genotypic solutions for the challenge of developing the same phenotype (phosphonate hydrolysis). This highlights the role of epistasis, the importance of interactions between amino acid sidechains within proteins, for the prediction of the effects of key mutations and their evolutionary destination. Furthermore, the evolutionary trajectories of the four enzymes reveal this evolutionary contingency – the dependence of mutation on those preceding it. For example, in the KE07 evolution, the mutation K146T in R6, which appears to enrich the active configuration of active site, has no effect on R2 activity, which does not have the I7D mutation (Chapter 3). Similarly, in AiiA evolution, mutation S20F in R4, which appears to optimize the paraoxon binding mode, has no effect on wild type activity because the wild-type protein does not have the early mutations V69G and F64C that predispose the active site to allow the phenylalanine sidechain to affect substrate binding. VIM2 and NDM1 appear to show similar contingency, whereby the early mutation W93G in NDM1 dramatically changed the conformation of loop L3 and constrains the effects of later mutations, while in contrast, the early mutation D223A in VIM2 appears to have led to the disassembly of two β/α-domains – obviously constraining the effects of later mutations in that trajectory also. Molecular epistasis and evolutionary contingency have been shown in other systems (15, 16), suggesting that this may be a general aspect of enzyme evolution.

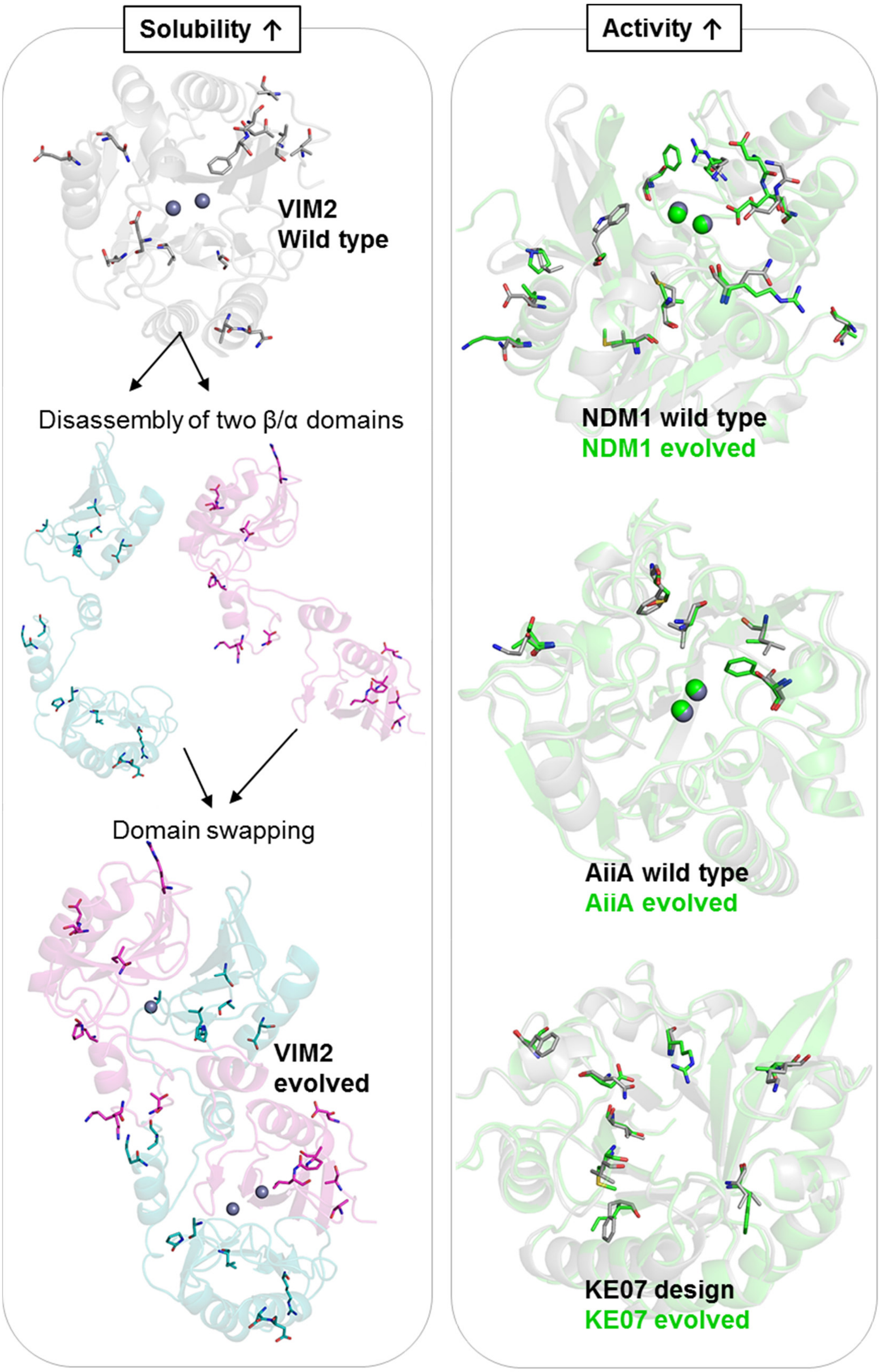


Figure 8. 3 | Structural changes of main chains during directed evolutions of β/α-fold enzymes. VIM2 evolution mainly to improve solubility is shown in left side panel. Evolutions of NDM1, AiiA, KE07 enzymes mainly to improve activity are shown in right side panel. Wild type (grey) and most evolved variant (green) are superimposed. Zinc ions are shown as spheres and mutated residues are shown as sticks. PDB ID: 1KO3 (VIM2-WT), 5D2T (KE07-WT), 5D38 (KE07-R7-2), 5EH9 (AiiA-WT), 5EHT (AiiA-R4), 3SPU (NDM1-WT), 5K4N (NDM1-R10).

8.4 Future Work

The studies described in this thesis have primarily focused on understanding how molecular evolution occurs in model systems so that we can learn the rules of molecular design and the principles that underlie the structure-dynamics-function relationship in enzymes. The studies themselves are essentially complete, although further work on the domain swapping that occurs in VIM2 could be necessary to identify the point at which the dimer originates and the relative stability of the metal-ion bound and apo-enzymes. The most important considerations for future work are how these results will benefit two important aspects of molecular evolution: protein design and natural evolution. The results shown here conflict with the primary methods and prevailing paradigm around enzyme design: huge changes in activity can be accomplished with (in some cases) no alteration to the active site, rather by optimizing a new configuration of the inner shell residues involved in catalysis and substrate binding. This suggests a new direction focussed on mapping minor conformations that might be responsible for different or improved activities and stabilizing these conformations by remote mutations. Secondly, we can begin to apply the principles discussed here to our understanding of natural evolution. In particular the work on evolutionary contingency can teach us a lot about how one organism might be predisposed to evolving resistance to toxic chemicals, whereas another might not. The obvious application for this is in the management of insecticide, herbicide and antibiotic resistance in insects, weeds and bacteria – by sequencing a variety of organisms we may be able to identify resistance-prone species. This work could also have some impact upon our understanding of diease-related polymorphisms in humans, particularly in terms of how epistasis could control how a mutation could have a catastrophic effect in one individual, but no effect in another. Altogether, this work has deepened our understanding of important processes in molecular evolution and design.

8.5 References

1. Meyer MP, Tomchick DR, Klinman JP (2008) Enzyme structure and dynamics affect hydrogen tunneling: The impact of a remote side chain (I553) in soybean lipoxygenase-1. *Proc Natl Acad Sci* 105(4):1146–1151.
2. Basran J, Sutcliffe MJ, Scrutton NS (1999) Enzymatic H-transfer requires vibration-driven extreme tunneling. *Biochemistry* 38(10):3218–3222.
3. Loveridge E, Tey L, Allemann R (2010) Solvent effects on catalysis by E. coli dihydrofolate reductase. *J Am Chem Soc* 132:1137–1143.
4. Masgrau L, et al. (2006) Atomic description of an enzyme reaction dominated by proton tunneling. *Science* 312(2006):237–241.
5. Glowacki DR, Harvey JN, Mulholland AJ (2012) Taking Ockham's razor to enzyme dynamics and catalysis. *Nat Chem* 4(3):169–176.
6. Preiswerk N, et al. (2014) Impact of scaffold rigidity on the design and evolution of an artificial Diels-Alderase. *Proc Natl Acad Sci* 111(22):8013–8018.
7. Giger L, et al. (2013) Evolution of a designed retro-aldolase leads to complete active site remodeling. *Nat Chem Biol* 9(8):494–498.
8. Blomberg R, et al. (2013) Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature* 503(7476):418–421.
9. Khersonsky O, et al. (2011) Optimization of the in-silico-designed Kemp eliminase KE70 by computational design and directed evolution. *J Mol Biol* 407(3):391–412.
10. Jiménez-Osés G, et al. (2014) The role of distant mutations and allosteric regulation on LovD active site dynamics. *Nat Chem Biol* 10(6):431–436.
11. Ram Prasad B, Warshel A (2011) Prechemistry versus preorganization in DNA replication fidelity. *Proteins: Struct, Funct, Bioinf* 79(10):2900–2919.
12. Osuna S, Jiménez-Osés G, Noey EL, Houk KN (2015) Molecular dynamics explorations of active site structure in designed and evolved enzymes. *Acc Chem Res* 48(4):1080–1089.
13. Frushicheva MP, Cao J, Warshel A (2011) Challenges and advances in validating enzyme design proposals: The case of kemp eliminase catalysis. *Biochemistry* 50(18):3849–3858.
14. Harms MJ, Thornton JW (2013) Evolutionary biochemistry: revealing the historical and physical causes of protein properties. *Nat Publ Gr* 14(8):559–571.
15. Shah P, McCandlish DM, Plotkin JB (2015) Contingency and entrenchment in protein evolution under purifying selection. *Proc Natl Acad Sci* 112(25):E3226–E3235.
16. Harms MJ, Thornton JW (2014) Historical contingency and its biophysical basis in glucocorticoid receptor evolution. *Nature* 512(7513):203–207.

Appendices

A1. Recipes

Pre-mixtures	Recipes
Borate buffer 20X (DNA running buffer)	8 g of NaOH, 45 g of Boric acid (pH 8.0)
Bromophenol Blue Loading Dye 6X (DNA dye)	63 % (v/v) glycerol, 0.025 % (w/v) bromophenol blue, 0.025 % (w/v) xylene cyanol FF, 50 mM EDTA in water
Bromophenol Blue Loading Dye 4X (Protein dye)	25 mg bromophenol blue, 5 ml glycerol, 150 mg Tris, 1 g (w/v) SDS, 1 ml β-mercaptoethanol in 10 ml water
Coomasie blue protein staining solution	400 ml Methanol, 100 ml acetic acid, 1 Phast Gel Blue R tablet (0.1 % coomassie stain)
LB (Luria broth)	10 g trypton, 5 g yeast extract, 10 g NaCl in 1 L water
LB-Agarose plate	10 g trypton, 5 g yeast extract, 10 g NaCl in 1 L water / Autoclave and cool to about 40 ℃ / Add antibiotics and pour into plates under lamina flow.
MES buffer 10X (SDS-PAGE running buffer)	60.6 g Tris, 97.6 g MOPS, 10 g SDS, 3 g EDTA in 1 L water
MOPS buffer 10X (SDS-PAGE running buffer)	60.6 g Tris, 104.6 g MOPS, 10 g SDS, 3 g EDTA in 1 L water
TB (Terrific broth)	12 g trypton, 24 g yeast extract, 2.31 g KH_2PO_4, 12.54 g K2HPO4 g, 4 - 10 ml glycerol in 1 L water

A2. Suppliers of equipment

Equipment	Suppliers
16/200 Size exclusion column – Superdex 75	GE Healthcare
16/26 HiPrep Desalting coulumn	G.E. Healthcare
AKTA FPLC	G.E. Healthcare
Amicon Ultra-15 Centrifugal Filter	Millipore
ASB270BT Autoclave	Astell Scientific
Bio-Rad Mini Sub Cell GT	Bio-Rad

Bolt Mini Gel Tank	Novex
Cary Eclipse Spectrophotometer	Agilent
Cary Spectrophotometer	Varian
CO8000 Cell Density Meter	WPA Biowave
Crystal plate with three well (Intelli-plate)	Art Robbins Instruments
French Pressure Cell Press	SLM Instruments
HisTrap FF Column - 5ml	G.E. Healthcare
MJ Mini Personal Thermal Cycler	Bio-Rad
NanoDrop ND-1000 Spectrophotometer	Thermo
Orion ROSS Combination pH Electrode	Orion Pacific
SpectraMax M2 Microplate Reader	Molecular Devices
UV transilluminator (312 nm)	Hanimax Statesman
Veriti 96-well Thermal Cycler	Applied Biosystems
VX22G (R15A, R9A roters)	Hitachi

A3. Suppliers of chemicals and reagents

Chemicals and reagents	Suppliers
2-Log DNA ladder	New England Biolabs
2-Methyl-2,4-pentanediol (MPD)	Sigma-Aldrich
Acetonitrile	Sigma-Aldrich
Agar	Bacto laboratories
Agarose	Bio-Rad
Ampicilin sodium salt	Astral Scientific
BIS-TRIS	Sigma-Aldrich
BIS-TRIS PROPANE	Sigma-Aldrich
Bromophenol Blue	Sigma-Aldrich
BugBuster Protein Extraction Reagent	Novagen
Dithiothreitol	Sigma-Aldrich
dNTPs	New England Biolabs
EDTA	Sigma-Aldrich
Ethanol	Merck
Glycerol	Merck
Glycerol	Merck
HCl	Merck

HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)	Astral Scientific
Imidazole	Alfa Aesar
IPTG	Merck
K_2HPO_4	Ajax Finechem
Kanamycin	Merk
KH_2PO_4	Ajax Finechem
Low Range Protein ladder	Invitrogen
Methanol	Merck
$MnCl_2$	Ajax Finechem
2-(N-morpholino)ethanesulfonic acid (MES)	Sigma-Aldrich
$MgSO_4$	Sigma-Aldrich
$MgCl_2$	Sigma-Aldrich
NaCl	Sigma-Aldrich
NaF	Ajax Finechem
NaOH	Ajax Finechem
Paraoxon-ethyl	Fluka
PEG10000	Sigma-Aldrich
PEG3350	Sigma-Aldrich
PEG400	Sigma-Aldrich
PEG4000	Sigma-Aldrich
PEG8000	Sigma-Aldrich
Phast Gel Blue R tablet	Sigma
Phusion Polymerase	New England Biolabs
4-Nitrophenyl phenylphosphonate	Sigma-Aldrich
RedSafe DNA Stain	Chembio
Shirimp Alkaline Phosphatase	New England Biolabs
Sodium dodecyl sulfate (98.5%)	Sigma-Aldrich
TRIS BASE (ULTRA PURE)	Astral Scientific
Trypton	Bacto Laboratories
Tris(2-carboxyethyl)phosphine (TCEP)	Sigma-Aldrich
Wizard SV Gel and PCR Clean-Up System	Promega
Yeatst Extract	Bacto Laboratories
$ZnCl_2$	BDH Laboratories
β-Mercaptoethanol	Sigma-Aldrich

A4. Parameters of proteins from ProtParam tool

Enzyme name	Evolution round	Molecular weight (Da)	Theoretical pI	Ext. coefficients (280 nm) $M^{-1}cm^{-1}$	Ext. coefficients (280 nm) Abs 0.1%
KE07	1	29116.4	6.01	18450	0.634
	3	29073.2	5.90	18450	0.635
	4	29182.3	5.68	18450	0.632
	5	29213.4	5.90	18450	0.632
	6	29155.2	5.58	18450	0.633
	7	29070.1	5.72	19940	0.686
	7-2	29197.3	5.79	18450	0.632
AiiA	WT	28471.4	4.94	21890	0.769
	4	28422.3	4.88	21890	0.770
NDM1	WT	25931.2	5.88	27960	1.078
	10	25959.4	6.19	22460	0.865
VIM2	WT	25854.8	4.92	29910	1.157
	10	25827.9	5.19	29910	1.158

The physical and chemical parameters were calculated using ProtParam tool (http://web.expasy.org/protparam/).

A5. Output files for quantum cluster models in Chapter 5

Designed-like model

```
1\1\GINC-R1250\FTS\RM062X\Gen\C42H64N12O12\ROOT\07-Dec-2013\0\\#M062X/
 gen 6D SCF=Tight INT(grid=ultrafine) OPT=(TS,calcfc,noeigentest,maxcyc
 =250,Loose,MaxStep=10) IOP(2/17=4) maxdisk=2684354560\\des_ts1.freq\\0
 ,1\H,-1.15139,7.94951,2.99285\C,-1.0429,6.82879,2.97418\O,-1.97086,6.1
 438,2.14274\H,0.23825,5.50298,-2.81084\C,-0.0341,4.61393,-2.48883\C,0.
 61692,3.41703,-3.13737\C,0.04346,2.54193,-4.04701\N,0.90795,1.55045,-4
 .38727\C,2.07556,1.69642,-3.73342\C,1.95686,2.8907,-2.89974\C,3.04709,
 3.27942,-2.10943\C,4.21399,2.49252,-2.15564\C,4.30689,1.35736,-2.97981
 \C,3.23626,0.93506,-3.77195\H,7.9154343124,2.5790235738,1.336888499\C,
 7.2320998552,1.7526906982,1.1046389404\C,5.7826907267,2.2434220487,1.0
 955697834\C,4.8242700861,1.114727523,0.7328174595\C,3.3393601084,1.474
 596179,0.7160645966\O,3.021495201,2.5547928783,1.3557202051\O,2.532270
 6344,0.7249849148,0.1592944779\C,7.639210669,1.2070465153,-0.256339860
 2\O,7.3056704064,1.7450418117,-1.3005686851\N,8.4059194452,0.080968482
 1,-0.2420164575\H,8.7140725687,-0.3073349755,-1.1235686821\H,6.88563,-
 2.26746,-0.83936\C,6.08326,-2.33538,-1.58868\C,5.48852,-3.72691,-1.757
 55\O,6.1451,-4.726,-1.49074\N,4.24561,-3.79073,-2.24464\H,4.7866,-1.70
 155,3.51892\C,4.01679,-2.36174,3.08118\C,3.31614,-1.65481,1.87291\C,1.
 96584,-2.29898,1.48167\C,1.89748,-3.44905,0.70412\C,0.65486,-4.02831,0
 .35239\C,-0.55084,-3.42574,0.76731\O,-1.82105,-4.00631,0.40793\C,-0.48
```

```
372,-2.30096,1.54125\C,0.74882,-1.72991,1.87062\C,4.39568,-3.7579,2.66
882\O,4.87235,-4.00692,1.54163\N,4.15878,-4.65001,3.61592\H,-4.29321,-
3.0819,4.6045\C,-3.52533,-2.7242,4.09881\C,-4.02324,-1.98965,2.85409\N
,-3.50856,-2.19822,1.6227\C,-4.18764,-1.44412,0.72386\N,-5.1383,-0.777
59,1.37727\C,-5.0841,-1.10615,2.68687\H,-5.41703,-4.84202,0.23544\C,-5
.5912,-5.13546,-0.89646\N,-6.27041,-4.05044,-1.55847\C,-5.62545,-3.058
4,-2.12205\N,-4.31034,-3.07025,-2.07791\N,-6.28038,-2.08409,-2.73401\H
,-2.63995,1.98763,3.67761\C,-1.70145,2.06846,4.24595\N,-1.41887,3.5131
4,4.54978\H,-1.32405,6.61526,3.98424\H,-0.06722,6.56477,2.62313\H,-2.4
4809,6.77946,1.60446\H,-1.07305,4.55087,-2.73688\H,0.27351,4.59373,-1.
46422\H,-0.95083,2.63526,-4.43122\H,0.50389,0.69586,-4.06115\H,2.9943,
4.14936,-1.48864\H,5.05131,2.7655,-1.54802\H,3.30782,0.06013,-4.38374\
H,5.22143,0.80232,-3.00098\H,5.5238389784,2.6457360659,2.0797516332\H,
5.6869014944,3.0698919813,0.3856865172\H,4.9460074786,0.299173114,1.45
94789118\H,5.0587699271,0.689465791,-0.2495294993\H,7.373635455,0.9880
261368,1.8794552874\H,5.48048,-1.84609,-0.85244\H,6.15965,-1.9273,-2.5
7489\H,3.92309,-2.86828,-2.4571\H,3.64195,-4.20648,-1.5644\H,3.08351,-
0.66514,2.20655\H,3.96381,-1.79788,1.03333\H,2.80328,-3.90615,0.36426\
H,0.76176,-0.8247,2.44097\H,0.63018,-4.9257,-0.22976\H,-1.3864,-1.8518
8,1.89959\H,-2.52619,-3.39108,0.62203\H,3.31246,-2.37854,3.88656\H,3.1
8084,-4.54467,3.79636\H,-3.17064,-1.96887,4.7685\H,-2.85902,-3.48494,3
.74918\H,-5.73846,-0.75109,3.45545\H,-5.78695,-0.1377,0.96526\H,-3.993
89,-1.39312,-0.32717\H,-6.18941,-6.01407,-1.01921\H,-4.57386,-5.13024,
-1.22805\H,-7.26993,-4.05468,-1.58891\H,-3.7892,-2.32769,-2.49864\H,-3
.83209,-3.82184,-1.6236\H,-7.27964,-2.09917,-2.77034\H,-5.77581,-1.333
06,-3.15995\H,-1.76653,1.55115,5.18031\H,-0.99494,1.75207,3.50734\H,-2
.19485,4.07292,4.25909\H,-1.27682,3.62533,5.53322\H,-0.59693,3.80236,4
.05911\C,-1.7161225168,0.8698102528,-1.1606345233\C,-2.9617592262,1.36
31217818,-0.7227373205\C,-3.0240584753,2.2806325048,0.3048902308\C,-1.
7988380874,2.6658699024,0.8622087949\C,-0.5547967033,2.201310135,0.404
4164418\C,-0.4882812106,1.2734814322,-0.6354853186\N,-1.7411279656,-0.
1418196842,-2.2010733863\H,-3.8725372466,1.0230093599,-1.2022310394\H,
-3.9716093055,2.6768255566,0.6537808102\O,-1.6449626031,3.4828290034,1
.8932968105\C,0.4359595738,2.8389895201,1.2667452896\H,0.4699390653,0.
8766451896,-0.9680795518\N,-0.1688220086,3.5783730956,2.1341642174\O,-
0.7036426955,-0.4751505908,-2.735920975\O,-2.838348808,-0.6348016146,-
2.4971504834\H,1.7862949195,2.7046491746,1.2937438518\H,4.25755,-5.376
8,2.98776\H,8.6697160624,-0.3861988531,0.6116537585\O,-7.60101,0.78626
,-0.05777\C,-8.94355,1.08955,-0.88293\O,-9.57592,-0.1848,-1.52508\C,-9
.56023,2.36951,-1.05797\H,-10.61913,2.29603,-0.92285\H,-9.33985,2.7534
7,-2.03208\H,-9.16816,3.0406,-0.32257\\Version=EM64L-G09RevC.01\State=
1-A\HF=-3196.9628894\RMSD=3.749e-09\RMSF=6.185e-06\Dipole=-2.4702016,1
.6468998,6.5005241\Quadrupole=-95.6049511,39.7333615,55.8715896,28.748
9967,10.99201,42.9303483\PG=C01 [X(C42H64N12O12)]\\@
```

Evolved-like model

```
1\1\GINC-R1836\FTS\RM062X\Gen\C42H64N12O12\ROOT\11-Dec-2013\0\\#M062X/
 gen 6D SCF=Tight INT(grid=ultrafine) OPT=(TS,calcfc,noeigentest,maxcyc
 =250,Loose,MaxStep=10) IOP(2/17=4) maxdisk=2684354560\\evo_ts1.freq\\0
 ,1\C,-1.735112,-5.421901,-3.691826\O,-2.812662,-4.785607,-3.090277\H,-
 0.843289,-6.212371,2.275087\C,-1.117472,-5.069011,2.318903\C,-0.88248,
 -4.438731,0.985408\C,0.184599,-4.654116,0.12658\N,0.055217,-3.908458,-
 1.002221\C,-1.070383,-3.174806,-0.961428\C,-1.730312,-3.469836,0.30297
 8\C,-2.91792,-2.836699,0.608222\C,-3.443193,-1.923805,-0.315783\C,-2.7
 91088,-1.657073,-1.514322\C,-1.588024,-2.281656,-1.860019\C,6.65315262
 42,-1.7078321549,-1.9715450708\C,5.5039622281,-2.70941022,-1.860770210
 3\C,6.3639758881,-0.4687672119,-1.1297913146\O,5.5571927519,-0.4937834
 221,-0.2158390055\N,7.039227892,0.6630946232,-1.4816063515\H,6.9147961
 156,1.5205295004,-0.9493674329\C,6.153832,1.870401,1.724869\C,6.020726
 ,3.328176,2.151278\O,7.006611,4.105351,2.162053\N,4.792202,3.722718,2.
```

```
495287\H,5.199616,1.928247,-2.848878\C,4.430866,2.546843,-2.391222\C,3
.737881,1.79908,-1.271738\C,2.360762,2.363045,-0.92058\C,2.206598,3.60
0221,-0.283892\C,0.942331,4.090888,0.034399\C,-0.161434,3.352229,-0.32
2957\O,-1.426702,3.77338,-0.05205\C,-0.036134,2.158141,-0.939023\C,1.2
10931,1.667981,-1.269052\C,4.98857,3.864557,-1.881855\O,5.903098,3.869
406,-1.048469\N,4.500393,4.979564,-2.423522\H,4.911066,5.971567,-1.888
163\C,-3.421304,3.933418,-3.23598\C,-4.014805,3.009214,-2.162008\N,-3.
424659,2.794696,-0.9666\C,-4.160342,1.921024,-0.245729\N,-5.244916,1.5
9049,-0.975307\C,-5.199197,2.272944,-2.150223\H,-5.06813,6.199332,0.48
8537\C,-5.380276,6.183244,1.759338\N,-5.716542,4.829433,2.143727\C,-4.
826008,3.912904,2.501626\N,-3.526688,4.196461,2.490493\N,-5.248721,2.6
96711,2.831819\H,-2.900896,-0.991981,-4.194478\C,-1.855935,-1.26775,-4
.766828\N,-2.136299,-2.658927,-5.210582\H,-9.529036,-2.091751,-0.79375
1\C,-8.64998,-2.122794,-0.139664\C,-8.350191,-0.717567,0.410251\O,-9.2
31865,-0.069946,1.074687\O,-7.19521,-0.265449,0.184463\H,-0.911692,-5.
369137,-3.010515\H,-1.638621,-5.022109,-4.679683\H,-2.657797,-3.838311
,-3.076305\H,-0.465156,-4.574,3.007691\H,-2.175262,-5.054631,2.479205\
H,0.99994,-5.317837,0.325691\H,0.816852,-3.260534,-1.003597\H,-3.42537
1,-3.038072,1.528459\H,-4.362638,-1.422954,-0.095029\H,-1.094532,-2.07
051,-2.785614\H,-3.222951,-0.95201,-2.193513\H,5.3784365068,-2.9965940
374,-0.8139798324\H,5.7655398426,-3.6172900134,-2.4124518726\H,6.82712
08781,-1.4154864054,-3.0158281543\H,7.7682281342,0.6452583362,-2.17907
11877\H,5.942131,1.287094,2.596579\H,5.548844,1.762431,0.848982\H,4.80
5618,4.697071,2.719686\H,4.485488,3.201459,3.29164\H,4.344558,1.944813
,-0.402462\H,3.563833,0.810829,-1.643235\H,3.07313,4.177756,-0.037952\
H,1.292417,0.742195,-1.799261\H,0.832141,5.023849,0.546668\H,-0.910231
,1.5872,-1.173309\H,-1.402852,4.683886,0.251287\H,3.714134,2.738978,-3
.162094\H,4.847327,4.970031,-3.361331\H,-2.750245,4.589244,-2.721669\H
,-3.053807,3.289751,-4.007689\H,-5.93724,2.249486,-2.924573\H,-5.96223
2,0.950902,-0.698737\H,-3.924111,1.555666,0.731897\H,-4.57371,6.485775
,2.393975\H,-6.297542,6.734079,1.760491\H,-6.681701,4.567713,2.139439\
H,-4.590032,1.99514,3.103724\H,-6.22545,2.483292,2.808357\H,-3.215516,
5.104844,2.211209\H,-2.860855,3.501284,2.761325\H,-1.045617,-1.313542,
-4.069488\H,-1.350124,-3.011866,-5.717778\H,-2.303903,-3.236249,-4.411
495\H,-2.943477,-2.664225,-5.800861\H,-7.763046,-2.430635,-0.653003\H,
-9.011074,-2.690712,0.69216\H,-1.674144,-0.317874,-5.254811\C,1.718314
6255,-2.9771995396,3.7804222337\C,0.6925355977,-2.5404560808,4.6349499
798\C,-0.2938701291,-1.6930202504,4.1671514094\C,-0.2465335439,-1.2678
95571,2.8188161476\C,0.8202525837,-1.7161250795,1.9841351604\C,1.80165
59411,-2.5761511878,2.4478590865\N,2.7227906974,-3.8882572548,4.307964
0748\H,0.685455652,-2.8907575199,5.6609372055\H,-1.0988685364,-1.36247
2644,4.8148307676\O,-1.092227382,-0.4896231435,2.2289644963\C,0.585795
0147,-1.1209986048,0.6764189263\H,2.5940088183,-2.9329749824,1.7973947
544\N,-0.4369990036,-0.408664658,0.6169982313\O,3.5898454096,-4.295227
7243,3.5490325345\O,2.6488564468,-4.207497046,5.4889334347\H,1.4488385
876,-1.2732785106,-0.6398718444\C,4.176352164,-2.1231040737,-2.3736060
616\C,3.0292058264,-2.2566784509,-1.3838360294\O,2.0769779133,-1.35676
9692,-1.5475845882\H,4.275221232,-1.0624783691,-2.6298354834\H,3.85122
61441,-2.6278389746,-3.2932866304\O,2.9700212592,-3.154573623,-0.55933
99675\H,-2.024027,-6.35491,-3.67959\H,-4.120217,4.710122,-3.382254\H,7
.5865779291,-2.1547733074,-1.6057642026\H,7.079727,1.533057,1.592423\\
Version=EM64L-G09RevC.01\State=1-A\HF=-3196.9662348\RMSD=7.375e-09\RMS
F=4.717e-05\Dipole=-0.4197693,5.2118562,-7.9219935\Quadrupole=-128.975
8577,43.6528751,85.3229826,-48.0473019,-28.3492052,74.2216023\PG=C01 [
X(C42H64N12O12)]\\@
```