

Exploring the emergence of pesticide degradation in the α/β hydrolase superfamily

Juan David F. Schnettler Fernández

Darwin College &
Department of Biochemistry,
University of Cambridge



Under the supervision of Prof. Florian Hollfelder

This thesis is submitted for the degree of
Doctor of Philosophy

May 2021

To Lorenz and Vitus

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared below. It is not substantially the same as any work that has already been submitted before for any degree or other qualification except as declared in the preface and specified in the text. It does not exceed the prescribed limit of 60 000 words specified by the Degree Committee for the Faculty of Biology.

Collaborations and external contributions:

- Dr Tomasz S. Kamiński designed the layouts and manufactured the master wafers for the microfluidic chips used in chapters 2 and 5 in the clean room of the Cambridge Nanoscience Centre.
- Dr Eleanor Campbell set up crystallisation trials and obtained the X-ray structures of the P91 homologues discussed in Chapters 4, 5, and 6.
- Oskar James Klein synthesised substrates 5–10 used for the linear free-energy relationships in Chapters 3 and 4 and performed the specificity kinetics in Chapter 3 in the context of a summer project.
- Alfie Elliott-Anderson screened point mutants and small libraries of *PkDLH* in the context of a Part II project, which contributed to the design of the large *PkDLH* library screened in Chapter 5.
- Rashid Khashiev designed and screened the loop-grafting library of *HsDLH* in Chapter 6 in the context of a Part III project.

Summary

J. David F. Schnettler Fernández:

Exploring the emergence of pesticide degradation in the α/β hydrolase superfamily

How do new enzymatic motifs emerge in evolution? Charting new mechanistic solutions for biocatalytic challenges is a core obstacle in the constant evolutionary adaptation of living organisms. If understood and harnessed, these processes hold the potential to revolutionise industrial processes in the chemical and pharmaceutical industry.

The recent massive release of new, man-made substance classes into the environment in the form of pesticides provides a dynamic testing ground to observe protein evolution in fast-forward mode. In this global experiment, we can observe the emergence of new enzymatic activities and compare the adaptability of different catalytic motifs. Among the xenobiotic substances used, phosphotriester pesticides range among the most toxic. Phosphotriester pesticides were designed to irreversibly inhibit the catalytic triad of synaptic acetylcholinesterase. Numerous enzymes from different evolutionary origins have rapidly evolved to hydrolyse phosphotriesters since. Interestingly, all known enzymes have convergently reached the same mechanistic solution by recruiting divalent cations as a cofactor for metal-ion catalysis.

The starting point of this study is the α/β hydrolase P91. This enzyme was recently identified in a functional metagenomic screening and displays weak promiscuous phosphotriesterase (PTE) activity. P91's PTE activity is remarkable because, in contrast to other currently known PTEs, it utilises a fundamentally different, metal-independent mechanism. At the same time, it has the same fold and a very similar active site configuration as acetylcholinesterase, the biochemical target of most organophosphate pesticides and nerve gases.

Summary

In this dissertation, I subject P91 to directed evolution in order to explore the evolvability of its catalytic motif to this new enzymatic activity (chapter 2). To be able to cover otherwise inaccessible sequence space, I use droplet microfluidics to screen large libraries of enzyme variants in high throughput. Within only two rounds of evolution, P91's phosphotriesterase activity could be increased by 400-fold. The improved enzyme displays a catalytic efficiency close to $10^6 \text{ M}^{-1}\text{s}^{-1}$ and a turnover rate of $> 10 \text{ s}^{-1}$. This activity is comparable to the catalytic efficiencies of many metal-dependent 'conventional' phosphotriesterases. These results highlight that enzymes can be evolved to hydrolyse phosphotriester pesticides using a catalytic Cys–His–Asp triad, independent of metal co-factors.

Detailed mechanistic analysis reveals that P91 is pre-disposed for the turnover of organophosphates due to the intrinsic suitability of its cysteine nucleophile (chapter 3). This residue is able to quickly break down the otherwise unresolvable intermediate providing the initial foothold for rounds of directed evolution to improve upon the rate-limiting initial formation of the covalent intermediate.

Further phylogenetic exploration of P91 homologues reveals that this promiscuous activity is widespread within the diene lactone hydrolase protein family, a cluster of proteins in the α/β hydrolase superfamily, which possess a Cys–His–Asp catalytic triad (chapter 4). By detailed structural analysis and targeted engineering I show that the determinant for the level of displayed PTE activity is not a specific arrangement of specificity-determining residues in the active site (chapter 5), but rather the length and flexibility of loops surrounding the active site (chapter 6). In summary, these results highlight the enormous potential for synergy by combining focussed, combinatorial libraries with high-throughput droplet microfluidics. In combination, these techniques can establish new mechanistic tracks and find efficient biocatalytic solutions that – despite strong natural selective pressure – were previously unseen in Nature.

Acknowledgements

I would like to thank my supervisor Florian Hollfelder for giving me the opportunity to work on this thesis and granting me freedom along the way, while providing support in all life situations. Thank you, Florian, for creating such a friendly, creative, and supportive lab atmosphere.

Throughout this PhD, I had wonderful co-workers and collaborators: I thank Pierre-Yves Colin, who introduced me to the lab and originally conceived the idea of evolving the enzyme P91. Tomek Kaminski was an immense help and quick problem solver in all microfluidics-related challenges. Eleanor Campbell performed magic in crystallising so many of the proteins that crossed our way. In this context, I would also like to thank Diamond Light Source for beamtime (proposal mx18548), and the staff of beamlines I03, I04, and I24 for assistance with crystal testing and data collection. Jamie Klein did amazing work in the fume hood and gave brilliant input regarding the mechanistic questions discussed in this work. Alfie Elliott-Anderson and Rashid Khashiev were great project students and a big support in conceptualising, cloning, and screening several of the libraries discussed in this work. A special thanks also goes to my collaborators at Princeton University, Michael Hecht and Mike Wang. The fruits of our collaboration did not go into this thesis but stimulated much of the thought behind it.

I owe a big thank you to all the people who helped to proof-read this dissertation and gave me valuable feedback (in alphabetical order): Godwin Aleku, Lise Boursinhac, Katrin Fischer, Maximilian Gantz, Josephin Holstein, Wolfgang Koch, Stefanie

Summary

Neun, and Cyril Statzer.

I would like to thank the Gates Cambridge Trust for funding my PhD studies, in particular for their family allowance and support of parental leave. I am also indebted to the German National Academic Foundation (Studienstiftung des deutschen Volkes) who co-funded the first year of my PhD and whose scholarly support and inspiration throughout my studies originally brought me on track to pursue this project.

Everything I learned during this time I learned from or with somebody. It is very hard to do justice to all the help, encouragement, and friendship I have received from members of the Hollfelder lab. To everybody in the lab, thank you so much for the time spent together!

Parenting while pursuing a PhD can feel challenging and the company of other parents is the best remedy. I want to thank all our friends and fellow parents with their respective hordes of children, especially from the Fawcett Court community, for having sweetened our time in Cambridge.

This thesis would not have gotten so far without the constant support of my parents-in-law Elisabeth and Thomas, and my sister-in-law Lucia, especially during lockdown.

Throughout all these years, I was sustained by the support and unconditional love from my parents Bernt and Jacqueline and my siblings Amanda, Alba, and Leo. Muchas gracias por vuestro amor.

My biggest thanks go to Agnes, my friend, lover, and partner in crime. Thank you for enduring, sustaining and cheering me up during this time in Cambridge. Thank you for being there, all three of you, Agnes, Lorenz, Vitus. Thank you for all the adventures along the way – this is only the beginning!

Contents

Summary	iii
List of Figures	xi
List of Tables	xv
List of Abbreviations	xvii
1 Introduction	1
1.1 Xenobiotics as an evolutionary challenge	2
1.2 Organophosphate-degrading enzymes	5
1.2.1 Discovery	5
1.2.2 Evolutionary origins	6
1.2.3 Models for protein evolution	7
1.2.4 Potential applications	9
1.3 Navigating sequence space	11
1.3.1 Protein design	11
1.3.2 Directed evolution	11
1.3.3 Upscaling screening throughput	12
1.3.4 <i>In vitro</i> compartmentalisation	12
1.3.5 Droplet microfluidics	13
1.4 New mechanistic terrain	15
1.4.1 P91 – an unusual phosphotriesterase	15
1.4.2 The α/β hydrolase fold	16
1.4.3 Adaptability of different catalytic motifs	19

Contents

1.5	Structure and aims of this thesis	22
2	Directed evolution of the promiscuous phosphotriesterase P91	25
2.1	Introduction	26
2.2	Results and discussion	26
2.2.1	Mutational scanning	26
2.2.2	Droplet screening identifies highly improved variants	28
2.2.3	An evolved P91 variant rivals the efficiencies of metal-dependent phosphotriesterases	36
2.3	Conclusion and outlook	41
3	The catalytic mechanism of P91	43
3.1	Introduction	44
3.2	Results and discussion	44
3.2.1	Intermediate breakdown is not rate-limiting	44
3.2.2	Nucleophile exchange	45
3.2.3	Specificity analysis	48
3.2.4	Linear free-energy relationships	49
3.2.5	Speculations on the molecular basis of k_2 acceleration	53
3.3	Conclusion and outlook	55
4	Phylogenetic exploration of P91 homologues	59
4.1	Introduction	60
4.1.1	The dienelactone hydrolase (DLH) protein family	60
4.2	Results and Discussion	61
4.2.1	Exploration of P91's surrounding sequence space	61
4.2.2	Sampling the transition-state adaptation	74
4.3	Conclusion and outlook	77
5	In search of specificity-determining positions	79
5.1	Introduction	80
5.2	Results and discussion	80
5.2.1	The extended hexad	80

5.2.2	Experimental determination of specificity-determining positions in <i>PkDLH</i>	82
5.3	Conclusion and outlook	88
6	Proteinometrics	91
6.1	Introduction	92
6.2	Results and discussion	92
6.2.1	Structural comparison	92
6.2.2	P91 is mutationally robust to a large loop insertion	98
6.2.3	Grafting a P91 loop into <i>HsDLH</i> increases its phosphotriesterase activity	100
6.3	Conclusion and outlook	105
7	Final conclusions and perspective	107
8	Materials and Methods	113
8.1	Chemicals and materials	114
8.2	Synthesis of LFER substrates	114
8.3	Cloning	116
8.3.1	Single-site mutagenesis	116
8.3.2	Multiple-site saturation libraries	117
8.4	Microfluidic droplet screening	119
8.4.1	Design and fabrication of master wafers for microfluidic devices	119
8.4.2	Fabrication of microfluidic devices	120
8.4.3	Preparation of cells for encapsulation	121
8.4.4	Compartmentalisation of cells into microdroplets	121
8.4.5	Fluorescence-assisted droplet sorting (FADS)	122
8.4.6	DNA recovery from microdroplets	123
8.5	Microtiter plate screening	124
8.6	Protein expression	124
8.7	Protein purification	125
8.7.1	Strep-Tactin affinity chromatography	126
8.7.2	Ni-NTA affinity chromatography	126

Contents

8.7.3 Buffer exchange and protein quantification	127
8.8 Kinetics	127
8.8.1 Steady-state kinetics	127
8.8.2 Transient-state kinetics	129
8.9 Multiple sequence alignments and phylogenetic trees	131
8.10 Sequence similarity networks	131
8.11 Protein crystallisation and X-ray structure determination	132
8.11.1 Protein crystallisation	132
8.11.2 Data collection and structure solution	132
8.12 Structural models	133
8.13 B-factor analysis	133
9 References	135
10 Appendix	167
10.1 Amino acid sequences of studied proteins	167
10.2 Pre-steady state kinetics	172
10.3 Steady-state kinetics	176
10.4 Crystal structures: Data collection details and refinement statistics . .	187

List of Figures

1.1	Overview of organophosphate structures	3
1.2	Mechanism of action of organophosphate pesticides	4
1.3	Organophosphate hydrolase activity has convergently evolved in different protein folds	6
1.4	Droplet microfluidics for high-throughput enzyme screening	14
1.5	Active site comparison between <i>BdPTE</i> and P91	16
1.6	Schematic topology of the α/β hydrolase fold	18
1.7	Comparison of ground and transition state geometries of carboxyesters and phosphoesters	20
2.1	Hydrolysis of FDDEP	27
2.2	Structure-guided mutational active-site scanning	28
2.3	Microfluidic chips designs for droplet generation and sorting with off-line droplet incubation	29
2.4	Library strategy for the directed evolution of P91	30
2.5	Microfluidic droplet screening assay	31
2.6	Screening of the multiple site-saturation library P91-A in microfluidic droplets by FADS	32
2.7	Design of microfluidic chips for on-chip droplet incubation	34
2.8	On-chip fluorescence measurements	35
2.9	P92-R2 rivals the efficiencies of engineered and naturally evolved metal-dependent phosphotriesterases	37
3.1	Overview of the assumed reaction scheme and the transition states	45

List of Figures

3.2	Wild-type P91 does not show biphasic burst kinetics	46
3.3	The largest effect of evolution is on intermediate formation	47
3.4	Overview of substrates	48
3.5	Comparison of reaction type specificity and leaving group preference	50
3.6	Linear free-energy relationships in enzymatic catalysis	51
3.7	Phosphotriester substrates used to construct the linear free-energy relationships	52
3.8	Brønsted analysis shows that the evolved variant P91-R2 accelerates intermediate formation by improved leaving group stabilisation	53
3.9	Active site comparison of wild-type and evolved P91	55
4.1	Sequence similarity networks (SSNs) of the α/β hydrolase superfamily and the dienelactone hydrolase protein family	62
4.2	Genomic context of P91	65
4.3	Reported substrates of DLH family proteins	67
4.4	Phosphotriesterase activity is widespread in the DLH protein family .	71
4.5	Substrates used to measure the promiscuity profiles of P91 and its DLH family homologues	75
4.6	Promiscuity profiles of the studied DLH family proteins	76
4.7	Brønsted analysis shows no differences in leaving group stabilisation for phosphotriesterase across non-evolved P91 homologues	77
5.1	Structural alignment of the active sites of all characterised DLH family proteins	83
5.2	Rationale and library design for the functional conversion of <i>PkDLH</i> into a phosphotriesterase	85
5.3	Library cloning strategy for the assembly of the binary <i>PkDLH</i> library	86
5.4	Enrichment analysis of the binary <i>PkDLH</i> library across droplet screening rounds	89
6.1	Structural comparison of P91 and <i>PkDLH</i>	94
6.2	Correlation of loop lengths and active site volumes with phosphotriesterase activity	95

6.3	B-factor comparison across DLH family protein structures	97
6.4	Average B-factors of cavity-forming loop regions correlate with phosphotriesterase activity	98
6.5	A P91 variant with a 18-residue loop insertion in the cavity-forming domain	99
6.6	Expression optimisation of <i>HsDLH</i>	101
6.7	Loop grafting increases the phosphotriesterase activity of <i>HsDLH</i> . .	104
8.1	Synthesis of substrates for linear free-energy relationship measurements	114
10.1	Stopped-flow reaction traces of P91 C118S variants with FDDEP . .	172
10.2	Determination of the phosphorylation rate k_2	173
10.4	Determination of the de-phosphorylation rate k_3	175
10.5	Michaelis-Menten plots of esterase and phosphotriesterase kinetics of P91-WT, P91-R1, and P91-R2	176
10.6	Michaelis-Menten plots for steady-state kinetics of P91-WT with linear-free energy relationship substrates	178
10.7	Michaelis-Menten plots for steady-state kinetics of P91-R2 with linear-free energy relationship substrates	179
10.8	Michaelis-Menten plots for steady-state kinetics of <i>ScDLH</i> with linear-free energy relationship substrates	180
10.9	Michaelis-Menten plots for steady-state kinetics of <i>m3DLH</i> with linear-free energy relationship substrates	181
10.10	Michaelis-Menten plots for steady-state kinetics of DLH family proteins with paraoxon-ethyl	182
10.11	Michaelis-Menten plots for steady-state kinetics of DLH family proteins with fluorescein di(diethylphosphate)	183
10.12	Michaelis-Menten plots for steady-state kinetics of DLH family proteins with <i>p</i> -nitrophenyl butyrate	184
10.13	Michaelis-Menten plots for steady-state kinetics of DLH family proteins with dihydrocoumarin	186

List of Tables

2.1	Steady-state catalytic parameters for phosphotriester hydrolysis by P91 variants	36
2.2	Comparison of catalytic efficiencies of phosphotriesterases	40
3.1	Microscopic rate constants	47
3.2	Properties of LFER substrates	52
4.1	Overview of all characterised P91 homologues	72
8.1	Summary of fragments combinatorally grafted from P91 and its homologues into <i>HsDLH</i>	119
10.1	Esterase and phosphotriesterase kinetic parameters of P91-WT, P91-R1, and P91-R2	177
10.2	Steady-state kinetic parameters of P91-WT with linear-free energy relationship substrates	178
10.3	Steady-state kinetic parameters of P91-R2 with linear-free energy relationship substrates	179
10.4	Steady-state kinetic parameters of <i>ScDLH</i> with linear-free energy relationship substrates	180
10.5	Steady-state kinetic parameters of <i>m3DLH</i> with linear-free energy relationship substrates	181
10.6	Paraoxonase activities of studied DLH family proteins	182
10.7	FDDEP activities of studied DLH family proteins	183
10.8	Esterase activities of studied DLH family proteins	185

List of Tables

10.9 Lactonase activities of studied DLH family proteins	186
10.10 Data collection details and refinement statistics for crystal structures of <i>m3DLH</i> , <i>ScDLH</i> , and <i>SjDLH</i>	187

List of Abbreviations

AADS	Absorbance-activated droplet sorting
ABH	α/β hydrolase superfamily
AChE	Acetylcholinesterase
AH	Amidohydrolase superfamily
AU	Arbitrary units
BChE	Butyrylcholinesterase
bp	base pairs
BP	β -Propeller superfamily
DLH	Dienelactone hydrolase
<i>E. coli</i>	<i>Escherichia coli</i>
FACS	Fluorescence-assisted cell sorting
FADS	Fluorescence-assisted droplet sorting
FDDEP	Fluorescein di(diethylphosphate)
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	High-performance liquid chromatography
MBL	Metallo- β -hydrolase superfamily
MOPS	3-(<i>N</i> -morpholino)propanesulfonic acid
MPD	2-Methyl-2,4-pentanediol
MPH	Methyl parathion hydrolase
OD	Optical density
OP	Organophosphates
PCR	Polymerase chain reaction
PDMS	Poly(dimethyl)siloxane
PEG	Polyethylene glycol

List of Abbreviations

PLL	Phosphotriesterase-like lactonase
PMSF	phenylmethanesulfonyl fluoride
PNP	<i>para</i> -Nitrophenol
PTE	Phosphotriesterase
PXN	Paraoxon-ethyl
RFU	Relative fluorescence units
rpm	Revolutions per minute
SSN	Sequence similarity network
SUMO	Small ubiquitin-like modifier
TEV	Tobacco etch virus
Tris	Tris(hydroxymethyl)aminomethane
TS	Transition state
WT	Wild-type

1 Introduction

How do enzymes acquire new functions? In this chapter, I will lay out why organophosphate hydrolysis is an excellent model to study the evolution of new enzymatic function and highlight the potential of droplet microfluidics to navigate sequence space and screen enzymes for new activities. Furthermore, I will introduce the main questions addressed in this thesis and the context in which they are embedded.

1.1 Xenobiotics as an evolutionary challenge

Large-scale pollution of nearly all ecosystems on Earth with anthropogenic substances is the signature of the current geological epoch, which has accordingly been proposed to be named the Anthropocene [1]. The exact definition and beginning of this new era are subject to fierce debate. Propositions for watershed moments to define its beginning range from the onset of human landscaping of entire ecosystems by slash-and-burn agriculture [2] to the first nuclear weapons tests, which deposited a globally traceable, permanent layer of radionuclides around the globe [3].

Among the most pervasive and toxic substances of human origin, whose presence in the environment demarcates the Anthropocene, are organophosphates (Figure 1.1). Most classes of organophosphates, such as phosphotriesters, are xenobiotic and entirely man-made. Originally discovered in Germany in the 1930s as potent nerve agents and insecticides, they were further developed into the chemical warfare agents tabun and sarin in the dawn of World War II [4]. Never employed in the war, the less potent analogues of these agents, mostly phosphotriesters, were repurposed and commercially launched as insecticides from 1947 [5], starting their discharge into the environment on a massive global scale. A further source of organophosphates in the environment are triaryl phosphates, which are widely used as plasticisers and flame retardants in plastics, construction materials, textiles, and lubricants [6, 7]. Widely shed by airplane and shipping emissions, airborne dust particles, and sewage discharge, they can be detected in nearly all environments on earth, including the polar ice caps [8, 9], marine sediments [10], indoor dust [11] and human urine samples [12].

The high acute toxicity of organophosphates is due to their covalent and irreversible inhibition of the catalytic triad that many enzymes display as their catalytic motif. The phosphorus atom acts as a strong electrophile and, with its tetrahedral ground state mimicking the tightly bound transition state of ester hydrolysis, is attacked by the catalytic nucleophile of the triad, mostly a serine (Figure 1.2). The formed

1.1 Xenobiotics as an evolutionary challenge

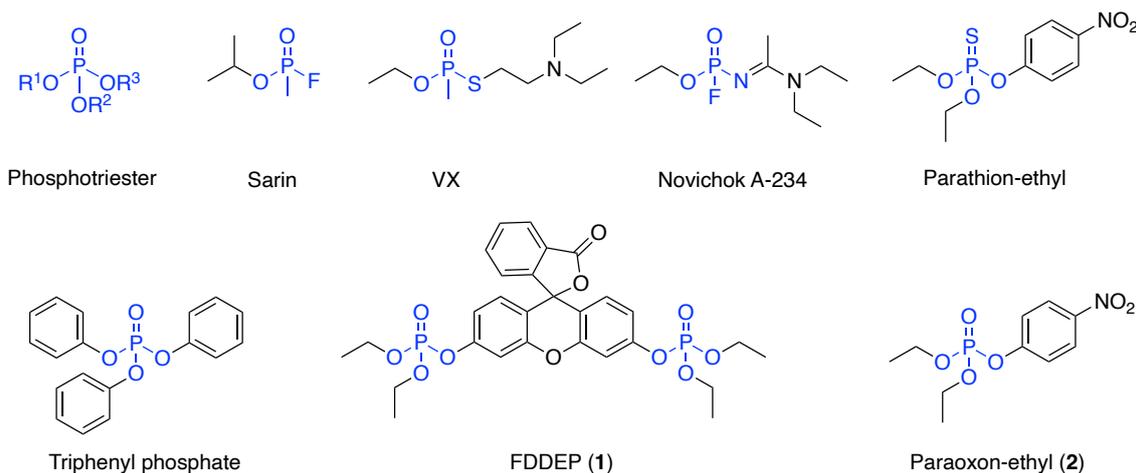


Figure 1.1: Overview of organophosphate structures. The organophosphate core structure is highlighted in blue. **Phosphotriesters** are phosphates with three ester groups. To act as acetylcholine inhibitor, at least one of the ester side groups is usually a good leaving group. **Sarin**, **VX**, and **Novichok** are examples of chemical warfare agents. **Parathion-ethyl** was the first commercialised insecticide. **Triphenyl phosphate** is an example for a triaryl phosphate, which are widely used as plasticisers and flame retardants. Fluorescein di(diethylphosphate) (**FDDEP**, **1**) is the fluorogenic model phosphotriester used for directed evolution in this study. **Paraoxon-ethyl** (**2**), also used in this study, is the actual toxic agent produced from parathion-ethyl by de-sulfurating native oxidases upon uptake.

phosphoryl intermediate is again tetrahedral (and therefore stabilised by enzymes that are adapted to tightly bind tetrahedral carbonyl transition states, see also Figure 1.7). It can only be hydrolysed off at very low rates, if at all, effectively inhibiting the catalytic triad motif irreversibly. Many enzymes are additionally prone to a side reaction called ageing in which one of the side groups of the organophosphate is hydrolysed off, leaving a negatively charged, hydrolysis-resistant intermediate behind (Figure 1.2). Although the majority of triad-bearing enzymes in the human body have been shown to bind organophosphates [13] the acute toxicity of these molecules is due to the inhibition of acetylcholinesterase (AChE). The primary physiological role of AChE is the termination of neurotransmission by rapid degradation of the neurotransmitter acetylcholine in the synaptic cleft. The high speed at which successive neurotransmission events need to happen at a synapse has put an immense evolutionary pressure for high turnover on this enzyme, making it one of the fastest enzymes known, acting at the diffusion limit [14]. Its inhibition therefore instantly derails synaptic transmission, leading to paralysis by seizures and subsequent death by respiratory arrest [15]. Furthermore, organophosphates also bind to

1 Introduction

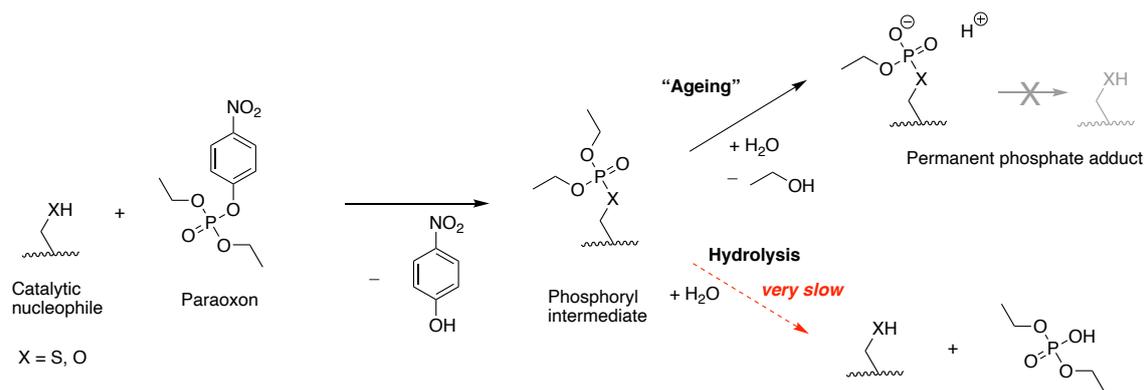


Figure 1.2: Mechanism of action of organophosphate pesticides. Organophosphate pesticides, here shown using paraoxon-ethyl as an example, near-irreversibly inhibit the catalytic triad of acetylcholinesterase and other enzymes by covalently binding to the catalytic nucleophile (mostly a serine). The phosphorylated intermediate is very stable and is only phosphorylated off again at negligible rates in most enzymes. Alternatively however, hydrolysis can lead to the de-alkylation of one of the alkoxy side chains of the phosphate, which, instead of removing the phosphate from the enzyme, positions a charge on the phosphate, making it highly resistant to further hydrolysis. This process is referred to as ‘ageing’.

a multitude of other physiological proteins that contribute to the long-term damage and secondary toxic effects of organophosphate poisoning [13].

Due to their high acute toxicity, phosphotriester pesticides are predominantly supplied as their corresponding phosphorothioate (e.g. parathion-ethyl), which is more stable and less toxic. The replacement of the P=O double bond with a P=S double bond strongly reduces the electrophilicity of the phosphorus atom and therefore its propensity to inhibit acetylcholinesterase. Upon uptake, phosphorothioates are desulfurated by native cytochrome P450 oxidases (e.g. CYP3A4) and converted into their active oxon form (e.g. paraoxon-ethyl) [16]. This oxidation is faster in insects than in vertebrates, forming the basis for the partly selective action of organophosphate pesticides.

1.2 Organophosphate-degrading enzymes

1.2.1 Discovery

The appearance of new, biologically active xenobiotic substances in the environment confronts organisms with a fundamental and ongoing challenge: How to adapt to new biocatalytic needs and degrade or exploit these substances? Therefore, the introduction of organophosphates into Nature provides a fertile testing ground to study the emergence of new catalytic function.

Only two decades after the market launch of the first organophosphate pesticide parathion, it was reported that the degradation of organophosphates in the soil is greatly accelerated by microorganisms [17]. Soon, bacteria were isolated which could hydrolytically cleave phosphotriesters and grow on such pesticides as a sole carbon and phosphate source [18–21]. This search culminated in the identification of an enzyme, phosphotriesterase from *Brevundimonas diminuta* (*BdPTE*), which is capable of hydrolysing the phosphotriester paraoxon and a range of other organophosphates [22–24]. Surprisingly, despite the very recent introduction of its substrate into Nature, only decades before, *BdPTE* displays very fast, diffusion-limited kinetics with catalytic efficiencies in the range of $\approx 10^8 \text{ M}^{-1}\text{s}^{-1}$, thus having evolved within shortest time to catalytic perfection [25].

Many additional organophosphate-degrading enzymes have been identified since. Strikingly, these newly evolved organophosphate hydrolases have arisen from diverse protein superfamilies with different fold topologies and evolutionary origins, such as the amidohydrolases [24], the pita-bread fold [26], the β -propellers [27, 28], and the metallo- β -lactamases [29] (Figure 1.3). In an astonishing showcase of rapid convergent evolution, all these enzymes have, despite their entirely different structural scaffolds, independently come to very similar mechanistic solutions, recruiting divalent cations as a cofactor for metal-ion catalysis [30] (Figure 1.5a). The naturally occurring metal cations in the active site include Zn^{2+} , Fe^{2+} , Cd^{2+} , Mn^{2+} , and Ca^{2+} ,

1 Introduction

but also Co^{2+} , Ni^{2+} , Mg^{2+} , and Ba^{2+} can be accepted [31]. Overall, metal-ion catalysis has emerged as a potent catalytic motif to address this new catalytic challenge.

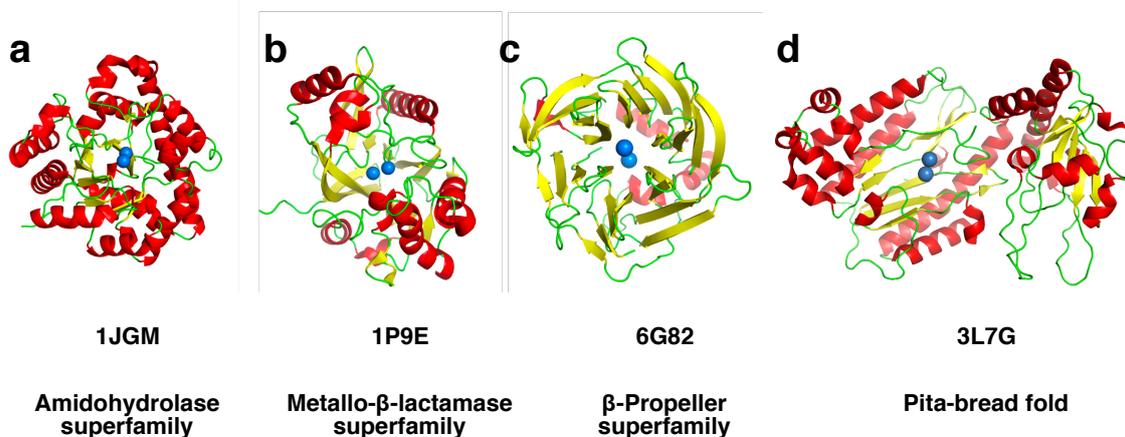


Figure 1.3: Organophosphate hydrolase activity has convergently evolved in different protein folds. Examples of organophosphate-degrading enzymes displaying different fold topologies. Note that all enzymes require divalent metal cations (blue spheres) in their active site for catalytic activity. **(a)** Phosphotriesterase (*BdPTE*) from *Brevundimonas diminuta* (PDB ID: 1JGM), a member of the amidohydrolase superfamily (also referred to as TIM barrel fold, Pfam: CL0036). **(b)** Methyl parathion hydrolase (MPH) from *Pseudomonas* sp. strain WBC-3 (PDB ID: 1P9E), a member of the metallo- β -lactamase superfamily (Pfam: PF00753). **(c)** Human Paraoxonase-1 (PON1) (PDB ID: 6G82), a member of the β -propeller superfamily (Pfam: CL0186), shows promiscuous phosphotriesterase activity but has presumably not evolved towards it. **(d)** Organophosphate anhydrolase/prolidase (OPAA) from *Alteromonas* sp. strain JD6.5 OPAA, (PDB ID: 3L7G) a protein displaying the 'pita-bread fold' (Pfam: PF00557).

1.2.2 Evolutionary origins

The rapid evolution of organophosphate-hydrolysing enzymes has since become a textbook example for fast acquisition of new enzymatic function. Their origin has become the subject of considerable research and consequently, phosphotriesterases have become some of the most studied and engineered proteins. Tawfik and co-workers have shown that *BdPTE* has presumably evolved from a closely homologous clade of quorum-quenching lactonases, which have accordingly been named phosphotriesterase-like lactonases (PLLs) and display low promiscuous phosphotriesterase activity [32]. Further studies revealed that PTE and PLLs can be interconverted with few mutational steps on the basis of their reciprocal promiscuity: *BdPTE* has been engineered into a proficient lactonase by deletion of a small

loop, creating a generalist enzyme with high phosphotriester and lactonase activities [33]. Conversely, PLLs from the amidohydrolase superfamily have been engineered into respectable phosphotriesterases in several instances [34–37]. An analogous adaptive evolution campaign toward phosphotriesterase activity has been done for a weakly promiscuous homologue of methyl parathion hydrolase (MPH) in the metallo- β -lactamase superfamily [38].

1.2.3 Models for protein evolution

The study of functional transitions in phosphotriesterases has been used to establish much of our current knowledge on the principles governing protein evolution.

In 1976, Roy Jensen postulated [39] that functional promiscuity (originally termed ‘substrate ambiguity’), i.e. activity towards non-native substrates which is not the result of adaptive evolution [40, 41], can give rise to new enzyme activities by providing a bridgehead towards new function. This notion, further articulated by Daniel Herschlag [40], has since been convincingly illustrated by the adaptive laboratory evolution of phosphotriesterase activity from promiscuous precursors [34–37, 42].

Laboratory evolution of the human promiscuous phosphotriesterase PON1 (from the β -propeller superfamily) first indicated that promiscuous enzymes can pioneer new activities by acquiring few fitness-endowing mutations without an initial strong negative trade-off on their original function [43]. Complete functional conversion of *BdPTE* into an arylesterase showed that the evolutionary path can pass through a ‘generalist’ intermediate before specialising towards the new function under diminishing returns and a strong negative trade-off [44]. Such ‘generalists’ can provide the starting points for further divergent adaptive evolution following gene duplication. However, generalist are likely selected for by strong evolutionary pressure [45] and indiscriminate improvement towards both native and new activity without a trade-off has also been demonstrated [46].

1 Introduction

The study of a small set of neutral variants of PON1 revealed drastic differences in their adaptive potential towards other promiscuous activities (termed ‘cryptic genetic variation’), proposing neutral drift as a promising approach when screening proteins for new function [47, 48]. Neutral drift of *Bd*PTE showed that the enzyme accumulated more mutations and was more evolvable in the presence of stabilising chaperones, suggesting that stability is a major constraint in protein evolution [49]. This notion has been supported by further studies on other proteins [50, 51]. In summary, mutational robustness allows the acquisition of neutral or deleterious mutations which alter the underlying ‘landscape’ of promiscuous activities, preparing the ground for later adaptive evolution under changing selective pressures [52, 53].

However, mutational robustness is not uniquely tied to stability but can also be conferred by flexibility and even hampered by excessive rigidity. Both order and disorder can promote evolvability [54, 55]. Structural and kinetic analysis of mutants of PTE homologues showed that conformational diversity promotes promiscuity and can therefore be a hallmark of evolvability [56]. Comparison of two closely related PTE homologues showed that catalysis is modulated by mutations remodeling the conformational landscape of the enzyme [57]. Likewise, it was shown that phosphotriesterase activity of an insect carboxylesterase is limited by unfavourable conformational disorganisation [58]. Structural examination of several intermediates of PTE along the evolutionary trajectory from phosphotriesterase to aryylesterase function firmly established that evolution can act through modulation of the conformational landscape of a protein and that protein dynamics is a major contributor to catalysis and specificity [59].

Further constraints of evolution were highlighted by the reverse evolution of an engineered aryylesterase-PTE back to phosphotriesterase activity. This special case of artificial convergent evolution led to functionally equivalent but genotypically incompatible sequences (*wild*PTE and *neo*PTE), demonstrating the ruggedness that fitness landscapes can display [60]. Similarly, the directed evolution of an inferred lactonase ancestor of MPH (from the metallo- β -lactamase superfamily) showed that

adaptive evolution is partly constrained by higher-order epistatic networks that pervade the entire structural scaffold and change in response to subtle changes in substrate polarity [42].

1.2.4 Potential applications for bioremediation and against organophosphate poisoning

Beyond their attraction as models for protein evolution, organophosphate-degrading enzymes also harbour potential applications in the bioremediation of contaminated areas and in the treatment of organophosphate poisoning.

Apart from their use as pesticides, plasticisers and flame retardants, organophosphates have been continuously refined and stockpiled for their use as chemical warfare agents. Despite the Geneva Protocol and the Chemical Weapons Convention, the organophosphate nerve agents sarin, cyclosarin and tabun have confirmedly been used in several instances in the armed conflicts between Iran and Iraq [61], and in Syria [62]. Sarin and VX have also been used in terrorist attacks in Japan [63, 64]. Furthermore, VX and Novichok agents have been used for several recent high-profile assassinations (or attempts), such as of Kim Jong-nam [64], Sergey Skripal [65], and Alexei Navalny [66].

But far beyond such attention-sparking cases, every single day, thousands of patients are admitted to hospital with organophosphate pesticide poisonings, mainly as a consequence of intentional self-poisoning [67]. In rural areas of the Global South, the high accessibility of highly toxic pesticides constitutes a major public health problem, making pesticide intoxication one of the primary means of suicide and causing $\approx 160\,000$ deaths every year [68–70].

The current treatment options for organophosphate poisoning are limited to mitigating primary symptoms using anticholinergics (e.g. atropin), oxime reactivators (e.g. pralidoxime), and anticonvulsants (e.g. diazepam) [15, 67], which are addition-

1 Introduction

ally constrained by a small therapeutic window. Furthermore, the therapeutic value of the most used oxime reactivator pralidoxime against pesticide poisoning has recently been put into question [71–73]. Finally, even if they can alleviate the acute effects of a cholinergic crisis, the currently available measures cannot prevent the onset of severe chronic disease and long-term disability which often afflict survivors of organophosphate poisoning.

Although the most effective measures to address this global problem will constitute regulatory interventions, limiting the access to highly toxic pesticides and banning the most hazardous ones for agricultural use [74], organophosphate-hydrolysing enzymes could represent a potent extension of the current therapeutic repertoire. Further to their use to bioremediate organophosphate-contaminated areas, enzymes have been proposed to be directly used as protective agents *in vivo*, a concept termed ‘catalytic bioscavenger’: Organophosphate-degrading enzymes could prevent intoxication by binding or hydrolysing the agent in the bloodstream before it can reach its target [75–78]. Although animal experiments have validated the concept [79, 80], currently available enzymes do not possess the required catalytic efficiency and stereoselectivity, especially towards warfare agents, that is projected to confer protection against intoxication [81]. In addition, enzymes of bacterial origin, such as the highly efficient phosphotriesterase from *Brevundimonas diminuta*, are limited by tolerability issues, although strategies such as PEGylation have been tested to reduce their immunogenicity [82]. Therefore, considerable effort has been invested in engineering efficient enzymes that directly act on G-series and V-series organophosphate warfare agents, such as sarin, tabun, soman, and VX [81, 83–86]. The development of prophylactic bioscavengers against chemical warfare agents, however, also raises ethical questions: The existence of a protective broad-spectrum antidote could increase the likeliness of actual deployment of nerve agents in armed conflicts – of which the most potent ones have never been used at scale – and fuel an arms race towards even more gruesome warfare agents.

1.3 Navigating sequence space

1.3.1 Protein design

Beyond the case of catalytic bioscavengers, the creation of proteins with new functions is immensely useful for numerous applications in chemical synthesis, medicine, or material science. However, we do not understand proteins down to the molecular detail and designing functional proteins rationally, from first principles, has been largely unsuccessful [87]. This is changing with the advent of computational design frameworks (such as Rosetta [88]) and lately machine learning. Impressive progress has been made in the *de novo* design of, e.g., potential therapeutic binders [89], self-assembling protein cages [90], membrane pores [91, 92], and biosensors [93]. Most success in protein design has been achieved in designing the adoption of a certain structural shape or interaction, often in response to a physical cue such as pH, light, or molecular recognition. In contrast, the design of proficient catalytic proteins has proven much more difficult, and the attained rate accelerations [94–96] range far below those achieved by naturally evolved enzymes [97].

1.3.2 Directed evolution

An alternative, far more successful approach to creating enzymes with new function has been mimicking natural evolution in a process called directed evolution [98, 99]. In this approach, departing from enzymes with pre-existing weak promiscuous starting activities, a library of variants is created by (partial) randomisation and subsequently screened for the desired new activity. However, as the sequence space to be sampled is enormous [100] most mutations are neutral or deleterious to the protein [101], and only few mutational paths may lead to a selective advantage [102]. Typically the examination of many clones and successive rounds of evolution are necessary in order to identify functionally improved variants. While Nature likely ‘tipples’ through evolution, making small mutational steps at a time by sin-

1 Introduction

gle substitutions, small insertions and deletions and recombination, and indulging in long detours by neutral drift [103], the time-pressured human protein engineer cannot afford working on the time scales of natural evolution.

1.3.3 Upscaling screening throughput

To defy the limitations of directed evolution, usually restricted by screening capacity and its associated costs, methods and strategies to navigate sequence space effectively are in need [104]. Conventional screening of enzymatic turnover can be upscaled in the laboratory using miniaturisation and automation of analytic procedures, e.g. HPLC and microtiter plate screening, bringing throughput to 10^3 – 10^4 variants per day, yet, with the associated cost scaling linearly. Far higher throughput can be attained using survival selections. However, selections are limited to enzymes whose activity can be coupled to a survival advantage or selectable trait, such as antibiotic resistance or metabolic function. *In vitro* selection using display platforms such as phage display [105], yeast display [106], ribosome display [107], mRNA display [108], or CIS display [109] overcomes these limitations. In combination with flow cytometry or affinity panning, display technologies can be used to screen a staggering 10^8 – 10^{12} variants in a single experiment and have been applied with overwhelming success to the evolution of therapeutic protein binders [110–113]. However, essentially being capture assays, display methods are largely unsuitable for the evolution of multiple-turnover enzymes.

1.3.4 *In vitro* compartmentalisation

A major breakthrough for enzymatic screening was brought by *in vitro* compartmentalisation [114], the encapsulation of reactions in surfactant-stabilised water-in-oil emulsions. The transcription and translation of genes in small aqueous compartments miniaturises reaction volume while linking phenotype (protein function) to

genotype (encoding gene). Initially carried out in polydisperse emulsions with high variation in droplet size, *in vitro* compartmentalisation can directly be applied to the directed evolution of nucleic-acid modifying enzymes [114–118]. This has been further developed into the concepts of compartmentalised self-replication [119] and compartmentalised self-tagging [120]. In combination with bead display [121, 122] or double emulsification [123–125] and subsequent flow cytometric sorting, *in vitro* compartmentalisation makes enzymatic screening amenable to the ultra-high throughput known from display methods [126]. Using such an approach, the high turnover of *BdPTE* could be further improved 63-fold to a stunning k_{cat} of $\approx 140\,000\text{ s}^{-1}$ [121]. However, the considerable variation in droplet diameters of polydisperse emulsions precludes the precise quantification of output signal needed for high-fidelity screening, an important requirement for evolution campaigns with more incremental improvements.

1.3.5 Droplet microfluidics

The introduction of microfluidic chips, fabricated by soft lithography, allowed the generation of highly monodisperse emulsions for quantitative assays (Figure 1.4a). In a monodisperse emulsion every droplet has the same volume, with very little variation, allowing precise optical readout and a multitude of downstream fluid manipulations [127]. The droplets can be generated at kHz frequency and serve as miniaturised reaction vessels that link phenotype (enzymatic activity, quantified by droplet absorbance or fluorescence) to genotype (gene variant, recovered by DNA extraction after sorting) (Figure 1.4b). Since complemented by fluorescence-activated droplet sorting (FADS, [128]), microfluidic droplet screening is emerging as a powerful tool to leverage the throughput of quantitative enzyme assays by orders of magnitude as compared to conventional microtiter plate screening. In microfluidic droplet sorting, monodisperse droplets are guided through a Y-shaped junction, flowing by default into a waste channel. An excitation laser is focussed on the junction, measuring the fluorescence of every single droplet. If a droplet surpasses a defined fluorescence threshold, it is electrophoretically deviated in passing by an

1 Introduction

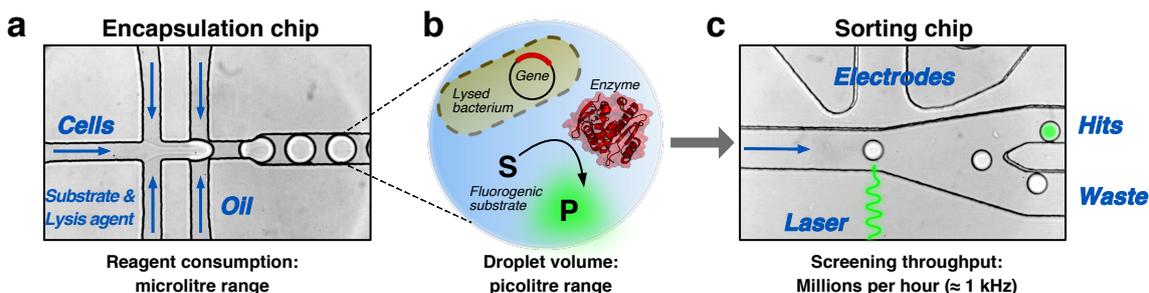


Figure 1.4: Droplet microfluidics for high-throughput enzyme screening. (a) A microfluidic flow-focussing chip, generated by soft lithography, can be used to generate monodisperse emulsions, where every droplet has the exact same diameter, at kHz frequencies. By using multiple input channels, substrate, lysis agent, and cells can be co-encapsulated, requiring only microlitre amounts. (b) The generated water-in-oil droplets are stabilised by a surfactant and serve as minaturised reaction vessels with picolitre volume. A bacterial cell expressing a library variant of an enzyme, can be lysed to release the enzyme and the encoding plasmid. The enzyme can then act on a fluorogenic substrate. The droplet compartment establishes the link between the phenotype (enzyme activity, quantified by fluorescence) with the genotype (plasmid encoding the enzyme variant). (c) After incubation, the droplets can be sorted according to their fluorescence on a separate sorting chip (fluorescence-activated droplet sorting, FADS), at throughputs of millions per hour. To this end, the droplets are guided through a Y-shaped junction with an excitation laser focussed on the droplet flow. When surpassing a certain threshold fluorescence value, a single droplet can be electrophoretically deviated by the electrodes away from the waste channel into the hit collection channel.

electrode into a hit channel from where it can be isolated (Figure 1.4c).

Combining microfluidic droplet generation and sorting, millions of monodisperse droplet compartments with picolitre volumes can quickly be generated, manipulated and sorted on a microfluidic chip, allowing the quantitative sampling of $\approx 10^7$ variants in a single day and with minimal use of reagents. This approach has already been applied to the laboratory evolution of horseradish peroxidase [129], sulfatase [130], amine dehydrogenase [131], aldolase [132, 133], and amine oxidase [134]. Monodisperse droplets can be manipulated on-chip in a multitude of further ways, including merging, splitting, picoinjection [135], and imaging in arrays [136]. Equally, the toolkit of detection modes for droplet sorting is constantly expanding, now also including absorbance-activated [137], and image-based sorting [138], among others [139]. Further droplet manipulations, such as monodisperse water-in-oil-in-water double encapsulation [140] or gel-shell formation [141] make enzymatic reactions amenable to flow cytometric analysis and sorting, and have been used to evolve a polymerase [142] and *BdPTE* [141]. In addition, droplet microflu-

idics is increasingly being applied in further fields such as strain engineering [143], 3D culture of stem cells [136], cancer diagnostics [144], and single-cell transcriptomics [145]. In summary, droplet microfluidics is emerging as a valuable tool for a variety of high-throughput screening applications, especially for the navigation of sequence space in the search for new enzymes [146].

One alternative approach in this quest for new biocatalysts is functional metagenomics: Microbial organisms harbour an immense wealth of genetic diversity, yet, the vast majority of microorganisms is to date unculturable. However, the functional potential of their genes (the ‘metagenome’ [147]) can be mined by extraction of DNA from the environment (e.g. from soil), followed by heterologous expression in a host organism (e.g. *E. coli*), and a functional assay [148]. As the chances of finding functional catalysts with this approach are minuscule, functional metagenomics also requires screening at high throughput. By using microfluidic droplets to screen bacteria, that were expressing environmental DNA, for phosphotriesterase and sulfatase activity, Colin and coworkers could identify new enzymes displaying these rare activities at weakly promiscuous levels – a feat that would not have been possible without the ultrahigh throughput conferred by droplet microfluidics [149].

1.4 New mechanistic terrain

1.4.1 P91 – an unusual phosphotriesterase

Among the new phosphotriesterases identified from this functional metagenomic screening in droplets was a promiscuous esterase called P91 [149]. Surprisingly, P91 achieves phosphotriester hydrolysis in absence of a metal cofactor by the formation and breakup of a covalent intermediate, mediated by a catalytic Cys–His–Asp triad (Figure 1.5b). P91’s ability to hydrolyse phosphotriesters is astonishing, given that phosphotriesters have been specifically designed to covalently inhibit the active-site nucleophile of similar catalytic triads, like the Ser–His–Glu triad of

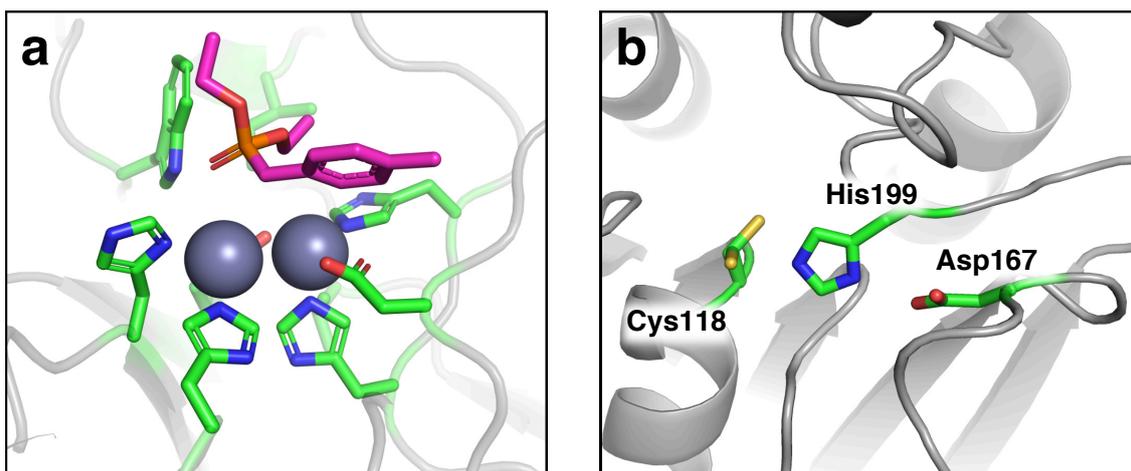


Figure 1.5: Active site comparison between *BdPTE* and P91. (a) Active site of *BdPTE* (PDB ID: 1DPM). Like all known organophosphate-hydrolysing enzymes, *BdPTE* uses metal ions (here Zn^{2+} , dark grey spheres) as a cofactor in the active site to coordinate and deprotonate a hydroxide and stabilise the transition state of the substrate (here a co-crystallised paraoxon analogue, in magenta). (b) Active site of P91 (PDB ID: 4ZI5). In contrast to all known organophosphate-hydrolysing enzymes, P91 can hydrolyse phosphotriesters in absence of a metal cofactor, by covalent catalysis, mediated by a catalytic Cys–His–Asp triad (C118, H199, D167; no co-crystal with paraoxon available). Note that the catalytic nucleophile C118 is present in two conformations.

synaptic acetylcholinesterase. In fact, being a member of the α/β hydrolase superfamily, P91 shares the same structural fold and a similar active site arrangement and is therefore a distant homologue of acetylcholinesterase. This makes P91 the first metal-independent enzyme showing significant turnover of phosphotriesters. This activity, however, is relatively low ($\frac{k_{cat}}{K_M} \approx 10^2 - 10^3 \text{ M}^{-1}\text{s}^{-1}$, depending on the leaving group) and promiscuous, meaning that it has not evolved under selective pressure. Its native substrate is not known but, given that its closest characterised homologue is a lactonase (see Chapter 4) and its highest identified activities are towards an ester and a lactone [149], it is probably an esterase or a lactonase.

1.4.2 The α/β hydrolase fold

P91 is a member of the α/β hydrolase superfamily (Pfam: CL0028), a widespread protein fold encompassing proteins sharing a common structural topology and evolutionary origin while displaying high functional versatility [150, 151]. The canonical fold of the α/β hydrolase superfamily consists of a β -sheet core composed of

eight parallel β -strands in which β -strands 4 and 3 are crossed over and strand 2 is antiparallel to the others (Figure 1.6a). The formed β -sheet is helically twisted. This core is surrounded by several α -helices and has a catalytic triad in topologically conserved positions of the arrangement. One strongly conserved feature of the fold is the ‘nucleophilic elbow’ which contains the nucleophile of the catalytic triad in a sharp turn between β -strand 5 and α -helix D. This sharp turn demands a pair of non-Ramachandran backbone dihedral angles, requiring small side chain residues around the nucleophile (positions -2, +2, and +3), which generates a characteristic primary sequence signature of G–X–S/C–X–G–G (with X being any residue) [150, 152].

The catalytic triad is commonly composed of the residues Ser–His–Asp (most esterases and lipases), Ser–His–Glu (acetylcholinesterase [153]) or Cys–His–Asp (di-enelactone hydrolases, including P91). In rare cases, the catalytic nucleophile can be an aspartate [154–156]. Some non-catalytic members of the α/β hydrolase superfamily have lost their catalytic triad but retain the characteristic fold topology (e.g. the synaptic cell adhesion protein neuroligin [157]). The catalytic triad is accompanied by a further conserved structural feature in the active site, the ‘oxyanion hole’, a positively charged pit which offsets the negative charge accumulating on the oxygen in the transition state. It is usually formed by main chain nitrogen atoms of the residue following the catalytic nucleophile and of a residue in the loop following β -sheet 3 (loop 3), which folds into close proximity of the catalytic nucleophile.

Beyond these core characteristics, proteins belonging to the α/β hydrolase superfamily can be extraordinarily diverse, both structurally and functionally. Although smaller enzymes lacking elements of this canonical fold have been identified [158–160], more commonly, this canonical core is heavily extended by further surrounding structural elements, often conferring substrate specificity by shaping the access to the active site. In acetylcholinesterase, for example, additional loops and helices form a deep and narrow gorge leading towards the catalytic triad. The lining of this gorge with numerous aromatic residues has been proposed to contribute to the extraordinary high catalytic efficiency of this enzyme via a mechanism called ‘aromatic

1 Introduction

guidance' [153]. The active site of many lipases is completely covered by a 'lid' loop domain in solution, getting only exposed upon interaction with a hydrophobic interface [161]. In the extreme case of proline endopeptidases, the pore of an entire fused seven-bladed β -propeller domain ensures that substrate peptide substrates reaching the active machinery do not exceed a certain length [162].

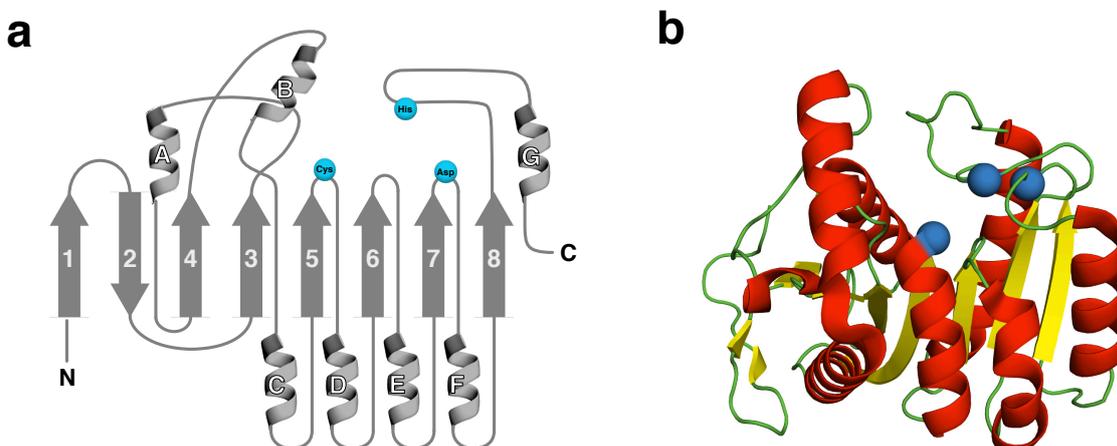


Figure 1.6: Schematic topology of the α/β hydrolase fold, using the example of diene lactone hydrolase from *Pseudomonas knackmussii* (*PkDLH*). **(a)** The canonical fold of the α/β hydrolase superfamily consists of a core of eight parallel beta-sheets (1–8) with a characteristic cross-over of β -sheets 4 and 3, and sheet 2 being antiparallel to the others. This core is surrounded by several α -helices (seven helices A–G in *PkDLH*) and has a catalytic triad (blue spheres) in topologically conserved positions of the arrangement (Cys–His–Asp in *PkDLH*). The nucleophile of the triad is positioned in a sharp turn between beta-sheet 5 and helix D, requiring non-Ramachandran angles and therefore inducing the characteristic sequence signature G–X–S/C–X–G–G. This figure was inspired by reference [163]. **(a)** Structure of diene lactone hydrolase from *Pseudomonas knackmussii* (*PkDLH*, PDB ID: 4U2B), an archetypical α/β hydrolase fold protein representing the canonical fold. The position of the catalytic triad residues C123–H202–D171 (in order of their hydrogen bond interactions) is highlighted in blue (respective α -carbons as blue spheres). Note the helical twist of the β -sheet core (yellow), which is a characteristic of the α/β hydrolase fold.

The functional diversity in the α/β hydrolase superfamily is vast. Mostly centered around hydrolytic reactions, it encompasses hydrolysis of carboxyesters [153], lactones [164], peptides [162], and epoxides [155]. Further catalysed reactions include hydrolytic C–C bond breaking [165], dehalogenation [154], transfer of functional groups [166], and production of light [156]. Organophosphate hydrolase activity, however, is not part of the huge functional repertoire of α/β hydrolase enzymes. On the contrary, as laid out initially, enzymes from the α/β hydrolase superfamily are uniquely susceptible to irreversible inhibition by organophosphates, with acetyl-

cholinesterase being the most prominent, but not the only one [13]. The only known naturally evolved α/β hydrolase that can escape phosphorylation at the catalytic nucleophile at significant rates is the G137D mutant of insect α -carboxylesterase E7 (*Lc α E7 G137D*), a close homologue of the cholinesterases, which evolved under selective pressure of insecticides in *Lucilia cuprina* and other insects [167, 168]. However, this enzyme is mainly characterised by high affinity to pesticides and its turnover rate for phosphotriesters is very low ($k_{cat} \approx 0.0014 \text{ s}^{-1}$) [58]. It therefore serves the fly as a stoichiometric scavenger against pesticides rather than acting as a fast-turnover enzyme. Similarly, human butyrylcholinesterase, a close homologue of synaptic acetylcholinesterase, has been engineered at the corresponding site (G117H) into a multiple-turnover enzyme, but also shows similar, slow turnover ($k_{cat} \approx 0.09 \text{ s}^{-1}$) [169, 170]. Further, it was shown that a bacterial homologue, *p*-nitrobenzyl esterase from *Bacillus subtilis*, bearing the homologous mutation to BChE G117H (*BspNBE A107H*) showed very slow reactivation and could be further evolved to reactivation rates of $\approx 1 \text{ h}^{-1}$ ($\approx 0.0003 \text{ s}^{-1}$) after treatment with paraoxon [171]. A further triple mutant of a snake acetylcholinesterase (from the banded krait *Bungarus fasciatus*) has been reported to display slow organophosphate hydrolysis, albeit with very low catalytic efficiencies in the range of 10^{-1} to $10^1 \text{ M}^{-1}\text{s}^{-1}$, depending on the substrate [172]. In summary, no enzyme in the α/β hydrolase superfamily has yet evolved to escape covalent inactivation by organophosphates at sufficiently high turnover rates to harness this binding interaction for efficient covalent catalysis, despite very high evolutionary pressure.

1.4.3 Adaptability of different catalytic motifs

But why did α/β hydrolases not adapt to the new catalytic challenge of phosphotriester degradation? Is this limitation intrinsic to the α/β hydrolase fold or their covalent catalytic mechanism?

The catalytic power of enzymes is staggering and mainly conferred by a strong discrimination between ground state and transition state of their catalysed reaction

1 Introduction

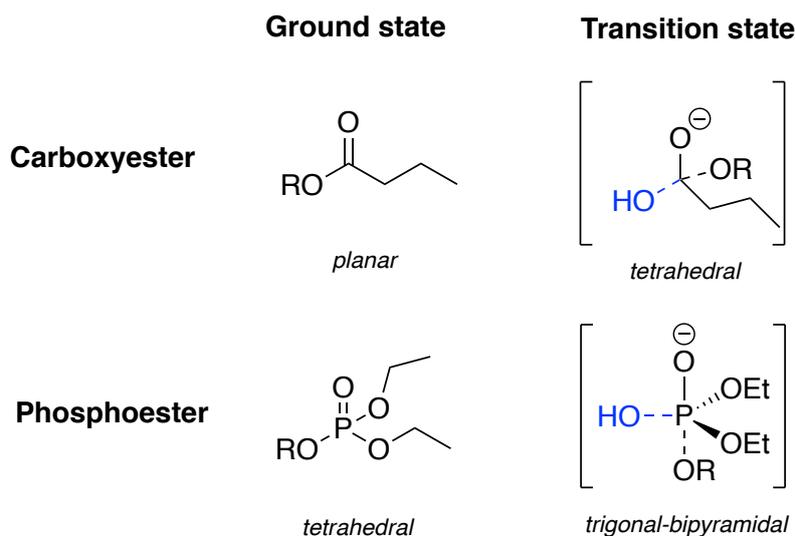


Figure 1.7: Comparison of ground and transition state geometries of carboxyesters and phosphoesters. Carboxyesters have a planar ground state and a tetrahedral transition state geometry. In contrast, phosphoesters have a tetrahedral ground state and a trigonal-bipyramidal transition state geometry. The attacking hydroxide is shown in blue. The geometrical similarity between the carboxyester transition state and the phosphoester ground state is a potential reason why the covalent phosphoryl-intermediate in acetylcholinesterase and other α/β hydrolase enzymes is so tightly stabilised and difficult to hydrolyse off again. Any esterase evolving towards phosphotriesterase function has to adapt from stabilising a tetrahedral to a trigonal-bipyramidal transition state.

[173, 174]. Enzymes exhibit very tight binding to the transition state, with affinities in the nanomolar range for promiscuous enzymes, the femto- to picomolar range for most evolved enzymes and, in the extreme case of orotidine 5'-phosphate decarboxylase, reaching the yoctomolar (10^{-24} M) range [97]. In addition, other catalytic tricks, like cofactors and catalytic groups, play a major role [175]. For this reason, when adapting from carboxyesterase to phosphotriesterase activity, enzymes have to alter the transition state geometry that they stabilise, from a tetrahedral (carboxyesters) to a trigonal-bipyramidal (phosphate esters) geometry (Figure 1.7).

The evolution of new mechanistic solutions to biocatalytic challenges is difficult and, given the vastness of sequence space, even Nature can only ever sample a small subset of all possibilities. It can therefore be deduced that not all potential catalytic motifs are harnessed by evolution but that Nature is 'lazy' and re-purposes existing catalytic strategies by tinkering with pre-existing motifs whenever possible. This notion is supported by large-scale bioinformatic analysis across 379 protein superfamilies which has shown that relatives from within one protein fold tend to

conserve their catalytic motifs while adapting to the catalysis of different substrates using the same chemistry [176].

But how adaptable are different catalytic mechanisms and motifs to such newly arising catalytic challenges? Functional versatility across all known enzymes concentrates in few very prominent folds, for example the Rossmann fold or the TIM-barrel fold. In effect, a quarter of all known folds carries out about 80 % of all known enzymatic functions [55]. The underlying reasons for this asymmetry in versatility, however, remain elusive.

One explanation, of course, could be differences in evolutionary times. Another – hypothetical – explanation, however, could be the degree of modularity of the fold. Some folds have arguably strong separation between structural scaffold and catalytic site, such as the highly modular β -propeller and TIM-barrel folds, which are characterised by a rigid scaffold, made up of repetitive elements, and seamed by flexible loops that contribute catalytic groups to the active site. If modularity is a versatility-promoting factor in a protein fold or catalytic strategy, is metal-ion catalysis intrinsically at advantage when evolving new catalytic activities, as compared to covalent catalysis? It has been shown that the functional diversity in TIM-barrel fold enzymes is mainly driven by the incorporation of a broad range of cofactors [177]. Thus, the recruitment of cofactors for catalysis could represent a more modular approach that might be more adaptable to new catalytic challenges, such as the evolution of phosphotriesterase activity. Arguably, with a metal-bound hydroxide ion as a potent nucleophile [174, 178], a relatively subtle re-arrangement of adjacent residues might be sufficient for existing metal-catalysed esterases to adapt to tight binding of a new transition state and thus evolve towards efficient organophosphate hydrolysis [179]. As in contrast to α/β hydrolase superfamily enzymes there is no covalent interaction involved, there is also no risk of entrapment. All this is speculation and therefore requires experimental evidence:

Can an α/β hydrolase catalyse phosphotriester hydrolysis as proficiently as metal-catalysed enzymes, in absence of a metal cofactor, mediated by covalent catalysis?

1.5 Structure and aims of this thesis

How adaptable are catalytic motifs to newly arising catalytic challenges? Taking the emergence of phosphotriesterase activity as a testing ground, I want to study the evolution of a pre-existing catalytic motif to a new function.

The promiscuous α/β hydrolase P91 uses a catalytic triad to hydrolyse an evolutionary new substrate, phosphotriesters, and is therefore an ideal test case for this evolutionary scenario. As, despite strong selective pressure, no similar mechanism has been observed at high efficiencies in Nature, in this work, I set out to explore whether a catalytic triad is evolvable to accommodate phosphotriesterase activity at high efficiency. I will specifically address the following questions:

- Does P91's promiscuous phosphotriesterase activity have the potential for substantial increase?
- If yes, can the α/β hydrolase fold accommodate phosphotriesterase activity in a metal-independent, covalently catalysed mechanism as effectively as established metal cofactor-assisted phosphotriesterases?
- What is the underlying mechanism? How does P91 escape covalent inactivation by phosphotriesters and achieve de-phosphorylation of its catalytic nucleophile? What is the mechanistic basis for the switch from single to multiple turnover?
- How unique is P91? Are there other enzymes exploiting this mechanism?
- What are the distinguishing features conferring phosphotriesterase activity to the α/β hydrolase fold? How could this mechanism be applied to other enzymes from the α/β hydrolase superfamily?

1.5 Structure and aims of this thesis

I will begin by subjecting P91 to directed evolution in microfluidic droplets (Chapter 2). I will then mechanistically analyse P91 and an evolved variant to elucidate the mechanism behind triad-mediated phosphotriester hydrolysis (Chapter 3). Departing from these insights, I will explore homologues of P91 that share similar features (chapter 4). By kinetic and structural comparison, I will attempt to carve out the distinguishing, activity-conferring features that enable this new catalytic motif to hydrolyse phosphotriesters. To test these emerging hypotheses, I will then apply protein engineering to non-active homologues of P91 (Chapters 5 and 6).

2 Directed evolution of the promiscuous phosphotriesterase P91

P91 is a promiscuous phosphotriesterase from the α/β hydrolase superfamily with an unprecedented, metal-independent catalytic motif. In this chapter, I subject P91 to directed evolution in microfluidic droplets to test the adaptability of its catalytic triad for phosphotriesterase activity.

External contributions:

Design and fabrication of delay-line chip wafers was performed by Dr Tomasz S. Kamiński.

2.1 Introduction

In the few decades since phosphotriester pesticides have been invented and shed into the environment, phosphotriesterase activity has evolved several times independently in different protein folds. Albeit coming from diverse evolutionary origins, all these enzymes share a very similar catalytic mechanism which uses a divalent metal ion as a cofactor. In contrast to these established phosphotriesterases, recently a new and unorthodox phosphotriesterase has been discovered in a functional metagenomic screening, P91 [149]. In contrast to all known phosphotriesterases, P91 achieves phosphotriester hydrolysis in absence of a metal co-factor, by forming a covalent intermediate with the catalytic nucleophile of its Cys–His–Asp triad. Surprisingly, and in contrast to other triad-bearing enzymes, such as acetylcholinesterase, P91 can hydrolyse this intermediate relatively quickly and thus act as a multiple turnover phosphotriesterase, reaching a $\frac{k_{cat}}{K_M}$ of $\approx 10^3 \text{ M}^{-1}\text{s}^{-1}$. Despite strong selective pressure, no similar mechanism has been observed at high efficiencies in nature. Here, I set out to test the evolvability of P91's catalytic mechanism towards efficient phosphotriester hydrolysis and thus examine the adaptability and evolutionary potential of a catalytic triad as compared to a metal-dependent motif. Testing this proposition experimentally requires a high throughput screening system. Therefore, I use microfluidic droplets as a quantitative, ultrahigh-throughput assay to screen large libraries of variants.

2.2 Results and discussion

2.2.1 Mutational scanning identifies the catalytic potential of residues near P91's active site

As previous directed evolution attempts of P91 using random mutagenesis (error-prone PCR) did not yield highly improved mutants (max. 2–3-fold improvement in

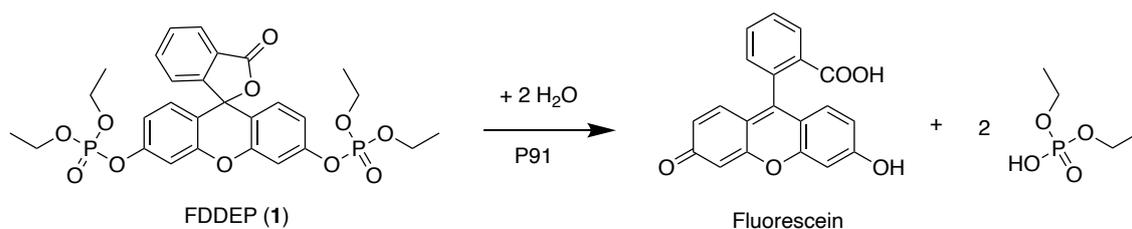


Figure 2.1: Hydrolysis of fluorescein di(diethylphosphate) (FDDEP, 1), the fluorogenic phosphotriester model substrate used for the directed evolution of P91. Upon hydrolysis of the phosphotriester groups, fluorescein is released and its fluorescence can be monitored at 490 nm (excitation) / 520 nm (emission).

$\frac{k_{cat}}{K_M}$) [180, 181], I first set out to explore the creation of a ‘smart’ library that would maximise chances of identifying improved P91 variants. In order to remodel P91’s active site to proficiently turn over phosphotriesters, I investigated how substitutions in P91’s active site cavity would affect its activity in a mutational scanning process. The protein mutability landscape concept [182] and other previous work [183] suggest that an exploration of catalytic potential can help to design a relatively small library that is amenable to plate screening. All 23 residues that comprise the first shell around the catalytic triad, lining the catalytic site within up to 12 Å of the active-site nucleophile C118, were individually randomised. I determined the maximal increase in lysate activity upon mutation by microtiter plate screening and ranked the residues according to the observed effects (Figure 2.2). Based on this screen for hydrolytic activity towards the fluorogenic model phosphotriester substrate 1 (fluorescein di(diethylphosphate), FDDEP; Figure 2.1) I identified the three positions with by far outstanding increases: A73, I211 and L214 (each increasing activity up to \approx 8–10-fold in cell lysate). These three residues are in direct vicinity with each other and line the upper lip of the active site (Figure 2.4). They are situated in loops which are partly covering the active site, highly variable in sequence and length within the DLH family. Based on the recognition of their roles as crucial for substrate specificity and substrate-induced activation [152, 184–186] I selected these residues for simultaneous randomisation, constructed a combinatorial library P91-A with a theoretical size of \approx 33 000 members (three residues, each partly randomised with the codon NNK which encodes 32 codons, thus yielding 32^3 variants on the nucleotide level) using the degenerate codon NNK and transformed it into *Escherichia coli*.

2 Directed evolution of the promiscuous phosphotriesterase P91

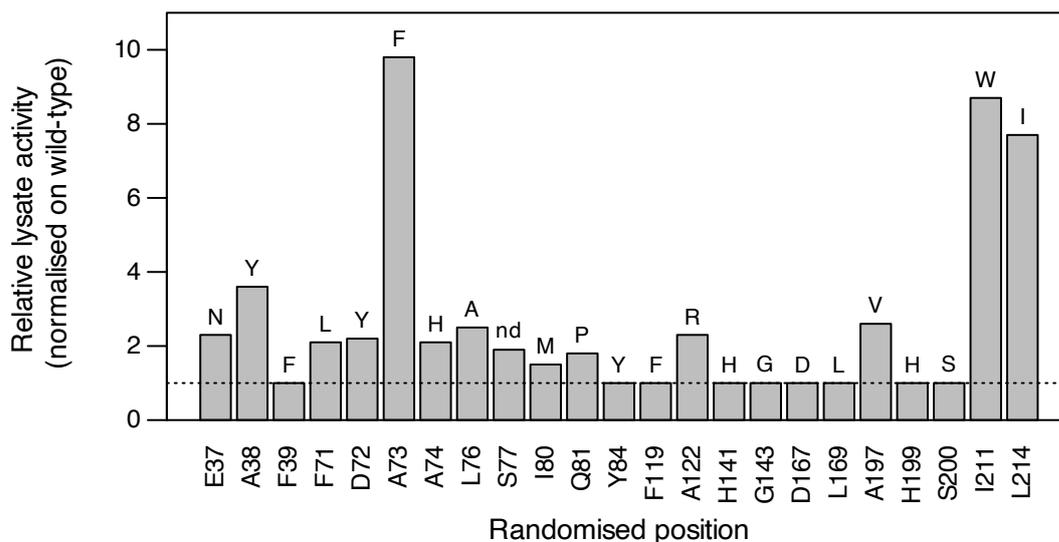


Figure 2.2: Structure-guided mutational active-site scanning. Positions lining the active site of P91 were individually randomised and all 19 variants in each position were screened for phosphotriesterase activity in microtiter plates using 1 μ M FDDEP. Over four-fold oversampling of the theoretical diversity at each position ensures that with high probability every single amino acid substitution appears in the screen. The bars show the lysate activity of the respective best-performing clone at each position (indicated in single-letter amino acid code), relative to the wild type. The dotted line indicates wild-type level activity. The substituting amino acid of the respective most active clone is indicated on top of the bar; nd: not determined. Note that the triad residues D167 and H199 were also included but did not show tolerance to mutation.

2.2.2 Droplet screening identifies highly improved variants

Round 1

To screen the library in microfluidic droplets, we compartmentalised *E. coli* cells expressing library members in a microfluidic chip (Figure 2.3a), generating monodisperse droplets (3 pL in round 1) at \approx kHz frequencies. On the chip, the enzymatic reaction was initiated by cell lysis upon droplet formation from three supply streams (containing the cell suspension, lysis reagents and fluorogenic phosphotriester substrate 1, and fluorinated oil with surfactant, respectively Figure 2.5). The bacterial suspension was diluted such that the average bacterial droplet occupancy was 20 %, resulting in less than 2 % multiple-cell encapsulation events. To exert selective pressure primarily on $\frac{k_{cat}}{K_M}$, we chose a low substrate concentration of 3 μ M (\approx 1/10 K_M). After cell encapsulation, droplets were collected in a long tubing and, after 2–3 h of

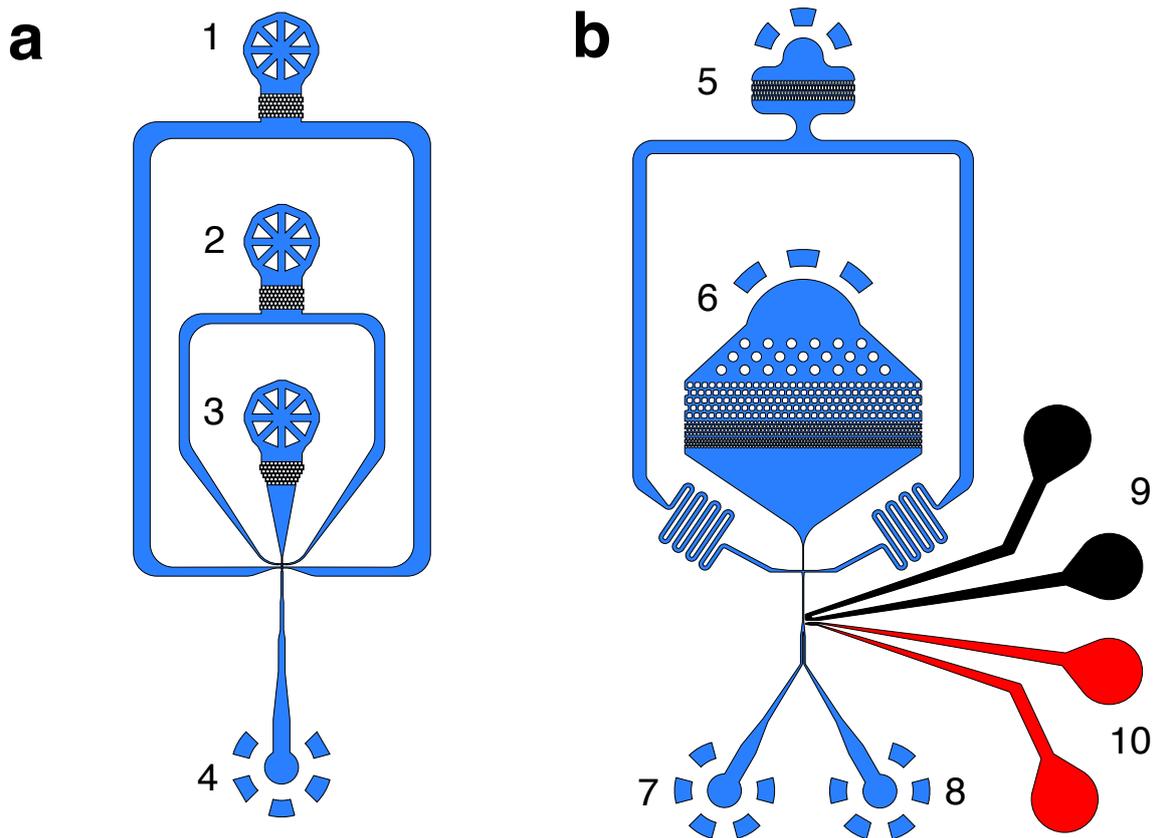


Figure 2.3: Microfluidic chip designs for droplet generation and sorting with off-line droplet incubation, as used for evolution round 1. (a) Flow-focussing chip (depth: 12 μm) for droplet generation with (1) oil/surfactant mixture inlet, (2) inlet for substrate/lysis agent mixture, (3) inlet for cell suspension, and (4) outlet for droplet collection. (b) Droplet sorting chip (depth: 20 μm) for fluorescence-activated droplet sorting with (5) inlet for spacing oil, (6) inlet for droplets, (7) waste outlet, (8) hit outlet, (9) ground electrode (+, black), and (10) signal electrode (-, red). This figure is adapted from reference [187].

2 Directed evolution of the promiscuous phosphotriesterase P91

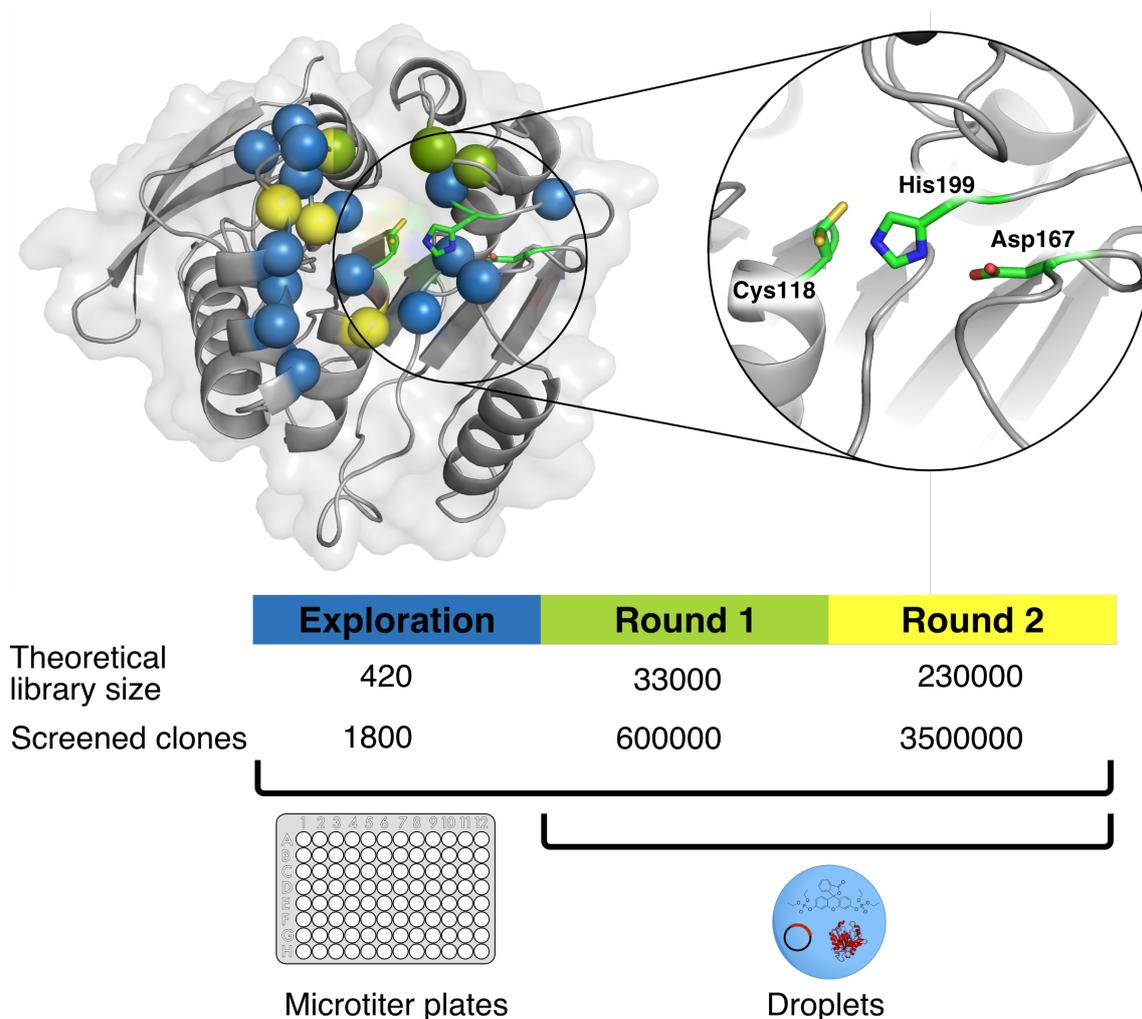


Figure 2.4: Library strategy for the directed evolution of P91. The active site of P91 was first mutationally explored (all spheres) by screening of single-site saturation libraries for phosphotriesterase activity in microtiter plates. A subset of these residues was then combined in round 1 (green spheres) and round 2 (yellow spheres) into combinatorial multiple-site saturation mutagenesis libraries which were screened in microfluidic droplets. In round 1, droplets were incubated off-chip whereas in round 2, droplets were incubated in a delay line on-chip in order to maintain selection stringency by shortening the reaction time. Blowout: Catalytic triad of P91, consisting of C118, H199, and D167. Note the two conformations of C118 in the structure, an inwards-pointing protected and an outwards-pointing active conformation.

incubation, re-injected into a separate sorting chip (Figure 2.3b) where they were screened and sorted according to their fluorescence. After sorting, the collected droplets were pooled, and the emulsion was broken. Plasmids were extracted and transformed back into *E. coli*.

In each sorting, $\approx 600\,000$ droplets were screened and $\approx 10\,000$ droplets selected,

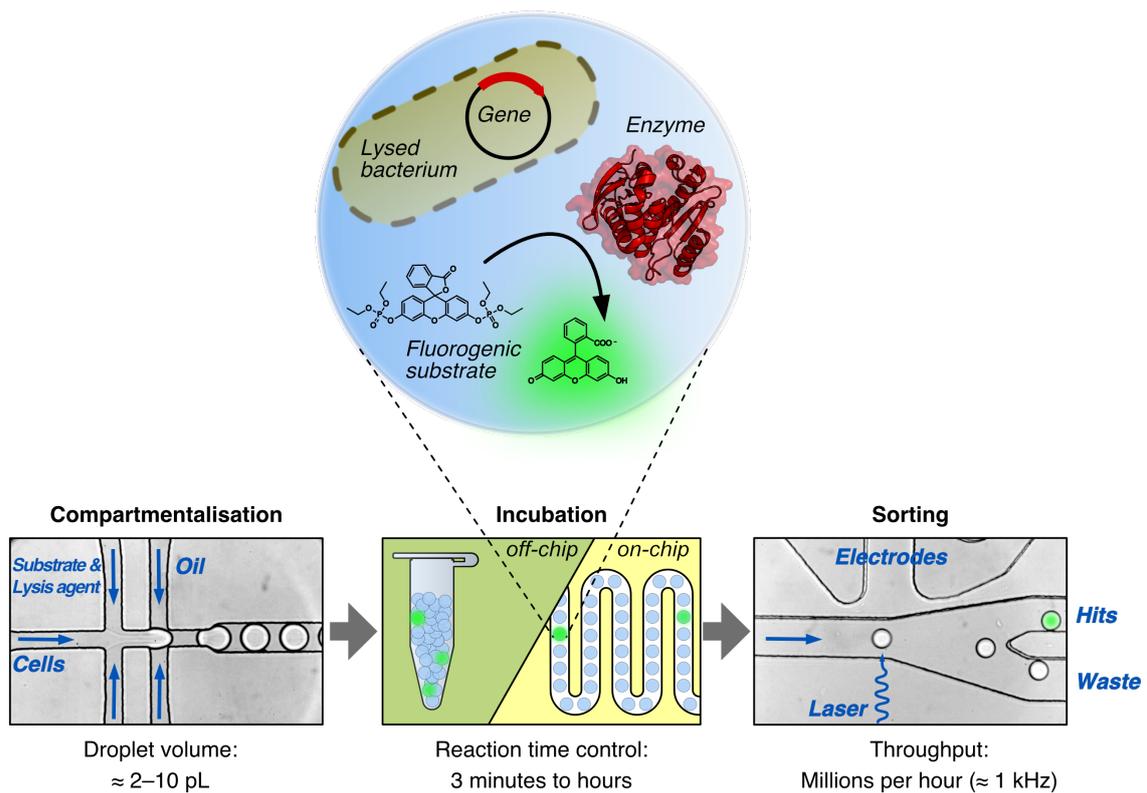


Figure 2.5: Microfluidic droplet screening assay. The microfluidic screening assay consists of three steps: (1) Encapsulation of bacterial cells together with a fluorogenic substrate and lysis agent into picolitre-sized aqueous droplets that are separated with fluorinated oil. The droplets serve as miniaturised reaction vessels that link phenotype (catalytic activity indicated by fluorescence) to genotype (gene sequence encoded on a plasmid). (2) Droplet incubation can be carried out in a delay line on-chip (for the range of minutes) or off-chip (for hours to weeks). (3) The droplets can then be sorted according to their fluorescence with an excitation laser that is focussed on the droplet flow along a Y-shaped junction. When surpassing a certain threshold fluorescence value, a single droplet can be electrophoretically deviated by the electrodes away from the waste channel into the hit collection channel. Additional chip details are given in Figures 2.3 and 2.7.

corresponding to the top $\approx 1.7\%$ most active droplets. These relatively permissive sorting conditions were chosen to avoid losing false negative clones due to phenotypic variation in enzyme expression on the single cell level, allowing cumulative enrichment of improved clones over several rounds of sorting. This enrichment process by droplet sorting was repeated two times (three times in total) in order to gradually narrow down the diversity of the library before proceeding to secondary screening in microtiter plates.

Following droplet screening, ≈ 90 randomly picked individual clones from each sorting (≈ 350 clones for the last sorting) were analysed in a lysate-based microtiter

2 Directed evolution of the promiscuous phosphotriesterase P91

plate screening, revealing a successive enrichment of active clones along the course of sorting (Figure 2.6b). The most active clones were sequenced, revealing that all sequenced clones had a tryptophan in position 211 and a valine in position 214. Position 73, in contrast, showed more diversity among enriched variants.

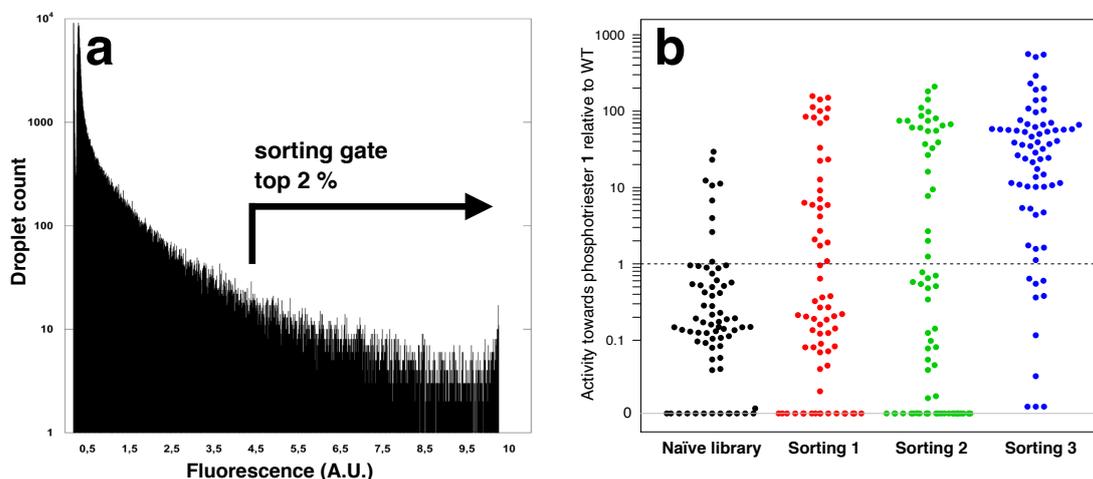


Figure 2.6: Screening of the multiple site-saturation library P91-A in microfluidic droplets by FADS. **a:** Droplet sorting histogram from sorting 1 illustrating the chosen sorting regime. After 2–3 h of incubation, the top 2 % of all droplets were sorted, corresponding to approximately the top 10 % of all droplets containing a bacterium, according to the Poisson distribution with an expected average number of bacteria encapsulated in one droplet of $\lambda = 0.2$. In each sorting, 500 000 droplets were screened and 10 000 droplets were sorted. The continuous shape of the fluorescence distribution reflects the distribution of activities in the library as well as phenotypic variation at the single cell level. **b:** Enrichment of improved variants over several rounds of droplet sorting. 70 clones were picked randomly and screened for lysate activity towards phosphotriester substrate 1 (FDDEP) in microtiter plates. The dashed line indicates the WT activity level, the grey line indicates no detectable activity. Improved variants are cumulatively enriched from sorting to sorting. After three sorting campaigns, variants with lysate activities increased by almost three orders of magnitude could be identified. Note that due to the dynamic range of the microtiter plate activity assay, the measurement of such high increases is relatively imprecise and needs to be quantified by kinetic analysis of purified proteins.

Round 2

Therefore, the library P91-B for the next round of evolution was built on the consensus sequence W211, V214 (dubbed P91-R1) and position 73 was re-included into the next library. In addition, three further residues, ranking next in terms of activity change upon individual mutation (Figure 2.2), were randomised in the sec-

ond library (A38, L76, and A122). Moreover, A38 and L76 are in direct vicinity to the residues already randomised in the first library, seaming the active site and thus making highly epistatic interactions accessible. In the canonical esterase mechanism, A38 contributes with its backbone nitrogen to stabilisation of the oxyanion and forms the so-called ‘oxyanion hole’ (Figure 3.1). A122 is directly below the active site nucleophile and was hypothesised to be involved in positioning the hydrolytic water molecule for the breakdown of the covalent intermediate. These positions were randomised using a mixture of the degenerate codons NDT, VHG, and TGG, which code for all canonical amino acids at approximately even representation and no stop codon (known as 22-codon trick [188]), yielding a theoretical diversity of 160 000 (20^4) on the amino acid level and $\approx 234\,000$ (22^4) on the nucleotide level.

To adjust for the shorter reaction time required for the improvement of mutants of P91-R1, we designed an integrated device on which droplet generation, incubation, and sorting were combined on a single chip [132, 134] (Figure 2.7) so that stringent sorting was possible in the second round. On this chip, the incubation time of the droplets was precisely controlled by the length of the delay line [189]. To optimise the length of the delay line for stringent sorting, we measured the reaction progress of the starting variant P91-R1 in a chip with a very long delay line (Figures 2.7a and 2.8a). The required delay line length for the library sorting device was then chosen such that sorting would occur in the early linear phase of the reaction, corresponding to a reaction time of ≈ 5 min (Figures 2.7b and 2.8b). In this set-up, droplet size was ≈ 11 pL. The droplet sorting was again repeated three times in total. In each sorting step, ≈ 1.5 – 3.5 million droplets were screened and $\approx 10\,000$ – $70\,000$ droplets were sorted. To balance throughput with accuracy, the average droplet occupancy was continuously reduced from 35 % in the first sort, to 20 % and then 10 % in the last sorting. Droplet sorting was again followed by a secondary microtiter plate screening (≈ 350 randomly picked clones) and sequencing of the most active clones.

2 Directed evolution of the promiscuous phosphotriesterase P91

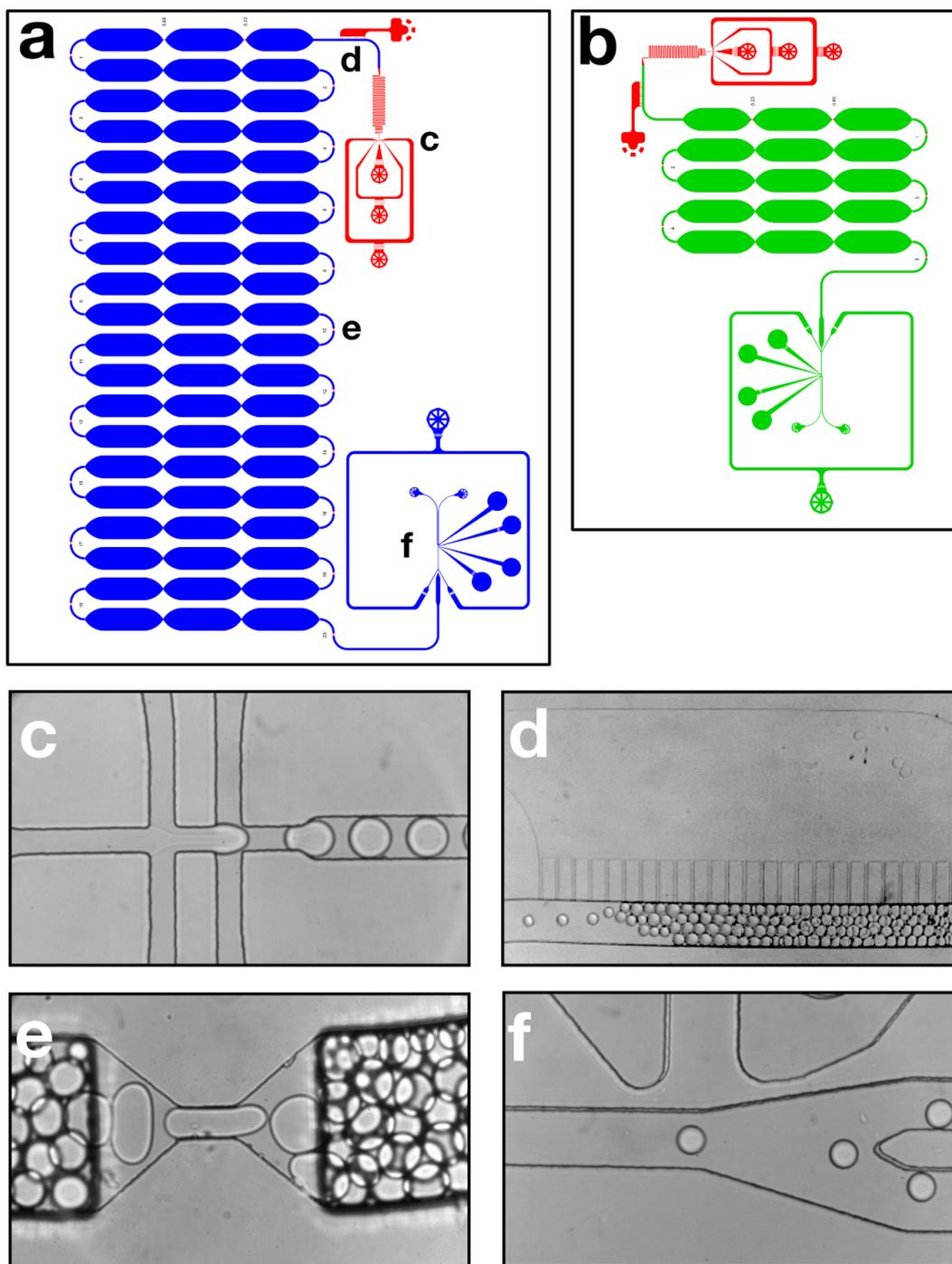


Figure 2.7: Design of microfluidic chips for on-chip droplet incubation. For short incubation times in evolution round 2, requiring on-chip incubation, an integrated chip was used. This chip combines a flow-focussing module, a delay line, and a sorting module on a single device. The chips are two-layered such that the deeper delay line reduces back-pressure. Red areas are 15 μm deep, green areas 28 μm , and blue areas 45 μm . **(a)** For initial stringency adjustments, a chip with a long delay line of 20 loops was used. **(b)** For library sorting, a shorter delay line consisting of five loops was used. **(c)** Monodisperse water-in-oil droplets are generated in a flow-focusing nozzle, co-encapsulating bacteria, the fluorogenic substrate and lysis agent. **(d)** An oil extractor ensures dense packing and equal incubation time for all droplets in the **(e)** delay line chamber. The delay line chambers possess mixing constrictions and shallow windows to monitor droplet fluorescence during incubation. **(f)** In the sorting junction droplets are electrophoretically sorted according to their fluorescence with up to kHz frequencies.

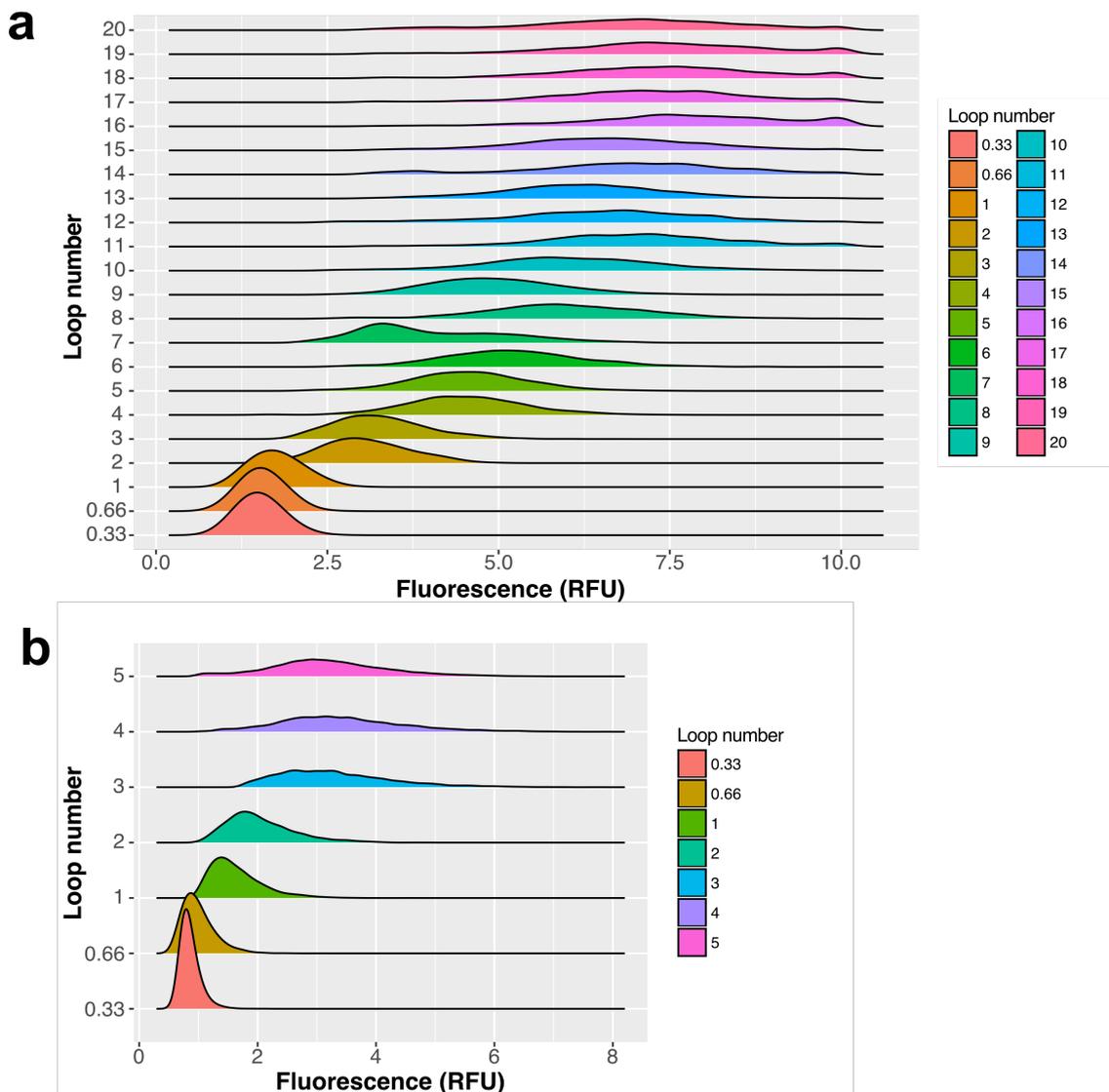


Figure 2.8: On-chip fluorescence measurements for the adjustment of reaction time and sorting stringency. Fluorescence distribution of 10 000 droplets in relative fluorescence units (RFU) at different points in the delay line. Fluorescence measurements were taken at the outer constrictions of the delay line loops (Figure 2.7e) in order to follow the progress of the reaction in droplets. The loop number indicates the place of the constriction along the delay line where the laser was placed for measurements and is a proxy of reaction time. 0.33 and 0.66 refer to measurements at constrictions within the first loop of the delay line (after a third and two thirds of the first loop length, respectively). With the chosen flow rates, 20 loops correspond to ≈ 28 min in the long chip and five loops correspond to ≈ 4.5 min incubation time in the shorter and shallower library sorting chip. For initial stringency adjustments, a chip with a long delay line of 20 loops (**a**) was designed and cells expressing the parent variant for round 2, P91-R1, were encapsulated and incubated in the delay line. Fluorescence distribution increases linearly in the early loops and the reaction saturates in the later loops. With the aim to sort the library within the early linear phase of the reaction, this extent of reaction progress provided the basis for choosing five loops for the library sorting chip (**b**).

2.2.3 An evolved P91 variant rivals the efficiencies of metal-dependent phosphotriesterases

Table 2.1: Steady-state catalytic parameters for phosphotriester hydrolysis by P91-WT, P91-R1, and P91-R2, measured with the substrate FDDEP in 50 mM HEPES-NaOH, 150 mM NaCl, 1 mM TCEP, pH 8.0 at 25 °C. Enzyme concentrations were: 0.2 μM for P91-WT, 2 nM for P91-R1, and 0.2 nM for P91-R2.

Enzyme variant	Mutations	k_{cat} (s^{-1}) ^a	K_M (μM) ^a	K_i (μM)	$\frac{k_{cat}}{K_M}$ ($\text{M}^{-1}\text{s}^{-1}$) ^b
P91-WT	—	0.054	27	470	$2.0 \cdot 10^3$
P91-R1	I211W, L214V	37	120	7.9	$3.0 \cdot 10^5$
P91-R2	A38L, A73E, L76V, I211W, L214V	15	20	14	$7.8 \cdot 10^5$

^a Due to strong substrate inhibition, estimates of k_{cat} and K_M for P91-R1 and P91-R2 are extrapolations and only $\frac{k_{cat}}{K_M}$ can be regarded as precise (see Michaelis-Menten plot, Figure 2.9a).

^b The values measured for P91-WT differ (< 5 -fold) from previously published values [149], which may be ascribed to differences in affinity tag, purification procedure, and buffer conditions.

The most improved variant, P91-R2, has five mutations compared to wild-type P91: A38L, A73E, L76V, I211W, and L214V. The kinetic characterisation of purified P91-R2 revealed a ≈ 400 -fold increase in $\frac{k_{cat}}{K_M}$ over wild-type (Figure 2.9a and Table 2.1), with $\frac{k_{cat}}{K_M} \approx 8 \cdot 10^5 \text{ M}^{-1}\text{s}^{-1}$ (Table 2.1). P91-R2 also shows a greater propensity for substrate inhibition, likely a consequence of adaptation to the low substrate concentrations (3 μM) used during screening. Despite only two rounds of evolution departing from a weakly promiscuous starting point, these catalytic parameters are above the median of most physiological enzymes shaped by long-term natural Darwinian evolution [190].

To put P91-R2's activity level into perspective, I compared P91-R2 with other organophosphate-degrading enzymes reported in the literature that have either evolved naturally (isolated from organophosphate-contaminated environments) or were evolved or engineered in the laboratory from promiscuous enzymes (Table 2.2).

Enzymatic activity has been found towards a wide variety of organophosphates: aryl dialkyl phosphotriesters (e.g., paraoxon-ethyl), triaryl phosphates (e.g. the widely used flame retardants and plasticiser triphenyl phosphate [191]), thiophosphates (e.g. parathion-ethyl), halogenated phosphoesters (e.g. DFP), phosphoramidates

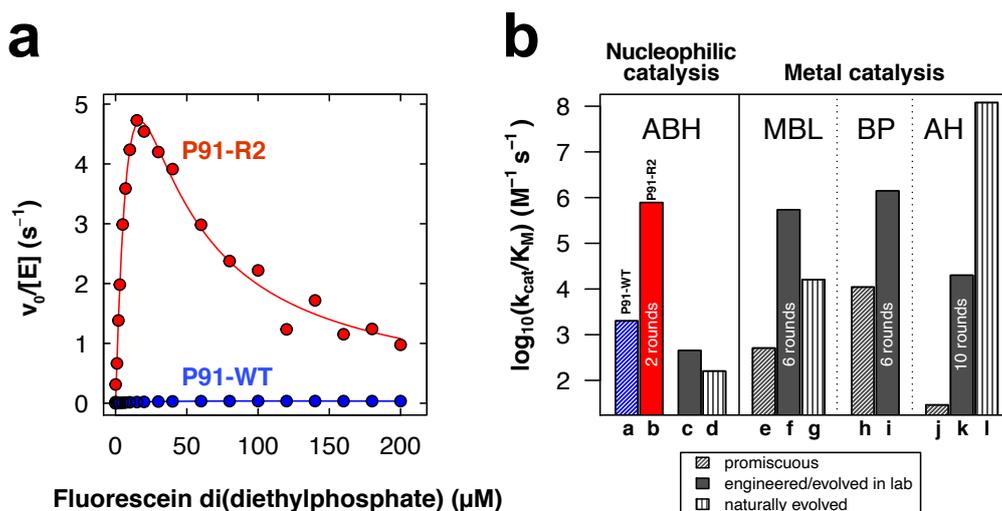


Figure 2.9: P92-R2 rivals the efficiencies of engineered and naturally evolved metal-dependent phosphotriesterases. (a) Michaelis-Menten plots for P91-WT (blue) and the evolved variant P91-R2 (red) for the hydrolysis of fluorescein di(diethylphosphate) (FDDEP, 1), measured in 50 mM HEPES-NaOH, 150 mM NaCl, 1 mM TCEP, pH 8.0 at 25 °C. v_0 , initial reaction velocity; $[E]$, initial enzyme concentration. Enzyme concentration was 0.2 μM for P91-WT and 0.2 nM for P91-R2. (b) Comparison of catalytic efficiencies of promiscuous (hatched), engineered (filled), and naturally evolved (lined) phosphotriesterases from different protein superfamilies: ABH, α/β -hydrolases; BP, β -propellers; MBL, metallo- β -lactamases; AH, amidohydrolases. P91-WT is shown in blue, the evolved variant P91-R2 in red. For enzymes that were evolved by directed evolution, the number of rounds is indicated in the bar. Annotation for the bar labels a to l and the respective references are detailed in Table 2.2. Substrates between enzymes differ in leaving group but are all diethyl-substituted phosphotriesters.

(e.g. tabun) and thiophosphonates (e.g. VX). Considerable effort has also been invested in engineering efficient enzymes that directly act on G-series and V-series organophosphate warfare agents, such as sarin, tabun, soman, and VX [78, 81, 83, 85, 86].

However, in this comparison, I restrict the substrates to diethyl-substituted phosphotriesters as these are among the most common organophosphate insecticides in agricultural use and, due to their high accessibility compared to highly regulated warfare agents, are used in most enzyme characterisations and directed evolution studies in the published literature. Thus, in my comparison, the substrates are all diethyl-substituted phosphotriesters and differ only in their leaving group (*p*-nitrophenol, fluorescein, and umbelliferone). When evolved variants showed higher

2 Directed evolution of the promiscuous phosphotriesterase P91

activity towards a different organophosphate substrate used in the respective study, this is additionally noted in Table 2.2.

From the α/β hydrolase superfamily, a few metal-independent, catalytic-triad bearing enzymes besides P91 are known to very slowly hydrolyse phosphotriesters, notably an engineered human cholinesterase (butyrylcholinesterase G117H) [169] and an insect carboxylesterase (*Lc α E7* G137D) isolated from pesticide-exposed blowflies [167]. However, rather than being veritable multiple turnover enzymes, these enzymes are characterised by tight substrate binding and relative resistance to an irreversible side-reaction called ageing, allowing very slow reactivation by de-phosphorylation. Thus, turnover rates ($k_{cat} \approx 10^{-2}$ to 10^{-3} s $^{-1}$) and therefore overall catalytic efficiencies are very low (Figure 2.9b, bars c & d, and Table 2.2).

Much more significant promiscuous or naturally evolved organophosphate hydrolase activity has been identified in enzymes from diverse protein superfamilies, such as the amidohydrolases [24], the pita-bread fold [26], the β -propellers [27, 28], and the metallo- β -lactamases [29], which all share a metal cofactor-dependent mechanism of degradation. An exhaustive overview of promiscuous and naturally evolved phosphotriesterases (with organism of origin and catalytic efficiencies towards paraoxon-ethyl or parathion-methyl) has previously been compiled [149].

The most prominent example among those enzymes is the phosphotriesterase found in *Brevundimonas diminuta* (*BdPTE*), which achieves catalytic efficiencies approaching the diffusion limit (Figure 2.9b, bar l, and Table 2.2) [25]. *BdPTE* has been the target of numerous directed evolution campaigns, mainly towards chemical warfare agents, but has also been further evolved towards paraoxon hydrolysis, reaching an even higher k_{cat} , however with low further increases in overall $\frac{k_{cat}}{K_M}$ [121].

A promiscuous homologue with the same protein fold (amidohydrolase superfamily), has been evolved for 10 rounds towards higher organophosphate hydrolase activity, reaching an efficiency of $2 \cdot 10^4$ M $^{-1}$ s $^{-1}$ for paraoxon-ethyl and $1 \cdot 10^6$ M $^{-1}$ s $^{-1}$ for a methyl phosphonate (Figure 2.9b, bars j & k, and Table 2.2) [192]. Promiscuous

phosphotriesterases from the β -propeller and the metallo- β -lactamase superfamilies have also been subjected to extensive laboratory evolution, reaching catalytic efficiencies in the range of 10^4 – 10^6 $\text{M}^{-1}\text{s}^{-1}$ within 6 rounds (Figure 2.9b, bars e, f & i and Table 2.2) [38, 43]. Overall, the outstanding catalytic efficiency of *Bd*PTE seems to be an exception rather than the norm for metal-catalysed phosphotriesterases, with most naturally evolved or artificially engineered enzymes displaying catalytic efficiencies in the range of 10^4 – 10^6 $\text{M}^{-1}\text{s}^{-1}$.

In summary, $\frac{k_{cat}}{K_M}$ of P91-R2 is similar to or even surpasses the efficiencies of many naturally evolved, metal-dependent phosphotriesterases and metal-catalysed phosphotriesterases that were engineered or further evolved in the lab from promiscuous precursors, typically requiring many more rounds of evolution than P91-R2 (Figure 2.9b). Notably, it is the first metal-independent phosphotriesterases reaching high catalytic efficiencies via a nucleophilic mechanism, mediated by a covalent intermediate.

Table 2.2: P91-R2 rivals the efficiencies of engineered and naturally evolved metal-dependent phosphotriesterases. Comparison of kinetic parameters of promiscuous, engineered, or naturally evolved phosphotriesterases from different protein superfamilies: ABH, α/β -hydrolases; BP, β -propellers; MBL, metallo- β -lactamases; AH, amidohydrolases. Substrates between enzymes differ in leaving group but are all diethyl-substituted phosphotriesters. *n.a.*; not applicable.

Label in Figure 2.9b	Enzyme	Protein superfamily	Catalytic mechanism	Organism of origin	Substrate	$\frac{k_{cat}}{K_M}$ ($10^3 \cdot M^{-1}s^{-1}$)	Relative increase in $\frac{k_{cat}}{K_M}$	Rounds ^a	Reference
a	P91-WT	ABH	nucleophilic	unknown	Fluorescein di(diethylphosphate)	2.0	<i>na</i>	<i>n.a.</i>	this work
b	P91-R2	ABH	nucleophilic	unknown	Fluorescein di(diethylphosphate)	780	400	2	this work
c	BChE G117H	ABH	nucleophilic	<i>Homo sapiens</i>	Echothiophate	0.45	<i>n.a.</i> ^b	1 ^b	[170]
d	Lc α E7 G137D	ABH	nucleophilic	<i>Lucilia cuprina</i>	Diethylumbelliferyl phosphate	0.16	<i>n.a.</i>	<i>n.a.</i>	[58]
e	AiiA-WT	MBL	metal	<i>Bacillus thuringiensis</i>	Paraoxon-ethyl	0.51	<i>n.a.</i>	<i>n.a.</i>	[38]
–	AiiA-WT	MBL	metal	<i>Bacillus thuringiensis</i>	Parathion-methyl ^c	70	<i>n.a.</i>	<i>n.a.</i>	[38]
f	AiiA-R6	MBL	metal	<i>Bacillus thuringiensis</i>	Paraoxon-ethyl	540	1 100	6	[38]
g	MPH	MBL	metal	<i>Pseudomonas sp.</i> WBC-3	Paraoxon-ethyl	16	<i>n.a.</i>	<i>n.a.</i>	[38]
h	rePON1-G3C9	BP	metal	<i>Oryctolagus cuniculus</i> ^d	DEPCyc	11	<i>n.a.</i>	<i>n.a.</i>	[193]
i	rePON1G3C93.2PC	BP	metal	<i>Oryctolagus cuniculus</i> ^d	DEPCyc	1 400	130	6	[43]
j	DrPLL-WT	AH	metal	<i>Deinococcus radiodurans</i>	Paraoxon-ethyl	0.029	<i>n.a.</i>	<i>na</i>	[192]
k	DrPLL.10	AH	metal	<i>Deinococcus radiodurans</i>	Paraoxon-ethyl	20	690	10	[192]
–	DrPLL.10	AH	metal	<i>Deinococcus radiodurans</i>	O-isopbutyl-O-4-nitrophenyl- <i>R_P</i> -methylphosphonate ^e	1 100	37 000	10	[192]
l	BdPTE	AH	metal	<i>Brevundimonas diminuta</i>	Paraoxon-ethyl	120 000	<i>n.a.</i>	<i>n.a.</i>	[192]
–	BdPTE	AH	metal	<i>Brevundimonas diminuta</i>	Paraoxon-ethyl	180 000	1.8 ^f	1	[121]

^a Number of laboratory evolution rounds, meaning successive iterations of diversification and screening.

^b A single substitution was rationally engineered into the wild-type precursor which has no detectable activity.

^c Substrate with highest final activity in this evolution campaign.

^d Engineered rabbit gene with contributions from the respective human, mouse, and rat genes through DNA shuffling during directed evolution.

^e Substrate with highest final activity and highest relative activity increase in this evolution campaign.

^f 1.8-fold increase in $\frac{k_{cat}}{K_M}$ as compared to a wild-type $\frac{k_{cat}}{K_M}$ of $9.9 \cdot 10^7 M^{-1}s^{-1}$, as measured in this study. In addition, the selected BdPTE variant displayed a 63-fold increase in k_{cat} .

2.3 Conclusion and outlook

By subjecting P91 to directed evolution, here I demonstrate that the α/β hydrolase P91, a distant relative of acetylcholinesterase, can be evolved to escape this covalent inhibition at such a high rate that it can effectively act as a high-turnover enzyme with a covalent mechanism, reaching catalytic efficiencies that rival those of its metal-dependent counterparts. Five mutations were sufficient to increase P91's phosphotriesterase activity by a factor of ≈ 400 , leveraging its catalytic efficiency close to $10^6 \text{ M}^{-1}\text{s}^{-1}$ and its turnover rate into the range of $> 10 \text{ s}^{-1}$, thus approaching the catalytic efficiencies of many metal-dependent, 'conventional' phosphotriesterases after only two rounds of evolution. This is the first instance of a catalytic triad efficiently turning over organophosphates at efficiencies comparable to naturally evolved, metal-dependent enzymes and defines a cysteine triad as an evolvable motif for phosphotriesterase activity. Consequently, this evolution campaign also reveals the theoretical potential of other α/β hydrolases to escape covalent inhibition at similar rates, encouraging efforts in engineering human enzymes, such as butyrylcholinesterase, into catalytic bioscavengers in order to treat organophosphate poisonings [77]. Promiscuous enzymes from different metal cofactor-dependent scaffolds, including the β -propeller, amidohydrolase, and metallo- β -lactamase superfamilies, have been evolved to high catalytic efficiencies, however only reaching comparable levels of activity as P91-R2 while requiring considerably more rounds of evolution (6–10 rounds) [38, 192, 193]. This directed evolution campaign also validates the idea that functional metagenomics can identify new (but often weak) catalysts as starting points in sequence space that are then further evolvable towards the desired function [146, 149].

What have droplets achieved so far? Regarding the role of droplet microfluidics for the directed evolution of enzymes, the intuitive belief that higher throughput equals higher improvements has not yet translated widely into highly improved enzymes. Very high improvements have been achieved using simple colorimetric agar plate assays and random mutagenesis (epPCR) libraries for directed evolution [194]. In

2 Directed evolution of the promiscuous phosphotriesterase P91

contrast, although numerous demonstrations of activity enrichments in microfluidic droplets as well as several full directed evolution campaigns are present in the literature [139], most evolved improvements do not surpass one order of magnitude [129, 130, 141, 195]. In fact, as the beneficial impact of a single mutation is usually limited, random mutagenesis mostly proceeds in small incremental steps over many rounds of evolution, provided the fitness landscape is smooth. Multiple rounds of evolution, however, are more straightforward with simpler assays, such as microtiter plate screening, as compared to the more elaborate droplet screenings, which only represent an enrichment step, followed by a secondary screening. The large phenotypic variation observed between droplets and the high number of false positives after a single sorting also suggest that – despite the conceptual beauty – droplet sorting is only a raw enrichment and only develops full power through repetitive re-sorting under permissive conditions. Therefore, the power of droplet screening comes into play when large combinatorial libraries with rare beneficial combinations of residues are screened. In this constellation, droplet microfluidics can swiftly screen through a vast pool of largely deleterious combinations (e.g. in a rugged fitness landscape) and identify rare, highly improved events. Using this strategy, this work achieves one of the highest single-round improvements in catalytic efficiency obtained by directed evolution in microfluidic droplets, only surpassed by the recent 960-fold improvement of an amine oxidase [134]. In addition, the precision of control over reaction time obtained by switching between off-chip and on-chip droplet incubation allowed to fine-tune selection stringency over a large range of catalytic efficiencies. In summary, the high throughput of microfluidic droplet sorting combined with focussed combinatorial library design allows leaps in sequence space, giving access to and establishing new mechanistic territory that, despite strong selective pressure, has not been exploited in nature.

3 The catalytic mechanism of P91

How does the evolved variant of P91 achieve efficient phosphotriester hydrolysis in absence of a metal co-factor? In this chapter, I kinetically analyse wild-type P91 and its evolved variant to elucidate its catalytic mechanism.

External contributions:

James Oskar Klein measured the esterase kinetics, synthesised the LFER substrates 5–10 and measured LFER kinetics for P91-WT as part of a summer project.

3.1 Introduction

The initial promiscuous phosphotriesterase activity of P91 is unexpected and more so the high activity of the variant P91-R2 obtained by directed evolution. Other α/β hydrolases such as acetylcholinesterase are fast at performing a nucleophilic attack on phosphotriesters but fail to resolve the formed phosphoryl intermediate. How does P91 achieve multiple turnover while other catalytic triads are arrested in single turnover? Here we set out to elucidate P91's mechanism by kinetic analysis.

The fatal reaction of phosphotriesters with many catalytic triads (e.g. Ser–His–Glu in acetylcholinesterase) consists of the fast and near-irreversible formation of a covalent adduct; thus, phosphorylation rate (k_2) \gg de-phosphorylation rate (k_3) and $k_3 \approx 0$. By analogy, we initially assumed that P91 would follow a similar reaction scheme, however with a more significant k_3 (Figure 3.1).

3.2 Results and discussion

3.2.1 Intermediate breakdown is not rate-limiting

To investigate the mechanistic basis for P91-R2's high catalytic efficiency as compared to its wild-type precursor as well as to irreversibly inhibited homologues (like acetylcholinesterase) we measured pre-steady state kinetics using paraoxon-ethyl (2) as substrate, following the two-step model (Figure 3.1) that is typical for textbook catalytic triads [174]. Time courses for P91-WT and its mutants showed monophasic kinetics that could be fitted to a single exponential increase in reaction product, even when measured in a rapid-mixing stopped-flow apparatus. So the rates of phosphorylation and de-phosphorylation, k_2 and k_3 were not accessible due to lack of an observable burst. Furthermore, no enzyme concentration-dependent signal offset could be detected which would hint towards a very fast burst within the dead

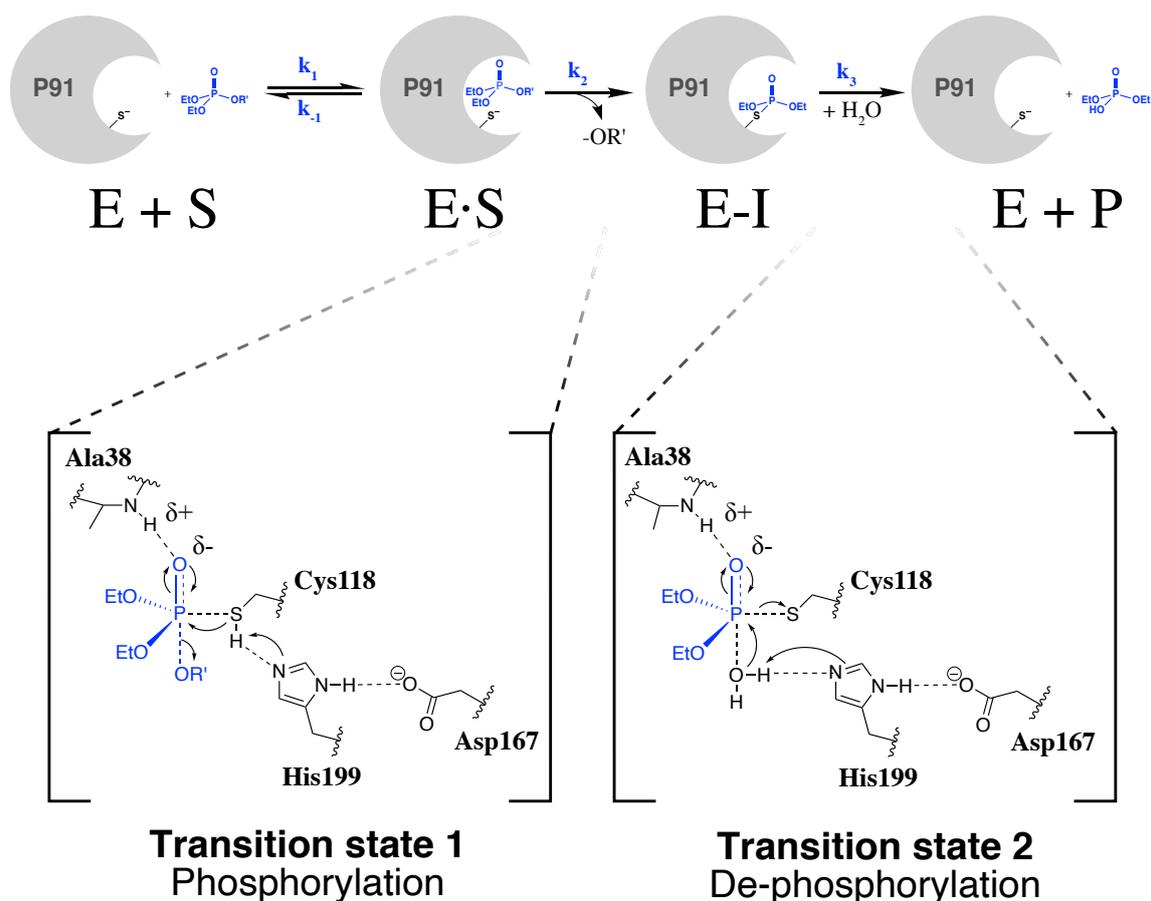


Figure 3.1: Overview of the assumed reaction scheme and the transition states. P91 hydrolyses phosphotriesters via a mechanism involving the formation (transition state 1, k_2) and breakdown (transition state 2, k_3) of a covalent intermediate (E-I). Both transition states have a trigonal-bipyramidal geometry of the phosphotriester during the transition state of the formation of the covalent adduct and the breakdown of the covalent intermediate. A38 contributes with its backbone nitrogen to stabilisation of the oxyanion and forms the so-called 'oxyanion hole'.

time of the stopped-flow instrument (≈ 1 ms) (Figure 3.2). We thus concluded that P91, in contrast to acetylcholinesterase, is rate-limited by the formation of the covalent adduct rather than by its breakdown.

3.2.2 Nucleophile exchange reveals acceleration of intermediate formation

We hypothesised that a nucleophile exchange from cysteine to serine in P91 (C118S) would change the rate-determining step in phosphotriester hydrolysis and thus make it possible to resolve the relative effects of the directed evolution campaign on k_2 and k_3 . We assumed that such a nucleophile change would have no effect on the

3 The catalytic mechanism of P91

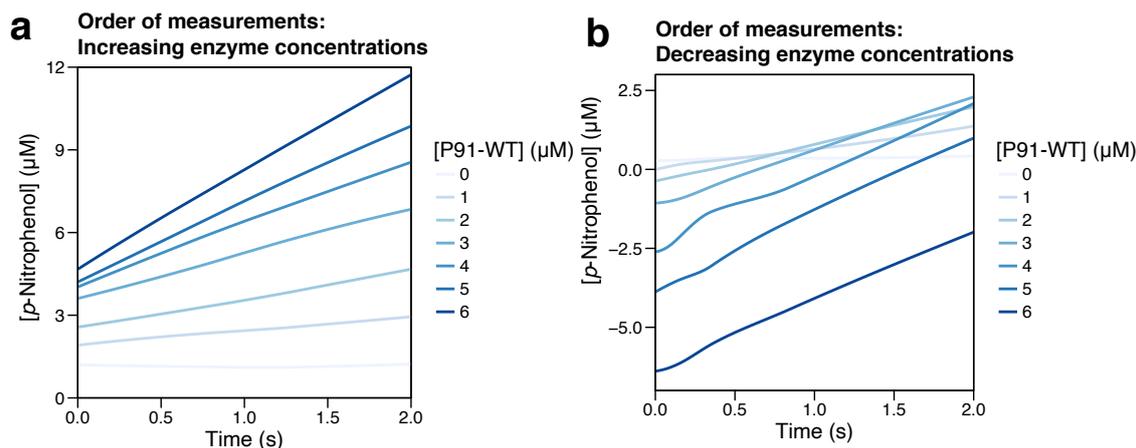


Figure 3.2: Wild-type P91 (containing a cysteine triad) does not show biphasic burst kinetics. Stopped-flow kinetic traces of P91-WT reacting with 2 mM paraoxon-ethyl, measured in 50 mM HEPES-NaOH, 150 mM NaCl, 1 mM TCEP, pH 8.0 at 25 °C at varying enzyme concentrations (0–6 μM, blue shades). Absorbance values are shown in units of corresponding concentration of released *p*-nitrophenol. **(a)** Time course measurements taken in order of *increasing* enzyme concentrations. **(b)** Time course measurements taken in order of *decreasing* enzyme concentrations. Although an absorbance offset is observable it does not depend on the enzyme concentration used but rather on the order of measurements (time-dependent). The absorbance offset is therefore probably consequence of chemical background hydrolysis, refuting the hypothesis that a very fast burst could be occurring in the dead time of the stopped-flow instrument (≈ 1 –5 ms).

catalytic contributions of other chemical groups in the active site, and thus, independent of the observed change in absolute rates, affect all variants in the same way. Therefore, relative differences in k_2 and k_3 between wild-type and evolved variants of P91 should therefore be transferrable from enzymes with a serine nucleophile to the native cysteine nucleophile enzymes.

Indeed, introducing a C118S mutation strongly slowed down overall phosphotriester turnover rates (k_{cat}) by ≈ 2 –3 orders of magnitude in all three variants and revealed burst-like pre-steady state kinetics (Figures 3.3a, 10.3, 10.1). These burst traces could be fit to a two-step model, describing a fast intermediate formation followed by its breakdown. Fitting of the burst kinetic traces to an exponential equation according to the two-step reaction scheme (Figure 3.1) allowed quantification of k_2 and k_3 (Figures 10.2 and 10.4). Notably, the rates of the initial burst phase (reflecting k_2) were so different between variants that P91-WT C118S could be measured using a conventional plate reader whereas the evolved variants needed to be measured in a stopped-flow apparatus (Figures 3.3a, 10.3, 10.1).

Overall, we observe that in the nucleophile-exchanged variants k_2 increases by ≈ 900 -fold from the P91-WT background to the evolved P91-R2 background (Figure 3.3b), approximately corresponding to the increase in $\frac{k_{cat}}{K_M}$ in the cysteine enzymes, where k_2 is rate-limiting. In contrast to that, k_3 changes little and remains within the same order of magnitude, revealing that selective pressure during directed evolution acted on intermediate formation.

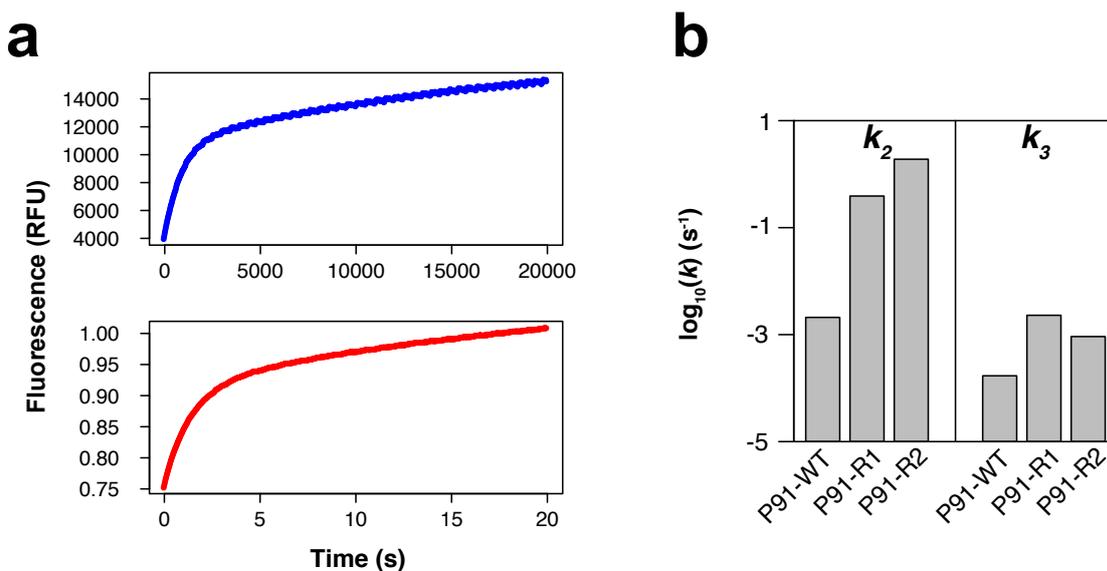


Figure 3.3: The largest effect of evolution is on intermediate formation. (a) Examples of traces of kinetic bursts with nucleophile-exchanged P91 variants (C118S). Reaction time course of P91-WT C118S (blue, enzyme concentration 10 μM) and P91-R2 C118S (red, enzyme concentration 1 μM) with 100 μM FDDEP (1) in 50 mM HEPES-NaOH, 150 mM NaCl, pH 8.0 at 25 $^{\circ}\text{C}$. Note the different time scales for the different variants, allowing the use of a spectrophotometric microplate reader for the wild-type-derived enzyme while requiring the use of a stopped-flow instrument for the evolved variant. (b) Individual rate constants of nucleophile-exchanged P91 variants (C118S) were determined from burst kinetics. The phosphorylation rate (k_2) was measured with FDDEP (1) and de-phosphorylation rate (k_3) was measured with paraoxon-ethyl (2). While k_2 changed by almost three orders of magnitude over evolution, k_3 remained approximately within the same order of magnitude.

Table 3.1: Microscopic rate constants for the formation (k_2) and the breakdown (k_3) of the covalent intermediate with nucleophile-exchanged (C118S) variants of P91. k_2 was determined with substrate 1 (FDDEP), k_3 was determined with substrate 2 (paraoxon-ethyl, PXN).

Enzyme variant	k_2 (s^{-1})	k_3 (s^{-1})	$\frac{k_3}{K_M}$ (PXN, $\text{M}^{-1}\text{s}^{-1}$)
P91-WT C118S	$2.4 \cdot 10^{-3}$	$1.7 \cdot 10^{-4}$	0.11
P91-R1 C118S	$3.9 \cdot 10^{-1}$	$2.3 \cdot 10^{-3}$	1.8
P91-R2 C118S	1.9	$9.2 \cdot 10^{-4}$	1.7

3.2.3 Specificity analysis with the native cysteine triad confirms a large acceleration in intermediate formation

In order to validate the effects observed in the nucleophile-exchanged enzyme variants in the native cysteine-triad enzymes, we determined their specificity profiles: As changes in transition state geometry or in the leaving group affect k_2 and k_3 differently, changes in reaction type specificity and leaving group preference can dissect the differential effects of the directed evolution campaign on intermediate formation and breakup. While reaction-type specificity is determined by the enzyme's adaptation to the transition state geometry of a reaction (k_2 and k_3), leaving-group preference is only determined by the rates of formation of the Michaelis complex (k_{-1}/k_1) and nucleophilic attack on the substrate (k_2) (Figure 3.1). Hence, major changes in intermediate hydrolysis (k_3) would only be reflected in reaction type specificity and not in leaving-group preference. In contrast, changes in the rate of intermediate formation (k_2) would affect both reaction type specificity and leaving group preference.

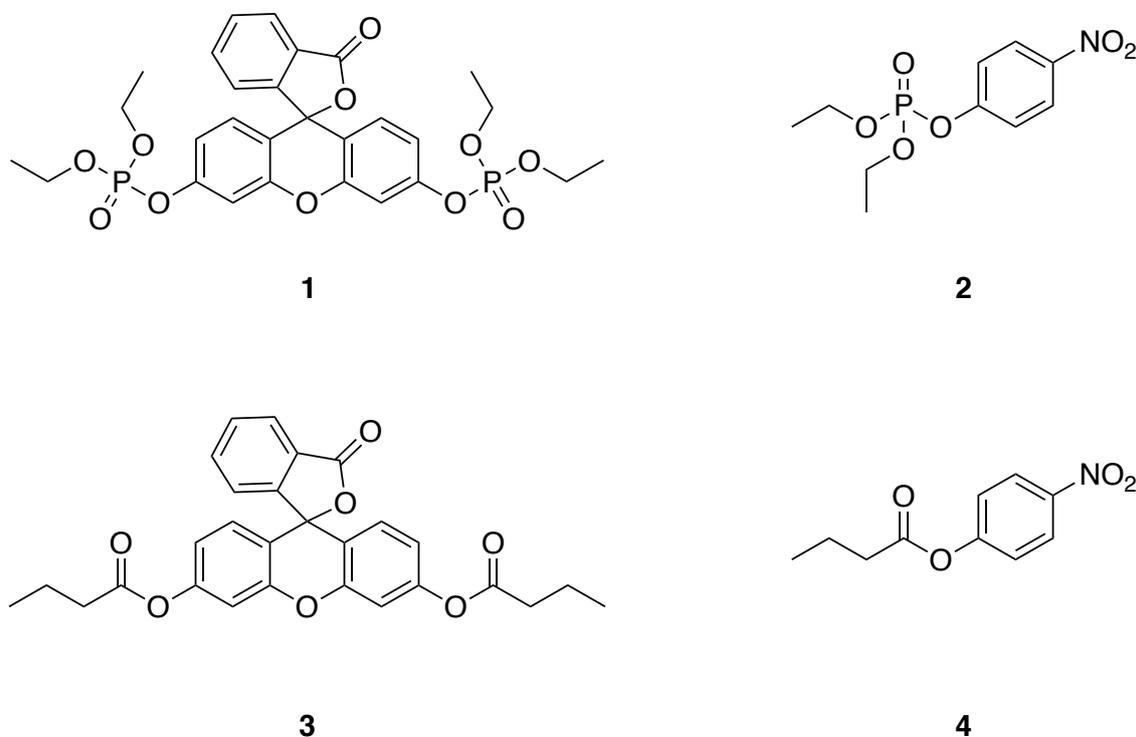


Figure 3.4: Overview of the substrates used for the specificity analysis. 1: Fluorescein di(diethylphosphate) (FDDEP); 2: Paraoxon-ethyl (PXN); 3: Fluorescein dibutylate (FDB); 4: *p*-Nitrophenyl butyrate.

In addition to its phosphotriesterase activity (trigonal-bipyramidal transition state), P91-WT has comparable catalytic parameters for the hydrolysis of carboxyesters (tetrahedral transition state). Thus, we measured arylbutylesterase and phosphotriesterase activity for two different leaving groups, *p*-nitrophenol and fluorescein, respectively (Figure 3.4). The starting point, wild-type P91, is a slightly better carboxyesterase than phosphotriesterase for both leaving groups. It also has a slight preference for fluorescein over *p*-nitrophenol as a leaving group. We observe that over the course of the directed evolution campaign, P91 specialises for both the fluorescein leaving group and for phosphotriester hydrolysis. Remarkably, this specialisation towards phosphotriesterase function and away from carboxyesterase function is strongly leaving-group dependent. The increase in phosphotriesterase activity is much more pronounced with fluorescein (≈ 400 -fold) than with *p*-nitrophenol (≈ 4 -fold) as a leaving group. Carboxyesterase activity, in contrast, decreased roughly equally with both leaving groups. The preference for fluorescein as a leaving group is therefore linked to the phosphotriester transition state, and thus rules out a simple binding effect (e.g. decrease of K_D). This finding is consistent with a significant increase in k_2 rather than in k_3 and confirms that the previous rate measurements in the nucleophile-exchanged variants are representative of a mechanism in which directed evolution strongly accelerated intermediate formation in P91.

3.2.4 Linear free-energy relationships: Intermediate formation is accelerated by improved leaving group stabilisation

To sample the nature of the transition state, we constructed a linear free energy relationship (LFER) for both the wild-type enzyme and the evolved enzyme. An LFER can be constructed by a quantitative assessment of the sensitivity of the rate of the reaction to the pK_a of the nucleophile or the leaving group, also called Brønsted analysis. The slope of this linear relationship for a range of substrates with different leaving-group pK_a values, β_{LG} , is a measure of charge accumulation during the transition state. In absence of acid-base catalysis, this parameter is directly linked to the extent of bond-formation and can indicate whether the transition state

3 The catalytic mechanism of P91

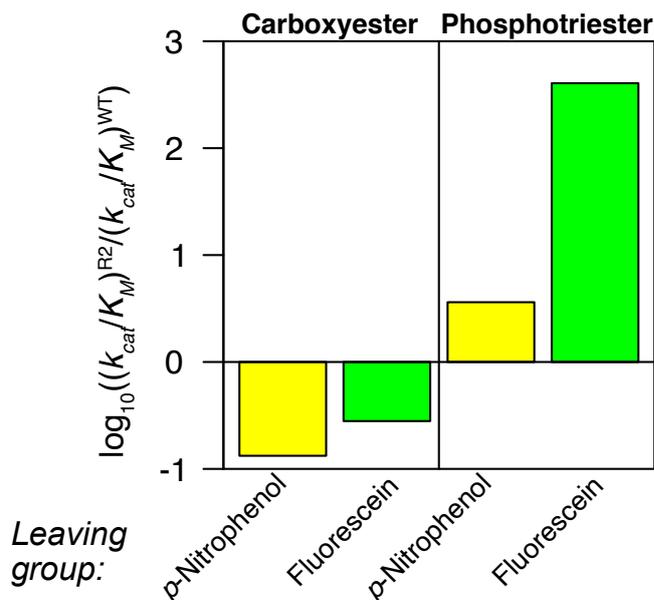


Figure 3.5: Comparison of reaction type specificity and leaving group preference for P91-WT and P91-R2 indicates that the main difference is intermediate formation. The relative change in catalytic efficiency ($\frac{k_{cat}}{K_M}$) from P91-WT to P91-R2 was measured for carboxyesterase and phosphotriesterase activity with two different leaving groups, *p*-nitrophenol (yellow) and fluorescein (green). Over the course of the directed evolution campaign, P91 specialises for both, phosphotriester substrates as well as to the fluorescein leaving group. The degree of specialisation for phosphotriesterase activity over carboxyesterase activity strongly depends on the identity of the leaving-group, suggesting that the major adaptation must affect the transition state stabilisation during the formation of the intermediate (transition state 1, k_2). The reduction in carboxyesterase activity is not comparably leaving group-specific, excluding a simple binding effect.

is rather educt- or product-like and inform about the order of bond-making and bond-breaking in the transition state (concerted or sequential). In the presence of an enzyme active site, however, the accumulating charge can be partly neutralised by catalytic groups, thereby reducing the dependence of the second order rate constant of the reaction on the pK_a of the leaving group. Brønsted analysis of an enzymatic reaction therefore measures the apparent charge accumulation on the oxygen of the leaving group and quantifies whether the offset of this charge is more or less efficient in the active site [174]. Comparing β_{LG} values of different enzymes therefore allows conclusions on which catalytic interactions are improved in the active site (Figure 3.6).

We determined k_{cat} and $\frac{k_{cat}}{K_M}$ for P91-WT and the evolved variant P91-R2 with a series of six paraoxon derivatives which differ in the pK_a of their leaving group,

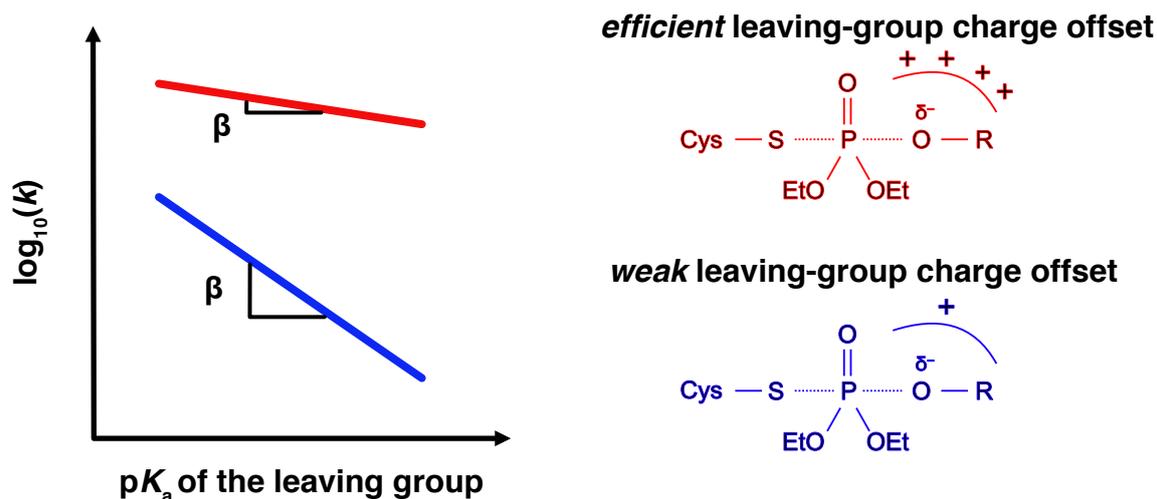


Figure 3.6: Linear free-energy relationships (LFERs) in enzymatic catalysis. By measuring the rate constant of the reaction with a range of similar substrates with different leaving-group pK_a values, it is possible to quantify the sensitivity of the reaction to charge accumulation in the transition state. The slope of this linear relationship, β_{LG} , quantifies whether the offset of this charge is more or less efficient in the active site. Comparing β_{LG} values of different enzymes therefore allows conclusions on which catalytic interactions are improved in the active site. This figure was inspired by a conference talk given by Dr Bert van Loo.

covering a pK_a range from 5.9 to 9.1 (substrates 2 and 5–10, Figure 3.7 and Table 3.2). The resulting Brønsted coefficients β_{LG} are almost identical between both catalytic parameters, k_{cat} and $\frac{k_{cat}}{K_M}$. This is the case for both enzymes and, as any significant rate-determining influence of the intermediate hydrolysis (k_3) would result in disparate effects on k_{cat} and $\frac{k_{cat}}{K_M}$ (equations 8.5 and 8.6), this is consistent with the previous finding that intermediate formation is the rate-limiting step of the reaction. We determined the Brønsted coefficient β_{LG} to be -1.1 for P91-WT ($\frac{k_{cat}}{K_M}$: $\beta_{LG} = -1.1$, k_{cat} : $\beta_{LG} = -0.95$, Figure 3.8a). In contrast, linear free-energy relationship for P91-R2 is less steep, with a β_{LG} of only -0.55 ($\frac{k_{cat}}{K_M}$: $\beta_{LG} = -0.55$, k_{cat} : $\beta_{LG} = -0.68$, Figure 3.8b). In summary, this difference in the Brønsted coefficient indicates that the acceleration of intermediate formation in P91-R2 is, at least in part, mediated by the enzyme's increased ability to offset the leaving group charge during the transition state.

As the nucleophile in P91's first transition state is a thiolate, these values are not directly comparable to the values reported for the metal-catalysed enzymatic reaction (PON1: $\beta_{LG} \approx -1.6$ [196], *BdPTE*: $\beta_{LG} \approx -1.84$ for non-diffusion-limited substrates

3 The catalytic mechanism of P91

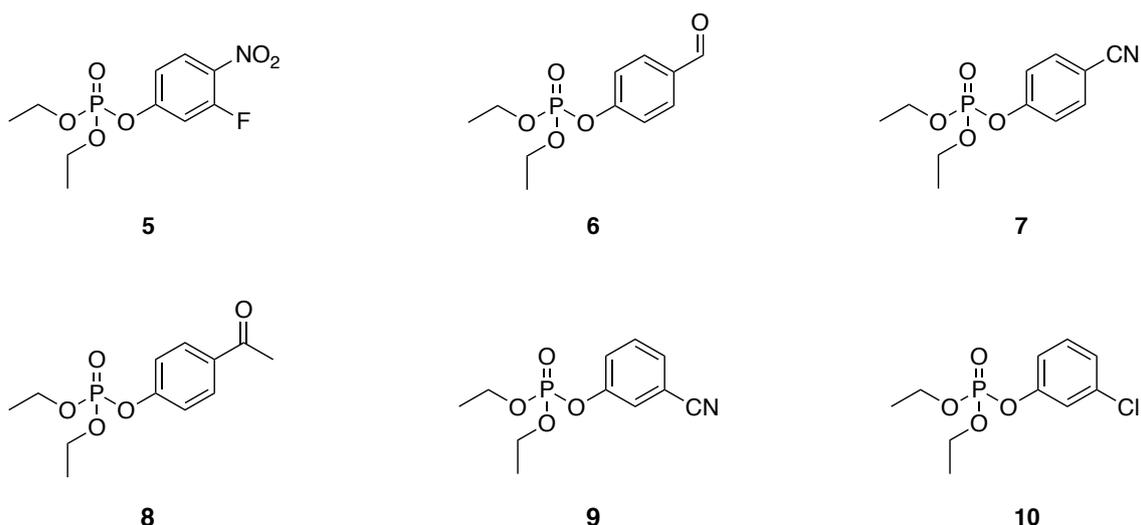


Figure 3.7: Phosphotriester substrates used to construct the linear free-energy relationships in order of increasing leaving-group pK_a . **5:** 3-Fluoro-4-nitrophenyl diethylphosphate; **6:** 4-Formylphenyl diethylphosphate; **7:** 4-Cyanophenyl diethylphosphate; **8:** 4-Acetylphenyl diethylphosphate; **9:** 3-Cyanophenyl diethylphosphate; **10:** 3-Chlorophenyl diethylphosphate.

Table 3.2: Properties of phosphotriester substrates (paraoxon-ethyl derivatives) used for the linear free-energy relationships.

Substrate	Leaving group	pK_a	Detection wavelength (nm)	Extinction coefficient (M^{-1})
5	3-fluoro-4-nitrophenol	5.94	390	10 473.2
2	4-nitrophenol	7.14	405	10 038.1
6	4-hydroxybenzaldehyde	7.66	330	12 483.8
7	4-cyanophenol	7.95	275	6 721.7
8	4-hydroxyacetophenone	8.05	320	7 242.1
9	3-cyanophenol	8.61	295	1 246.2
10	3-chlorophenol	9.12	276	796.0

[25]) and the non-enzymatic hydroxide-catalysed reaction ($\beta_{LG} \approx -0.44$), where the nucleophile is an oxyanion. Still, the fact that β_{LG} is higher in P91 than in solution indicates a hydrophobic active site (as compared to free solvent) which generates an electrostatic incompatibility, leading to more charge accumulating in the leaving group. Directed evolution improved stabilisation of the transition state by removing this clash in P91-R2, leveraging β_{LG} to a value comparable to the reaction in solution.

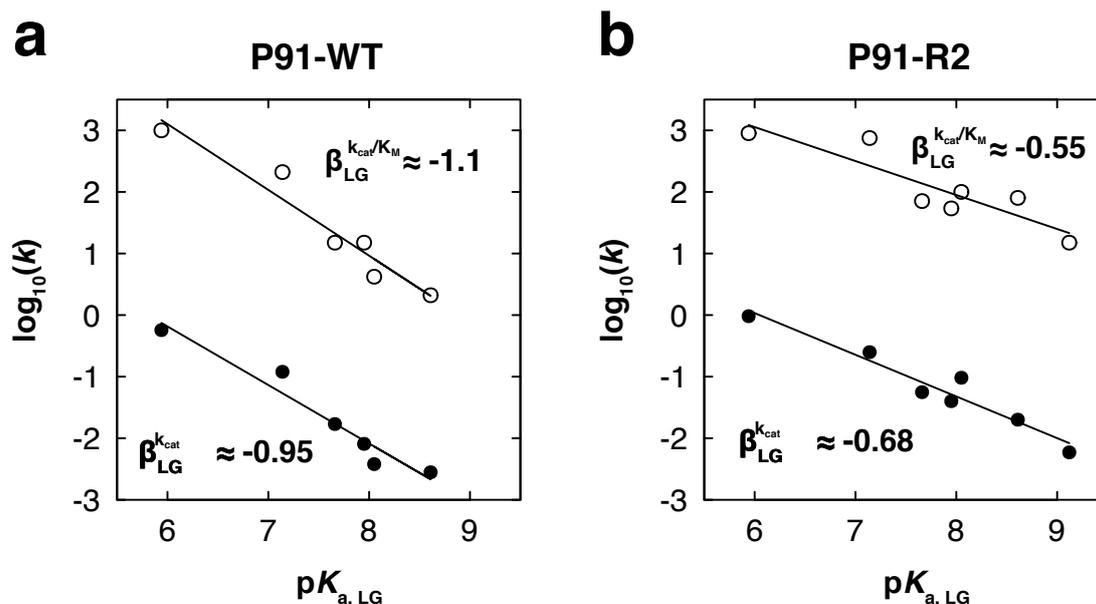


Figure 3.8: Brønsted analysis shows that the evolved variant P91-R2 accelerates intermediate formation by improved leaving group stabilisation. The Brønsted plots show the linear free-energy relationship between the rate of hydrolysis of paraoxon derivatives and the pK_a of the leaving group for P91-WT and P91-R2. Filled dots: k_{cat} in s^{-1} . Open circles: $\frac{k_{cat}}{K_M}$ in $M^{-1}s^{-1}$. **(a)** Brønsted plot for P91-WT: $\frac{k_{cat}}{K_M}$: $\beta_{LG} = -1.1$, $R^2 = 0.93$; k_{cat} : $\beta_{LG} = -0.95$, $R^2 = 0.94$. **(b)** Brønsted plot for P91-R2: $\frac{k_{cat}}{K_M}$: $\beta_{LG} = -0.55$, $R^2 = 0.80$; k_{cat} : $\beta_{LG} = -0.68$, $R^2 = 0.94$. As the slope of the linear fits (β_{LG}) is very similar for both kinetic parameters (k_{cat} and $\frac{k_{cat}}{K_M}$) intermediate formation (k_2) must be the rate-limiting step. The lower β_{LG} of P91-R2 (-0.55) as compared to P91-WT (-1.1) indicates that the evolved variant has adapted to offsetting the charge which accumulates on the leaving group during the transition state.

3.2.5 Speculations on the molecular basis of k_2 acceleration

A crystal structure of both, P91-WT and P91-R2 with a bound substrate analogue could help to elucidate the molecular basis of this effect. Despite various crystallisation efforts, no diffraction crystal of the evolved variant P91-R2 could be obtained. Likewise, no P91-WT structure with a substrate analogue could be obtained, neither by soaking nor by co-crystallisation. Although the rationalisation of occurred mutations is often very difficult and the molecular basis of directed evolution outcomes are seldom elucidated, one can nevertheless speculate on the effects of the identified mutations:

To estimate the molecular basis of P91-R2's mutations on the differential acceleration of k_2 and k_3 , one has to consider the differences between the two transition

3 *The catalytic mechanism of P91*

states undergone in the reaction (Figure 3.1). During the formation of the covalent adduct (transition state 1, TS1) the cysteine thiol is polarised and deprotonated while the charge developing on the leaving group needs to be stabilised. During the hydrolysis of the covalent adduct (transition state 2, TS2: a water molecule needs to be positioned, polarised, and deprotonated. The residues with the highest individual and combined effects upon mutation are I211 and L214 (Figure 2.2 and Table 2.1), located in a loop that is partly covering the active site entrance. The main chain of the triad histidine H199 is positioned through polar contact with the L214 backbone, which itself is in polar contact with the backbone of I211. Therefore, backbone rearrangements of either of these two residues may subtly tweak the positioning of the triad H199 and thus influence both the deprotonation of the triad nucleophile C118 (in TS1) and the deprotonation of water (in TS2). Assuming that, in analogy to the canonical ester triad mechanism in close homologues, the oxyanion formed during the transition state is stabilised by backbone nitrogen of A38 (Figure 3.1), H199 is also in the correct opposite apical position to facilitate leaving group charge offset. This way, re-positioning H199 could also influence leaving group charge offset and thus make a major contribution to accelerating k_2 , as observed in the I211W, L214V double mutant P91-R1 (Figure 3.3b). Mutating A73 alone has a similar, large effect. As A73 is on the opposing loop but in close contact with I211, any A73 mutation to bigger residue would likely introduce steric clash with I211 and thus indirectly reposition H199 as well.

In summary, all three residues with the highest effect upon mutation are located in loops partly covering the active site (loop 5 and loop 14) and are in contact with each other. This network of hydrogen bond reaches down to the triad histidine via backbone interactions and could through this path tune the polarising interactions in the triad's charge-relay system.

In addition, a comparison of the structure of P91-WT and a structural model of P91-R2 (in an AlphaFold2/ColabFold rendition [197, 198]) suggests a reshaping of the active site volume (Figure 3.9). The introduction of the large, bulky tryptophan (I211W) and the slight enlargement near the catalytic triad by the L214V mutation

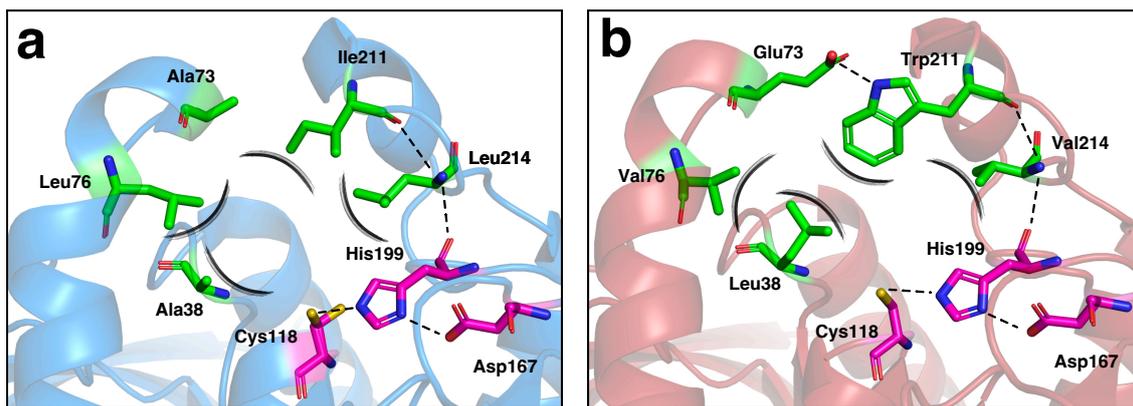


Figure 3.9: Active site comparison of wild-type and evolved P91. (a) Positions randomised for the screening are shown in the structure of P91-WT (blue). The catalytic triad is shown in magenta and positions chosen for mutation in green. Hydrogen bonds are highlighted as dotted lines, differences in bulky protrusions into the active site as brushy lines. (b) The mutations in P91-R2 (red) are shown in a structure created by AlphaFold2/ColabFold [197, 198].

reshape the active site cavity. This arrangement is rigidified by a new hydrogen bond between E73 and W211, potentially amounting to a tighter fit. Furthermore, the newly introduced W211 might additionally stabilise the developing charge on the leaving group during the transition state through pi-stacking interactions with the fluorescein leaving group.

3.3 Conclusion and outlook

How does P91 overcome covalent inactivation by phosphotriesters? Starting from the assumption that P91's catalytic triad would be innately limited by the difficult de-phosphorylation step, we tried to measure microscopic rate constants of the reaction. To our surprise, we find that P91-WT is already pre-disposed with a high de-phosphorylation rate and rate-limited by the nucleophilic attack on the substrate. Pre-steady state kinetics show that P91 is limited by interacting with the phosphotriester in the first place. Exchange of the catalytic nucleophile from a cysteine into a serine changes the rate-limiting step and makes the microscopic rate constants quantifiable. Consequently, we find that the evolved variant P91-R2 achieves its high efficiencies by accelerating the initial formation of the intermediate, rate con-

3 The catalytic mechanism of P91

stant k_2 . This is confirmed by a specificity analysis of both the wild-type and the evolved P91 variant in their native cysteine-bearing form, also excluding a simple binding effect. The specialisation for fluorescein as a leaving group and the increase in substrate inhibition also indicate that the evolved variant P91-R2 was specifically shaped by the substrate conditions that were used in the screening (“you get what you screen for” [199]). Brønsted analysis finally shows that P91-R2 achieves the acceleration of the initial phosphotriester binding by removing a charge incompatibility in the first transition state.

We argue that P91’s pre-disposition for high de-phosphylation rates can partly be explained by its cysteine triad. The de-phosphylation rate of all the very few known serine triad hydrolases (see Section 1.4.2) is in the (very low) range of $\approx 10^{-4}$ – 10^{-2} s $^{-1}$, which is in good agreement with the rates observed for the nucleophile-exchanged P91 variants (which bear a serine instead of the cysteine). The turnover rate of the evolved P91-R2 variant (with a cysteine) in contrast is two to four orders of magnitude higher than of the serine enzymes (≈ 15 s $^{-1}$, although this is an extrapolation due to the substrate inhibition). Conversely, the rate of intermediate formation k_2 is reported to be 1.3 s $^{-1}$ for the serine enzyme *Lc* α E7 and, inferring from k_{cat} , about two orders of magnitude lower in the cysteine enzyme P91-WT. However, as k_2 also depends on the leaving group, it can be significantly tweaked (e.g. by stabilising the leaving group charge), thus making further acceleration possible, as demonstrated in P91-R2 with an estimated k_2 of 15 s $^{-1}$ (assuming $k_{cat} \approx k_2$).

These differences in rate of nucleophilic attack and hydrolysis of the intermediate could be explained due to the intrinsic difference in pK_a between cysteine and serine. The rate of phosphorylation decreases with the pK_a of the nucleophile ($\beta_{nuc} = 0.3$ – 0.6 for oxyanions and amines) [200]. Assuming a similar β_{nuc} in the range of ≈ 0.5 for thiols, the rate of nucleophilic attack (phosphylation rate, k_2) is expected to be about two to three orders of magnitude slower with a cysteine than with a serine nucleophile ($\Delta pK_a (\approx 5.6) \times \beta_{nuc} (\approx 0.5) = 2$ – 3 orders of magnitude). This applies conversely to the aptitude as leaving group and thus the rate of hydrolysis upon hydroxide attack, and is in rough agreement with the differences between serine and

cysteine enzymes described above. In energetic terms, introducing a cysteine as a nucleophile increases the ground state energy of the intermediate and thus lowers the barrier to hydrolysing it.

The difference in pK_a between serine and cysteine in different catalytic triads should also have consequences on the susceptibility for a (fatal) side reaction, hydrolytic de-alkylation of the phosphotriester adduct, dubbed ageing (Figure 1.2): Upon water/hydroxide attack at the phosphorylated intermediate, the transition state collapses with preferential loss of the best leaving group. In the case of a serine triad, one of the two ethoxy side groups of the phosphotriester is prone to leave instead of the serine, resulting in a negatively charged adduct which is resistant to further hydrolysis. As the cysteine thiol group is a much better leaving group than any of the phosphate's ethoxy side groups, P91 should be much more resistant to ageing than serine triad enzymes.

This finding raises the question whether acetylcholinesterase and its homologues would become proficient phosphotriesterases when exchanging their catalytic serine for a cysteine. Nucleophile exchange in a triad is usually deleterious to the native activity but has been shown to be rescuable by directed evolution [201, 202].

Overall, the kinetic analysis of P91 and its evolved homologue suggests that other enzymes with a cysteine triad could possess significant phosphotriesterase activity, a possibility that demands further exploration.

4 Phylogenetic exploration of P91 homologues

After having searched the local sequence space around P91 by directed evolution, in this chapter, I phylogenetically explore the more distant sequence space in the α/β hydrolase superfamily around P91 that has already been charted by natural evolution.

External contributions in this chapter:

Crystallisation trials, X-ray data acquisition and analysis were performed by Dr Eleanor Campbell.

4.1 Introduction

In the previous chapter I have demonstrated that high-turnover phosphotriester hydrolysis is not restricted to metal-dependent catalytic motifs. By evolving P91 into a proficient, metal-independent phosphotriesterase I have shown that a catalytic Cys-His-Asp triad can accommodate this new activity at catalytic efficiencies matching those of metal-dependent enzymes. Still, none of the naturally evolved organophosphate hydrolases uses the metal-independent, nucleophilic catalytic mechanism of P91.

P91's activity is even more unexpected as other, homologous enzymes from the α/β hydrolase superfamily, among which acetylcholinesterase is the most prominent, are covalently inhibited by organophosphates at the catalytic nucleophile of their triad. Furthermore, although the starting activity of wild-type P91 is weak, it is already non-negligible ($\frac{k_{cat}}{K_M} \approx 10^2 \text{ M}^{-1}\text{s}^{-1}$), with a turnover rate ($k_{cat} \approx 10^{-1} \text{ s}^{-1}$) that surpasses those of homologous insect carboxylesterases ($k_{cat} \approx 10^{-3} \text{ s}^{-1}$) which also have a catalytic triad and evolved under strong selection pressure of organophosphate insecticides [58]. It therefore still remains unclear which features of P91 confer its promiscuous activity and enable it to turn over phosphotriesters in the first place. What are the features that convert an α/β hydrolase from a no-turnover or single-turnover enzyme into a multiple-turnover enzyme? Is P91 an exception, and if yes, what makes it special and distinguishes it from its inhibition-susceptible homologues? Conversely, if this does not hold true, how common is phosphotriesterase activity in the α/β hydrolase superfamily?

4.1.1 The diene lactone hydrolase (DLH) protein family

Within the α/β hydrolase superfamily, P91 resides in a sequence cluster which is annotated as the diene lactone hydrolase family (DLH) (Figure 4.1). DLH family proteins are characterised by a Cys–His–Asp triad and are the archetype of the

canonical α/β hydrolase fold. The annotation of the DLH family is based on a single extensively functionally characterised enzyme, dienelactone hydrolase from *Pseudomonas knackmussii* (*PkDLH*), whose native activity is the hydrolysis of dienelactone intermediates in the degradation of halogenated aromatic compounds in the β -keto adipate pathway (also known as halocatechol pathway) [164, 203–205]. However, sequence-based functional annotation very often stands on thin experimental evidence and therefore often falls short of representing the true functional diversity of uncharacterised proteins [206, 207]. Apart from *PkDLH*, functional characterisation in the DLH family is scarce. A homologue of *PkDLH* has been functionally characterised as acting on the same substrate in a soil bacterium from the *Bradyrhizobium* genus [208]. One further, rare example of a functionally characterised DLH family protein is the enzyme TropI from the fungus *Talaromyces stipitatus* which hydrolyses an organic acid anhydride intermediate in the biosynthesis of the natural product stipitatic acid [209] (Figure 4.3).

As shown in the previous chapter, P91 is pre-disposed for fast phosphoester intermediate hydrolysis due to its Cys–His–Asp triad. Given that this catalytic triad is not unique to P91, how high is the ‘promiscuity potential’ towards organophosphate hydrolysis in the α/β hydrolase superfamily or, more specifically, in the dienelactone hydrolase family?

4.2 Results and Discussion

4.2.1 Exploration of P91’s surrounding sequence space

In order to probe the sequence space surrounding P91, several putative proteins with high sequence similarity to P91 were cloned and expressed in *Escherichia coli*, purified, and tested for phosphotriesterase activity. Additionally, several other homologues from the DLH family which are further away in sequence space were also tested for PTE activity: all three homologues of which a structure is publicly

4 Phylogenetic exploration of P91 homologues

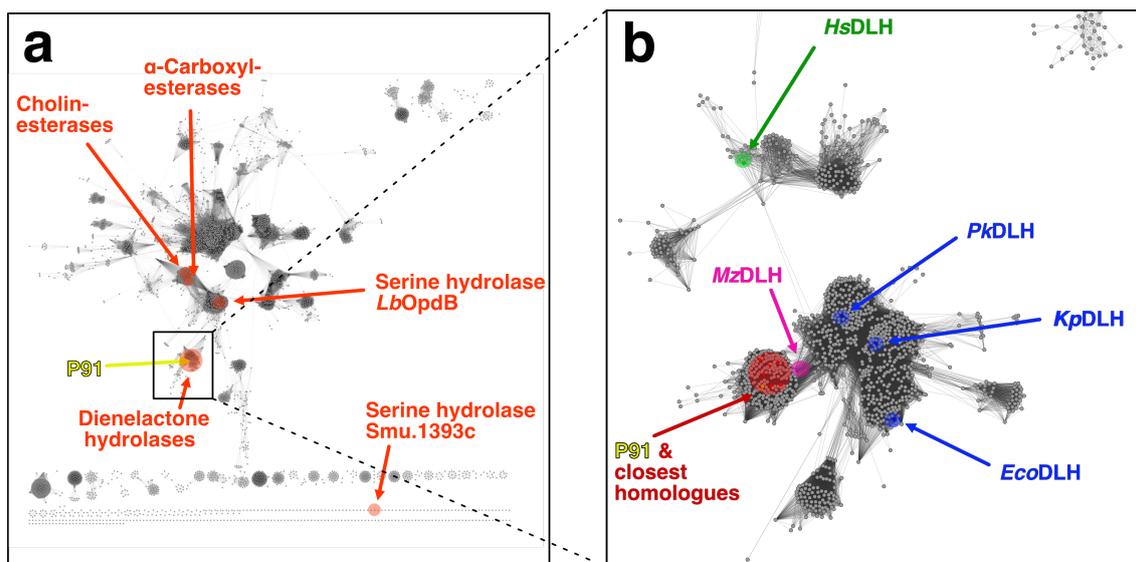


Figure 4.1: Sequence similarity networks (SSNs) of the α/β hydrolase superfamily and the dienelactone hydrolase protein family. (a) SSN of the α/β hydrolase superfamily (Pfam: CL0028) containing 5 062 sequences. Nodes with an alignment score ≥ 12 are connected by an edge, corresponding to a median sequence identity of $\approx 27\%$. Enzymes with reported organophosphate hydrolase activity are highlighted in orange, P91 is highlighted in yellow in the cluster of the dienelactone hydrolase protein family (box). **(b)** A separate SSN of the dienelactone hydrolase (DLH) protein family (Pfam: PF01738), containing 1 992 sequences. Nodes with an alignment score ≥ 25 are connected by an edge, corresponding to a median sequence identity of $\approx 32\%$. In this SSN, separate small clusters and unconnected nodes (singletons) were removed for clarity. All enzymes characterised in this study are highlighted: P91 (yellow), P91's closest homologues (red), an enzyme identified by sequence similarity of a neighbouring gene fragment (*MzDLH*, magenta), DLH family proteins with a publicly available structure (*PkDLH*, *EcoDLH*, *KpDLH*, blue), and the only homologue of P91 in the human genome (*HsDLH*, green).

available, one homologue with a similar adjacent gene, and the closest homologue from the human genome. Furthermore, two more distant homologues with reported phosphotriesterase activity from the literature were also purified and tested (Figure 4.1).

P91's direct vicinity in sequence space: *ScDLH*, *SfDLH*, *SifDLH*, *SaDLH*, *m1DLH*, *m3DLH*, and *HmDLH*

Of the closest known homologues of P91, as identified in the NCBI non-redundant nucleotide database using BLAST [210], five were selected for characterisation. These sequences are all annotated as DLH family proteins and originate from the gammaproteobacteria *Haliera mangrovi*, *Sinimarini bacterium flocculans*, *Solimonas fluminis*, *Solimonas cavernae*, and *Solimonas aquatica* and were called *HmDLH*, *SifDLH*, *SfDLH*, *ScDLH*, and *SaDLH*, respectively (Table 4.1). Two further close homologues were identified from metagenomic genome assemblies of uncharacterised bacteria and selected for characterisation, *m1DLH* from the MGnify database [211] and *m3DLH* from the MarDB database [212]. The sequence identity of those homologues to P91 ranges from 57–66 % (E -values 10^{-99} – 10^{-116}). Interestingly, the underlying native function does not seem to be highly conserved in the genera *Haliera*, *Sinimarini bacterium*, or *Solimonas*, as closely related bacteria within each of these genera lack comparably close homologues to P91. For example, the closest P91 homologues in the genomes of the respective type species of their genera, *Haliera salexigens* and *Solimonas soli* show only 34 % identity (E -value $6 \cdot 10^{-39}$) and 23 % identity (E -value $1 \cdot 10^{-6}$), respectively. This might indicate that the functional context in which P91 and its close homologues are embedded is linked to a certain ecological niche. Notably, all these bacteria were isolated from aquatic environments, such as freshwater, sea water, or marine sediments (Table 4.1).

4 Phylogenetic exploration of P91 homologues

A homologue inferred from genomic context similarity: *MzDLH*

P91 was originally identified from a functional metagenomic screening of short fragments of environmentally extracted DNA [149]. The DNA fragment on which P91 is located is approximately 3000 bp long and comprises three further genes, two of which flank the *P91* gene in the same orientation (inserts *A* and *B* in Figure 4.2, lengths 576 bp and 1170 bp, respectively). Both can be annotated as Major Facilitator Superfamily (MFS) transporters by sequence homology. One of the closest homologues of the downstream gene is a 4-hydroxybenzoate transporter (36 % sequence identity, E -value $9 \cdot 10^{-81}$). 4-Hydroxybenzoate is a precursor molecule that is fed into the β -keto adipate pathway [205], indicating that P91 might be involved in degradation of aromatic compounds via this pathway, just as *PkDLH*, the only functionally characterised homologue of P91 (see below). On the 3'-end of the original DNA fragment, a short, truncated C-terminal gene fragment is encoded on the opposite strand (insert *C* in Figure 4.2, length 126 bp). Its sequence matches with 88 % identity (E -value $7 \cdot 10^{-17}$) with the C-terminus of a predicted NADP-specific glutamate dehydrogenase in the genome of the marine gammaproteobacterium *Marinobacter zhejiangensis*, [213] isolated from marine sediment.

Following the hypothesis that a similar genomic context could indicate a similar function, for example the processing of a similar native substrate, and therefore imply a similar pattern of promiscuous activity, the closest homologue of P91 in the *Marinobacter zhejiangensis* genome was selected for testing. After deduction of an N-terminal signal peptide, this DLH family protein, called *MzDLH*, shares 34 % identity with P91 (E -value $2 \cdot 10^{-46}$).

Homologues with a publicly available structure: *PkDLH*, *KpDLH*, and *EcoDLH*

Moving further away in sequence space beyond the direct proximity to P91, I sought to test all DLH family proteins for phosphotriesterase activity for which there is a

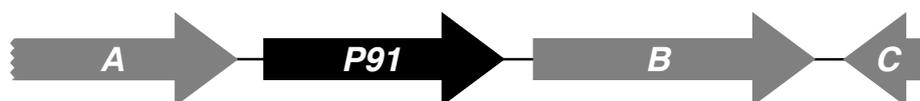


Figure 4.2: Genomic context of P91. The DNA fragment on which P91 was identified by a functional metagenomic screening [149] contains three further potential genes. Genes *A* and *B* can be annotated as Major Facilitator Superfamily (MFS) transporters, whereas *C* matches the C-terminal fragment of a NADP-specific glutamate dehydrogenase. The sequence similarity of gene fragment *C* was the basis for the identification of *MzDLH*.

protein structure available, in order to facilitate the analysis of potential structure-function relationships. Apart from P91, there are three structures of DLH family proteins publicly available on the Protein Data Bank (PDB): *PkDLH*, *KpDLH*, and *EcoDLH*.

PkDLH from the bacterium *Pseudomonas knackmussi* is the eponymous enzyme of the dienelactone hydrolase family and at the same time the archetype of the α/β hydrolase superfamily fold. It is the DLH enzyme studied in most detail, and the only one for which both the native function and several structures have been elucidated. Isolated from sewage water, *PkDLH* hydrolyses 4-carboxymethylenebut-2-en-4-olide (or dienelactone), an intermediate in the catabolic degradation of halogenated aromatics in the β -keto adipate pathway (Figure 4.3) [203–205].

With 18 structures publicly available on the PDB, *PkDLH* has been extensively studied structurally and subjected to directed evolution towards activity on long-chain esters [186]. In most studies, the oxidation-prone cysteine nucleophile of the catalytic triad has been mutated into a serine, slowing down the hydrolysis of the acyl-enzyme intermediate and consequently changing the native lactonase into isomerase activity [214, 215]. *PkDLH* is not very closely related to P91 on the protein sequence level and displays only 25 % identity (E -value $2 \cdot 10^{-10}$).

Structures are available for two further DLH family proteins, albeit in absence of reported functional characterisation. These proteins are from the genomes of *Kleb-*

4 Phylogenetic exploration of P91 homologues

siella pneumoniae (*KpDLH*, PDB ID: 3F67) and *Escherichia coli* (*EcoDLH*, PDB ID: 4ZV9), both ubiquitous gammaproteobacteria, and were also included into the panel of characterised enzymes. *PkDLH*, *KpDLH* and *EcoDLH* are only relatively distantly related to P91, with sequence identities of 24 % (E -value $4 \cdot 10^{-14}$) and 31 % identity (E -value $1 \cdot 10^{-12}$), respectively (Figure 4.1b).

A human homologue: *HsDLH*

The closest homologue of P91 and *PkDLH* in the human genome has been called JS-1 and carboxymethylenebutenolidase (CMBL) in previous studies (UniProtKB: Q96DG6). It also represents the only protein from the DLH family in the human genome. As its native function remains unknown and its activity on carboxymethylenebutenolide has not been tested, only inferred from homology, I will refer to it as the dienelactone hydrolase-like protein from *Homo sapiens* (*HsDLH*), in line with my naming of the other studied P91 homologues of unknown native function. Being overexpressed in about 50 % of esophageal squamous cell carcinoma cases, its overexpression has been linked to tumorigenesis [216]. *HsDLH* displays promiscuous esterase activity and has been identified as one of several enzymes involved in the activation of the antihypertensive prodrug olmesartan medoxomil and similar prodrugs by cleavage of an ester bond [217, 218] (Figure 4.3). Interestingly, it shares this promiscuous substrate with the human paraoxonase PON-1, which is also involved in the activation of olmetarsan medoxomil [219]. *HsDLH* is cytosolically expressed in the liver and the intestine and has previously been heterologously expressed and purified from *E. coli* [217]. *HsDLH* is relatively distantly related to P91 as compared to the other DLH family proteins tested, with only 20 % sequence identity (alignment bit score: 35.0, E -value: $5 \cdot 10^{-7}$, Figure 4.1b).

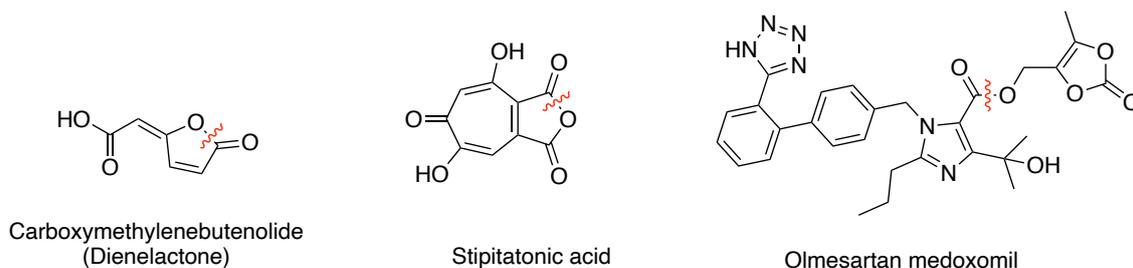


Figure 4.3: Reported substrates of DLH family proteins. The scissile bond is highlighted in red. Carboxymethylenebutenolide (a dienelactone) is hydrolysed by *PkDLH* as an intermediate in the catabolic degradation of halogenated aromatics in the β -ketoacid pathway [203, 204]. Stipitonic acid is hydrolysed by the fungal DLH family enzyme *TropI* in the biosynthesis of the natural product stipitonic acid [209]. Olmesartan medoxomil is one of several ester prodrugs which are hydrolytically activated in the human body by *HsDLH* [217].

Other homologues from the α/β hydrolase superfamily with reported phosphotriesterase activity

Looking beyond the DLH family, I searched the literature for further proteins from the α/β hydrolase superfamily which have been reported to display organophosphate hydrolase activity, apart from the well-studied point mutants of human butyrylcholinesterase [169] and insect carboxylesterase [168].

Very slow multi-phasic hydrolysis of the organophosphorothiate pesticide malathion has also been reported for human hepatic carboxylesterases, which are members of the α/β hydrolase superfamily with a catalytic serine [220]. Catalytic efficiency seems to be governed by a very low de-phosphorylation step, with steady-state turnover numbers in the range of $\approx 10^{-2}$ – 10^{-3} s $^{-1}$, comparable to the well-studied insect carboxylesterase *Lc α E7*. However, in this publication, kinetic experiments were carried out with a limited range of substrate concentrations and poorly defined enzyme samples from human biopsies, leaving a more detailed investigation to be desired.

Islam *et al.* reported a serine hydrolase from the α/β hydrolase superfamily (but not within the DLH family) from *Lactobacillus brevis* WCP902, *LbOpdB*, isolated from kimchi, that is able to degrade several phosphorothioate pesticides (diethyl-substi-

4 Phylogenetic exploration of P91 homologues

tuted phosphotriesters with a P=S double bond), albeit without detailed kinetic or mechanistic characterisation [221].

Jacob *et al.* report a promiscuous phosphotriesterase from the α/β hydrolase superfamily, identified by in silico screening of structures from the PDB, docked against the organophosphate chemical warfare agent VX, in an effort to identify new scaffolds for potential new catalytic bioscavenger against organophosphate poisoning [222]. This serine hydrolase from *Streptococcus mutans*, Smu.1393c, was reported to display considerable levels of paraoxonase activity ($\frac{k_{cat}}{K_M} \approx 5 \cdot 10^3 \text{ M}^{-1}\text{s}^{-1}$), however in a metal-dependent fashion (co-expression with 500 μM ZnCl_2 was required for enzymatic activity). A weak, metal-independent activity against the phosphorothioate pesticide omethoate ($\frac{k_{cat}}{K_M} \approx 2 \cdot 10^{-1} \text{ M}^{-1}\text{s}^{-1}$) was also reported.

Consequently, the two bacterial enzymes *LbOpdB* and Smu.1393c were added to the panel of analysed enzymes.

Among the well-studied α/β hydrolase enzymes displaying PTE activity, the human butyrylcholinesterase (BChE) G117H mutant stands out as having the highest k_{cat} (up to $\approx 0.09 \text{ s}^{-1}$ for paraoxon [170]), thus almost reaching the activity level of P91, although further values reported in the literature vary between 0.008 s^{-1} and 0.02 s^{-1} [169, 223]. Butyrylcholinesterase is a close homologue of acetylcholinesterase found in blood serum and has, due to its broader substrate spectrum, been proposed as a candidate bioscavenger against organophosphate poisoning [224]. Its native function is not entirely elucidated but it has been shown to be involved in the regulation of the hunger and aggression hormone ghrelin [225]. Native human BChE has three disulfide bridges and does not solubly express in *E. coli*. However, computational sequence optimisation, called PROSS, has recently been introduced to increase the stability of proteins and make them amenable to bacterial expression. The proof-of-principle of PROSS was shown by stabilising acetylcholinesterase [226] and lately also been applied to butyrylcholinesterase [224]. Bacterial expression of BChE G117H was therefore achieved from a PROSS-optimised sequence (hBChE-7) with a disulfide-bridge-promoting thioredoxin A fu-

sion tag in an *E. coli* strain optimised for the formation of disulfide bridges (T7 SHuffle Express).

Promiscuous phosphotriesterase activity is widespread in the DLH protein family

All the enzymes specified above were purified and tested for activity on the phosphotriester substrates FDDEP (1) and paraoxon-ethyl (2). The reported activity of the two serine hydrolases *LbOpdB* and *Smu.1393c* could not be confirmed, with none of the two enzymes displaying activity towards the tested substrates. However, this could also be attributed to the difference in substrates used (phosphotriester instead of phosphorothioate). BChE G117H did not show significant activity on paraoxon-ethyl but low activity on FDDEP could be detected, with $k_{cat} \approx 4 \cdot 10^{-4} \text{ s}^{-1}$. The lack of activity on paraoxon and the low turnover rate with FDDEP is not in line with the values reported in the literature (see above). One possible explanation is that the used PROSS construct differs from the human sequence by 47 mutations, and although mutations were excluded from the active site, any of these could affect the phosphotriesterase activity.

Surprisingly, all but two of the DLH family enzymes show some level of turnover of both or one of the two substrates. Reaction progress curves follow a single exponential product increase (no visible burst kinetics on a timescale of minutes to hours) so that Michaelis-Menten kinetics could be measured (Figures 4.4a and Table 4.1). Only for *PkDLH* and *EcoDLH* no activity could be detected.

In general, paraoxonase activity levels follow the pattern of sequence similarity to P91, with higher activities at higher similarities. The two closest homologues to P91, *m3DLH* and *HmDLH*, even surpass the catalytic efficiency of P91 by one to two orders of magnitude, with $\frac{k_{cat}}{K_M}$ of 5 900 and 10 000 $\text{M}^{-1}\text{s}^{-1}$, respectively. All enzymes with paraoxonase activity also show activity towards the fluorogenic phosphotriester substrate FDDEP (1). Due to the lower pK_a of fluorescein ($pK_a \approx 6.4$) as compared to *p*-nitrophenol ($pK_a \approx 7.1$), fluorescein acts as a better leaving group,

4 Phylogenetic exploration of P91 homologues

making the phosphotriester FDDEP more reactive than paraoxon. Indeed, the two least active phosphotriesterases among the tested enzymes, *KpDLH* and *HsDLH*, only show activity towards FDDEP but not towards paraoxon. Notably, *HsDLH*, which is even more distant to P91 in sequence space than the non-active homologues *PkDLH* and *EcoDLH*, does display some very weak activity towards FDDEP ($k_{cat} \approx 1.6 \cdot 10^{-4} \text{ s}^{-1} \approx 0.58 \text{ h}^{-1}$) and therefore represents, together with the well-known metal-dependent paraoxonase PON1 [28], the second human phosphotriesterase to be reported.

Crystallisation trials were set up for all characterised enzymes showing phosphotriesterase activity and crystal structures were obtained for *m3DLH* (PDB ID: 7JKA), *ScDLH* (PDB ID: 7JIZ, *to be released*), *SfDLH* (PDB ID: 7JOP, *to be released*), and *SaDLH* (to be submitted to the PDB), doubling the number of DLH family proteins with an available crystal structure. All structures are very similar to the structure of P91 and share the same principal active site features, including the presence of two alternative conformations of the active-site cysteine. In the structure of *SaDLH*, the electron density of the triad cysteine indicates a tetrameric adduct. This has previously been observed in a structure of *PkDLH* and identified as a sulfonate adduct (R-SO₃H) formed through oxidation of the cysteine [164, 227].

In order to visualise their phylogenetic relationships, all DLH family proteins were mapped into a sequence-similarity network, revealing that most active homologues fall into a close cluster together with P91, while the more distantly related enzymes *KpDLH*, *PkDLH*, and *HsDLH* each fall into separate clusters (SSN, Figures 4.1b and 4.4c). Interestingly, despite their close homology, catalytic efficiencies for phosphotriesterase activity within the sequence cluster around P91 still span across at least four orders of magnitude, raising the questions which features in the otherwise very similar enzymes determine this gradient in activity.

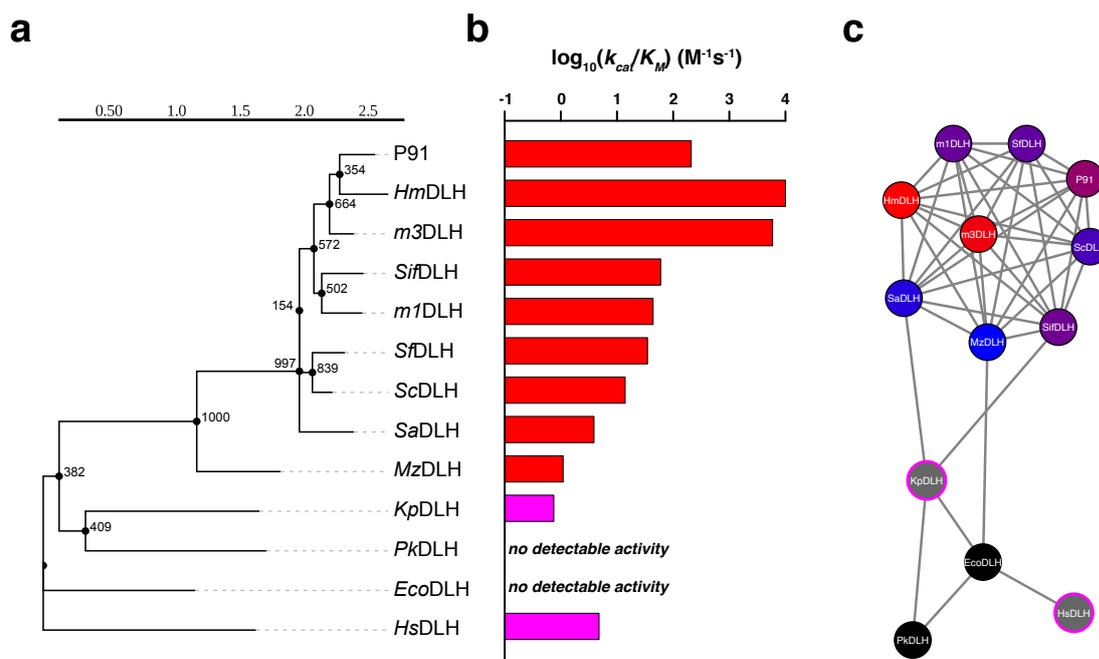


Figure 4.4: Phosphotriesterase activity is widespread in the DLH protein family.

Of all tested P91 homologues from the DLH protein family, only two (*PkDLH*, *EcoDLH*) show no detectable activity on the two phosphotriester substrates 1 (paraoxon-ethyl) or 2 (FDDEP). **(a)** Phylogenetic tree of all tested DLH family proteins. Bootstrap values are given for $N = 1000$. **(b)** Barplot of catalytic efficiencies for phosphotriester hydrolysis ($\log_{10}(\frac{k_{cat}}{K_M})$). Bars displaying values for paraoxon (1) are shown in red, bars displaying values for FDDEP (2) are shown in magenta. **(c)** Sequence similarity network of all tested DLH family proteins. Nodes with an alignment score ≥ 12 are connected by an edge. Nodes are coloured according to their phosphotriesterase activity ($\log_{10}(\frac{k_{cat}}{K_M})$) for paraoxon) from blue (low activity) to red (high activity). Enzymes displaying no activity are coloured in black (*PkDLH*, *EcoDLH*), enzymes with marginal activity (no activity on paraoxon, but low activity on the phosphotriester substrate FDDEP) are shown in grey with a magenta outline (*KpDLH*, *HsDLH*).

Table 4.1: Overview of all characterised P91 homologues, in order of sequence similarity to P91. Alignments were done using blastp at default parameters. Kinetic values marked with an asterisk were measured with substrate 1 (FDDEP), all others with substrate 2 (paraoxon-ethyl). For butyrylcholinesterase (hBChE-7 G117H), no significant sequence similarity could be found using the blastp multiple sequence alignment.

Enzyme	Sequence identity to P91 (%)	Alignment bit score S'	Organism of origin	Source environment	Phosphotriesterase kinetics		
					k_{cat} (s ⁻¹)	K_M (mM)	$\frac{k_{cat}}{K_M}$ (M ⁻¹ s ⁻¹)
P91	100	486	unknown (metagenomic) [149, 228]	thermophilic ripe compost (Groningen, The Netherlands)	0.12	0.58	210
<i>m3DLH</i>	66	318	<i>Halieaceae</i> bacterium SZUA-439 (metagenomic)	marine sulfide sediment from hydrothermal vent (South Atlantic Ridge)	6.5	1.1	5900
<i>HmDLH</i>	60	291	<i>Haliea</i> sp. SAOS-164 (<i>Haliea mangrovi</i>)	mangrove soil (Goa, India)	14	1.4	10000
<i>ScDLH</i>	59	288	<i>Solimonas</i> sp. K1W22B-7 (<i>Solimonas cavernae</i>)	water in a karst cave (Guizhou, China)	0.11	7.6	14
<i>m1DLH</i>	57	283	unknown (metagenomic)	seawater from aquaculture (Yantai, China)	0.20	4.6	44
<i>SjDLH</i>	57	282	<i>Sinimarinibacterium flocculans</i> [229]	offshore surface seawater (South China Sea)	0.59	9.8	60
<i>SjDLH</i>	58	280	<i>Solimonas fluminis</i> [230]	Han river (Seoul, South Korea)	0.19	5.6	35
<i>SaDLH</i>	57	275	<i>Solimonas aquatica</i> [231]	freshwater spring (Kaohsiung, Taiwan)	0.014	3.5	3.9
<i>MzDLH</i>	34	142	<i>Marinobacter zhejiangensis</i> [213]	marine sediment (East China Sea)	0.006	5.6	1.1
<i>KpDLH</i>	24	56.2	<i>Klebsiella pneumoniae</i>	soil/sewage water	0.00003*	0.040*	0.75*
<i>EcoDLH</i>	31	51.6	<i>Escherichia coli</i>	human gut	<i>no detectable activity</i>		
<i>PkDLH</i>	25	45.4	<i>Pseudomonas knackmussii</i> [232, 233]	sewage treatment plant (Göttingen, Germany)	<i>no detectable activity</i>		
<i>HsDLH</i>	20	35.0	<i>Homo sapiens</i>	liver/gut cell cytosole	0.00016*	0.033*	4.8*
hBChE-7 G117H	–	–	<i>Homo sapiens</i>	blood plasma	0.00041*	0.035*	12*

A native phosphotriester-like substrate?

The activity levels displayed by *m3DLH* and *HmDLH* are very high for promiscuous activities and come close to the median activity levels of naturally evolved, secondary metabolism enzymes (median $\frac{k_{cat}}{K_M}$ of $7 \cdot 10^4 \text{ M}^{-1}\text{s}^{-1}$, median k_{cat} of 2.5 s^{-1}) [190]. Could this be the result of natural selection rather than accidental promiscuity and reflect the presence of a paraoxon-like, naturally occurring native substrate, e.g. in aquatic environments? All newly identified phosphotriesterases stem from aquatic environments (apart from the human enzyme *HsDLH*). In such environments, metal-dependent organophosphate hydrolysis has previously been identified in eukaryotic organisms, namely squids (DFPase in *Loligo vulgaris* [234]) and mussels (OPAA in *Rangia cuneata* [235] and *Mytilus edulis* [236]) as well as an unidentified enzyme in sweetwater fish livers [237] (potentially a PON1 orthologue). Although organophosphates inactivate a wide range of enzymes, acute toxicity and therefore strongest selection pressure stems from the inhibition of acetylcholinesterase in the nervous system. Therefore, it is surprising that high phosphotriesterase activity predominantly evolved in prokaryotic organisms, which lack a nervous system. However, many environments are phosphate-limited, and the ability to break down pesticides into inorganic phosphate might bring a selective advantage. *HmDLH* comes from an environment of periodically flooded mangrove soil whereas *m3DLH* was isolated from a bacterium living in sulfide sediment from a hydrothermal deep vent on the South Atlantic ridge. Both environments are often phosphate-limited due to the accumulation of arsenic [238, 239]. Due to its similarity to phosphate, high arsenate concentrations can severely limit phosphate bioavailability [240] and could exert high selective pressure to evolve strategies of sourcing inorganic phosphate from potential anthropogenic organophosphate pollutants (such as traces of organophosphate pesticides, plasticisers and flame retardants). Beyond anthropogenic organophosphates, one known naturally occurring organophosphate acetylcholinesterase inhibitor is guanitoxin (formerly anatoxin-a(S)), a complex phosphodiester produced by sweetwater cyanobacteria that is known to cause mass poisonings among wildlife animals and cattle [241–243]. However, guanitoxin is a charged phospho-

4 Phylogenetic exploration of P91 homologues

diester and, due to its different transition state charge, unlikely to cross-induce promiscuous phosphotriesterase activity.

4.2.2 Sampling the transition-state adaptation

Promiscuity profiles

In order to address the question, whether the identified phosphotriesterase activity is the result of evolutionary adaptation to a similar, native substrate, a mechanistic approach was pursued. Evolutionary pressure induces an adaptation to a specific transition state geometry, as previously observed in P91-R2. Therefore, the transition state adaptation of all newly identified phosphotriesterases was surveyed.

To hydrolyse organophosphates, esterases have to change their adaptation from stabilising the tetrahedral transition state of carboxyesters to the trigonal-bipyramidal transition state of phosphate esters (Figure 1.7). For most deeply studied phosphotriesterases, evidence points towards an origin as lactonases [32, 42, 196] and also P91 has promiscuous lactonase activity and is closely related to a proven lactonase. The transition state of lactones is not a perfect tetrahedron due to distortion by the ring strain and therefore more organophosphate-like: one angle is bigger than the tetrahedron angle of 109.5° , closer to the a triangular angle of 120° . It can therefore be speculated that lactonases are already pre-evolved to stabilise a more organophosphate-like transition state, providing an explanation for the high prevalence of ex-lactonases among all known phosphotriesterases.

Three different transition states can therefore be represented with the three different substrates carboxyester (tetrahedron), lactone (distorted tetrahedron), and phosphotriester (trigonal-bipyramid). To test whether there is a correlation or trade-off between the phosphotriesterase, carboxyesterase, and lactonase activities, indicating a predisposition or specialisation for one of these substrate classes, the activities

of the homologues towards these substrates were measured. For all studied P91 homologues that act on paraoxon, activity was also measured towards the carboxyester *p*-nitrophenyl butyrate. Lactonase activity was additionally measured for a subset of the homologues (*Mz*DLH, *Sa*DLH, *Sc*DLH, *m1*DLH) with the substrate dihydrocoumarin (Figures 4.5 and 4.6).

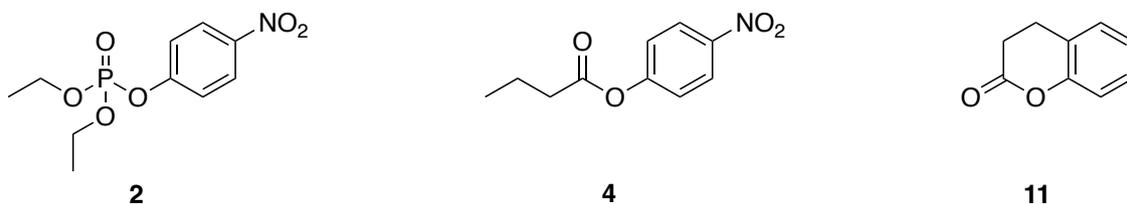


Figure 4.5: Substrates used to measure the promiscuity profiles of P91 and its DLH family homologues. The phosphotriester paraoxon-ethyl (**2**), the carboxyester *p*-nitrophenyl butyrate (**4**), and the lactone dihydrocoumarin (**11**), each representing a different transition state geometry.

In all enzymes, carboxyesterase activity turned out to be the highest activity and phosphotriesterase activity always the lowest activity. The relationship between these activities is scattered and no correlation or trade-off between esterase activity and phosphotriesterase activity is discernible (Figure 4.6a). Lactonase activities lie in between carboxyesterase and phosphotriesterase activity and give a similar, scattered picture in terms of trade-off, although more measurements are needed with the lactone to clearly define or refute a correlation (Figure 4.6b and c).

In summary, the spread in phosphotriesterase activity that can be observed among the studied DLH family proteins does not proportionally translate into activities towards other substrates. This indicates that the reactivity differences are not governed by differences in adaptation to a specific transition state geometry but might be due to other factors, such as leaving group preference or binding.

Brønsted analysis

In order to further sample the adaptation to the phosphotriester transition state, three enzymes with large differences in phosphotriesterase activity were chosen:

4 Phylogenetic exploration of P91 homologues

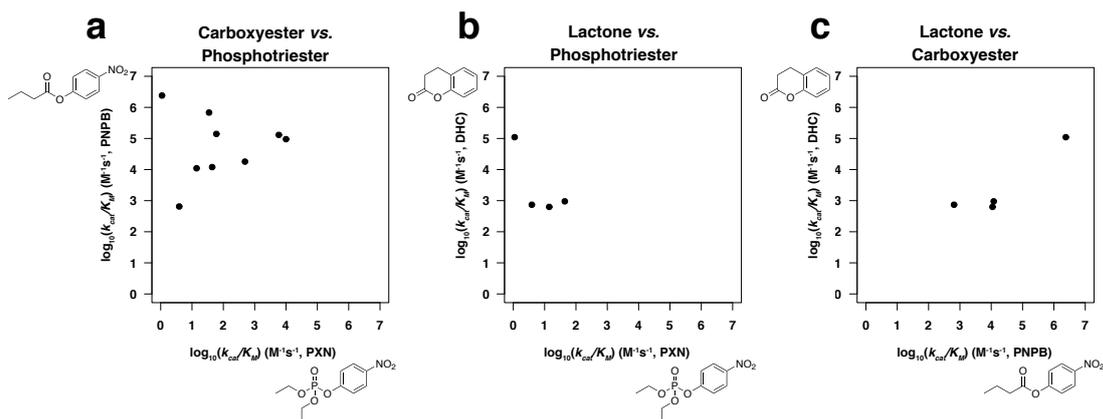


Figure 4.6: Promiscuity profiles of the studied DLH family proteins displaying paraoxonase activity. Steady-state kinetics were measured for phosphotriester 2, carboxyester 4, and lactone 11. Lactonase activity was only measured for a subset of the enzymes (*MzDLH*, *SaDLH*, *ScDLH*, *m1DLH*). **(a)** Carboxyesterase versus phosphotriesterase activity. **(b)** Lactonase versus phosphotriesterase activity. **(c)** Lactonase versus carboxyesterase activity. For all enzymes, carboxyesterase is the highest and phosphotriesterase the lowest activity. None of the promiscuity relationships follows a distinct correlation or trade-off, indicating that the observed reactivity differences are not governed by differences in adaptation to a specific transition state geometry but might be due to other factors, such as leaving group preference or binding. For a definite conclusion regarding the lactonase activities more data points are needed.

ScDLH (low), P91 (medium), and *m3DLH* (high). For those enzymes, in order to establish a linear free-energy relationship, steady-state kinetics were measured towards a range of paraoxon analogues with differing leaving group pK_a (substrates 2 and 5–10), as previously done for P91-WT and P91-R2 (Section 3.2.4). In the resulting Brønsted plot, for both *ScDLH* and *m3DLH*, the slope of the linear fits (β_{LG}) is very similar for both kinetic parameters (k_{cat} and $\frac{k_{cat}}{K_M}$, Figure 4.7), confirming that intermediate formation must be the rate-limiting step, as previously shown for P91-WT and P91-R2. Interestingly, and in contrast to the evolved variant P91-R2, all β_{LG} values are very similar across the different enzymes ($\beta_{LG} \approx -1$), despite the fact that they differ in phosphotriesterase activity by ≈ 400 -fold. In other words, although the difference in activity between *ScDLH* and *m3DLH* is as large as between P91-WT and P91-R2, this is not reflected in their leaving-group stabilisation. Hence, it can be concluded that leaving group stabilisation does not contribute to the observed differences in phosphotriesterase activity. This finding confirms that the identified phosphotriesterase activities in those enzymes are fortuitous, promiscuous activities and, as already indicated by the promiscuity profile above, not the result of adaptation to the specific transition state under evolutionary pressure.

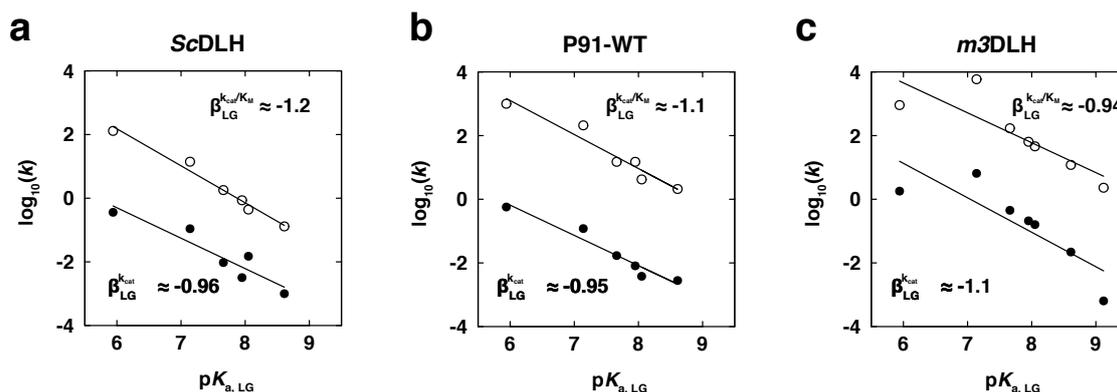


Figure 4.7: Brønsted analysis shows no differences in leaving group stabilisation for phosphotriesterase across non-evolved P91 homologues. The Brønsted plots show the linear free-energy relationship between the rate of hydrolysis of paraoxon derivatives and the pK_a of the leaving group for ScDLH, P91-WT, and *m3DLH* (in order of their phosphotriesterase activity), measured with phosphotriester substrates 2 and 5–10. Filled dots: k_{cat} in s^{-1} . Open circles: $\frac{k_{cat}}{K_M}$ in $M^{-1}s^{-1}$. **(a)** Brønsted plot for ScDLH: $\frac{k_{cat}}{K_M}$: $\beta_{LG} = -1.2$, $R^2 = 0.98$; k_{cat} : $\beta_{LG} = -0.96$, $R^2 = 0.87$. **(b)** Brønsted plot for P91-WT: $\frac{k_{cat}}{K_M}$: $\beta_{LG} = -1.1$, $R^2 = 0.93$; k_{cat} : $\beta_{LG} = -0.95$, $R^2 = 0.94$. **(c)** Brønsted plot for *m3DLH*: $\frac{k_{cat}}{K_M}$: $\beta_{LG} = -0.94$, $R^2 = 0.72$; k_{cat} : $\beta_{LG} = -1.1$, $R^2 = 0.72$. As the slope of the linear fits (β_{LG}) is very similar for both kinetic parameters (k_{cat} and $\frac{k_{cat}}{K_M}$) intermediate formation (k_2) must be the rate-limiting step. All β_{LG} values are very similar between the different enzymes, despite the fact that they differ in phosphotriesterase activity by ≈ 400 -fold. Hence, leaving group stabilisation does not contribute to the observed differences in phosphotriesterase activity.

4.3 Conclusion and outlook

Departing from the mechanistic insight that P91 might be pre-disposed for promiscuous phosphotriesterase activity due to its Cys–His–Asp triad, I explored homologues of P91 from the α/β hydrolase superfamily with the same catalytic motif. Indeed, it was found that promiscuous phosphotriesterase activity is widespread in members of the dienelactone hydrolase protein family. Among the identified new phosphotriesterases, two enzymes (*HmDLH* and *m3DLH*) display even higher paraoxonase activity than P91. In addition, the human homologue of P91, named *HsDLH*, also displays very low promiscuous activity on the phosphotriester substrate FDDEP, making it one further potential starting point for the development of a catalytic bioscavenger. The high activity levels of some enzymes propose the

4 *Phylogenetic exploration of P91 homologues*

question whether they are results of evolutionary adaptation to a naturally occurring organophosphate (be it of natural or anthropogenic origin). However, systematic inspection of the transition state adaptation of all homologues by mapping their promiscuity profile and measuring linear-free energy relationships refute this hypothesis. As the phosphotriesterase activity level does not correlate with specific adaptation to phosphotriester transition state, the identified phosphotriesterase activities are presumably coincidental and promiscuous.

Beyond the enzymes identified here, these findings encourage the further exploration of similar enzymes. As illustrated by the identification of the new promiscuous phosphotriesterase *MzDLH* from genomic context similarity, this promiscuous activity could still be linked to a functional context in the β -ketoacid pathway. Although this catabolic pathway is conserved in prokaryotes, organisation and distribution of the involved genes is highly variable [205], a clustering of the phosphotriesterase-activity displaying enzymes in a functional sub-branch of the pathway would not necessarily be apparent upon visual inspection. A combination of structural information and genomic context has previously been used to predict function in absence of experimental data, beyond relying on bare sequence similarity [244]. A systematic bioinformatic survey of the genomic context of all studied enzymes and the annotations of the surrounding genes in each direction, e.g. supported by visualisation in a sequence similarity network, could help to reveal such concealed patterns and shine a light on the native function of these enzymes.

Beyond the DLH protein family, Cys–His–Asp triads also occur in an entirely different fold, in the cysteine proteases. The catalytic triad of TEV protease, for example, has convergently evolved towards the same Cys–His–Asp motif – however, in reversed (mirrored) orientation [245]. Although potentially facing high hurdles due to other incompatibilities, the bold protein engineer might ask whether TEV protease is evolvable towards phosphotriesterase activity.

5 In search of specificity-determining positions

Although phosphotriesterase activity is widespread in the dienelactone hydrolase protein family, activity levels vary over many orders of magnitude. What determines these differences in activity? In this chapter, I apply protein engineering and droplet screening to a non-active homologue of P91 with the aim of carving out the molecular basis for covalently catalysed organophosphate hydrolysis in metal-independent enzymes.

External contributions:

Alfie Elliott-Anderson screened point mutants and small libraries of *PkDLH* in the context of a Part II project, which contributed to the design of the large *PkDLH* library screened in this chapter.

5.1 Introduction

In the previous chapter, a set of new promiscuous phosphotriesterases was identified which all use the same nucleophilic catalysis mechanism as P91. As a broad pattern, among those new enzymes levels of activity correlate with sequence similarity to P91. However, within the cluster of high similarity to P91, catalytic efficiencies still span across four orders of magnitude, without obvious correlation to sequence similarity.

What are the determinants for high promiscuous phosphotriesterase activity on the molecular level? Here, I depart from the hypothesis that a specific arrangement of residues in the active site, complementary to the Cys–His–Asp triad, could be key in stabilising the new phosphotriester hydrolysis transition state. The emergence of metal-dependent phosphotriesterases has been re-capitulated in the laboratory by evolving weakly promiscuous homologues of naturally evolved phosphotriesterases into more proficient enzymes (Figure 2.9b) [34, 36, 38, 43, 192]. However, these directed evolution campaigns were started from homologues with measurable promiscuous activities and exploited an established metal cofactor-dependent mechanism.

In the following, I aimed at addressing the hypothesis stated above by targeted mutagenesis and combinatorial ‘active-site grafting’.

5.2 Results and discussion

5.2.1 The extended hexad

In all four previously structurally characterised DLH family proteins (P91, *PkDLH*, *KpDLH*, and *EcoDLH*), the active-site cysteine of the triad is present in two conformations in the crystal structure, one pointing inwards and one pointing outwards

(Figure 5.1). Structural and computational evidence has been presented to demonstrate that in *PkDLH* this dual conformation is part of a mechanism of substrate-induced activation, hypothesised to protect the oxidation-prone cysteine when not interacting with the substrate. Binding of the dienelactone substrate to residue R206 in the lid loop (loop 15) which partly covers the active site, breaks a chain of hydrogen bond bridges, resulting in the active-site cysteine C123 to swing from the inwards-pointing ('inactive') into the outwards-pointing ('active') conformation [246, 247].

In P91 however, the residue corresponding to R206, D203, carries the opposite charge and is not part of a similar chain of hydrogen bonds reaching the catalytic cysteine C118. This renders substrate-induced activation, at least mediated by the same chain of interactions as in *PkDLH*, unlikely in P91. The inwards-pointing conformation of C118 in P91 is stabilised by three neighbouring residues: E37, S200, and H141 (Figure 5.1). Mutating E37 or H141 into an alanine does not fully abolish phosphotriesterase activity but makes P91 more susceptible to losing activity upon treatment with alkylating reagents, which proposes a role of these residues in protecting C118 from oxidation [248].

The residues stabilising the inwards-pointing conformation of the catalytic cysteine are not completely conserved across the different studied homologues. Following the hypothesis that the subtle positioning of the two alternative conformations and the activation of the thiolate by these three residues, referred to as the 'extended hexad', could be a key determining feature of phosphotriesterase activity, these residues were systematically compared between all characterised homologues (Figure 5.1).

In P91 and all other identified highly active enzymes, the alternative conformation of the cysteine nucleophile is stabilised by a Glu/Ser/His arrangement. In the enzymes displaying none or only marginal activities, this consensus is broken. Notably, *SaDLH* is the only enzyme from the close sequence cluster around P91, which does not conform to this consensus, having an aspartate instead of a glutamate in the cor-

5 *In search of specificity-determining positions*

responding position. At the same time, the turnover rate of *Sa*DLH falls off by one order of magnitude as compared to the very similar, closely P91-related homologues (see Table 4.1).

To test whether this deviation in the ‘extended hexad’ arrangement could be the reason for decrease or absence of phosphotriesterase activity, these residues were harmonised with the P91 consensus in two enzymes, *Sa*DLH and *Pk*DLH. However, neither *Sa*DLH D36E nor *Pk*DLH Y145H displayed increased phosphotriesterase activity. This does not refute the hypothesis that a specific ‘extended hexad’ arrangement is necessary for promiscuous phosphotriesterase activity, but indicates that it is not sufficient.

5.2.2 Experimental determination of specificity-determining positions in *Pk*DLH

Therefore, I extended the focus beyond the direct arrangement of triad conformations towards the entire first and second shell. As droplet microfluidics allow to fully oversample defined combinatorial libraries, it is possible to draw conclusions from both, the presence and the absence of active variants in a library, thus making it possible to test hypotheses by protein engineering. By grafting parts of P91’s active site into a very similar but inactive scaffold and screening the resulting combinatorial library for active variants, it would be possible to test a large combinatorial set of different alternative residue arrangements and thus re-trace phosphotriesterase activity-conferring features in the DLH protein family.

As the archetypical and only well-characterised member of the DLH protein family, with known structure and native function but no detectable phosphotriesterase activity, *Pk*DLH was chosen as the starting point of this directed evolution campaign.

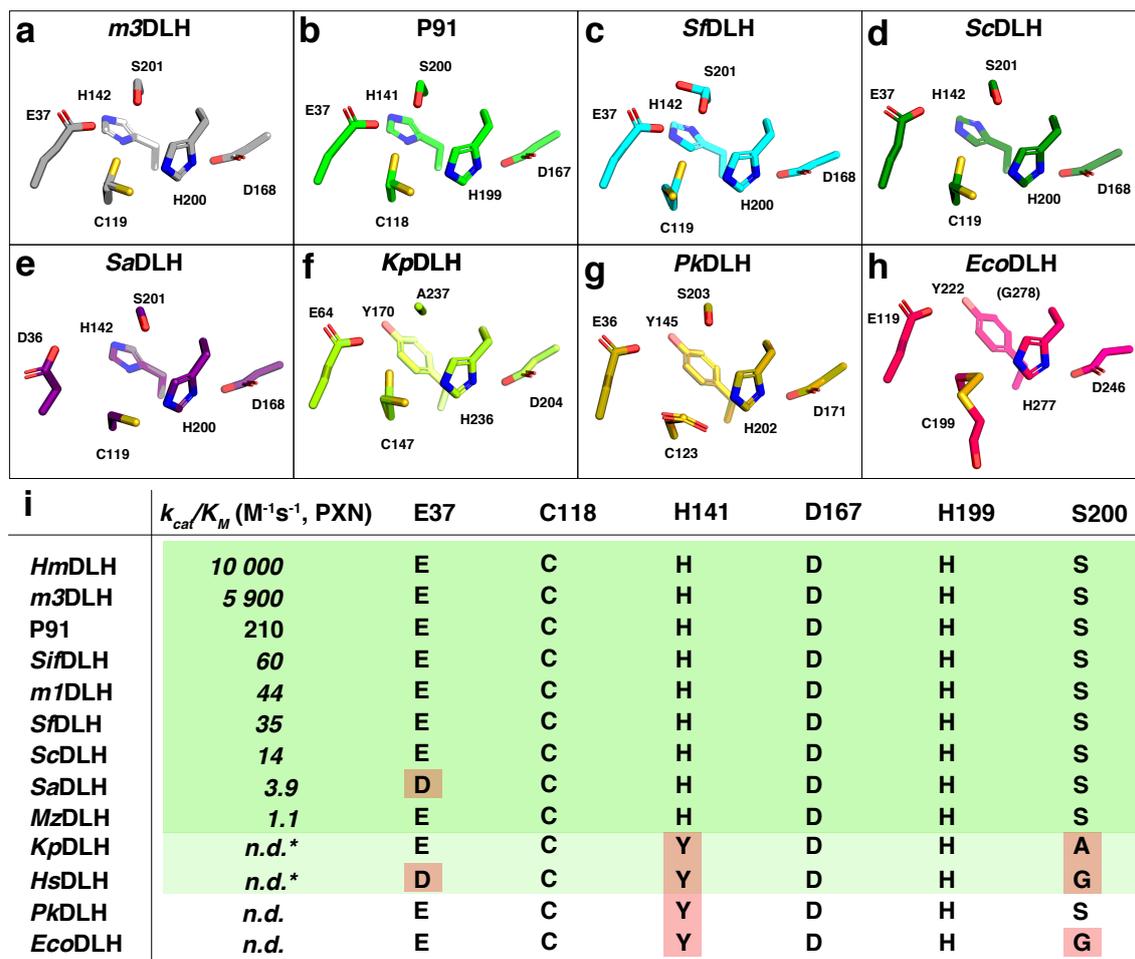


Figure 5.1: (a)–(h) Structural alignment of the active sites of all characterised DLH family proteins, arranged in decreasing phosphotriesterase activity from top left (a) to bottom right (h). All structural arrangements are shown in identical orientations, based on structural superposition of the publicly available (*P91*, *KpDLH*, *PkDLH*, *EcoDLH*) and the newly determined structures of DLH family proteins (*m3DLH*, *SfDLH*, *ScDLH*, *SaDLH*). The three residues of the Cys–His–Asp triad are shown as well as the three residues which are presumably stabilising the alternative conformation of the triad cysteine, together referred to as the 'extended hexad'. Note that in (e), (g), and (h) the two alternative cysteine conformations are not reflected in the structure but are visible in the original electron density. The cysteine C123 in *PkDLH* is oxidised and the apparent covalent mercaptoethanol adduct of C199 in *EcoDLH* is a false interpretation of the electron density which presumably represents two alternative cysteine conformations. The table in panel (i) shows the paraoxonase activities and the residues of the 'extended hexad' for all characterised DLH family proteins, including enzymes for which there is no structure available. Enzymes with promiscuous phosphotriesterase activity are highlighted in green, while enzymes with marginal activity on only one of the two tested phosphotriester substrates are highlighted in faint green. Residues mismatching the corresponding residues in *P91* are highlighted in red. *n.d.*, no detectable phosphotriesterase activity; *n.d.**, no detectable activity on paraoxon, but marginal activity on FDDEP.

5 In search of specificity-determining positions

Library design Despite their low similarity on the sequence level, *PkDLH* and P91 are very similar on the structural level (RMSD \approx 1.7). The main structural difference between *PkDLH* and P91 and its closer homologues is the length of the active-site flanking α -helix B and loops 5 and 14 (Figure 1.6a). However, in the active site, most first and second shell residues are positionally conserved and their exact counterpart can be identified by structural superposition. To determine a set of potentially specificity-determining residues, I compared the arrangement of residues within or near the active site in *PkDLH* and P91 and its active homologues. For homologues without an available crystal structure, a structural model was generated using the SWISS-MODEL server [249] in order to identify the homologous residues.

The criteria for selection of residues were:

1. Proximity to the catalytic triad: residues within a 12 Å radius of the active-site cysteine
2. Interaction with the ‘extended hexad’: residues which have (side chain or main chain) contacts with the catalytic triad or any of the three residues that interact with the alternative conformation of the catalytic nucleophile in P91 (‘extended hexad’)
3. Differential consensus: residues that are identical or chemically similar in P91 and its active homologues but different in *PkDLH*

Using these criteria, the number of positions could be narrowed down to a set of 20 potential specificity-determining residues (Figure 5.2).

In order to test the possibility that all these residues together conferred activity, a mutant of *PkDLH* was constructed bearing all 20 mutations, equivalent to grafting the entire P91 active-site arrangement. However, introducing all 20 mutations simultaneously into *PkDLH* made the protein insoluble, even in presence of the N-terminal solubility tag SUMO, and no significant phosphotriesterase activity could

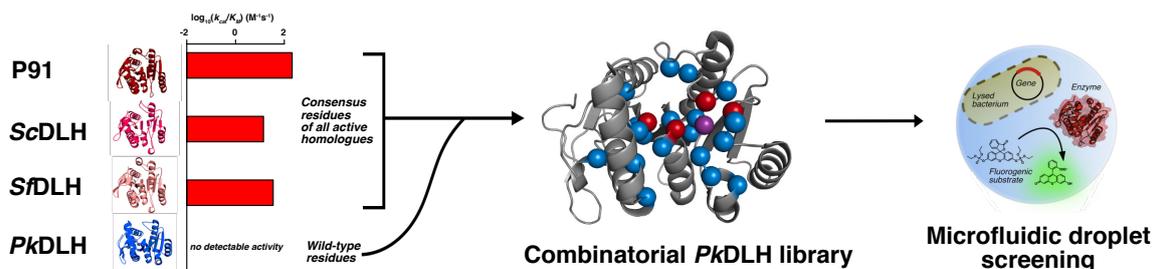


Figure 5.2: Rationale and library design for the functional conversion of *PkDLH* into a phosphotriesterase. Taking advantage of the high structural similarity despite low sequence similarity, *PkDLH* and all homologues with high promiscuous phosphotriesterase activity were structurally superimposed (only P91, ScDLH, and SfDLH shown) and all residues in the active site that are interacting with the catalytic triad or other key residues (red spheres in the structure) were matched across structures. Positions for which there was a consensus across all active homologues, while being different in *PkDLH*, were then included into a combinatorial *PkDLH* library. In this library, the 20 selected residues (blue spheres in the structure) can either have the wild-type side chain or the consensus side chain from the active homologues, resulting in a theoretical diversity of $\approx 1 \cdot 10^6$ variants. This large ‘binary’ library was then screened for phosphotriesterase activity in microfluidic droplets.

be measured from the bacterial lysate. As a consequence, a combinatorial approach was needed.

Complete randomisation of 20 residues exceeds the throughput of any available screening method by far (theoretical diversity: $20^{20} \approx 1 \cdot 10^{26}$). Assuming an upper throughput limit of $10^6 - 10^7$ for microfluidic droplet screening and 10^3 for (academic-scale) microtiter plate screening, even partial simultaneous randomisation of so many residues is out of reach. As a result, droplet screening is limited to a library size of five completely randomised residues (or six residues using a reduced amino acid alphabet with the codon NDT).

Therefore, a library was devised in which every of the selected positions can consist in either the wild-type *PkDLH* residue or the consensus residue that is present in the homologues with phosphotriesterase activity. The resulting ‘binary’ library has a theoretical size of $2^{20} \approx 1 \cdot 10^6$, which is screenable in droplets.

To construct the library on a high-copy plasmid, accommodating binary mutations at 20 different sites, the *PkDLH* gene was divided into 7 fragments. In an exten-

5 In search of specificity-determining positions

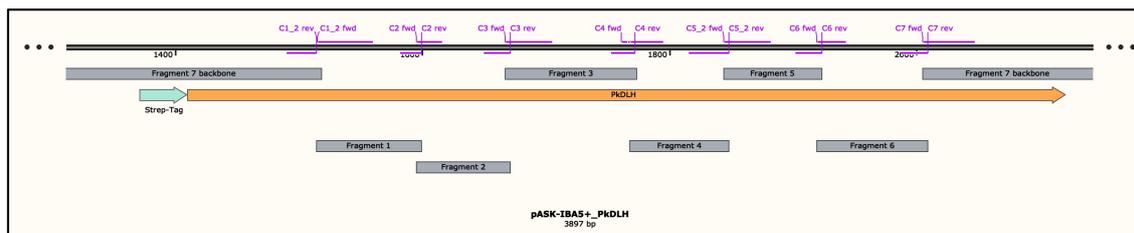


Figure 5.3: Library cloning strategy for the assembly of the binary *PkDLH* library. The *PkDLH* gene (orange) was divided into 7 fragments (grey) which were created by PCR, thus introducing the mutations at the primer annealing sites (purple). The short fragments were then assembled separately and, after limited-cycle PCR amplification, ligated into the large backbone fragment.

sion of the approach used in building the second-round library of P91 (Chapter 2), the fragments were created by PCR with mutagenic primers, introducing the mutations at the primer annealing sites (Figure 5.3). The individual fragments were then assembled by Golden Gate Assembly with the type-IIs restriction enzyme *Bsa*I. As one-pot assembly was not efficient enough, yielding only 10^3 – 10^4 colonies, the six short fragments were ligated separately and amplified by limited-cycle PCR in order to avoid PCR bias. The resulting insert was then size-selected by agarose gel electrophoresis and ligated into the large pASK-IBA5+ backbone fragment, yielding a library of $> 3 \cdot 10^7$ transformants and < 0.5 % non-insert background. Sequencing of randomly picked clones confirmed that the majority (20/24) assembled into complete genes without fragment duplications or deletions.

Library screening The *PkDLH* library was screened in microfluidic droplets for activity on the fluorogenic model phosphotriester 1 (FDDEP), as previously P91 in chapter 2 (Figure 2.5). The library was transformed into *E. coli* cells which, after expression, were encapsulated as single bacterial cells into picoliter-sized monodisperse droplets on a microfluidic chip (Figure 2.3) together with lysis agent and substrate. Subsequently, the droplets were injected into a fluorescence sorting chip and screened according to their fluorescence at a throughput of ≈ 0.5 – 1 kHz. Droplets with the highest fluorescence signal were sorted into the hit channel and their DNA was recovered by transformation into highly electrocompetent *E. coli* 10G cells. As *PkDLH* has no detectable starting activity, even an evolved variant would probably only display low activity levels close to the detection limit. In order to max-

imise the sensitivity of the assay, several changes were introduced in comparison to the evolution campaign of P91: The library was expressed in the expression strain BL21(DE3) instead of the DNA recovery strain *E. coli* 10G, thus maximising the protein concentration in the droplets, at the cost of introducing an additional transformation step between sorting rounds. Substrate concentration was elevated from the previously used 3 μM to 50 μM and the incubation time was extended to three days.

In the first sorting, six million droplets were screened in total and the top 2.5 % most fluorescent droplets were sorted and recovered. These relatively permissive conditions were chosen in order to increase the chances of recovering weakly active clones. To purge false positives and allow cumulative enrichment of potentially active clones, the droplet sorting was repeated twice (three screenings in total) using the respective hit fraction of the preceding screening. After droplet sorting, \approx 500 clones were randomly picked and analysed by microtiter plate screening. In addition, one 96-well plate (84 clones) of each, the unsorted library and the enriched libraries after every sorting round, were analysed in parallel.

Analysis of screening hits The microtiter plate screening revealed no strong enrichment over the three rounds of sorting. Only in the last sorting an enrichment of active clones becomes apparent, to activity levels of approximately 4-fold over the wild-type background (lysate of *PkDLH* wild type-expressing bacteria) (Figure 5.4a). However, when comparing the clones to lysate of bacteria bearing the empty pASK-IBA5+ vector without insert, which were also included as controls on the microtiter plates, the activity levels are reduced down to background level (Figure 5.4b). The empty vector has a higher activity level without than with insert, consistently across all tested plates. One potential explanation for this observation could be the absence of a transcription terminator in pASK-IBA5+. Therefore, induction in the absence of an insert leads to transcriptional read-through and subsequently to higher expression of the two backbone proteins, which are encoded in the same orientation, the ampicillin resistance gene TEM-116 and the Tet repressor protein. If any of those

5 *In search of specificity-determining positions*

two proteins has a very weak promiscuous phosphotriesterase activity this would increase the background hydrolysis rate in absence of a plasmid insert.

Nevertheless, as the sorted output library displays a slight enrichment as compared to the unsorted input library, the 15 most active clones were sequenced. Nine of the 15 most active clones had frameshifts leading to premature stop codons and were consequently translated as short peptides. The absence of such truncated sequences in the clones sequenced from the unsorted library suggest that these were actively enriched. Enrichment of apparently non-functional short peptide inserts displaying no enzymatic activity has also been observed in functional metagenomic screenings [Stefanie Neun, manuscript in preparation]. One potential mode of action could be that these overexpressed, quasi-random peptides act as transcriptional modulators, upregulating native *E. coli* host genes that show very low levels of promiscuity towards the activity of interest, thereby increasing the biological background of the reaction in these clones.

Six of the 15 sequenced clones, however, were full-length genes, bearing eight to eleven mutations at the designated sites in the protein. These proteins were purified and tested for activity, yet no phosphotriesterase activity could be detected above chemical hydrolysis background level in the buffer.

5.3 Conclusion and outlook

The absence of genuine hits from this library could have several potential underlying reasons: Given that both, the size of the correctly assembled DNA library and the number of screened and cumulatively enriched variants sufficiently oversampled the theoretical diversity of variants (25-fold and 6-fold, respectively), it can be safely assumed that none of the tested combinations of 20 residues confers significant phosphotriesterase activity to *PkDLH*.

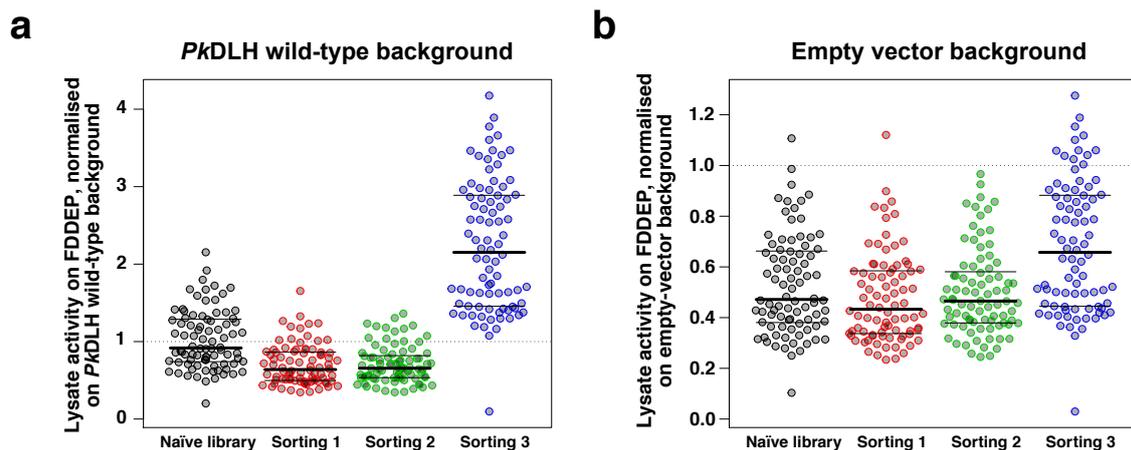


Figure 5.4: Enrichment analysis of the binary *PkDLH* library across droplet screening rounds reveals no strong enrichment for clones with increased phosphotriesterase activity. To quantify the enrichment of active clones through droplet screening, 84 clones were randomly selected, before screening and after each droplet screening round, and their activity towards the phosphotriester substrate 1 (FDDEP) was measured in a microtiter plate-based lysate assay. The black horizontal lines in each beeswarm plot indicate the first quartile, the median, and the third quartile of the data. (a) Normalisation on lysate of bacteria expressing the wild type *PkDLH* enzyme as a negative control. (b) The same data, normalised on lysate of bacteria with the empty vector as a negative control.

One potential reason could be that potentially function-conferring residue arrangements might not be captured in the screened libraries due to incompatibility with the *PkDLH* scaffold. Even if an activity-conferring combination of residues is found within the library, its effect could be limited by stability trade-offs, requiring further compensatory mutations (a different selection of residues for the library) to increase stability and remove compatibility clashes. Indeed, stability might be one of the main constraints on protein evolvability [49, 50].

Another limitation to the identification of active variants could be the sensitivity of the droplet assay. Assuming a droplet volume of 2 pL and a lower detection limit of 2.5 nM fluorescein [130, 149], $\approx 3\,000$ turnovers are needed to detect activity in a droplet. In practice, promiscuous phosphotriesterases with catalytic efficiencies down to $\frac{k_{cat}}{K_M} \approx 50 \text{ M}^{-1}\text{s}^{-1}$ have been identified in a metagenomic droplet screening under comparable conditions [149]. As *PkDLH* does not display any starting activity, it is unclear how large the leap in activity would have to be in order to pass the lower

5 In search of specificity-determining positions

detection limit threshold of the droplet assay.

Overall, however, the absence of hits from the designed library could also imply that the initial underlying hypothesis, that a specific arrangement of catalytic residues around a Cys–His–Asp triad is sufficient to confer phosphotriesterase activity, is too simplistic. Phosphotriesterase activity in DLH family proteins might be determined by higher-order properties, beyond simple arrangements of active-site residues. Therefore, a more holistic approach, taking the secondary and tertiary structure level into account, could elucidate the factors that govern promiscuous phosphotriesterase activity among the studied enzymes.

6 ‘Proteinometrics’ – In search of higher-order determinants of phosphotriesterase activity

After previously having shown that phosphotriesterase activity in the DLH protein family cannot solely be conferred by a specific arrangement of active-site residues, in this chapter, I expand the analysis to broader structural properties and systematically compare features such as active-site volumes and loop flexibilities. I apply these findings to engineer higher phosphotriesterase activity into a human homologue of P91 by loop grafting.

External contributions:

The loop grafting into *HsDLH* was performed by Rashid Khashiev in the context of a Part III project.

6.1 Introduction

In Chapter 4, I showed that phosphotriesterase activity is widespread among enzymes from the diene lactone hydrolase protein family. However, the levels of this promiscuous activity span across several orders of magnitude without discernible correlation to sequence similarity. As shown in the previous chapter 5, this gradient in activity is not governed by a specific set or arrangement of active site residues. Therefore, as a next step, I hypothesised that the observed promiscuity could be determined by higher-order structural properties that vary among these enzymes. This hypothesis is in alignment with the observation that functional diversity in the α/β hydrolase superfamily is accompanied by a large diversity of alternative loop lengths and arrangements around the core topology of conserved β -sheets and α -helices regulating access to the active site [152]. Thus, here I turn the focus beyond single residue positions to the structural level.

6.2 Results and discussion

6.2.1 Structural comparison

Detailed inspection of all available structures reveals small but systematic differences on the secondary and tertiary structure level (Figure 6.1). Firstly, helix E is developed to differing degrees across the different homologues. While this helix is entirely absent in P91, forming a disordered loop (Figure 6.1a), it is a bent α -helix in *PkDLH* (Figure 6.1b). All other structures fall in between those two cases, often having a shorter, interrupted helix or a loop with a very short helical element.

The main systematic differences, however, lie in the length and structure of the two structural features which shape the access to the active site: Loop 14 and the transition of helix B to loop 5. In all DLH family proteins, helix B has a kink and abruptly

changes direction in the middle, forming a pillar left (as seen in the orientation depicted in Figure 6.1) to the catalytic triad before transitioning into loop 5, forming the left half of the active-site cavity. While in P91, helix B is shorter, bends over the active site and disorders earlier into loop 5 (Figure 6.1a), in *PkDLH* this helix is long and straight (Figure 6.1b). Loop 14 lies on the opposite side, forming the ‘lid’ and right half of the active site cavity. In P91, loop 14 is long and has a short helical element (Figure 6.1a), whereas in *PkDLH* this loop is significantly shorter (Figure 6.1b).

In summary, these two features result in a deeper, more closed active site cavity in P91 and a shallower, more open cavity in *PkDLH*. Upon visual inspection, the marginally active and inactive homologues *KpDLH* and *EcoDLH* fall into the *PkDLH*-like structural type, whereas the active homologues *m3DLH*, *ScDLH*, and *SyDLH* fall into the P91-like type. Unfortunately, in the structure of *SaDLH*, which is of lower quality than the other obtained structures, precisely these two features are not well resolved. To further compare these differences in a more quantitative way, I then compared intrinsic physical parameters of the structures and correlated them with the observed differences in promiscuous activity.

Loop lengths

In a multiple sequence alignment of all studied homologues, a very noticeable element is the introduction of different gap lengths to adjust for the variable length in the regions surrounding the active site, especially loop 14. Plotting the length of loop 14, starting from the triad histidine to the beginning of helix G (Figure 1.6a) against the $\frac{k_{cat}}{K_M}$ for paraoxon hydrolysis supports this first impression, revealing that all active homologues have a loop length between 18 and 22 residues whereas the inactive homologues *PkDLH* and *EcoDLH* have much shorter loops with 11 and 8 residues, respectively. Interestingly, the marginally active homologue *HsDLH*, which only shows activity on the more reactive phosphotriester substrate FDDEP (1) but not on paraoxon (2), falls in between those two groups with an intermediate loop

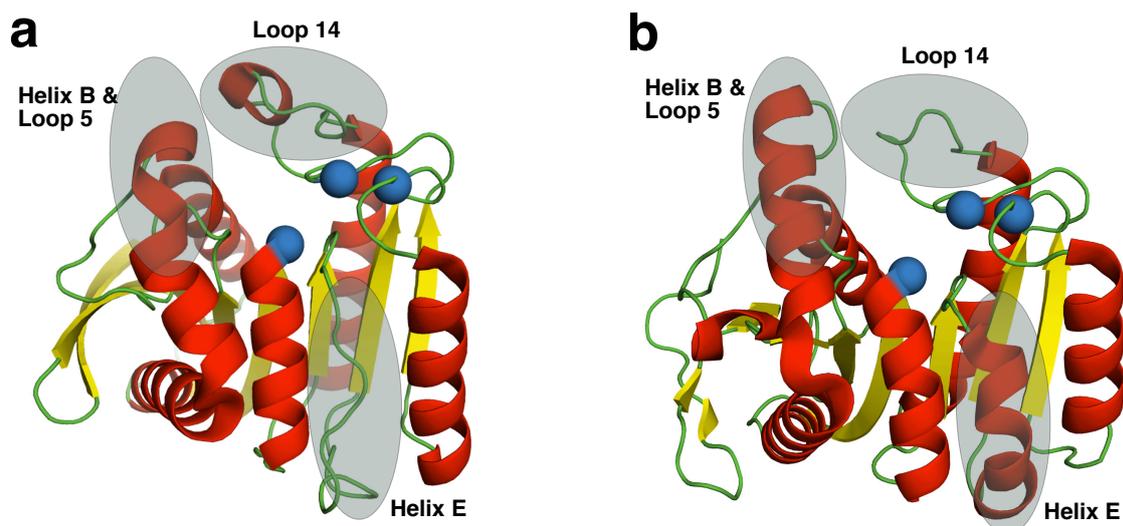


Figure 6.1: Structural comparison of (a) P91 and (b) *PkDLH* reveals systematic differences. (a) Schematic structure of P91. In P91, helix B is short, bends over the active site and disorders early into loop 5. In contrast, loop 14 is long and has a short helical element. The canonical helix E is missing and present as a disordered loop. (b) Schematic structure of *PkDLH*. In *PkDLH*, helix B is long and straight, whereas loop 14 is short and the canonical helix E is present. As loop 14 and helix B/loop 5 form the ‘lid’ domain above the active site, these differences result in a deeper, more closed active site cavity in P91 and a shallower, more open cavity in *PkDLH*.

length of 15 residues (Figure 6.2a). However, given that the active homologues are closer related to each other than to the inactive homologues, it is unclear whether the connection between loop length and phosphotriesterase activity is merely due to evolutionary linkage or due to a functional link.

Active site volumes

How does loop length relate to improved catalysis on a molecular level? I hypothesised that the different loop lengths could remodel the active site towards different cavity volumes and thus different abilities to promiscuously accommodate non-native substrates of varying size and transition state requirements. Active site effective volume has previously been correlated to and used as a predictor of the substrate promiscuity levels of esterases [185]. In addition, a directed evolution campaign of a promiscuous phosphonate hydrolase revealed active site volume as one of the primary evolved higher-order properties [194]. Therefore, I measured the active

site volumes of all available structures of DLH family proteins using two publicly available cavity analysis tools, Caver Web 1.1 [250] and CASTp 3.0 [251]. Visual inspection revealed that Caver identified the active site cavity more reliably than CASTp, which often misidentified the active site pocket. Caver uses the Fpocket algorithm for pocket detection and volume quantification, which relies on alpha sphere clustering, an implementation of Voronoi tessellation [252]. Cavity volume measurements from Caver varied strongly between different chains of the same structure or, in case of *PkDLH*, where several structures are available, across different structures. Multiple measurement values for the same protein were averaged and plotted against phosphotriesterase activity (Figure 6.2b). Active site volumes of the studied structures vary between ≈ 400 and 1300 \AA^3 but show no correlation with phosphotriesterase activity. Therefore, although their relatively shallow active sites might preclude accurate quantification of the cavity volume in *PkDLH*, *EcoDLH*, and *KpDLH*, and add imprecision to the plot, I concluded that active site volume is not a determinant or predictor of phosphotriesterase activity.

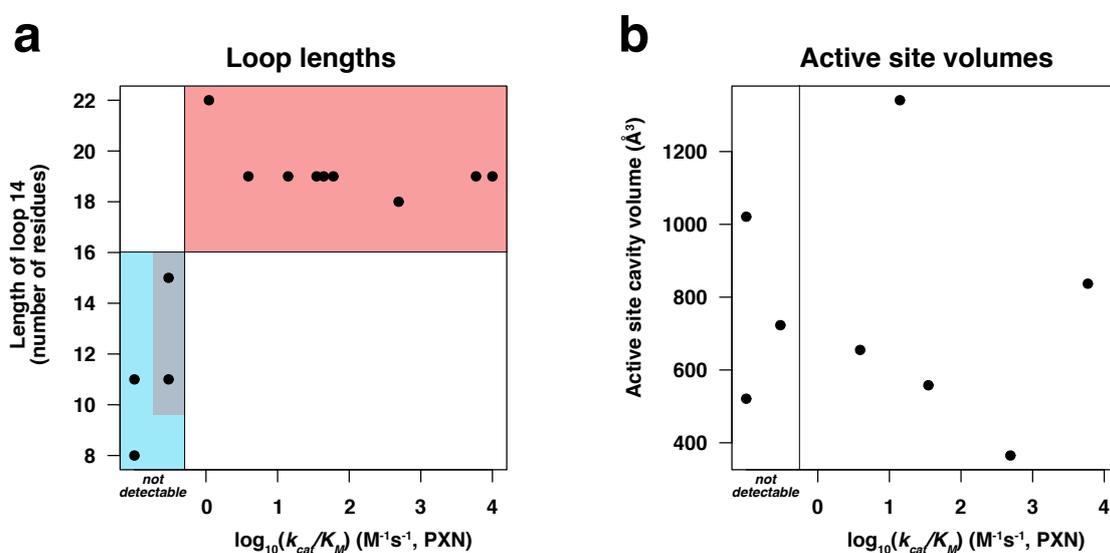


Figure 6.2: Correlation of loop lengths and active site volumes with phosphotriesterase activity. (a) The length of loop 14 correlates with phosphotriesterase activity. The number of residues in loop 14, as measured from the catalytic triad histidine to helix G, was plotted against the phosphotriesterase activity ($\log_{10}(\frac{k_{cat}}{K_M})$ towards paraxon, PXN) for every P91 homologue. While all active homologues have long loops of 18 residues or more (red sector), the inactive homologues have shorter loops (blue sector). The variants displaying marginal activity on the ore reactive phosphotriester FDDEP (*HsDLH*, *KpDLH*) lie in-between in loop length (grey sector). (b) Active site volumes do not correlate with phosphotriesterase activity. Active site volumes were measured for all P91 homologues of which as structure is available and plotted against their paraxonase activity ($\log_{10}(\frac{k_{cat}}{K_M})$).

B-factors

While protein dynamics are an integral determinant of the efficiency of enzyme catalysis [58, 253], a protein crystal structure is restricted to representing a single, static conformation. However, indications on flexibility and the presence of alternative conformations in solution are contained in the B-factors of the structure. The B-factor, also called the temperature factor, Debye-Waller factor, or atomic displacement parameter, describes the attenuation of X-ray scattering at a specific atom due to thermal motion in the crystal lattice [254]. High B-factors in a structural region indicate higher flexibility and the presence of alternative conformations in solution. However, the interpretation of B-factors has to be approached with care as their specific values are a result of the structure resolution, crystal packing and contacts, as well as of the refinement procedure used to obtain the structure. Being so strongly influenced by structure-specific parameters, B-factors have to be normalised to be comparable between different structures (see Section 8.13) [254].

I inspected normalised B-factors in all the available structures of DLH family proteins. Interestingly, in the three proteins with the highest promiscuous phosphotriesterase activity, *m3DLH*, P91, and *SjDLH*, B-factors are highest in loop and helix regions surrounding the active site (Figure 6.3a–c). This local accumulation of apparent conformational heterogeneity is less pronounced in proteins of lower promiscuous activity (Figure 6.3d–f) and absent in non-active homologues (Figure 6.3g and h). Analysis of the B-factor profile of active homologues reveals that B-factor values show local maxima in the regions of loop 14 and the transition of helix B and loop 5, which form the ‘lid’ domain shaping the entrance of the active site. In inactive homologues, however, this pattern is completely absent (Figure 6.3i and j). Averaging the normalised B-factor for each of these sections and for every protein reveals a clear correlation between phosphotriesterase activity level and lid domain flexibility (Figure 6.4).

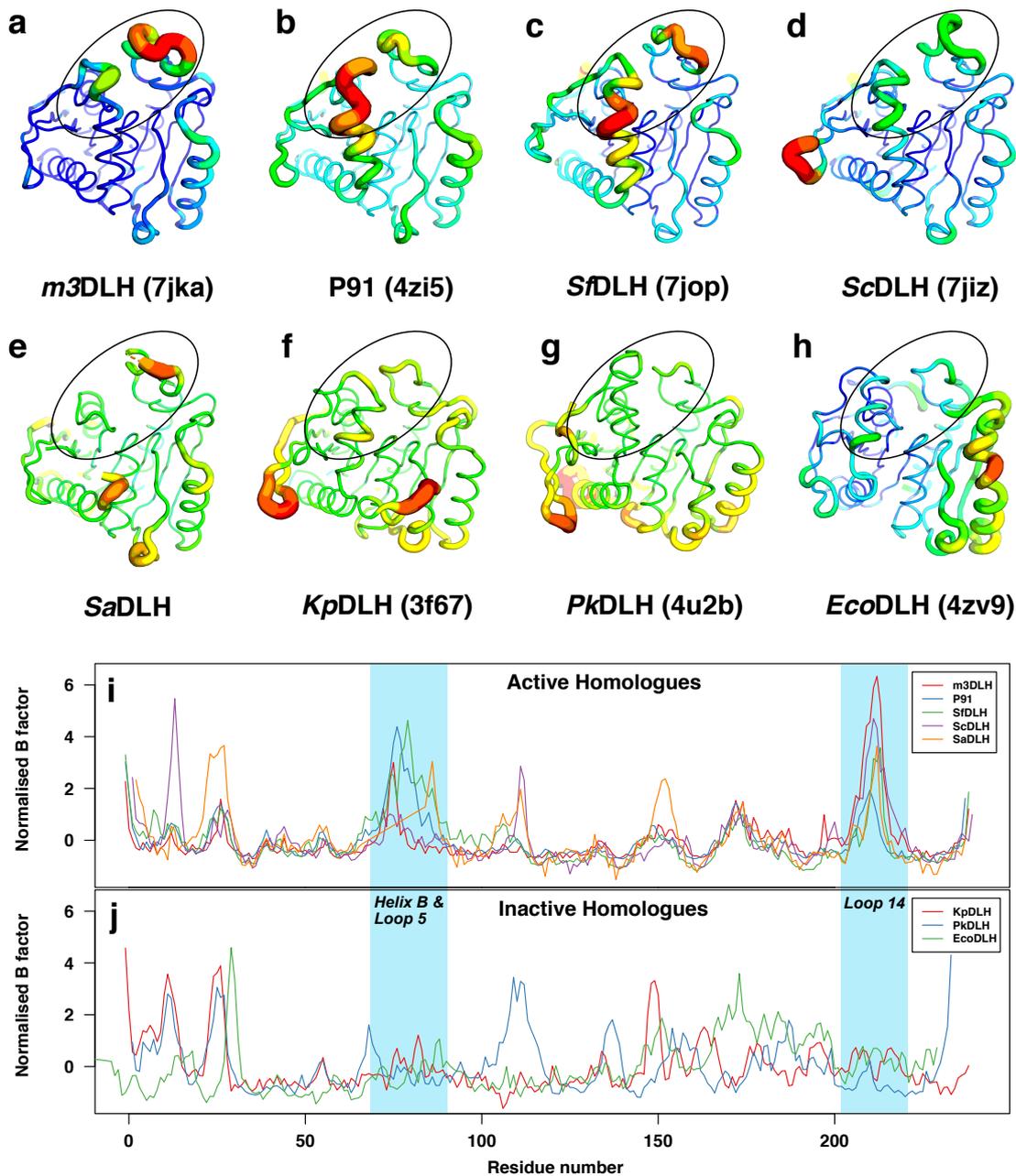


Figure 6.3: B-factor comparison across DLH family protein structures. (a)–(h) B-factor putty representation of all available crystal structures from the dienelactone hydrolase superfamily in the order of their phosphotriesterase activity. Thickness and colour represent normalised B-factors (B'), a proxy of flexibility, of the α -carbon at each residue (blue/thin: low B factor; red/thick: high B factor). Phosphotriesterase activity of homologues decreases from left to right and from top to bottom (in alphabetical order of the labels a–h). (i) & (j) Comparison of normalised main chain B-factors from DLH family proteins (i) with promiscuous phosphotriesterase activity (*m3DLH*, *P91*, *S \bar{f} DLH*, *ScDLH*, *SaDLH*) and (j) without (*EcoDLH*, *PkDLH*) or only marginal activity (*KpDLH*). The structural features covering the active site, the C-terminal end of helix B and the adjacent stretch of loop 5, as well as loop 14, are highlighted in a black circle (a–h) and in a blue box (i & j). DLH family proteins with promiscuous phosphotriesterase activity show higher flexibility in these loops. For *PkDLH*, PDB structure 4u2b was used. The sequence of *EcoDLH* was N-terminally truncated by 65 residues to align with its homologues. Note that structures 7jka, 7jop, 7jiz are on the PDB but not publicly released yet.

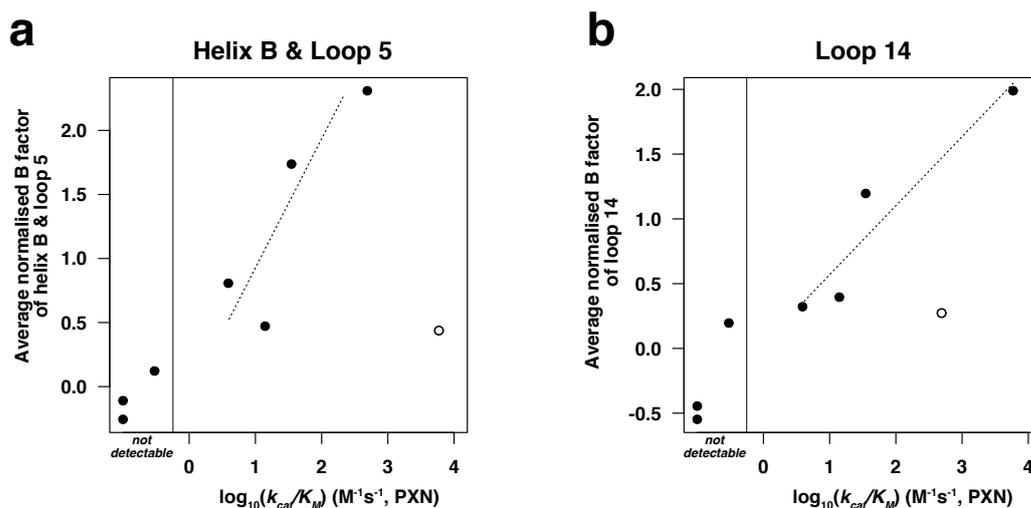


Figure 6.4: Average B-factors of cavity-forming loop regions correlate with phosphotriesterase activity. Normalised B-factors (B') of α -carbon atoms in all DLH family proteins with an available crystal structure were averaged over the two regions that partly cover the active site cavity. Two outliers (*m3DLH* in (a) and P91 in (b)) are highlighted in empty circles. Note that the value for the marginally active variant *KpDLH* (no activity on paraoxon but low activity on FDDEP) was placed right of the non-active values in the 'not detectable' box to indicate its tendency in activity despite a measurable value for paraoxonase activity. **(a)** Average normalised B-factors of the left cavity-forming domain, formed by the transition of helix B and loop 5, correlate well with phosphotriesterase activity ($R^2 = 0.76$). **(b)** Average normalised B-factors of the left cavity-forming domain, formed by loop 14, also correlate well with phosphotriesterase activity ($R^2 = 0.90$). Both correlations indicate that the flexibility of the 'lid domain' formed by those two loops is a determinant of promiscuous phosphotriesterase activity.

6.2.2 P91 is mutationally robust to a large loop insertion

Given the high diversity of additional structural elements inserted in loops found in naturally occurring enzymes, it seems plausible that the core of the α/β hydrolase fold is intrinsically robust against outer loop modifications. Anecdotal evidence for this intrinsic robustness is provided by the isolation of a functional P91 variant bearing a large insertion. This variant of P91 was identified as the most active clone when screening a small library in microtiter plates in which residue D72 was completely randomised (section 2.2.1). However, sequencing of this clone revealed that instead of bearing a mutation at the intended site, this variant had a long insertion at this position, consisting of a duplication of two short DNA sequence regions af-

ter and before the site. This results in an 18-residue insertion of a new sequence in the transition between helix B and loop 5, which is precisely one of the highly flexible regions forming the active site cavity in P91 and other active homologues (Figure 6.5b)(Figure 6.5a and b). Re-streaking, re-sequencing and subsequent purification and kinetic characterisation revealed that the increased activity was genuine and could be attributed to the variant with the large insertion. Not only is this variant stable, its k_{cat} towards paraoxon is also ≈ 2 -fold higher than of the wild-type (Figure 6.5c). This is surprising, as insertions alter the length of the backbone and therefore tend to have more drastic effects on protein structure beyond small local effects. It has been shown that even short insertions and deletions are on average much more deleterious than substitutions [101]. The isolation of this resilient, insertion-bearing P91 variant is therefore an astounding demonstration of the potential role of loop modifications in the acquisition of new promiscuous functions.

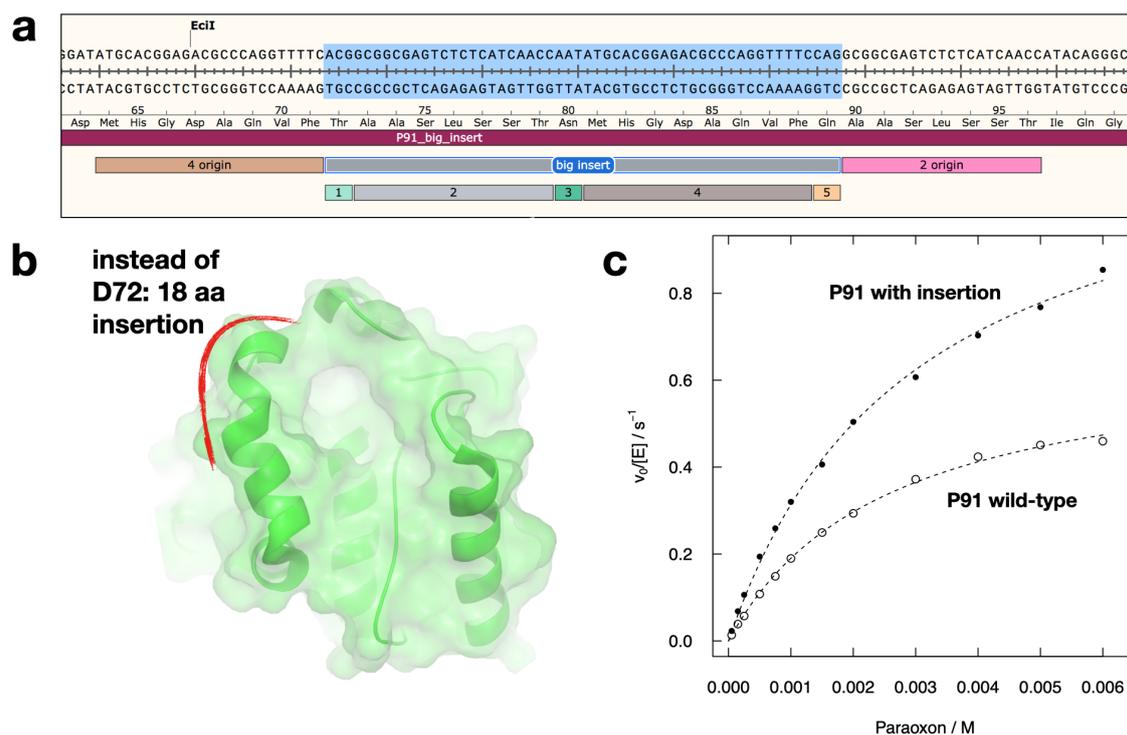


Figure 6.5: A P91 variant with a 18-residue loop insertion in the cavity-forming domain is stable and displays higher phosphotriesterase activity than the wild-type enzyme. (a) The insertion consists of a duplication of two adjacent sequence motifs, resulting in a 18-residue insertion of new sequence after F71. **(b)** Frontal view of P91 with the active site cavity in the centre. The insertion lies at the junction between helix B and loop 5 (marked in red) which partly covers the active site entrance. **(c)** The insertion variant has a ≈ 2 -fold higher k_{cat} for paraoxon than the wild type.

6.2.3 Grafting a P91 loop into *HsDLH* increases its phosphotriesterase activity

Encouraged by the observation of P91's robustness to large-scale loop modifications, we devised an alternative engineering strategy to the single residue-focused 'active-site arrangement grafting' tested in Section 5.2.2, this time including the exchange of entire loop elements from a more active enzyme into a less active homologue.

For this campaign, we chose *HsDLH* as a starting point, as – in contrast to the previously chosen *PkDLH* – it displays some very weak starting activity on the fluorogenic model phosphotriester FDDEP (but not on the less reactive substrate paraoxon), thus making any increase in activity precisely quantifiable, while simultaneously being relatively far away in sequence space from P91 (Figure 4.1b). In addition, there is an interest in *HsDLH* as it could represent a potential starting point for the development of a catalytic bioscavenger against organophosphate poisoning. Being a native human protein – albeit cytosolic – fewer tolerability issues would be expected upon intravenous administration as compared to bacterial enzymes such as P91 or the extremely efficient *BdPTE*.

The drawback of *HsDLH* as a starting point is the lack of an available crystal structure and its poorer expression yield in *E. coli*. When originally expressed and purified as an N-terminal StrepII-tag fusion, only $\approx 50\%$ of the protein were solubly expressed, the remainder being present in inclusion bodies in the insoluble fraction of the lysate. As stability is a major constraint in protein evolution (Section 1.2.3), we tested the solubility of a range of different fusion constructs of *HsDLH* in order to maximise soluble expression (Figure 6.6). We included StrepII-tag and a TEV protease-cleavable version of it, a construct including the solubility enhancer SUMO (small ubiquitin-like modifier [255], on two different vector backbones, pRSF and pASK-IBA5+), and the thioredoxin construct used to express butyrylcholinesterase (Section 4.2.1), all N-terminal. Among all tested constructs, the originally used

StrepII-tag and the construct including the solubility enhancer SUMO performed best, both with solubilities of $\approx 50\%$. As any kind of mutation is likely to decrease stability, the solubility-tag construct was chosen on the assumption that the solubility-enhancing effect of the SUMO-tag could buffer the potential deleterious effect of newly introduced mutations.

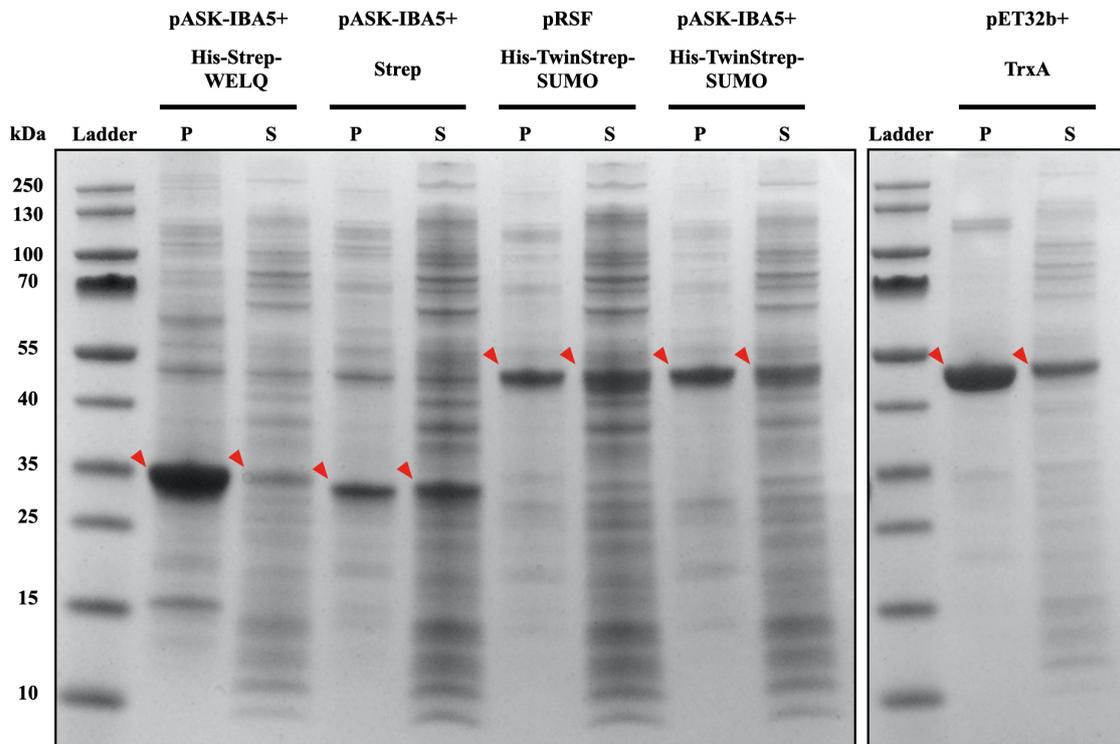


Figure 6.6: Expression optimisation of *HsDLH*. In order to identify a suitable fusion construct for engineering, *HsDLH* was cloned into a range of vectors with different N-terminal tags and expressed in *E. coli* BL21(DE3). After expression and cell lysis, the lysate was separated into soluble (S) and insoluble fractions (P) and analysed by SDS-PAGE. Red arrows indicate the *HsDLH* band. Constructs with a Strep-tag and a His-TwinStrep-SUMO tag showed the highest soluble expression of $\approx 50\%$. Due to its higher versatility regarding affinity purification and the solubility-enhancing tag, the His-TwinStrep-SUMO construct on the pASK-IBA+ vector was chosen for this evolution campaign. P: pellet fraction (insoluble lysate fraction); S: cleared lysate (soluble lysate fraction)

Starting from the SUMO-tagged *HsDLH* fusion construct, a small combinatorial library was created. Building on the assumption that a minimal residue arrangement around the catalytic triad might still be necessary but not sufficient, we designed a library in which larger sections of the enzyme would be swapped from the comparatively active enzyme P91 into *HsDLH*. In an effort to obtain a crystal structure

6 Proteinometrics

of *HsDLH*, crystallisation trials were performed but did not yield diffracting crystals. Therefore, homology models were generated using the SWISS-MODEL web-server [249] (based on the structures of *EcoDLH* and *KpDLH*) and the *de novo* deep learning-based structure prediction method DMPfold [256] on the PSIPRED web server [257]. Although such generic structural models might be inaccurate on the single residue level and vary in the precise positioning of flexible loop regions and side chains, the models are accurate enough for a structural alignment with P91 and other homologues in order to assign homologous regions and residues. The phosphotriesterase activity of *HsDLH* on the substrate FDDEP could not be detected in droplets. However, the activity was detectable above background in a microtiter plate lysate assay. Therefore, a small library of a size screenable in microtiter plates was devised and therefore four peptide fragments were chosen to be grafted into *HsDLH* (Table 8.1, Figure 6.7a):

- A fragment containing the acidic residue stabilising the inwards-pointing conformation of the cysteine nucleophile (D50 in *HsDLH*, E37 in P91), as well as a residue contributing to the oxyanion hole (I51 in *HsDLH*, A38 in P91).
- The transition between helix B and loop 6, which appears to be in contact with loop 14 in both structural models of *HsDLH*.
- A stretch of loop 10, which forms the bottom of the active site cavity where the attacking water or hydroxide molecule might be positioned.
- The entire loop 14, which partly covers the active site cavity and whose flexibility correlates with phosphotriesterase activity across different P91 homologues (Section 6.2.1).

The grafted peptide sequences were sourced from the most active homologues: *HmDLH*, *m3DLH*, P91-WT, and two evolved P91 variants, and cloned together with the corresponding *HsDLH* sequences into *HsDLH* by cassette mutagenesis and Golden Gate Assembly, generating a small combinatorial library with a theoretical

diversity of 80 variants.

The library was expressed in *E. coli* and lysate of 264 clones was screened in microtiter plates at a substrate concentration of 200 μ M FDDEP, oversampling the theoretical library diversity by a factor of 3.3, ensuring that with a ≈ 96 % probability all variants of the library have been assayed. The 18 most active clones, displaying 2.5–3.9-fold increased lysate activities, were sequenced, revealing eight different genotypes. The four most active genotypes were then purified and their steady-state kinetic parameters were determined alongside those of the corresponding wild-type construct. Of the four tested variants, two showed improved kinetics, with 3.2-fold and 7.8-fold increase in $\frac{k_{cat}}{K_M}$, respectively, mainly conferred by increase in k_{cat} (Figure 6.7b).

Interestingly, in the most improved variant *HsDLH-1E11* (7.8-fold increase in $\frac{k_{cat}}{K_M}$), the entire loop 14 is replaced by the corresponding loop from the evolved P91 variant P91-R2. In addition, the acidic residue stabilising the catalytic nucleophile in the inwards-pointing position is exchanged from an aspartate to a glutamate (D50E). The identification of a loop grafted-*HsDLH* with significantly increased activity confirms the initial hypothesis that phosphotriesterase activity might be co-determined by architecture of the active-site surrounding loops.

It is notable that the activity of wild-type *HsDLH* is ≈ 13 -fold lower in the SUMO-tag fusion construct than in the initially tested StrepII-tagged construct (Table 4.1). In order to test whether the activity increase found in *HsDLH-1E11* would translate into overall higher activities in the more active StrepII construct, the tag was exchanged accordingly. The resulting StrepII-*HsDLH-1E11* clone, however, did not express solubly, precluding any kinetic measurement. This finding indicates that the loop exchange places a destabilising burden on the solubility of the enzyme and confirms the previously made choice of screening the library with a solubility-enhancing tag.

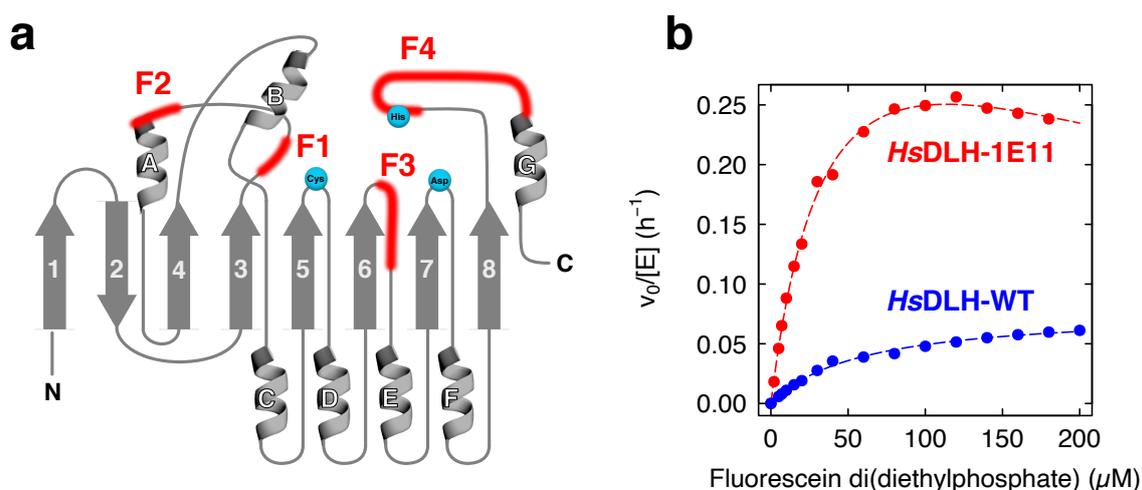


Figure 6.7: Loop grafting increases the phosphotriesterase activity of *HsDLH*. (a) Schematic representation of the fragments (red) grafted into the small combinatorial *HsDLH* library. Fragment 1 (F1) contains the acid residue stabilising the inwards-pointing conformation of the cysteine nucleophile, as well as a residue contributing to the oxyanion hole. Fragment 2 (F2) encompasses the transition between helix B and loop 6, which is in contact with loop 14 in both structural models of *HsDLH*. Fragment 3 (F3) forms the bottom of the active site cavity where the attacking water or hydroxide molecule might be positioned. Fragment 4 (F4) encompasses the entire loop 14, which partly covers the active site cavity and whose flexibility correlates with phosphotriesterase activity across different P91 homologues. (b) The variant *HsDLH*-1E11 which was isolated from the combinatorial library and contains fragment 1 (F1) from P91-WT and fragment 4 (F4, the entire loop 14) from P91-R2, has a ≈ 8 -fold increased catalytic efficiency towards the phosphotriester substrate FDDEP as compared to the wild-type.

6.3 Conclusion and outlook

Systematic comparison of structural features across P91 homologues revealed distinct differences between enzymes displaying and enzymes lacking phosphotriesterase activity. The major difference lies in the appearance of two structural domains partly covering the active site cavity as a ‘lid’. This lid is formed by the transition between loop 5 and helix B and the opposite-lying loop 14. Conspicuously, the length of loop 14 varies across the studied enzymes and correlates with phosphotriesterase activity, indicating that an insertion in this loop boosts promiscuous activity. The role of insertions and deletions (InDels) in evolution is understudied, partly also because it is more difficult to generate InDel libraries for laboratory evolution as compared to substitution libraries, although this has recently been amended [101]. It has been shown in a few marked cases – notably in the evolution of phosphotriesterase activity in the amidohydrolase superfamily – that functional conversion in enzymes can be driven by loop insertions as key specificity switches [32, 36, 101, 258]. Hypothesising that different loop lengths might mediate different levels of phosphotriesterase activity via variable active site volumes, as found previously [185, 194], this parameter was also compared. However, no correlation between active site volume and phosphotriesterase activity could be found. Further inspection of the B-factors in the available structures revealed a marked increase in these loop regions for the enzymes displaying high phosphotriesterase activity. In the enzymes for which a structure is available, flexibility of the ‘lid’ domain correlates well with phosphotriesterase activity levels, indicating that the loops surrounding the active site could be activity-conferring determinants. Testing this hypothesis by combinatorial loop grafting from highly active P91 homologues into the marginally active human homologue *HsDLH* yielded a ≈ 8 -fold improved chimera which incorporated the entire lid-forming loop 14 from the highly active variant P91-R2. This finding also converges with the fact that the key mutations in the directed evolution of P91, conferring ≈ 200 -fold increased activity, lie in loop 14. Similarly, cross-wise loop grafting (corresponding to loop 3 in P91 and *HsDLH*) from a lipase into a cutinase and vice versa, both members of the α/β hydrolase superfamily and ho-

6 Proteinometrics

mologues of the studied enzymes, was tolerated and yielded chimeras with inverted enantioselectivity [184].

Loop grafting into *HsDLH*, an enzyme with already low soluble expression in *E. coli* in the wild type, seems to be constrained by the stability burden of such an intrusive structural modification. The hypothesis could therefore further be validated with a more robust (i.e. higher soluble expression in *E. coli*) and more P91-like enzyme which at the same time displays relatively low phosphotriesterase activity. *SaDLH*, *MzDLH* and potentially *ScDLH* fulfill these criteria and could therefore represent promising starting points for a future, similar loop grafting approach.

Complementary to the established notion that stability is of paramount importance for evolvability [49–51], the role of protein flexibility and dynamics for catalytic function has recently shifted into focus [57, 253, 259–261]. While it has been shown that conformational disorganisation can limit the catalytic efficiency of enzymes [57, 58], conformational diversity also promotes promiscuity [56] and can therefore probe bridgeheads for new activities. Evolution can then reduce conformational heterogeneity, removing unproductive conformational sub-states [253, 260, 261]. In this context, a systematic comparison of thermophilic and psychrophilic homologues of P91, which have evolved towards contrasting flexibility demands, could shed further light on the role of loop flexibility in promiscuous phosphotriesterase activity and further on the general interplay between flexibility, stability, and evolvability. Regarding the toolset for this new ‘dynamics-centered’ angle, B-factor analysis has previously been used to define target residues in the directed evolution of enzymes for increased thermostability [262]. However, B-factor analysis has pitfalls [254] and conformational sampling will therefore advance the study of enzymatic catalysis by using molecular dynamics simulations and NMR spectroscopy.

7 Final conclusions and perspective

In this chapter, I summarise the findings from the previous chapters and apply them to the initial question regarding the acquisition of new enzymatic function and the adaptability of catalytic motifs. Finally, I give a perspective on the potential of droplet microfluidics in the further quest for new catalysts and catalytic motifs.

7 Final conclusions and perspective

Understanding biocatalytic innovation – the acquisition of new catalytic activities by enzymes – is not only of fundamental scientific interest, but can also advance the repurposing of enzymes for medical, chemical and biotechnological applications. How evolvable are different catalytic motifs towards newly arising catalytic challenges? The introduction of xenobiotic phosphotriesters as pesticides into Nature provides an excellent test case to study this question. While in Nature significant phosphotriesterase activity has only evolved in metal cofactor-assisted enzymes [30], this study departs from an entirely different, metal-independent enzyme, P91, that shows weak starting activity on phosphotriesters.

P91 is an unorthodox, recently discovered, promiscuous phosphotriesterase from the α/β hydrolase superfamily. In contrast to all known phosphotriesterases, it does not require a metal-cofactor and hydrolyses organophosphates with a catalytic triad via a covalent intermediate [149]. Using a combination of semi-rational, combinatorial libraries and high-throughput microfluidic droplet screening, I increased its phosphotriesterase activity by a factor of ≈ 400 close to a $\frac{k_{cat}}{K_M}$ of $\approx 10^6 \text{ M}^{-1}\text{s}^{-1}$, leveraging it to the efficiencies of naturally evolved phosphotriesters which use the conventional metal cofactor-catalysed mechanism. The success of this adaptation suggests that a catalytic triad is an evolvable functional motif that is set up for efficient catalysis, even though such adaptation has never been observed in Nature.

Regarding the role of droplet microfluidics for the discovery of new enzymatic function, this work promotes a particular strategy: Apart from only two studies [133, 134], no directed evolution campaign in microfluidic droplets has reported improvements beyond one order of magnitude, contrary to the intuitive assertion that higher throughput is generally better for directed evolution. In fact, the widely used approach of random mutagenesis, e.g. using epPCR, mostly brings improvements in small, incremental increases. Random mutagenesis will therefore likely rather benefit from more successive rounds of evolution (diversification and screening) than from higher single-round screening throughput. This strategy is more circuitous to realise with droplet screening, which only represents an enrichment step upstream of a secondary assay, than with less elaborate assays, e.g. microtiter plate screen-

ing. In addition, notwithstanding the conceptional beauty of droplet sorting, the high phenotypic variation observed in this screening campaign highlights that the precision and enrichment power of a single sorting round can be limited. However, this is readily overcome by successive rounds of re-sorting. Therefore, the full strengths of droplet microfluidics come into play when combined with large combinatorial libraries. While Nature and random-mutagenesis laboratory evolution are confined to ‘tiptoeing’ through smooth fitness landscapes in small incremental steps [103], the combination of multiple-site saturation mutagenesis with the high throughput of droplet screening allows to make ‘leaps’ in sequence space and cross more rugged fitness landscapes to reach otherwise inaccessible new fitness peaks. The successful evolution of P91 within only two rounds is a powerful demonstration of this potential and firmly establishes droplet microfluidics as a tool for functional sequence exploration.

But how does P91 achieve efficient catalysis with its catalytic triad, while similar enzymes, such as synaptic acetylcholinesterase, are irreversibly inhibited by phosphotriesters? The serine triad of acetylcholinesterase is quick to form a phosphorylation adduct but fails to then resolve this intermediate. In this study, exchange of the catalytic nucleophile, specificity analysis, and pre-steady state kinetics show that, surprisingly, P91 is not limited by hydrolysing the phosphoryl intermediate from its cysteine (this step is fast), but rather by the initial formation of the adduct. Directed evolution therefore acted by accelerating the formation of this intermediate. Brønsted analysis shows that the evolved P91 variant achieves this by offsetting charge accumulation on the leaving group during the transition state. The likely reason for the high de-phosphorylation rate of P91 is that the low pK_a of the cysteine nucleophile destabilises the intermediate (by being a better leaving group than serine), suggesting that cysteine triads are intrinsically predisposed to display promiscuous phosphotriesterase activity. In this way, a cysteine triad can act as a ‘launchpad’ for the evolution of higher phosphotriesterase activity.

Given this predisposition of cysteine triads for fast de-phosphorylation, how common is phosphotriesterase activity in similar proteins with a catalytic triad? By sys-

7 Final conclusions and perspective

tematically exploring the sequence space surrounding P91 – mostly uncharacterised enzymes from the dienelactone hydrolase protein family as well as other α/β hydrolases – I discovered twelve new enzymes with phosphotriesterase activity, indicating that it is in fact widespread in enzymes with a Cys–His–Asp triad. In this context, a human enzyme, named *HsDLH*, was also discovered to display very low, promiscuous phosphotriesterase activity, adding another potential starting point for the development of a catalytic bioscavenger against organophosphate poisoning. The identification of so many new phosphotriesterases, with catalytic efficiencies reaching up to $\approx 10^4 \text{ M}^{-1}\text{s}^{-1}$, predominantly from aquatic environments, poses the question whether this activity is linked to a similar native substrate. Do yet unknown organophosphates, similar to the phosphotriesters tested, naturally occur in the environment? However, interrogating the promiscuity profile of the enzymes studied here as well as measuring the transition state adaptation by Brønsted analysis revealed that none of the newly identified enzymes displays specific adaptation to the trigonal-bipyramidal organophosphate hydrolysis transition state. This indicates that the activity displayed by those enzymes is in fact promiscuous, and is not the result of adaptive evolution that was guided by a similar organophosphate substrate.

Among the newly characterised enzymes, phosphotriesterase activity levels span across four orders of magnitude, even within a closely related sequence cluster. What are the determinants for these differences? I first hypothesised that phosphotriesterase activity is conferred by a distinct arrangement of specificity-determining residues around the catalytic triad, e.g. by achieving polarisation of the nucleophile in a unique way. I then tested this hypothesis by systematic inspection of active-site residues across homologues and subsequent harmonisation of residue arrangements across enzymes by protein engineering. However, neither small scale (targeted active site mutants) nor large scale residue grafting (entire combinatorial active site residue shuffling) converted a weakly active or non-active homologue into a multiple-turnover phosphotriesterase. Although potentially limited by stability trade-offs or an incomplete choice of the potentially involved residues, the absence of functional conversion indicates that the initial hypothesis of a specificity-

determining, singular residue arrangement is too simplistic.

Therefore, as an alternative hypothesis, I considered that phosphotriesterase activity might be determined by higher-order protein properties beyond the single residue level. Systematic inspection of structure-level properties revealed a striking correlation between flexibility of two loop regions that together form a partial 'lid' on the active site cavity. I then experimentally tested this hypothesis by grafting loops from more active homologues into a marginally active homologue in a combinatorial fashion. Indeed, this approach led to the isolation of a chimeric *HsDLH* variant with increased phosphotriesterase activity bearing a large 'lid' loop from the evolved P91 variant, in line with the previously identified correlation between loop architecture and promiscuous activity. This finding supports the previously stated concept that conformational diversity exacerbates efficient catalysis [253, 259] but simultaneously promotes promiscuity [56, 57], providing a gateway for the evolution of new catalytic function.

Overall, the successful evolution of P91 as well as the identification of many new phosphotriesterases with this mechanism demonstrates that functional transitions are possible in catalytic motifs without precedent in Nature. While Nature is 'lazy' and takes the quickest (most probable) path by tinkering with modular and pre-existing catalytic motifs to adapt to new catalytic requirements, this work encourages us to take new paths. With high-throughput droplet screening as a vehicle to navigate sequence space, we can now explore structural space unsampled by Nature and would be able, for example, to search for new catalytic solutions in *de novo* designed protein libraries, which have only been designed for stable folding [263, 264].

8 Materials and Methods

8.1 Chemicals and materials

All chemicals were purchased from Sigma-Aldrich and all biological reagents from New England Biolabs, unless otherwise stated. The fluorogenic phosphotriester model substrate fluorescein di(diethylphosphate) (FDDEP, 1) was synthesised by Mark F. Mohamed as previously described [141]. The plasmid pASK-IBA5+ with the *StrepII-P91* insert was provided by Pierre-Yves Colin, who modified it from the commercial plasmid pASK-IBA5+ (IBA Life Science). The plasmid pRSF used during construct solubility optimisation was obtained from Philipp Knyphausen, who modified it, after previous modification by Stefanie Jonas, from the commercial plasmid pRSFDuet-1 (Novagen). The solubility tag His-TwinStrep-SUMO was excised from a plasmid obtained from Stefanie Neun, which was previously modified by Joachim DeJonghe from the commercial pSOL plasmid (Lucigen). The solubility tag TrxA-His was excised from the plasmid pET32b+ obtained from the Sarel Fleishman lab via Addgene (<https://www.addgene.org/83917/>) [265].

8.2 Synthesis of LFER substrates

These syntheses were done by James Oskar Klein.

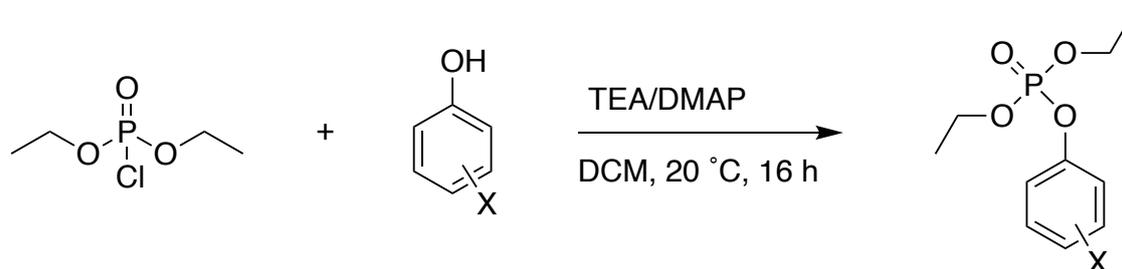


Figure 8.1: Synthesis of substrates for linear free-energy relationship measurements. TEA: triethylamine, DMAP: dimethylaminopyridine, DCM: dichloromethane.

Synthesis procedures and pK_a values were adapted from reference [196]. In brief,

a phenol derivative (1 g) and diethyl phosphochloridate (1.2 equivalents) were dissolved in dichloromethane (DCM). Base (triethylamine (TEA) for all but 3-fluoro-4-nitrophenyl, where dimethylaminopyridine (DMAP) was used) was added dropwise and the reaction mixture stirred at room temperature (20–23 °C) overnight (\approx 16 h). Reaction progress was followed by thin layer chromatography (silica gel 60, EtOAc:Hex, 2:1). The reaction mixture was washed with HCl (100 mM, 50 mL), saturated Na_2HCO_3 (50 mL) and NaCl brine (50 mL, pH \approx 7). Products contaminated by starting material were purified by flash chromatography (silica gel 60, EtOAc:Hex 2:1). Products were characterized by ^1H -NMR. All NMR data were collected at 298 K using Bruker Avance spectrometers with ^1H resonance frequencies of 400 MHz. Chemical shifts (δH) are reported in parts per million (ppm), to the nearest 0.01 ppm and are referenced to the residual non-deuterated solvent peak. Coupling constants (J) are reported in Hertz (Hz) to the nearest 0.1 Hz. Data are reported in the order: (i) chemical shift, (ii) multiplicity (s = singlet; d = doublet; t = triplet; q = quartet; m = multiplet; or as a combination of these, e.g. dd, dt etc.), (iii) coupling constant(s) and (iv) integration. Peak integrals were used to produce correction factors for residual solvent contamination.

3-Fluoro-4-nitrophenyl diethyl phosphate (5)

^1H NMR (399.6 MHz, CDCl_3): δ (ppm) 8.14 (t, $^3J_{\text{HH}}/^4J_{\text{HF}} = 8.8$ Hz, 1H), 7.24 (dd, $^3J_{\text{HF}} = 11.5$, $^4J_{\text{HH}} = 2.4$ Hz, 1H), 7.20 (m, 1H), 4.29 (m, 4H), 1.41 (td, $^3J_{\text{HH}} = 7.1$ Hz, $^4J_{\text{HP}} = 1.1$ Hz, 6H).

4-Formylphenyl diethyl phosphate (6)

^1H NMR (399.6 MHz, CDCl_3): δ (ppm) 10.00 (s, 1H), 7.92 (d, $^3J_{\text{HH}} = 8.5$ Hz, 2H), 7.41 (d, $^3J_{\text{HH}} = 8.5$ Hz, 2H), 4.27 (m, 4H), 1.39 (td, $^3J_{\text{HH}} = 7.1$ Hz, $^4J_{\text{HP}} = 1.0$ Hz, 6H).

4-Cyanophenyl diethyl phosphate (7)

^1H NMR (399.6 MHz, CDCl_3): δ (ppm) 7.68 (dd, $^3J_{\text{HH}} = 8.9$, $^4J_{\text{HP}} = 0.5$ Hz, 2H), 7.37 (dd, $^3J_{\text{HH}} = 8.9$ Hz, $^4J_{\text{HP}} = 0.9$ Hz, 2H), 4.26 (m, 4H), 1.39 (td, $^3J_{\text{HH}} = 7.1$ Hz, $^4J_{\text{HP}} = 1.1$ Hz, 6H).

8 Materials and Methods

4-Acetylphenyl diethyl phosphate (8)

^1H NMR (399.6 MHz, CDCl_3): δ (ppm) 7.99 (d, $^3J_{HH} = 8.4$ Hz, 2H), 7.34 (dd, $^3J_{HH} = 8.9$, $^4J_{HP} 0.9$ Hz, 2H), 4.26 (m, 4H), 2.61 (s, 3H), 1.39 (td, $^3J_{HH} = 7.1$ Hz, $^4J_{HP} 1.1$ Hz, 6H).

3-Cyanophenyl diethyl phosphate (9)

^1H NMR (399.6 MHz, CDCl_3): δ (ppm) 7.56–7.45 (m, 4H), 4.26 (m, 4H), 1.40 (td, $^3J_{HH} = 7.1$ Hz, $^4J_{HP} 1.1$ Hz, 6H).

3-Chlorophenyl diethyl phosphate (10)

^1H NMR (399.6 MHz, CDCl_3): δ (ppm) 7.30 (d, $^3J_{HH} = 8.2$ Hz, 1H), 7.27 (m, 1H), 7.20–7.14 (m, 2H), 4.25 (m, 4H), 2.61 (s, 3H), 1.38 (td, $^3J_{HH} = 7.1$ Hz, $^4J_{HP} 1.0$ Hz, 6H).

8.3 Cloning

Libraries were constructed on the high-copy number (≈ 800 copies/cell), anhydro-tetracyclin-inducible plasmid pASK-IBA5plus (IBA Life Sciences, Germany) bearing an N-terminal StrepII-tag. For high-yield purification, the StrepII-tag was exchanged for a $6\times$ His-tag using the Golden Gate Assembly method laid out below. PCRs, restriction digestions, ligations and transformations were done according to the manufacturer's instructions.

8.3.1 Single-site mutagenesis

Single-site saturation libraries were constructed using the 22-codon trick [188]: Primers bearing the degenerate codons NDT, VHG and TGG were mixed in the ratio 12:9:1, in order to achieve balanced amino acid representation while avoiding stop codons, allowing to sufficiently oversample the diversity of a single randomised

position with a single 96-well microtiter plate (N: any base; D: any base but C; V: any base but T; H: any base but G). The gene was amplified from the randomised position in a whole-plasmid PCR using Q5 DNA Polymerase and digested and re-circularised in a single step of Golden Gate Assembly using BsaI-HFv2 and T4 DNA ligase [266]. Single-site mutagenesis was carried out according to the same principle.

8.3.2 Multiple-site saturation libraries

P91 – Round 1 (“P91-A”) To construct multiple site-saturation libraries, the P91 gene was sub-divided into suitable fragments such that each randomised position was covered by a primer with the degenerate codon NNK. The fragments were created individually by PCR and then assembled by a further assembly PCR. Forward primers for fragment creation contained degenerate codons as described in section 4.2 and ≈ 20 base pair homology arms to the neighbouring fragments with an annealing temperature of 55 °C. For assembly PCR, the individual fragment PCR products were pooled and amplified with outer primers by PCR with Phusion DNA Polymerase at the annealing temperature of the homology arms. The resulting PCR product was column-purified (DNA Clean & Concentrator-5 Kit; Zymo Research) and digested with NheI and HindIII. After gel purification, the library insert was ligated into accordingly digested pASK-IBA5plus vector backbone with T4 DNA Ligase at 16 °C for 14 h. For plasmid amplification, the column-purified ligation product was electroporated into highly electrocompetent *E. coli* (*E. coli* 10G Elite; Lucigen). After overnight incubation, the bacterial carpet was scratched off the agar plate with 6 mL of LB medium and the DNA was extracted with a plasmid isolation kit (GeneJET Plasmid Miniprep Kit; Thermo Fisher). The quality of the library was assessed by measurement of the peak heights in the sequencing chromatogram of a pooled sample as previously described [267].

8 Materials and Methods

P91 – Round 2 (“P91-B”) For the second round of P91 evolution, the library was created by whole-plasmid PCR with degenerate primers containing a type-II restriction site overhang (BsaI: GGTCTCN) to create sticky ends for ligation. At the desired position, forwards primers contained the mutagenic trinucleotides NDT, VHG and TGG and were used as a 12:9:1 mixture, previously described as the ‘22-codon trick’ [188]. After whole-plasmid PCR with Phusion DNA Polymerase (ThermoFisher) or Q5 DNA Polymerase the PCR mix was digested with DpnI to selectively remove wild-type template. Subsequently, the amplicons were column-purified (DNA Clean & Concentrator-5 Kit; Zymo Research) and digested with BsaI to create sticky ends. After a further column purification, amplicons were self-ligated with T4 DNA Ligase. The ligation product was amplified and purified in the same way as the round 1 library.

Combinatorial active site substitution library of *PkDLH* 20 residues in and around the active site were chosen according to the criteria described in Section 5.2.2. Each of these positions was cloned in a binary fashion, either corresponding to the *PkDLH* residue or the corresponding residue in P91 and its homologues, such that the library consisted of all permutations of the following substitutions:

A34F, Q35P, I37A, M44A, Y64H, W88I, M93W, Y122F, L124F, G143T, Y144F, Y145H, V147G, G148L, E150P, H172P, F173L, G201V, A205T, R206D.

The *PkDLH* gene was then split into seven fragments such that all residues to be mutated could be clustered together on the overhangs of mutagenic primers (Figure 5.3). After creation by PCR, the individual fragments were then assembled by Golden Gate Assembly with the type-IIs restriction enzyme BsaI, as describe above for the P91 libraries. As one-pot assembly was not efficient enough, yielding only 10^3 – 10^4 colonies, the six short fragments were ligated separately and amplified by limited-cycle PCR in order to avoid PCR bias. The resulting insert was then size-selected by agarose gel electrophoresis and ligated into the large pASK-IBA5+ backbone fragment, yielding a library of $> 3 \cdot 10^7$ transformants and < 0.5 % non-insert background. Sequencing of randomly picked clones confirmed that the majority (20/24) assembled into complete genes without fragment duplications or

deletions.

Loop grafting into *HsDLH* The combinatorial loop grafting library was constructed using cassette mutagenesis. The single fragments (Table 8.1) were ordered as nucleotide oligomers with *Bsa*I-overhangs, annealed into duplexes and subsequently assembled by Golden Gate Assembly into the *HsDLH* gene, in analogous fashion to the libraries of P91 described above.

Table 8.1: Summary of fragments combinatorally grafted from P91 and its homologues into *HsDLH*.

Fragment	Region in <i>HsDLH</i>	Introduced mutations	Origin
F1	Q49–I51 (QDI)	PEA	residues 36–38 in P91-WT, <i>HmDLH</i> , <i>m3DLH</i>
		PEL	residues 36–38 in P91-R2
		PEG	residues 36–38 in P91-GGRGWV
F2	Q55–T59 (QLPNT)	NDHA	residues 42–45 in P91, <i>HmDLH</i> , <i>m3DLH</i>
F3	Y155–V158 (YGIV)	HGGL	residues 141–144 in P91, <i>HmDLH</i>
F4	S208–P228 (SGQTHGFVHRKREDPCSPADKP)	GNALHSFTDPEADTRGMDGLAYDAR	residues 195–209 in <i>HmDLH</i>
		GNAHSFTDPAADAHGMAGLAYEPL	residues 196–220 in <i>m3DLH</i>
		GNAVHSFTDPLAGSHGIPGLAYDAT	residues 195–219 in P91-WT
		GNAVHSFTDPLAGSHGWPVAYDAT	residues 195–219 in P91-R2

8.4 Microfluidic droplet screening

8.4.1 Design and fabrication of master wafers for microfluidic devices

Note: This step was performed by Tomasz S. Kamiński.

The channel layout for the microfluidic chips was designed using AutoCAD (Autodesk, USA) and printed out on a film photomask (Micro Lithography Services, UK). The master wafer was then photolithographically fabricated in a clean room. In brief, a silicon wafer (Prime 3 Inch Czochralski Silicon wafer, thickness = 381 ± 20 μm , orientation (100), one-side polished; Microchemicals, Germany) was covered with a thin layer of SU-8 2000 series photoresist material using a spin coater (SPIN150i spin coater, Polos by SPS, Germany). Depending on the channel height, various resists with different viscosities were used (SU-8 2015, SU-8 2025, SU-8

8 *Materials and Methods*

2050, and SU-8 2075). Next, the wafer was soft-baked on a hot plate and the channel pattern was subsequently patterned into the master by photolithography using a mask aligner (MJB4 mask aligner; Süss MicroTec, Germany). In subsequent steps, the wafer was post-baked and developed in propylene glycol monomethyl ether acetate. Next, the wafer was hard-baked for 10 min at 200 °C and the heights of the structures were measured using a profilometer (Veeco Dektak 6M Stylus Surface Profilometer; Bruker, USA). Finally, the chip was silanised by deposition of pure trichloro(1H,1H,2H,2H- perfluorooctyl)silane (2 µL) in close proximity to the wafer placed in a Petri dish which was kept for 30 min in the vacuum chamber at 20 mbar to generate vapours of silane.

8.4.2 **Fabrication of microfluidic devices**

For the fabrication of a microfluidic chip, the master wafer was coated with a mixture of poly(dimethyl)siloxane (PDMS) and curing agent (Sylgard 184 Silicone Elastomer Kit, Dow Chemical Company, USA) in a 10:1 ratio (w/w). After degassing and curing at 65 °C for \approx 3 h, the PDMS device was removed from the master and holes for tubing connections were punched using a 1 mm biopsy punch with a plunger (Kai Medical, Japan). The device was then attached to a microscope 1 mm thick glass slide (flow-focusing device) or a 0.13 mm thin glass cover slip (sorting device) by treatment with oxygen plasma (Femto plasma system; Diener Electronic, Germany) for 30 s, followed by a baking step of \approx 20 min incubation at 65 °C. For hydrophobic treatment of the channel surface, a freshly prepared and filtered solution of 1 % (v/v) trichloro(1H,1H,2H,2H- perfluorooctyl)silane in fluorinated oil (Novec HFE-7500, Fluorochem, UK) was injected into the channels, followed by approximately 45 min of incubation at 65 °C.

8.4.3 Preparation of cells for encapsulation

Escherichia coli cells (E. cloni 10G Elite; Lucigen, USA) were transformed with 2.5 μL of library plasmids, yielding 10^6 – 10^7 colonies after overnight incubation on agar plates, as determined by serial dilution. Transformed cells were induced with anhydrotetracycline (final concentration 200 ng/mL; IBA Life Sciences, Germany) and incubated for expression for \approx 14 h in 20 mL LB medium at 20 °C and 220 rpm shaking. After expression, the cells were washed five times with buffer (100 mM MOPS-NaOH, 150 mM NaCl, pH 8.0) and diluted to $\text{OD}_{600} = 1.0$. A 200 μL aliquot of the suspension was diluted 1:2 with 100 μL droplet assay buffer (100 mM MOPS-NaOH, 150 mM NaCl, pH 8.0, cOMplete EDTA-free protease inhibitor (one tablet per 50 mL; Roche, Switzerland) and 100 μL Percoll (a silica nanoparticle solution to reduce cell-cell adhesion and prevent sedimentation of cells in the syringe; Cytiva, USA) to a final bacterial density of $\text{OD}_{600} = 0.5$. For a droplet diameter volume of \approx 3 pL the resulting final bacterial density of $\text{OD}_{600} = 0.25$ resulted in 16 % of droplets with a single bacterium and 3 % of droplets containing two or more bacteria. All solutions (except for cell suspensions) were previously filtered with 0.2 μm PTFE syringe filters (Acrodisc CR 13 mm syringe filters, PALL Life Sciences, USA) to avoid clogging of microfluidic channels.

8.4.4 Compartmentalisation of cells into microdroplets

Monodisperse water-in-oil microdroplets were generated with a microfluidic flow-focusing device. The device was connected via PE tubing (0.38 mm inner diameter, 1.09 mm outer diameter; Adtech Polymer Engineering, UK) to glass syringes (100 μL and 1 mL; SGE Analytics, Australia), which were driven by syringe pumps (neMESYS, Cetoni, Germany). Fluorocarbon oil (3M Novec HFE-7500) containing 0.5 to 2 % (w/w) surfactant (008-FluoroSurfactant; RAN Biotechnologies, USA) served as oil phase. The two aqueous streams were supplied with the cell solution and with a 3 μM substrate solution containing lysis agents ($0.7\times$ BugBuster protein

8 *Materials and Methods*

extraction reagent, Merck Millipore; 60 kU/mL rLysozyme, Novagen) in droplet assay buffer, respectively. Droplet formation was monitored using an inverted microscope (SP98I, Brunel Microscopes, UK) with a high-speed camera (Phantom Miro eX4, Vision Research, USA). For long incubation times in evolution round 1, requiring off-chip incubation, flow rates of 50 $\mu\text{L/h}$ for the aqueous phases and 500 $\mu\text{L/h}$ for the oil phase were used to generate droplets with a volume of 3 pL at rates of 0.5 to 3 kHz. Droplets were collected into a long PE tubing (0.38 mm ID, 1.09 mm OD; Adtech Polymer Engineering, UK) which was closed with a syringe needle after collection. For short incubation times in evolution round 2, requiring on-chip incubation, an integrated chip was used, combining a flow-focussing module, a delay line, and a sorting module on a single device. A detailed description of the chip design can be found in the Appendix. On this chip, flow rates were 7.5 $\mu\text{L/h}$ for the aqueous phases, 25 $\mu\text{L/h}$ for the oil phase, ≈ -10 $\mu\text{L/h}$ for the oil extractor, and ≈ 300 $\mu\text{L/h}$ for the spacing oil, resulting in a droplet volume of ≈ 11 pL.

8.4.5 **Fluorescence-assisted droplet sorting (FADS)**

Optics and electronics of the microfluidic on-chip sorting device were set-up as previously described [187, 268]. After incubation at room temperature, droplets were reinjected from the collection tubing into the sorting device at a rate of 10 to 25 $\mu\text{L/h}$. To enable precise sorting of single droplets, the distance between the droplets was increased by injection of spacing oil (Novec HFE-7500, 3M, USA) into the device at a flow rate of 100 to 300 $\mu\text{L/h}$. The asymmetric Y-shaped junction in the device ensures that all droplets automatically flow into the waste channel, unless deviated by an electrical pulse into the sorting channel. A 488-nm laser was focused 100 μm upstream of the sorting junction through a 40 \times microscope objective (UPlanFLN, Olympus, Japan) for fluorophore excitation and the emitted fluorescent light was collected and amplified using photomultiplier tubes (PMM02, Thorlabs, USA). Whenever the fluorescence peak reached a user-defined threshold, an electric field was applied by the two electrodes on the sorting device, attracting the highly fluorescent droplet towards the narrower sorting channel (Figure 1). Droplets

were sorted into a collection tube that was pre-filled with 100 μL nuclease-free water.

8.4.6 DNA recovery from microdroplets

Plasmids from sorted droplets were recovered by de-emulsification with 1H,1H,2H,2H-perfluoro-1-octanol (PFO, Alfa Aesar, USA) and subsequent column purification and electroporation into highly electrocompetent *E. coli* cells (*E. coli* 10G Elite, Lucigen, USA) as previously described [187]. In brief, after droplet sorting, 400 ng of salmon sperm DNA (Invitrogen, USA) were added to the collection tube. Subsequently, PFO (approx. half the volume of the collected oil volume; e.g. 400 μL PFO for 800 μL oil in the collection tube) was added. The tube was vortexed for ≈ 60 s and centrifuged for 1 min at 14000 g. The top aqueous layer was completely removed and transferred into a new tube. The oil phase remaining in the collection tube was re-extracted by adding 100 μL nuclease-free water, 400 ng salmon sperm DNA, 100 μL PFO, vortexing for 1 min and subsequent centrifugation for 1 min at 14000 g. The top aqueous layer was completely removed and united with the previously removed aqueous layer. The united aqueous fractions usually contain a small remaining bottom oil phase. This remaining oil phase was then extracted by addition of 100 μL PFO, 1 min of vortexing, and 1 min of centrifugation at 1400 g. The top aqueous phase was then removed without any traces of oil phase and the plasmid DNA was column-purified using a DNA Clean & Concentrator-5 Kit (Zymo Research, USA) according to the manufacturer's instructions, with the following modifications: The ratio of binding buffer to sample volume was 6:1. The wash buffer was carefully rinsed along the walls. After the second wash step, the flow-through was removed and the column was centrifuged once more for 1 min to dry. For elution, 6 μL of pre-warmed (50 $^{\circ}\text{C}$) elution buffer were used. The eluted DNA was then electroporated into highly electrocompetent *E. coli* cells (*E. coli* 10G Elite, Lucigen, USA) according to the manufacturer's instructions. Throughout the DNA extraction procedure, low-DNA-binding tubes (1.5 mL DNA LoBind tube; Eppendorf, Germany) and low-retention tips (Axygen Maxymum Recovery Filter

8 Materials and Methods

Tips; Corning, USA) were used.

8.5 Microtiter plate screening

To quantify the lysate activity of P91 variants, individual colonies were picked and grown in 96-deep-well plates in 500 μ L Luria-Bertani (LB) medium with 100 μ L/mL carbenicillin at 37 °C/1 050 rpm for \approx 14 h. Subsequently, 10 μ L of overnight cultures were used to inoculate 490 μ L of medium for expression cultures in 96-well deep-well plates which were grown at 37 °C/1 050 rpm for \approx 2 h until OD₆₀₀ \approx 0.5. Expression was then induced with anhydrotetracycline (final concentration 200 ng/mL; IBA Life Sciences, Germany) and carried out for 14 h at 20 °C and 1 050 rpm shaking. Cells were pelleted by centrifugation at 3 320 rcf for 60 min, the supernatant was then discarded, and cells were lysed with 100 μ L lysis buffer (50 mM HEPES-NaOH, 150 mM NaCl, pH 8.0, 60 kU/mL rLysozyme, 1 \times BugBuster) for 20 min at 20 °C and 1 050 rpm shaking. Cell lysates were diluted 1:20 or 1:400 in assay buffer (50 mM HEPES-NaOH, 150 mM NaCl, pH 8.0). For the reaction, 190 μ L of the phosphotriester substrate paraoxon-ethyl (100 μ M) or FDDEP in assay buffer (3 μ M) were added to 10 μ L aliquots of the diluted lysate in microtiter plates and the formation of fluorescein or *p*-nitrophenol was recorded in a plate reader (Infinite M200, Tecan, Switzerland) for 15 min at a wavelength of 405 nm for *p*-nitrophenol and at an excitation wavelength of 480 nm and an emission wavelength of 520 nm for fluorescein.

8.6 Protein expression

Plasmids isolated from single colonies were used to transform BL21(DE3) cells (Exception: BChE was expressed in Origami B cells). Expression cultures were then inoculated by a similar ‘plating’ method as previously described [269, 270]. In brief, a

8.7 Protein purification

dense lawn of freshly transformed BL21(DE3) cells was directly scraped into 500 mL TB medium containing 100 $\mu\text{g}/\text{mL}$ carbenicillin in a 2-L baffled erlenmeyer flask. The cells were grown for ≈ 60 min at 37 °C and 200 rpm shaking before being induced with anhydrotetracyclin (final concentration 200 ng/mL). Protein was then expressed at 20 °C/200 rpm for 18 to 20 h. Cells were harvested by centrifugation at 4 000 rcf for 10 min, the supernatant was discarded, and the dry pellet was stored at -80 °C.

8.7 Protein purification

Proteins were purified by affinity chromatography. All constructs were initially cloned with an N-terminal StrepII-tag (MASWSHPQFEKGA) which allows single-step affinity capture at very high purities on Strep-tactin resin. For high-yield purification, as required for transient-state kinetics, the StrepII-tag was exchanged for an N-terminal 6 \times His-tag (MHHHHHGGG) for Ni-NTA affinity chromatography. For the combinatorial loop-grafting into *HsDLH*, an N-terminal His-TwinStrep-SUMO-tag was cloned into the construct for higher solubility. hBChE-7 G1187H was expressed with an N-terminal TrxA-His-tag (origin: pET32b(+)) for the promotion of disulfide-bridge formation and higher solubility. Proteins with these tags were equally purified by Ni-NTA affinity chromatography.

The sequence of the His-TwinStrep-SUMO-tag is:

```
MGSSHHHHHSSGLVPRGSHMASWSHPQFEKGGGSGGGSGGSAWSHPQFEKMSDSEVNQEAKPE  
VKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFKRQKEMDSLRFlyDGIRIQADQT  
PEDLMEDNDIIEAHREQIGGS
```

The sequence of the TrxA-His-tag is:

```
MSDKIIHLTDDSFDTDLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQN  
PGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSGHMMMMHHHSSGLVP  
RSGMKETAAAKFERQHMDSPDLGTDDDDKA
```

8.7.1 Strep-Tactin affinity chromatography

For purification of StrepII-tagged proteins, the cell pellet was resuspended in lysis buffer (50 mM HEPES-NaOH, 150 mM NaCl, pH 8.0, 1 mM TCEP, 0.5–1 mg/mL lysozyme, 0.1 % Triton X-100, 0.01 % (= 25 units/mL) benzonase nuclease) and rolled for 30–60 min at room temperature (≈ 23 °C). Afterwards, the lysate was cleared by centrifugation at 20 000 rcf/4 °C for 20 min and the soluble fraction was loaded onto an equilibrated Strep-Tactin gravity flow column (Strep-Tactin Agarose Resin, IBA Life Sciences, Germany). The protein on the column was washed with 5×1 column volume of wash buffer (50 mM HEPES-NaOH, 150 mM NaCl, pH 8.0, 1 mM TCEP) and eluted with $6-8 \times 0.5$ column volumes of elution buffer (50 mM HEPES-NaOH, 150 mM NaCl, pH 8.0, 1 mM TCEP, 2.5 mM *d*-desthiobiotin).

8.7.2 Ni-NTA affinity chromatography

For purification of 6 \times His-tagged proteins, the cell pellet was resuspended in lysis buffer (50 mM HEPES-NaOH, 150 mM NaCl, pH 8.0, 1 mM TCEP, 20 mM imidazole, 0.5–1 mg/mL lysozyme, 0.1 % Triton X-100, 0.01 % (= 25 units/mL) benzonase nuclease) and rolled for 30–60 min at room temperature (≈ 23 °C). Afterwards, the lysate was cleared by centrifugation at 20 000 rcf/4 °C for 20 min and the soluble fraction was loaded onto an equilibrated Ni-NTA gravity flow column (Super Ni-NTA Agarose Resin, Neo Biotech, France). The protein on the column was washed with 5×2.5 column volumes of wash buffer (50 mM HEPES-NaOH, 150 mM NaCl, pH 8.0, 1 mM TCEP, 20 mM imidazole) and eluted with 5×0.5 column volumes of elution buffer (50 mM HEPES-NaOH, 150 mM NaCl, pH 8.0, 1 mM TCEP, 250 mM imidazole).

8.7.3 Buffer exchange and protein quantification

Following affinity chromatography, the eluate was concentrated with a spin concentrator (Vivaspin 10 000 kDa MWCO, Sartorius, Germany) and subsequently exchanged into eluent-free assay buffer (50 mM HEPES-NaOH, 150 mM NaCl, pH 8.0, 1 mM TCEP) using PD 10 desalting columns (Cytiva, USA). Enzyme purity was controlled by SDS-PAGE and concentrations were determined by measurement of absorption at 280 nm on a NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific, USA), using an extinction coefficient calculated with the ProtParam web tool (<https://web.expasy.org/protparam>) [271].

8.8 Kinetics

8.8.1 Steady-state kinetics

For kinetic measurements, enzyme concentrations were kept at least 10–100-fold lower than the lowest substrate concentration. Substrate concentrations were chosen to span ≈ 10 -fold below and above K_M , as far as not limited by substrate solubility. Optimal starting enzyme concentration (E_0) and substrate concentration ranges were determined for each variant and substrate combination by empirical sampling. Substrates were pre-dissolved in DMSO at stocks of 200-fold the final concentration, in order to ensure constant DMSO concentration (0.5 %) across all substrate concentrations, as DMSO content was found to influence the catalytic parameters while being required as a co-solvent for substrate stocks. Upon measurement, aliquots of these substrate stocks were diluted 1:100 in assay buffer (50 mM HEPES-NaOH, 150 mM NaCl, pH 8.0, 1 mM TCEP), of which 100 μ L were subsequently mixed with 100 μ L of 2-fold concentrated enzyme solution in microtiter plate wells. The progress of the reaction was monitored by absorbance (at a wavelength of 405 nm for *p*-nitrophenol) or fluorescence (at an excitation wavelength of

8 Materials and Methods

480 nm and an emission wavelength of 520 nm for fluorescein) in a spectrophotometric microplate reader (Tecan Infinite 200PRO, Tecan, Switzerland) at 25 °C. The initial rates were extracted by linear fit of the first measurements (at < 10 % progress of the reaction) for each substrate concentration and normalised with an extinction coefficient determined from a calibration curve. In order to determine the Michaelis-Menten parameters k_{cat} , K_M , and (in case of substrate inhibition) K_i , the data was fitted to the following equation using the non-linear fitting function `nls()` in R [272]:

$$\frac{v}{[E_0]} = \frac{k_{cat} \cdot [S]}{K_M + [S]} \quad (8.1)$$

or, in the case of substrate inhibition,

$$\frac{v}{[E_0]} = \frac{k_{cat} \cdot [S]}{K_M + [S] + \frac{[S]^2}{K_i}} \quad (8.2)$$

where v is the initial rate of the reaction, $[E_0]$ is the initial enzyme concentration, and $[S]$ is the substrate concentration.

The kinetics of P91 and its homologues were subject to considerable variation between purification batches (up to \approx 3–4-fold), potentially due to slight differences in expression and purification times affecting the oligomeric state of the enzyme or the oxidation state of the cysteine nucleophile. To increase the precision of enzyme quantification for kinetics, stoichiometric active site titration was tested. However, the widely used covalent catalytic triad inhibitor phenylmethylsulfonyl fluoride (PMSF) did not react with P91. Accordingly, best efforts were made to minimise the variation between expression, purification and storage conditions and to keep the enzyme in a reducing environment at all times. Nevertheless, to ensure a maximum of comparability, whenever kinetic values of two variants are directly compared to each other on a linear scale (i.e. not only looking at orders of magnitude on a logarithmic scale), the two variants were expressed, purified, and assayed in parallel (e.g. P91-WT and P91-R2; P91-WT and P91-ins; *HsDLH* and *HsDLH*-1E11).

Comparing previously by Pierre-Yves Colin [149, 248] measured kinetics values of

P91-WT for paraoxon-ethyl to the ones measured in this work, it stands out that, while the overall $\frac{k_{cat}}{K_M}$ values are very similar, both k_{cat} and K_M values are approximately one order of magnitude higher in the previous work than in this work. Beyond differences in buffer, affinity tag and purification procedure, the content of the co-solvent DMSO likely contributes to this difference. While 0.5 % DMSO were present in the measurements reported in this work, increasing the DMSO content to 2.5 % resulted in a very similar $\frac{k_{cat}}{K_M}$, however increasing k_{cat} by ≈ 10 -fold, while simultaneously increasing K_M by the same factor, overall replicating the pattern described above.

8.8.2 Transient-state kinetics

Fast transient-state kinetics for the hydrolysis of phosphotriesters FDDEP and paraoxon-ethyl were measured with a SX20 stopped-flow spectrophotometer (Applied Photophysics, UK) at the same temperature, in the same buffer and with the same substrate concentrations as the steady-state kinetics. Measurement traces (Figure 10.1) were fitted to the following exponential burst equation:

$$F = A \cdot (1 - e^{-k_{obs} \cdot t}) + B \cdot t + C \quad (8.3)$$

where F is the measured absorbance or fluorescence, t is the time, A is the amplitude of the burst, B is the slope of the second phase of the reaction and C is the offset. The observed rate k_{obs} was then fitted to the following equation to determine k_2 (Figure 3.3b, Table 3.1):

$$k_{obs} = \frac{k_2 \cdot [S]}{K_M + [S]} \quad (8.4)$$

For the double-substituted phosphotriester FDDEP, the apparent k_2 as measured from the burst phase represents the hydrolysis of one of the two phosphotriester groups on the substrate and thus the release of the mono-phosphorylated reaction product, fluorescein mono(diethylphosphate), which has an unknown extinction coefficient. However, the observed burst amplitude (0.4–1.0, thus corresponding to the value expected for fluorescein release) indicates that most of the substrate

8 Materials and Methods

fluorescence is generated by the release of this first group, thus allowing the use of a fluorescein calibration curve for the approximate quantification of k_2 .

The de-phosphorylation rate k_3 was separately determined with the mono-substituted substrate paraoxon-ethyl, in order to exclude potential obfuscating effects due to the complex downstream kinetics of the double substituted substrate FDDEP (where the original double-substituted substrate FDDEP and the increasingly formed, mono-substituted intermediate FMDEP compete for turnover). As reaction with paraoxon-ethyl forms the same diethyl-phosphate adduct intermediate, k_3 with FDDEP is identical to k_3 with paraoxon-ethyl. Assuming a reaction model with one reversible binding step and two irreversible steps (Figure 3.1), the turnover number k_{cat} and the catalytic efficiency $\frac{k_{cat}}{K_M}$ can be described as:

$$k_{cat} = \frac{k_2 \cdot k_3}{k_2 + k_3} \quad (8.5)$$

$$\frac{k_{cat}}{K_M} = \frac{k_2 \cdot k_1}{k_2 + k_{-1}} \quad (8.6)$$

Given that in this case $k_2 \ll k_3$, the term for k_{cat} can be simplified to:

$$k_{cat} \approx k_3 \quad (8.7)$$

Thus, the initial slope of the second phase of the burst reaction with paraoxon-ethyl was plotted against substrate concentration and fitted to the Michaelis-Menten equation to determine k_3 (Figure 10.3, Table 3.1). In cases where the burst is less pronounced (as in P91-R1 C118S), k_{cat} is an upper estimate of k_3 while the real value of k_3 might be lower.

8.9 Multiple sequence alignments and phylogenetic trees

Multiple sequence alignments were done using Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) at default settings [273]. For the generation of the phylogenetic tree shown in Figure 4.4, a multiple sequence alignment was generated using T-coffee Espresso (<http://tcoffee.crg.cat/apps/tcoffee/do:espresso>) [274], which takes structural information into account. This multiple sequence alignment was used as input for PhyML 3.0 (<http://www.atgc-montpellier.fr/phyml/>) [275]. To choose a substitution model, automatic model selection with Smart Model Selection (SMS) was used with the AIC criterion [276]. Tree search was initiated with BioNJ and tree improvement was done using nearest-neighbour interchanges (NNI). Bootstrapping was performed with $N = 1000$.

8.10 Sequence similarity networks

All sequence similarity networks (SSNs) were generated using the EFI-EST web resource (<https://efi.igb.illinois.edu/efi-est/>) [277, 278]. Edge calculation for the SSNs was done by all-by-all BLAST alignment of the input sequences using an E -value of 10^{-5} and excluding UniProt-defined fragments. For Figure 4.1 the number of nodes in the networks was reduced by using a fractioning parameter of $N = 1000$ (a, α/β hydrolase superfamily) and $N = 40$ (b, DLH protein family). This selects every N^{th} sequence in the family, while keeping all sequences with Swiss-Prot annotations. Sequences were additionally trimmed to the domain boundaries containing the predefined input Pfam family. SSNs were visualised using Cytoscape 3.4.0 in the yFiles Organic Layout [279].

8.11 Protein crystallisation and X-ray structure determination

Crystallisation trials, X-ray data collection, and solution of structures were done by Dr Eleanor Campbell.

8.11.1 Protein crystallisation

Purified variants were concentrated to 10–20 mg/mL in 50 mM HEPES-NaOH, 150 mM NaCl, pH 8.0. Screens were optimised by sitting-drop vapour diffusion at 19 °C. Protein solutions were dispensed using a Mosquito Protein Crystallisation System (sptlabtech Ltd.; UK), with final drop volumes of 300 nL (1:2 ratio protein to crystallisation buffer).

Crystals of *m3DLH* grew in a solution of 0.1 M sodium citrate (pH 5.6), 30 % (w/v) PEG 4K, and 0.2 M ammonium acetate. Crystals of *SaDLH* grew in a solution of 0.1 M HEPES (pH 7.5) and 4.3 M NaCl. Crystals of *ScDLH* grew in a solution of 0.1 M Tris (pH 7.0), 20 % (w/v) PEG 3K, and 0.2 M calcium acetate. Crystals of *SjDLH* grew in a solution 0.1 M Bicine (pH 9.0) and 65 % (v/v) 2-methyl-2,4-pentanediol (MPD).

8.11.2 Data collection and structure solution

All samples were snap frozen in liquid nitrogen with no additional cryoprotection, and diffraction data were collected at Diamond Light Source (beamlines i03, i04, and i24) at 0.987 Å, 80 K. Diffraction data were indexed and integrated using XDS [280] and merged using AIMLESS as implemented within the CCP4 program suite [281]. Data resolution cut-offs were chosen using half dataset correlation coefficients, described by Karplus and Diederichs [282]. Structures were solved by molec-

ular replacement (MOLREP) [283] using the structure of a putative diene lactone hydrolase from *Klebsiella pneumoniae* (KpDLH, PDB ID: 3F67) as the input model. Refinement was performed using REFMAC [284] and phenix.refine [285], and manual rebuilding was done in Coot [286].

8.12 Structural models

When crystal structures could not be obtained, homology models were generated using the SWISS-MODEL webserver (<https://swissmodel.expasy.org/>) [249]. In addition, for HsDLH the *de novo* deep learning-based structure prediction method DMPfold [256] on the PSIPRED web server [257] was used (<http://bioinf.cs.ucl.ac.uk/psipred/>). For the comparison of wild-type and evolved P91, a structural model of P91-R2 was generated using AlphaFold2/ColabFold [197, 198].

8.13 B-factor analysis

For comparison across different structures, B-factor were normalised according to the following equation:

$$B' = \frac{B - \bar{B}}{\sigma} \quad (8.8)$$

where B' is the normalised B-factor, B is the raw B-factor of any given α -carbon atom in the structure, \bar{B} is the average B-factor of all α -carbon atoms in the structure and σ is the corresponding standard deviation.

B-factors were analysed using the Bio3d package in R [287].

9 References

1. Lewis, S. L. & Maslin, M. A. Defining the Anthropocene. *Nature* **519**, 171–180. doi:10.1038/nature14258 (2015).
2. Ruddiman, W. F. The Anthropocene. *Annual Review of Earth and Planetary Sciences* **41**, 45–68. doi:10.1146/annurev-earth-050212-123944 (2013).
3. Waters, C. N., Zalasiewicz, J. A., Williams, M., Ellis, M. A. & Snelling, A. M. A Stratigraphical Basis for the Anthropocene? *Geological Society, London, Special Publications* **395**, 1–21. doi:10.1144/SP395.18 (2014).
4. Harris, R. & Paxman, J. *A Higher Form of Killing: The Secret History of Chemical and Biological Warfare* 301 pp. (Random House Trade Paperbacks, New York, 2002).
5. Matolcsy, G., Nádasy, M. & Andriska, V. *Pesticide Chemistry* (Elsevier, 1989).
6. Reemtsma, T., Quintana, J. B., Rodil, R., García-López, M. & Rodríguez, I. Organophosphorus Flame Retardants and Plasticizers in Water and Air I. Occurrence and Fate. *TrAC Trends in Analytical Chemistry* **27**, 727–737. doi:10.1016/j.trac.2008.07.002 (2008).
7. Van der Veen, I. & de Boer, J. Phosphorus Flame Retardants: Properties, Production, Environmental Occurrence, Toxicity and Analysis. *Chemosphere* **88**, 1119–1153. doi:10.1016/j.chemosphere.2012.03.067 (2012).
8. Cheng, W. *et al.* Detection and Distribution of Tris(2-Chloroethyl) Phosphate on the East Antarctic Ice Sheet. *Chemosphere* **92**, 1017–1021. doi:10.1016/j.chemosphere.2013.03.038 (2013).

9 References

9. Salamova, A., Hermanson, M. H. & Hites, R. A. Organophosphate and Halogenated Flame Retardants in Atmospheric Particles from a European Arctic Site. *Environmental Science & Technology* **48**, 6133–6140. doi:10.1021/es500911d (2014).
10. Ma, H. *et al.* Correction of a Pathogenic Gene Mutation in Human Embryos. *Nature*. doi:10.1038/nature23305 (2017).
11. Marklund, A., Andersson, B. & Haglund, P. Organophosphorus Flame Retardants and Plasticizers in Air from Various Indoor Environments. *Journal of Environmental Monitoring* **7**, 814–819. doi:10.1039/B505587C (2005).
12. Van den Eede, N. *et al.* Age as a Determinant of Phosphate Flame Retardant Exposure of the Australian Population and Identification of Novel Urinary PFR Metabolites. *Environment International* **74**, 1–8. doi:10.1016/j.envint.2014.09.005 (2015).
13. Glaros, T. *et al.* Discovery of Treatment for Nerve Agents Targeting a New Metabolic Pathway. *Archives of Toxicology* **94**, 3249–3264. doi:10.1007/s00204-020-02820-4 (2020).
14. Hasinoff, B. B. Kinetics of Acetylthiocholine Binding to Electric Eel Acetylcholinesterase in Glycerol/Water Solvents of Increased Viscosity Evidence for a Diffusion-Controlled Reaction. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **704**, 52–58. doi:10.1016/0167-4838(82)90131-5 (1982).
15. Eddleston, M., Buckley, N. A., Eyer, P. & Dawson, A. H. Management of Acute Organophosphorus Pesticide Poisoning. **371**, 11 (2008).
16. Buratti, F. M. *et al.* Kinetic Parameters of OPT Pesticide Desulfuration by C-DNA Expressed Human CYPs. *Environmental Toxicology and Pharmacology* **11**, 181–190. doi:10.1016/S1382-6689(02)00010-8 (2002).
17. Sethunathan, N. & MacRae, I. C. Persistence and Biodegradation of Diazinon in Submerged Soils. *Journal of Agricultural and Food Chemistry* **17**, 221–225. doi:10.1021/jf60162a042 (1969).

18. Siddaramappa, R., Rajaram, K. P. & Sethunathan, N. Degradation of Parathion by Bacteria Isolated from Flooded Soil. *Applied Microbiology* **26**, 846–849 (1973).
19. Sethunathan, N. & Yoshida, T. A Flavobacterium Sp. That Degrades Diazinon and Parathion. *Canadian Journal of Microbiology*. doi:10.1139/m73-138 (1973).
20. Munnecke, D. M. & Hsieh, D. P. H. Microbial Decontamination of Parathion and P-Nitrophenol in Aqueous Media. *Applied Microbiology* **28**, 212–217 (1974).
21. Cook, A. M., Daughton, C. G. & Alexander, M. Phosphorus-Containing Pesticide Breakdown Products: Quantitative Utilization as Phosphorus Sources by Bacteria. *Applied and Environmental Microbiology* **36**, 668–672 (1978).
22. Serdar, C. M., Gibson, D. T., Munnecke, D. M. & Lancaster, J. H. Plasmid Involvement in Parathion Hydrolysis by *Pseudomonas Diminuta*. *Applied and Environmental Microbiology* **44**, 246–249 (1982).
23. Serdar, C. M. & Gibson, D. T. Enzymatic Hydrolysis of Organophosphates: Cloning and Expression of a Parathion Hydrolase Gene from *Pseudomonas Diminuta*. *Bio/Technology* **3**, 567–571. doi:10.1038/nbt0685-567 (1985).
24. Dumas, D. P., Caldwell, S. R., Wild, J. R. & Raushel, F. M. Purification and Properties of the Phosphotriesterase from *Pseudomonas Diminuta*. *The Journal of Biological Chemistry* **264**, 19659–19665 (1989).
25. Caldwell, S. R., Newcomb, J. R., Schlecht, K. A. & Raushel, F. M. Limits of Diffusion in the Hydrolysis of Substrates by the Phosphotriesterase from *Pseudomonas Diminuta*. *Biochemistry* **30**, 7438–7444. doi:10.1021/bi00244a010 (1991).
26. DeFrank, J. J. & Cheng, T. C. Purification and Properties of an Organophosphorus Acid Anhydrase from a Halophilic Bacterial Isolate. *Journal of Bacteriology* **173**, 1938–1943. doi:10.1128/jb.173.6.1938-1943.1991 (1991).

9 References

27. Scharff, E. I., Koepke, J., Fritsch, G., Lücke, C. & Rüterjans, H. Crystal Structure of Diisopropylfluorophosphatase from *Loligo Vulgaris*. *Structure* **9**, 493–502. doi:10.1016/S0969-2126(01)00610-4 (2001).
28. Harel, M. *et al.* Structure and Evolution of the Serum Paraoxonase Family of Detoxifying and Anti-Atherosclerotic Enzymes. *Nature Structural & Molecular Biology* **11**, 412–419. doi:10.1038/nsmb767 (2004).
29. Dong, Y.-J. *et al.* Crystal Structure of Methyl Parathion Hydrolase from *Pseudomonas* Sp. WBC-3. *Journal of Molecular Biology* **353**, 655–663. doi:10.1016/j.jmb.2005.08.057 (2005).
30. Bigley, A. N. & Raushel, F. M. Catalytic Mechanisms for Phosphotriesterases. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1834**, 443–453. doi:10.1016/j.bbapap.2012.04.004 (2013).
31. Wang, L. & Sun, Y. Engineering Organophosphate Hydrolase for Enhanced Biocatalytic Performance: A Review. *Biochemical Engineering Journal* **168**, 107945. doi:10.1016/j.bej.2021.107945 (2021).
32. Afriat, L., Roodveldt, C., Manco, G. & Tawfik, D. S. The Latent Promiscuity of Newly Identified Microbial Lactonases Is Linked to a Recently Diverged Phosphotriesterase. *Biochemistry* **45**, 13677–13686. doi:10.1021/bi061268r (2006).
33. Afriat-Jurnou, L., Jackson, C. J. & Tawfik, D. S. Reconstructing a Missing Link in the Evolution of a Recently Diverged Phosphotriesterase by Active-Site Loop Remodeling. *Biochemistry* **51**, 6047–6055. doi:10.1021/bi300694t (2012).
34. Hawwa, R., Larsen, S. D., Ratia, K. & Mesecar, A. D. Structure-Based and Random Mutagenesis Approaches Increase the Organophosphate-Degrading Activity of a Phosphotriesterase Homologue from *Deinococcus Radiodurans*. *Journal of Molecular Biology* **393**, 36–57. doi:10.1016/j.jmb.2009.06.083 (2009).
35. Meier, M. M. *et al.* Molecular Engineering of Organophosphate Hydrolysis Activity from a Weak Promiscuous Lactonase Template. *Journal of the American Chemical Society* **135**, 11670–11677. doi:10.1021/ja405911h (2013).

36. Hoque, M. A. *et al.* Stepwise Loop Insertion Strategy for Active Site Remodeling to Generate Novel Enzyme Functions. *ACS Chemical Biology* **12**, 1188–1193. doi:10.1021/acscchembio.7b00018 (2017).
37. Jacquet, P. *et al.* Rational Engineering of a Native Hyperthermostable Lactonase into a Broad Spectrum Phosphotriesterase. *Scientific Reports* **7**, 16745. doi:10.1038/s41598-017-16841-0 (2017).
38. Yang, G., Hong, N., Baier, F., Jackson, C. J. & Tokuriki, N. Conformational Tinkering Drives Evolution of a Promiscuous Activity through Indirect Mutational Effects. *Biochemistry* **55**, 4583–4593. doi:10.1021/acs.biochem.6b00561 (2016).
39. Jensen, R. A. Enzyme Recruitment in Evolution of New Function. *Annual Review of Microbiology* **30**, 409–425. doi:10.1146/annurev.mi.30.100176.002205 (1976).
40. O'Brien, P. J. & Herschlag, D. Catalytic Promiscuity and the Evolution of New Enzymatic Activities. *Chemistry & biology* **6**, R91–R105 (1999).
41. Khersonsky, O. & Tawfik, D. S. Enzyme Promiscuity: A Mechanistic and Evolutionary Perspective. *Annual Review of Biochemistry* **79**, 471–505. doi:10.1146/annurev-biochem-030409-143718 (2010).
42. Yang, G. *et al.* Higher-Order Epistasis Shapes the Fitness Landscape of a Xenobiotic-Degrading Enzyme. *Nature Chemical Biology* **15**, 1120–1128. doi:10.1038/s41589-019-0386-3 (2019).
43. Aharoni, A. *et al.* The 'evolvability' of Promiscuous Protein Functions. *Nature Genetics* **37**, 73–76. doi:10.1038/ng1482 (2005).
44. Tokuriki, N. *et al.* Diminishing Returns and Tradeoffs Constrain the Laboratory Optimization of an Enzyme. *Nature Communications* **3**, 1257. doi:10.1038/ncomms2246 (2012).
45. Kaltenbach, M., Emond, S., Hollfelder, F. & Tokuriki, N. Functional Trade-Offs in Promiscuous Enzymes Cannot Be Explained by Intrinsic Mutational Robustness of the Native Activity. *PLoS genetics* **12**, e1006305. doi:10.1371/journal.pgen.1006305 (2016).

9 References

46. Van Loo, B. *et al.* Balancing Specificity and Promiscuity in Enzyme Evolution: Multidimensional Activity Transitions in the Alkaline Phosphatase Superfamily. *Journal of the American Chemical Society* **141**, 370–387. doi:10.1021/jacs.8b10290 (2019).
47. Amitai, G., Gupta, R. D. & Tawfik, D. S. Latent Evolutionary Potentials under the Neutral Mutational Drift of an Enzyme. *HFSP Journal* **1**, 67–78. doi:10.2976/1.2739115/10.2976/1 (2007).
48. Gupta, R. D. & Tawfik, D. S. Directed Enzyme Evolution via Small and Effective Neutral Drift Libraries. *Nature Methods* **5**, 939–942. doi:10.1038/nmeth.1262 (2008).
49. Tokuriki, N. & Tawfik, D. S. Chaperonin Overexpression Promotes Genetic Variation and Enzyme Evolution. *Nature* **459**, 668–673. doi:10.1038/nature08009 (2009).
50. Bloom, J. D., Labthavikul, S. T., Otey, C. R. & Arnold, F. H. Protein Stability Promotes Evolvability. *Proceedings of the National Academy of Sciences* **103**, 5869–5874. doi:10.1073/pnas.0510098103 (2006).
51. Besenmatter, W., Kast, P. & Hilvert, D. Relative Tolerance of Mesostable and Thermostable Protein Homologs to Extensive Mutation. *Proteins: Structure, Function, and Bioinformatics* **66**, 500–506. doi:10.1002/prot.21227 (2007).
52. Wagner, A. Neutralism and Selectionism: A Network-Based Reconciliation. *Nature Reviews Genetics* **9**, 965–974. doi:10.1038/nrg2473 (2008).
53. Paaby, A. B. & Rockman, M. V. Cryptic Genetic Variation: Evolution's Hidden Substrate. *Nature Reviews Genetics* **15**, 247–258. doi:10.1038/nrg3688 (2014).
54. Tokuriki, N. & Tawfik, D. S. Protein Dynamism and Evolvability. *Science* **324**, 203–207. doi:10.1126/science.1169375 (2009).
55. Tóth-Petróczy, A. & Tawfik, D. S. The Robustness and Innovability of Protein Folds. *Current Opinion in Structural Biology* **26**, 131–138. doi:10.1016/j.sbi.2014.06.007 (2014).

56. Hiblot, J., Gotthard, G., Elias, M. & Chabriere, E. Differential Active Site Loop Conformations Mediate Promiscuous Activities in the Lactonase SsoPox. *PLoS ONE* **8**, e75272. doi:10.1371/journal.pone.0075272 (2013).
57. Jackson, C. J. *et al.* Conformational Sampling, Catalysis, and Evolution of the Bacterial Phosphotriesterase. *Proceedings of the National Academy of Sciences* **106**, 21631–21636. doi:10.1073/pnas.0907548106 (2009).
58. Mabbitt, P. D. *et al.* Conformational Disorganization within the Active Site of a Recently Evolved Organophosphate Hydrolase Limits Its Catalytic Efficiency. *Biochemistry* **55**, 1408–1417. doi:10.1021/acs.biochem.5b01322 (2016).
59. Campbell, E. *et al.* The Role of Protein Dynamics in the Evolution of New Enzyme Function. *Nature Chemical Biology* **12**, 944–950. doi:10.1038/nchembio.2175 (2016).
60. Kaltenbach, M., Jackson, C. J., Campbell, E. C., Hollfelder, F. & Tokuriki, N. Reverse Evolution Leads to Genotypic Incompatibility despite Functional and Active Site Convergence. *eLife* **4**. doi:10.7554/eLife.06492 (2015).
61. Haines, D. D. & Fox, S. C. Acute and Long-Term Impact of Chemical Weapons: Lessons from the Iran-Iraq War. *Forensic Science Review* **26**, 97–114 (2014).
62. John, H. *et al.* Fatal Sarin Poisoning in Syria 2013: Forensic Verification within an International Laboratory Network. *Forensic Toxicology* **36**, 61–71. doi:10.1007/s11419-017-0376-7 (2018).
63. Morita, H. *et al.* Sarin Poisoning in Matsumoto, Japan. *Lancet (London, England)* **346**, 290–293. doi:10.1016/s0140-6736(95)92170-2 (1995).
64. Nakagawa, T. & Tu, A. T. Murders with VX: Aum Shinrikyo in Japan and the Assassination of Kim Jong-Nam in Malaysia. *Forensic Toxicology* **36**, 542–544. doi:10.1007/s11419-018-0426-9 (2018).
65. Vale, J. A., OBE, T. C. M. & CBE, R. L. M. Novichok: A Murderous Nerve Agent Attack in the UK. *Clinical Toxicology* **56**, 1093–1097. doi:10.1080/15563650.2018.1469759 (2018).

9 References

66. Steindl, D. *et al.* Novichok Nerve Agent Poisoning. *The Lancet* **0**. doi:10.1016/S0140-6736(20)32644-1 (2020).
67. Eddleston, M. & Chowdhury, F. R. Organophosphorus Poisoning: The Wet Opioid Toxidrome. *The Lancet* **397**, 175–177. doi:10.1016/S0140-6736(20)32749-5 (2021).
68. Gunnell, D. & Eddleston, M. Suicide by Intentional Ingestion of Pesticides: A Continuing Tragedy in Developing Countries. *International Journal of Epidemiology* **32**, 902–909 (2003).
69. Gunnell, D., Eddleston, M., Phillips, M. R. & Konradsen, F. The Global Distribution of Fatal Pesticide Self-Poisoning: Systematic Review. *BMC Public Health* **7**. doi:10.1186/1471-2458-7-357 (2007).
70. Utyasheva, L. & Eddleston, M. Prevention of Pesticide Suicides and the Right to Life - the Intersection of Human Rights and Public Health Priorities. *The Journal of Human Rights*. doi:10.1080/14754835.2020.1850241 (2021).
71. Blumenberg, A. *et al.* Utility of 2-Pyridine Aldoxime Methyl Chloride (2-PAM) for Acute Organophosphate Poisoning: A Systematic Review and Meta-Analysis. *Journal of Medical Toxicology* **14**, 91–98. doi:10.1007/s13181-017-0636-2 (2018).
72. Kharel, H., Pokhrel, N. B., Ghimire, R. & Kharel, Z. The Efficacy of Pralidoxime in the Treatment of Organophosphate Poisoning in Humans: A Systematic Review and Meta-Analysis of Randomized Trials. *Cureus* **12**. doi:10.7759/cureus.7174 (2020).
73. Worek, F., Thiermann, H. & Wille, T. Organophosphorus Compounds and Oximes: A Critical Review. *Archives of Toxicology*. doi:10.1007/s00204-020-02797-0 (2020).
74. Lee, Y. Y. *et al.* The Cost-Effectiveness of Banning Highly Hazardous Pesticides to Prevent Suicides Due to Pesticide Self-Intoxication across 14 Countries: An Economic Modelling Study. *The Lancet Global Health* **9**, e291–e300. doi:10.1016/S2214-109X(20)30493-9 (2021).

75. Nachon, F., Brazzolotto, X., Trovaslet, M. & Masson, P. Progress in the Development of Enzyme-Based Nerve Agent Bioscavengers. *Chemico-Biological Interactions* **206**, 536–544. doi:10.1016/j.cbi.2013.06.012 (2013).
76. Worek, F., Wille, T., Koller, M. & Thiermann, H. Toxicology of Organophosphorus Compounds in View of an Increasing Terrorist Threat. *Archives of Toxicology* **90**, 2131–2145. doi:10.1007/s00204-016-1772-1 (2016).
77. Goldsmith, M. & Ashani, Y. Catalytic Bioscavengers as Countermeasures against Organophosphate Nerve Agents. *Chemico-Biological Interactions* **292**, 50–64. doi:10.1016/j.cbi.2018.07.006 (2018).
78. Bigley, A. N. & Raushel, F. M. The Evolution of Phosphotriesterase for Decontamination and Detoxification of Organophosphorus Chemical Warfare Agents. *Chemico-Biological Interactions* **308**, 80–88. doi:10.1016/j.cbi.2019.05.023 (2019).
79. Worek, F. *et al.* Efficacy of the rePON1 Mutant IIG1 to Prevent Cyclosarin Toxicity in Vivo and to Detoxify Structurally Different Nerve Agents in Vitro. *Archives of Toxicology* **88**, 1257–1266. doi:10.1007/s00204-014-1204-z (2014).
80. Jackson, C. J. *et al.* Use of OpdA, an Organophosphorus (OP) Hydrolase, Prevents Lethality in an African Green Monkey Model of Acute OP Poisoning. *Toxicology* **317**, 1–5. doi:10.1016/j.tox.2014.01.003 (2014).
81. Gupta, R. D. *et al.* Directed Evolution of Hydrolases for Prevention of G-Type Nerve Agent Intoxication. *Nature Chemical Biology* **7**, 120–125. doi:10.1038/nchembio.510 (2011).
82. Novikov, B. N., Grimsley, J. K., Kern, R. J., Wild, J. R. & Wales, M. E. Improved Pharmacokinetics and Immunogenicity Profile of Organophosphorus Hydrolase by Chemical Modification with Polyethylene Glycol. *Journal of Controlled Release* **146**, 318–325. doi:10.1016/j.jconrel.2010.06.003 (2010).
83. Goldsmith, M. *et al.* Evolved Stereoselective Hydrolases for Broad-Spectrum G-Type Nerve Agent Detoxification. *Chemistry & Biology* **19**, 456–466. doi:10.1016/j.chembiol.2012.01.017 (2012).

9 References

84. Cherny, I. *et al.* Engineering V-Type Nerve Agents Detoxifying Enzymes Using Computationally Focused Libraries. *ACS Chemical Biology* **8**, 2394–2403. doi:10.1021/cb4004892 (2013).
85. Goldsmith, M. *et al.* Catalytic Efficiencies of Directly Evolved Phosphotriesterase Variants with Structurally Different Organophosphorus Compounds in Vitro. *Archives of Toxicology* **90**, 2711–2724. doi:10.1007/s00204-015-1626-2 (2016).
86. Goldsmith, M. *et al.* Overcoming an Optimization Plateau in the Directed Evolution of Highly Efficient Nerve Agent Bioscavengers. *Protein Engineering, Design and Selection* **30**, 333–345. doi:10.1093/protein/gzx003 (2017).
87. Lupas, A. N. What I Cannot Create, I Do Not Understand. *Science* **346**, 1455–1456. doi:10.1126/science.aaa2721 (2014).
88. Leman, J. K. *et al.* Macromolecular Modeling and Design in Rosetta: Recent Methods and Frameworks. *Nature Methods*, 1–14. doi:10.1038/s41592-020-0848-2 (2020).
89. Cao, L. *et al.* De Novo Design of Picomolar SARS-CoV-2 Miniprotein Inhibitors. *Science* **370**, 426–431. doi:10.1126/science.abd9909 (2020).
90. Bale, J. B. *et al.* Accurate Design of Megadalton-Scale Two-Component Icosahedral Protein Complexes. *Science* **353**, 389–394. doi:10.1126/science.aaf8818 (2016).
91. Xu, C. *et al.* Computational Design of Transmembrane Pores. *Nature* **585**, 129–134. doi:10.1038/s41586-020-2646-5 (2020).
92. Vorobieva, A. A. *et al.* De Novo Design of Transmembrane β Barrels. *Science* **371**. doi:10.1126/science.abc8182 (2021).
93. Quijano-Rubio, A. *et al.* De Novo Design of Modular and Tunable Protein Biosensors. *Nature* **591**, 482–487. doi:10.1038/s41586-021-03258-z (2021).
94. Jiang, L. *et al.* De Novo Computational Design of Retro-Aldol Enzymes. *Science* **319**, 1387–1391. doi:10.1126/science.1152692 (2008).
95. Röthlisberger, D. *et al.* Kemp Elimination Catalysts by Computational Enzyme Design. *Nature* **453**, 190–195. doi:10.1038/nature06879 (2008).

96. Siegel, J. B. *et al.* Computational Design of an Enzyme Catalyst for a Stereoselective Bimolecular Diels-Alder Reaction. *Science* **329**, 309–313. doi:10.1126/science.1190239 (2010).
97. Radzicka, A. & Wolfenden, R. A Proficient Enzyme. *Science* **267**, 90–93. doi:10.1126/science.7809611 (1995).
98. Chen, K. & Arnold, F. H. Enzyme Engineering for Nonaqueous Solvents: Random Mutagenesis to Enhance Activity of Subtilisin E in Polar Organic Media. *Bio/Technology* **9**, 1073–1077. doi:10.1038/nbt1191-1073 (1991).
99. Arnold, F. H. Innovation by Evolution: Bringing New Chemistry to Life (Nobel Lecture). *Angewandte Chemie International Edition* **58**, 14420–14426. doi:10.1002/anie.201907729 (2019).
100. Maynard Smith, J. Natural Selection and the Concept of a Protein Space. *Nature* **225**, 563–564. doi:10.1038/225563a0 (1970).
101. Emond, S. *et al.* Accessing Unexplored Regions of Sequence Space in Directed Enzyme Evolution via Insertion/Deletion Mutagenesis. *Nature Communications* **11**, 3469. doi:10.1038/s41467-020-17061-3 (2020).
102. Weinreich, D. M., Delaney, N. F., DePristo, M. A. & Hartl, D. L. Darwinian Evolution Can Follow Only Very Few Mutational Paths to Fitter Proteins. *Science* **312**, 111–114. doi:10.1126/science.1123539 (2006).
103. Tracewell, C. A. & Arnold, F. H. Directed Enzyme Evolution: Climbing Fitness Peaks One Amino Acid at a Time. *Current Opinion in Chemical Biology* **13**, 3–9. doi:10.1016/j.cbpa.2009.01.017 (2009).
104. Sheludko, Y. V. & Fessner, W.-D. Winning the Numbers Game in Enzyme Evolution – Fast Screening Methods for Improved Biotechnology Proteins. *Current Opinion in Structural Biology* **63**, 123–133. doi:10.1016/j.sbi.2020.05.003 (2020).
105. Smith, G. P. Filamentous Fusion Phage: Novel Expression Vectors That Display Cloned Antigens on the Virion Surface. *Science* **228**, 1315–1317. doi:10.1126/science.4001944 (1985).

9 References

106. Boder, E. T. & Wittrup, K. D. Yeast Surface Display for Screening Combinatorial Polypeptide Libraries. *Nature Biotechnology* **15**, 553–557. doi:10.1038/nbt0697-553 (1997).
107. Hanes, J. & Plückthun, A. In Vitro Selection and Evolution of Functional Proteins by Using Ribosome Display. *Proceedings of the National Academy of Sciences* **94**, 4937–4942. doi:10.1073/pnas.94.10.4937 (1997).
108. Roberts, R. W. & Szostak, J. W. RNA-Peptide Fusions for the in Vitro Selection of Peptides and Proteins. *Proceedings of the National Academy of Sciences* **94**, 12297–12302. doi:10.1073/pnas.94.23.12297 (1997).
109. Odegrip, R. *et al.* CIS Display: In Vitro Selection of Peptides from Libraries of Protein–DNA Complexes. *Proceedings of the National Academy of Sciences* **101**, 2806–2810. doi:10.1073/pnas.0400219101 (2004).
110. McCafferty, J., Griffiths, A. D., Winter, G. & Chiswell, D. J. Phage Antibodies: Filamentous Phage Displaying Antibody Variable Domains. *Nature* **348**, 552–554. doi:10.1038/348552a0 (1990).
111. Winter, G., Griffiths, A. D., Hawkins, R. E. & Hoogenboom, H. R. Making Antibodies by Phage Display Technology. *Annual Review of Immunology* **12**, 433–455. doi:10.1146/annurev.iy.12.040194.002245 (1994).
112. Bradbury, A. R. M., Sidhu, S., Dübel, S. & McCafferty, J. Beyond Natural Antibodies: The Power of in Vitro Display Technologies. *Nature Biotechnology* **29**, 245–254. doi:10.1038/nbt.1791 (2011).
113. Smith, G. P. Phage Display: Simple Evolution in a Petri Dish (Nobel Lecture). *Angewandte Chemie International Edition* **58**, 14428–14437. doi:10.1002/anie.201908308 (2019).
114. Griffiths, A. D. & Duncan, A. R. Strategies for Selection of Antibodies by Phage Display. *Current Opinion in Biotechnology* **9**, 102–108. doi:10.1016/S0958-1669(98)80092-X (1998).
115. Ghadessy, F. J. *et al.* Generic Expansion of the Substrate Spectrum of a DNA Polymerase by Directed Evolution. *Nature Biotechnology* **22**, 755–759. doi:10.1038/nbt974 (2004).

116. Cohen, H. M., Tawfik, D. S. & Griffiths, A. D. Altering the Sequence Specificity of HaeIII Methyltransferase by Directed Evolution Using in Vitro Compartmentalization. *Protein Engineering, Design and Selection* **17**, 3–11. doi:10.1093/protein/gzh001 (2004).
117. D’Abbadie, M. *et al.* Molecular Breeding of Polymerases for Amplification of Ancient DNA. *Nature Biotechnology* **25**, 939–943. doi:10.1038/nbt1321 (2007).
118. Loakes, D., Gallego, J., Pinheiro, V. B., Kool, E. T. & Holliger, P. Evolving a Polymerase for Hydrophobic Base Analogues. *Journal of the American Chemical Society* **131**, 14827–14837. doi:10.1021/ja9039696 (2009).
119. Ghadessy, F. J., Ong, J. L. & Holliger, P. Directed Evolution of Polymerase Function by Compartmentalized Self-Replication. *Proceedings of the National Academy of Sciences* **98**, 4552–4557. doi:10.1073/pnas.071052198 (2001).
120. Pinheiro, V. B. *et al.* Synthetic Genetic Polymers Capable of Heredity and Evolution. *Science* **336**, 341–344. doi:10.1126/science.1217622 (2012).
121. Griffiths, A. D. & Tawfik, D. S. Directed Evolution of an Extremely Fast Phosphotriesterase by in Vitro Compartmentalization. *The EMBO Journal* **22**, 24–35. doi:10.1093/emboj/cdg014 (2003).
122. Lindenburg, L. & Hollfelder, F. “NAD-Display”: Ultrahigh-Throughput in Vitro Screening of NAD(H) Dehydrogenases Using Bead Display and Flow Cytometry. *Angewandte Chemie International Edition* **60**, 9015–9021. doi:10.1002/anie.202013486 (2021).
123. Aharoni, A., Amitai, G., Bernath, K., Magdassi, S. & Tawfik, D. S. High-Throughput Screening of Enzyme Libraries: Thiolactonases Evolved by Fluorescence-Activated Sorting of Single Cells in Emulsion Compartments. *Chemistry & Biology* **12**, 1281–1289. doi:10.1016/j.chembiol.2005.09.012 (2005).
124. Mastrobattista, E. *et al.* High-Throughput Screening of Enzyme Libraries: In Vitro Evolution of a β -Galactosidase by Fluorescence-Activated Sorting of Double Emulsions. *Chemistry & Biology* **12**, 1291–1300. doi:10.1016/j.chembiol.2005.09.016 (2005).

9 References

125. Körfer, G., Pitzler, C., Vojcic, L., Martinez, R. & Schwaneberg, U. In Vitro Flow Cytometry-Based Screening Platform for Cellulase Engineering. *Scientific Reports* **6**. doi:10.1038/srep26128 (2016).
126. Miller, O. J. *et al.* Directed Evolution by in Vitro Compartmentalization. *Nature Methods* **3**, 561–570. doi:10.1038/nmeth897 (2006).
127. Whitesides, G. M. The Origins and the Future of Microfluidics. *Nature* **442**, 368–373. doi:10.1038/nature05058 (2006).
128. Baret, J.-C. *et al.* Fluorescence-Activated Droplet Sorting (FADS): Efficient Microfluidic Cell Sorting Based on Enzymatic Activity. *Lab on a Chip* **9**, 1850. doi:10.1039/b902504a (2009).
129. Agresti, J. J. *et al.* Ultrahigh-Throughput Screening in Drop-Based Microfluidics for Directed Evolution. *Proceedings of the National Academy of Sciences* **107**, 4004–4009. doi:10.1073/pnas.0910781107 (2010).
130. Kintses, B. *et al.* Picoliter Cell Lysate Assays in Microfluidic Droplet Compartments for Directed Enzyme Evolution. *Chemistry & Biology* **19**, 1001–1009. doi:10.1016/j.chembiol.2012.06.009 (2012).
131. Gielen, F. *et al.* Ultrahigh-Throughput–Directed Enzyme Evolution by Absorbance-Activated Droplet Sorting (AADS). *Proceedings of the National Academy of Sciences* **113**, E7383–E7389. doi:10.1073/pnas.1606927113 (2016).
132. Obexer, R. *et al.* Emergence of a Catalytic Tetrad during Evolution of a Highly Active Artificial Aldolase. *Nature Chemistry*. doi:10.1038/nchem.2596 (2016).
133. Obexer, R., Pott, M., Zeymer, C., Griffiths, A. D. & Hilvert, D. Efficient Laboratory Evolution of Computationally Designed Enzymes with Low Starting Activities Using Fluorescence-Activated Droplet Sorting. *Protein engineering, design & selection: PEDS* **30**, 531. doi:10.1093/protein/gzx032 (2017).
134. Debon, A. *et al.* Ultrahigh-Throughput Screening Enables Efficient Single-Round Oxidase Remodelling. *Nature Catalysis* **2**, 740–747. doi:10.1038/s41929-019-0340-5 (2019).

135. Abate, A. R., Hung, T., Mary, P., Agresti, J. J. & Weitz, D. A. High-Throughput Injection with Microfluidics Using Picoinjectors. *Proceedings of the National Academy of Sciences* **107**, 19163–19166 (2010).
136. Mulas, C. *et al.* Microfluidic Platform for 3D Cell Culture with Live Imaging and Clone Retrieval. *Lab on a Chip*. doi:10.1039/D0LC00165A (2020).
137. Gielen, F. *et al.* Ultrahigh-Throughput–Directed Enzyme Evolution by Absorbance-Activated Droplet Sorting (AADS). *Proceedings of the National Academy of Sciences* **113**, E7383–E7389. doi:10.1073/pnas.1606927113 (2016).
138. Anagnostidis, V. *et al.* Deep Learning Guided Image-Based Droplet Sorting for on-Demand Selection and Analysis of Single Cells and 3D Cell Cultures. *Lab on a Chip*. doi:10.1039/D0LC00055H (2020).
139. Neun, S., Zurek, P. J., Kaminski, T. S. & Hollfelder, F. in *Methods in Enzymology* (ed Tawfik, D. S.) 317–343 (Academic Press, 2020). doi:10.1016/bs.mie.2020.06.002.
140. Zinchenko, A. *et al.* One in a Million: Flow Cytometric Sorting of Single Cell-Lysate Assays in Monodisperse Picolitre Double Emulsion Droplets for Directed Evolution. *Analytical Chemistry* **86**, 2526–2533. doi:10.1021/ac403585p (2014).
141. Fischlechner, M. *et al.* Evolution of Enzyme Catalysts Caged in Biomimetic Gel-Shell Beads. *Nature Chemistry* **6**, 791–796. doi:10.1038/nchem.1996 (2014).
142. Larsen, A. C. *et al.* A General Strategy for Expanding Polymerase Function by Droplet Microfluidics. *Nature Communications* **7**, 11235. doi:10.1038/ncomms11235 (2016).
143. Napiorkowska, M., Pestalozzi, L., Panke, S., Held, M. & Schmitt, S. High-Throughput Optimization of Recombinant Protein Production in Microfluidic Gel Beads. *Small* **17**, 2005523. doi:10.1002/smll.202005523 (2021).
144. Eduati, F. *et al.* A Microfluidics Platform for Combinatorial Drug Screening on Cancer Biopsies. *Nature Communications* **9**, 2434. doi:10.1038/s41467-018-04919-w (2018).

9 References

145. Klein, A. M. *et al.* Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. *Cell* **161**, 1187–1201. doi:10.1016/j.cell.2015.04.044 (2015).
146. Mair, P., Gielen, F. & Hollfelder, F. Exploring Sequence Space in Search of Functional Enzymes Using Microfluidic Droplets. *Current Opinion in Chemical Biology* **37**, 137–144. doi:10.1016/j.cbpa.2017.02.018 (2017).
147. Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J. & Goodman, R. M. Molecular Biological Access to the Chemistry of Unknown Soil Microbes: A New Frontier for Natural Products. *Chemistry & Biology* **5**, R245–R249. doi:10.1016/S1074-5521(98)90108-9 (1998).
148. Lorenz, P. & Schleper, C. Metagenome—a Challenging Source of Enzyme Discovery. *Journal of Molecular Catalysis B: Enzymatic* **19-20**, 13–19. doi:10.1016/S1381-1177(02)00147-9 (2002).
149. Colin, P.-Y. *et al.* Ultrahigh-Throughput Discovery of Promiscuous Enzymes by Picodroplet Functional Metagenomics. *Nature Communications* **6**, 10008. doi:10.1038/ncomms10008 (2015).
150. Ollis, D. L. *et al.* The Alpha/Beta Hydrolase Fold. *Protein Engineering* **5**, 197–211. doi:10.1093/protein/5.3.197 (1992).
151. Lenfant, N. *et al.* ESTHER, the Database of the α/β -Hydrolase Fold Superfamily of Proteins: Tools to Explore Diversity of Functions. *Nucleic Acids Research* **41**, D423–D429. doi:10.1093/nar/gks1154 (2013).
152. Carr, P. D. & Ollis, D. L. Alpha/Beta Hydrolase Fold: An Update. *Protein and Peptide Letters* **16**, 1137–1148 (2009).
153. Sussman, J. *et al.* Atomic Structure of Acetylcholinesterase from Torpedo Californica: A Prototypic Acetylcholine-Binding Protein. *Science* **253**, 872–879. doi:10.1126/science.1678899 (1991).
154. Verschueren, K. H. G., Seljée, F., Rozeboom, H. J., Kalk, K. H. & Dijkstra, B. W. Crystallographic Analysis of the Catalytic Mechanism of Haloalkane Dehalogenase. *Nature* **363**, 693–698. doi:10.1038/363693a0 (1993).

155. Rink, R., Fennema, M., Smids, M., Dehmel, U. & Janssen, D. B. Primary Structure and Catalytic Mechanism of the Epoxide Hydrolase from *Agrobacterium radiobacter* AD1*. *Journal of Biological Chemistry* **272**, 14650–14657. doi:10.1074/jbc.272.23.14650 (1997).
156. Loening, A. M., Fenn, T. D. & Gambhir, S. S. Crystal Structures of the Luciferase and Green Fluorescent Protein from *Renilla Reniformis*. *Journal of Molecular Biology* **374**, 1017–1028. doi:10.1016/j.jmb.2007.09.078 (2007).
157. Fabrichny, I. P. *et al.* Structural Analysis of the Synaptic Protein Neuroligin and Its β -Neurexin Complex: Determinants for Folding and Cell Adhesion. *Neuron* **56**, 979–991. doi:10.1016/j.neuron.2007.11.013 (2007).
158. Longhi, S., Czjzek, M., Lamzin, V., Nicolas, A. & Cambillau, C. Atomic Resolution (1.0 Å) Crystal Structure of *Fusarium Solani* Cutinase: Stereochemical analysis. Edited by R. Huber. *Journal of Molecular Biology* **268**, 779–799. doi:10.1006/jmbi.1997.1000 (1997).
159. Ghosh, D. *et al.* Determination of a Protein Structure by Iodination: The Structure of Iodinated Acetylxyln Esterase. *Acta Crystallographica Section D: Biological Crystallography* **55**, 779–784. doi:10.1107/S0907444999000244 (1999).
160. Van Pouderoyen, G., Eggert, T., Jaeger, K.-E. & Dijkstra, B. W. The Crystal Structure of *Bacillus Subtilis* Lipase: A Minimal α/β Hydrolase Fold Enzyme. *Journal of Molecular Biology* **309**, 215–226. doi:10.1006/jmbi.2001.4659 (2001).
161. Brzozowski, A. M. *et al.* A Model for Interfacial Activation in Lipases from the Structure of a Fungal Lipase-Inhibitor Complex. *Nature* **351**, 491–494. doi:10.1038/351491a0 (1991).
162. Fülöp, V., Böcskei, Z. & Polgár, L. Prolyl Oligopeptidase: An Unusual β -Propeller Domain Regulates Proteolysis. *Cell* **94**, 161–170. doi:10.1016/S0092-8674(00)81416-6 (1998).
163. Porter, J. L. *et al.* Directed Evolution of New and Improved Enzyme Functions Using an Evolutionary Intermediate and Multidirectional Search. *ACS Chemical Biology* **10**, 611–621. doi:10.1021/cb500809f (2015).

9 References

164. Pathak, D. & Ollis, D. Refined Structure of Dienelactone Hydrolase at 1.8 Å. *Journal of Molecular Biology* **214**, 497–525. doi:10.1016/0022-2836(90)90196-S (1990).
165. Wagner, U., Hasslacher, M, Griengl, H, Schwab, H & Kratky, C. Mechanism of Cyanogenesis: The Crystal Structure of Hydroxynitrile Lyase from *Hevea Brasiliensis*. *Structure* **4**, 811–822. doi:10.1016/S0969-2126(96)00088-3 (1996).
166. Mirza, I. A., Nazi, I., Korczynska, M., Wright, G. D. & Berghuis, A. M. Crystal Structure of Homoserine Transacetylase from *Haemophilus Influenzae* Reveals a New Family of α/β -Hydrolases, *Biochemistry* **44**, 15768–15773. doi:10.1021/bi051951y (2005).
167. Newcomb, R. D. *et al.* A Single Amino Acid Substitution Converts a Carboxylesterase to an Organophosphorus Hydrolase and Confers Insecticide Resistance on a Blowfly. *Proceedings of the National Academy of Sciences* **94**, 7464–7468. doi:10.1073/pnas.94.14.7464 (1997).
168. Campbell, P. M., Newcomb, R. D., Russell, R. J. & Oakeshott, J. G. Two Different Amino Acid Substitutions in the Ali-Esterase, E3, Confer Alternative Types of Organophosphorus Insecticide Resistance in the Sheep Blowfly, *Lucilia Cuprina*. *Insect Biochemistry and Molecular Biology* **28**, 139–150. doi:10.1016/S0965-1748(97)00109-4 (1998).
169. Lockridge, O. *et al.* A Single Amino Acid Substitution, Gly117His, Confers Phosphotriesterase (Organophosphorus Acid Anhydride Hydrolase) Activity on Human Butyrylcholinesterase[†]. *Biochemistry* **36**, 786–795. doi:10.1021/bi961412g (1997).
170. Zueva, I. V., Lushchekina, S. V., Daudé, D., Chabrière, E. & Masson, P. Steady-State Kinetics of Enzyme-Catalyzed Hydrolysis of Echothiophate, a P–S Bonded Organophosphorus as Monitored by Spectrofluorimetry. *Molecules* **25**, 1371. doi:10.3390/molecules25061371 (2020).
171. Legler, P. M., Boisvert, S. M., Compton, J. R. & Millard, C. B. Development of Organophosphate Hydrolase Activity in a Bacterial Homolog of Human

- Cholinesterase. *Frontiers in Chemistry* **2**. doi:10.3389/fchem.2014.00046 (2014).
172. Poyot, T. *et al.* Mutant of Bungarus Fasciatus Acetylcholinesterase with Low Affinity and Low Hydrolase Activity toward Organophosphorus Esters. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1764**, 1470–1478. doi:10.1016/j.bbapap.2006.07.008 (2006).
173. Jencks, W. P. in *Advances in Enzymology and Related Areas of Molecular Biology* 219–410 (John Wiley & Sons, Ltd, 1975). doi:10.1002/9780470122884.ch4.
174. Fersht, A. *Enzyme Structure and Mechanism* 2nd ed. 475 pp. (W.H. Freeman, New York, 1985).
175. Hollfelder, F., Kirby, A. J. & Tawfik, D. S. Off-the-Shelf Proteins That Rival Tailor-Made Antibodies as Catalysts. *Nature* **383**, 60–63. doi:10.1038/383060a0 (1996).
176. Furnham, N., Dawson, N. L., Rahman, S. A., Thornton, J. M. & Orengo, C. A. Large-Scale Analysis Exploring Evolution of Catalytic Machineries and Mechanisms in Enzyme Superfamilies. *Journal of Molecular Biology* **428**, 253–267. doi:10.1016/j.jmb.2015.11.010 (2016).
177. Goldman, A. D., Beatty, J. T. & Landweber, L. F. The TIM Barrel Architecture Facilitated the Early Evolution of Protein-Mediated Metabolism. *Journal of Molecular Evolution* **82**, 17–26. doi:10.1007/s00239-015-9722-8 (2016).
178. Buckingham, D. A. & Engelhardt, L. M. Metal Hydroxide Promoted Hydrolysis of Carbonyl Substrates. *Journal of the American Chemical Society* **97**, 5915–5917. doi:10.1021/ja00853a047 (1975).
179. Ben-David, M. *et al.* Catalytic Metal Ion Rearrangements Underline Promiscuity and Evolvability of a Metalloenzyme. *Journal of Molecular Biology* **425**, 1028–1038. doi:10.1016/j.jmb.2013.01.009 (2013).
180. Kuhlen, L. *Convergent Directed Evolution of a Novel Phosphotriesterase* Part III Dissertation (University of Cambridge, 2015).

9 References

181. Schnettler Fernández, J. D. F. *Directed Evolution of a Promiscuous Phosphotriesterase in Microdroplets* MPhil Dissertation (University of Cambridge, 2016).
182. Van der Meer, J.-Y. *et al.* Using Mutability Landscapes of a Promiscuous Tautomerase to Guide the Engineering of Enantioselective Michaelases. *Nature Communications* **7**, 10911. doi:10.1038/ncomms10911 (2016).
183. Bayer, C. D., van Loo, B. & Hollfelder, F. Specificity Effects of Amino Acid Substitutions in Promiscuous Hydrolases: Context-Dependence of Catalytic Residue Contributions to Local Fitness Landscapes in Nearby Sequence Space. *Chembiochem: A European Journal of Chemical Biology* **18**, 1001–1015. doi:10.1002/cbic.201600657 (2017).
184. Boersma, Y. L. *et al.* Loop Grafting of Bacillus Subtilis Lipase A: Inversion of Enantioselectivity. *Chemistry & Biology* **15**, 782–789. doi:10.1016/j.chembiol.2008.06.009 (2008).
185. Martínez-Martínez, M. *et al.* Determinants and Prediction of Esterase Substrate Promiscuity Patterns. *ACS Chemical Biology* **13**, 225–234. doi:10.1021/acscchembio.7b00996 (2018).
186. Porter, J. L., Collyer, C. A. & Ollis, D. L. Compensatory Stabilizing Role of Surface Mutations During the Directed Evolution of Dienelactone Hydrolase for Enhanced Activity. *The Protein Journal* **34**, 82–89. doi:10.1007/s10930-015-9600-7 (2015).
187. Neun, S., Kaminski, T. S. & Hollfelder, F. in *Methods in Enzymology* (eds Allbritton, N. L. & Kovarik, M. L.) 95–112 (Academic Press, 2019). doi:10.1016/bs.mie.2019.07.009.
188. Kille, S. *et al.* Reducing Codon Redundancy and Screening Effort of Combinatorial Protein Libraries Created by Saturation Mutagenesis. *ACS Synthetic Biology* **2**, 83–92. doi:10.1021/sb300037w (2013).
189. Frenz, L., Blank, K., Brouzes, E. & Griffiths, A. D. Reliable Microfluidic On-Chip Incubation of Droplets in Delay-Lines. *Lab on a Chip* **9**, 1344–1348. doi:10.1039/B816049J (2009).

190. Bar-Even, A. *et al.* The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. *Biochemistry* **50**, 4402–4410. doi:10.1021/bi2002289 (2011).
191. Xiang, D. F. *et al.* Interrogation of the Substrate Profile and Catalytic Properties of the Phosphotriesterase from *Sphingobium* Sp. Strain TCM1: An Enzyme Capable of Hydrolyzing Organophosphate Flame Retardants and Plasticizers. *Biochemistry* **54**, 7539–7549. doi:10.1021/acs.biochem.5b01144 (2015).
192. Meier, M. M. *et al.* Molecular Engineering of Organophosphate Hydrolysis Activity from a Weak Promiscuous Lactonase Template. *Journal of the American Chemical Society* **135**, 11670–11677. doi:10.1021/ja405911h (2013).
193. Aharoni, A. *et al.* Directed Evolution of Mammalian Paraoxonases PON1 and PON3 for Bacterial Expression and Catalytic Specialization. *Proceedings of the National Academy of Sciences* **101**, 482–487. doi:10.1073/pnas.2536901100 (2004).
194. Miton, C. M. *et al.* Evolutionary Repurposing of a Sulfatase: A New Michaelis Complex Leads to Efficient Transition State Charge Offset. *Proceedings of the National Academy of Sciences*, 201607817. doi:10.1073/pnas.1607817115 (2018).
195. Holstein, J. M., Gylstorff, C. & Hollfelder, F. Cell-Free Directed Evolution of a Protease in Microdroplets at Ultrahigh Throughput. *ACS Synthetic Biology*. doi:10.1021/acssynbio.0c00538 (2021).
196. Khersonsky, O. & Tawfik, D. S. Structure-Reactivity Studies of Serum Paraoxonase PON1 Suggest That Its Native Activity Is Lactonase. *Biochemistry* **44**, 6371–6382. doi:10.1021/bi047440d (2005).
197. Jumper, J. *et al.* Highly Accurate Protein Structure Prediction with AlphaFold. *Nature*, 1–11. doi:10.1038/s41586-021-03819-2 (2021).
198. Mirdita, M., Ovchinnikov, S. & Steinegger, M. ColabFold - Making Protein Folding Accessible to All. *bioRxiv*, 2021.08.15.456425. doi:10.1101/2021.08.15.456425 (2021).

9 References

199. You, L. & Arnold, F. H. Directed Evolution of Subtilisin E in *Bacillus Subtilis* to Enhance Total Activity in Aqueous Dimethylformamide. *Protein Engineering* **9**, 77–83. doi:10.1093/protein/9.1.77 (1996).
200. Khan, S. A. & Kirby, A. J. The Reactivity of Phosphate Esters. Multiple Structure–Reactivity Correlations for the Reactions of Triesters with Nucleophiles. *J. Chem. Soc. B* **0**, 1172–1182. doi:10.1039/J29700001172 (1970).
201. Shafee, T., Gatti-Lafranconi, P., Minter, R. & Hollfelder, F. Handicap–Recover Evolution Leads to a Chemically Versatile, Nucleophile-Permissive Protease. *ChemBioChem* **16**, 1866–1869. doi:10.1002/cbic.201500295 (2015).
202. Cen, Y. *et al.* Artificial Cysteine-Lipases with High Activity and Altered Catalytic Mechanism Created by Laboratory Evolution. *Nature Communications* **10**, 3198. doi:10.1038/s41467-019-11155-3 (2019).
203. Schmidt, E & Knackmuss, H. J. Chemical Structure and Biodegradability of Halogenated Aromatic Compounds. Conversion of Chlorinated Muconic Acids into Maleoylactic Acid. *Biochemical Journal* **192**, 339–347. doi:10.1042/bj1920339 (1980).
204. Ngai, K. L., Schlömann, M., Knackmuss, H. J. & Ornston, L. N. Dienelactone Hydrolase from *Pseudomonas* Sp. Strain B13. *Journal of Bacteriology* **169**, 699–703. doi:10.1128/jb.169.2.699-703.1987 (1987).
205. Harwood, C. S. & Parales, R. E. The Beta-Ketoadipate Pathway and the Biology of Self-Identity. *Annual Review of Microbiology* **50**, 553–590. doi:10.1146/annurev.micro.50.1.553 (1996).
206. Harrington, E. D. *et al.* Quantitative Assessment of Protein Function Prediction from Metagenomics Shotgun Sequences. *Proceedings of the National Academy of Sciences* **104**, 13913–13918. doi:10.1073/pnas.0702636104 (2007).
207. Schnoes, A. M., Brown, S. D., Dodevski, I. & Babbitt, P. C. Annotation Error in Public Databases: Misannotation of Molecular Function in Enzyme Superfamilies. *PLOS Computational Biology* **5**, e1000605. doi:10.1371/journal.pcbi.1000605 (2009).

208. Qu, Y. & Spain, J. C. Molecular and Biochemical Characterization of the 5-Nitroanthranilic Acid Degradation Pathway in Bradyrhizobium Sp. Strain JS329. *Journal of Bacteriology* **193**, 3057–3063. doi:10.1128/JB.01188-10 (2011).
209. Al Fahad, A., Abood, A., Simpson, T. J. & Cox, R. J. The Biosynthesis and Catabolism of the Maleic Anhydride Moiety of Stipitatononic Acid. *Angewandte Chemie (International Ed. in English)* **53**, 7519–7523. doi:10.1002/anie.201403450 (2014).
210. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic Local Alignment Search Tool. *Journal of Molecular Biology* **215**, 403–410. doi:10.1016/S0022-2836(05)80360-2 (1990).
211. Mitchell, A. L. *et al.* MGnify: The Microbiome Analysis Resource in 2020. *Nucleic Acids Research* **48**, D570–D578. doi:10.1093/nar/gkz1035 (2020).
212. Klemetsen, T. *et al.* The MAR Databases: Development and Implementation of Databases Specific for Marine Metagenomics. *Nucleic Acids Research* **46**, D692–D699. doi:10.1093/nar/gkx1036 (2018).
213. Huo, Y.-Y., Wang, C.-S., Yang, J.-Y., Wu, M. & Xu, X.-W. Marinobacter Mobilis Sp. Nov. and Marinobacter Zhejiangensis Sp. Nov., Halophilic Bacteria Isolated from the East China Sea. *INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY* **58**, 2885–2889. doi:10.1099/ijls.0.2008/000786-0 (2008).
214. Walker, I., J. Easton, C. & L. Ollis, D. Site-Directed Mutagenesis of Dienelactone Hydrolase Produces Dienelactone Isomerase. *Chemical Communications* **0**, 671–672. doi:10.1039/B000365O (2000).
215. Walker, I., Hennessy, J. E., Ollis, D. L. & Easton, C. J. Substrate-Induced Conformational Change and Isomerase Activity of Dienelactone Hydrolase and Its Site-Specific Mutants. *ChemBioChem* **13**, 1645–1651. doi:10.1002/cbic.201200232 (2012).
216. Fatima, S. *et al.* Transforming Capacity of Two Novel Genes JS-1 and JS-2 Located in Chromosome 5p and Their Overexpression in Human Esophageal

9 References

- Squamous Cell Carcinoma. *International Journal of Molecular Medicine* **17**, 159–170 (2006).
217. Ishizuka, T. *et al.* Human Carboxymethylenebutenolidase as a Bioactivating Hydrolase of Olmesartan Medoxomil in Liver and Intestine. *Journal of Biological Chemistry* **285**, 11892–11902. doi:10.1074/jbc.M109.072629 (2010).
218. Ishizuka, T. *Bioconversion of the Prodrug Olmesartan Medoxomil via Carboxymethylenebutenolidase and Paraoxonase* PhD Dissertation (Chiba University, 2013).
219. Ishizuka, T. *et al.* Paraoxonase 1 as a Major Bioactivating Hydrolase for Olmesartan Medoxomil in Human Blood Circulation: Molecular Identification and Contribution to Plasma Metabolism. *Drug Metabolism and Disposition* **40**, 374–380. doi:10.1124/dmd.111.041475 (2012).
220. Buratti, F. M. & Testai, E. Malathion Detoxification by Human Hepatic Carboxylesterases and Its Inhibition by Isomalathion and Other Pesticides. *Journal of Biochemical and Molecular Toxicology* **19**, 406–414. doi:10.1002/jbt.20106 (2006).
221. Islam, S. M. A. *et al.* Organophosphorus Hydrolase (OpdB) of *Lactobacillus Brevis* WCP902 from Kimchi Is Able To Degrade Organophosphorus Pesticides. *Journal of Agricultural and Food Chemistry* **58**, 5380–5386. doi:10.1021/jf903878e (2010).
222. Jacob, R. B., Michaels, K. C., Anderson, C. J., Fay, J. M. & Dokholyan, N. V. Harnessing Nature's Diversity: Discovering Organophosphate Bioscavenger Characteristics among Low Molecular Weight Proteins. *Scientific Reports* **6**. doi:10.1038/srep37175 (2016).
223. Millard, C. B., Lockridge, O. & Broomfield, C. A. Organophosphorus Acid Anhydride Hydrolase Activity in Human Butyrylcholinesterase: Synergy Results in a Somanase. *Biochemistry* **37**, 237–247. doi:10.1021/bi972057c (1998).
224. Brazzolotto, X., Igert, A., Guillon, V., Santoni, G. & Nachon, F. Bacterial Expression of Human Butyrylcholinesterase as a Tool for Nerve Agent Bioscav-

- engers Development. *Molecules* **22**, 1828. doi:10.3390/molecules22111828 (2017).
225. Chen, V. P. *et al.* Plasma Butyrylcholinesterase Regulates Ghrelin to Control Aggression. *Proceedings of the National Academy of Sciences* **112**, 2251–2256. doi:10.1073/pnas.1421536112 (2015).
226. Goldenzweig, A. *et al.* Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability. *Molecular Cell* **63**, 337–346. doi:10.1016/j.molcel.2016.06.012 (2016).
227. Reddie, K. G. & Carroll, K. S. Expanding the Functional Diversity of Proteins through Cysteine Oxidation. *Current Opinion in Chemical Biology* **12**, 746–754. doi:10.1016/j.cbpa.2008.07.028 (2008).
228. Gabor, E. M., de Vries, E. J. & Janssen, D. B. Efficient Recovery of Environmental DNA for Expression Cloning by Indirect Extraction Methods. *FEMS Microbiology Ecology* **44**, 153–163. doi:10.1016/S0168-6496(02)00462-2 (2003).
229. Zhang, X.-Q. *et al.* *Sinimaribacterium Flocculans* Gen. Nov., Sp. Nov., a Gammaproteobacterium from Offshore Surface Seawater. *International Journal of Systematic and Evolutionary Microbiology*, **65**, 3541–3546. doi:10.1099/ijsem.0.000452 (2015).
230. Lee, Y., Lee, B., Lee, K. & Jeon, C. O. *Solimonas Fluminis* Sp. Nov., Isolated from a Freshwater River. *International Journal of Systematic and Evolutionary Microbiology* **68**, 2755–2759. doi:10.1099/ijsem.0.002865 (2018).
231. Sheu, S.-Y., Cho, N.-T., Arun, A. B. & Chen, W.-M. Proposal of *Solimonas Aquatica* Sp. Nov., Reclassification of *Sinobacter Flavus* Zhou *et al.* 2008 as *Solimonas Flava* Comb. Nov. and *Singularimonas Variicoloris* Friedrich and Lipski 2008 as *Solimonas Variicoloris* Comb. Nov. and Emended Descriptions of the Genus *Solimonas* and Its Type Species *Solimonas Soli*. *International Journal of Systematic and Evolutionary Microbiology* **61**, 2284–2291. doi:10.1099/ijse.0.023010-0 (2011).

9 References

232. Dorn, E., Hellwig, M., Reineke, W. & Knackmuss, H. J. Isolation and Characterization of a 3-Chlorobenzoate Degrading Pseudomonad. *Archives of Microbiology* **99**, 61–70. doi:10.1007/BF00696222 (1974).
233. Stolz, A., Busse, H.-J. & Kämpfer, P. Pseudomonas Knackmussii Sp. Nov. *International Journal of Systematic and Evolutionary Microbiology*, **57**, 572–576. doi:10.1099/ijs.0.64761-0 (2007).
234. Hoskin, F. C. & Long, R. J. Purification of a DFP-Hydrolyzing Enzyme from Squid Head Ganglion. *Archives of Biochemistry and Biophysics* **150**, 548–555. doi:10.1016/0003-9861(72)90073-2 (1972).
235. Anderson, K. S., Sikorski, J. A. & Johnson, K. A. Evaluation of 5-Enolpyruvoylshikimate-3-Phosphate Synthase Substrate and Inhibitor Binding by Stopped-Flow and Equilibrium Fluorescence Measurements. *Biochemistry* **27**, 1604–1610. doi:10.1021/bi00405a032 (1988).
236. Noellgen, R. M. & Landis, W. G. Identification and Characterization of the Organophosphate Acid Anhydrase Activity of the Blue Mussel, *Mytilus Edulis*. *Comparative Biochemistry and Physiology Part C: Comparative Pharmacology* **101**, 615–623. doi:10.1016/0742-8413(92)90095-O (1992).
237. Hogan, J. W. & Knowles, C. O. Degradation of Organophosphates by Fish Liver Phosphatases. *Journal of the Fisheries Board of Canada*. doi:10.1139/f68-142 (1968).
238. Wu, G.-R., Hong, H.-L. & Yan, C.-L. Arsenic Accumulation and Translocation in Mangrove (*Aegiceras Corniculatum* L.) Grown in Arsenic Contaminated Soils. *International Journal of Environmental Research and Public Health* **12**, 7244–7253. doi:10.3390/ijerph120707244 (2015).
239. Price, R. E., Breuer, C., Reeves, E., Bach, W. & Pichler, T. Arsenic Bioaccumulation and Biotransformation in Deep-Sea Hydrothermal Vent Organisms from the PACMANUS Hydrothermal Field, Manus Basin, PNG. *Deep Sea Research Part I: Oceanographic Research Papers* **117**, 95–106. doi:10.1016/j.dsr.2016.08.012 (2016).
240. Elias, M. *et al.* The Molecular Basis of Phosphate Discrimination in Arsenate-Rich Environments. *Nature* **491**, 134–137. doi:10.1038/nature11517 (2012).

241. Matsunaga, S., Moore, R. E., Niemczura, W. P. & Carmichael, W. W. Anatoxin-a(s), a Potent Anticholinesterase from *Anabaena Flos-Aquae*. *Journal of the American Chemical Society* **111**, 8021–8023. doi:10.1021/ja00202a057 (1989).
242. Hyde, E. G. & Carmichael, W. W. Anatoxin-A(S), a Naturally Occurring Organophosphate, Is an Irreversible Active Site-Directed Inhibitor of Acetylcholinesterase (EC 3.1.1.7). *Journal of Biochemical Toxicology* **6**, 195–201. doi:10.1002/jbt.2570060305 (1991).
243. Fiore, M. F. *et al.* Guanitoxin, Re-Naming a Cyanobacterial Organophosphate Toxin. *Harmful Algae* **92**, 101737. doi:10.1016/j.hal.2019.101737 (2020).
244. Zhao, S. *et al.* Prediction and Characterization of Enzymatic Activities Guided by Sequence Similarity and Genome Neighborhood Networks. *eLife* **3**, e03275. doi:10.7554/eLife.03275 (2014).
245. Beveridge, A. J. A Theoretical Study of the Active Sites of Papain and S195C Rat Trypsin: Implications for the Low Reactivity of Mutant Serine Proteinases. *Protein Science* **5**, 1355–1365. doi:10.1002/pro.5560050714 (1996).
246. Cheah, E., Austin, C., Ashley, G. W. & Ollis, D. Substrate-Induced Activation of Dienelactone Hydrolase: An Enzyme with a Naturally Occurring Cys-His-Asp Triad. *Protein Engineering, Design and Selection* **6**, 575–583. doi:10.1093/protein/6.6.575 (1993).
247. Beveridge, A. J. & Ollis, D. L. A Theoretical Study of Substrate-Induced Activation of Dienelactone Hydrolase. *Protein Engineering, Design and Selection* **8**, 135–142. doi:10.1093/protein/8.2.135 (1995).
248. Colin, P.-Y. *From Metagenomes to Directed Evolution - Microfluidic Droplets Identify Novel Promiscuous Enzymes in Environmental Gene Libraries* (University of Cambridge, 2015).
249. Waterhouse, A. *et al.* SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Research* **46**, W296–W303. doi:10.1093/nar/gky427 (2018).

9 References

250. Stourac, J. *et al.* Caver Web 1.0: Identification of Tunnels and Channels in Proteins and Analysis of Ligand Transport. *Nucleic Acids Research* **47**, W414–W422. doi:10.1093/nar/gkz378 (2019).
251. Tian, W., Chen, C., Lei, X., Zhao, J. & Liang, J. CASTp 3.0: Computed Atlas of Surface Topography of Proteins. *Nucleic Acids Research* **46**, W363–W367. doi:10.1093/nar/gky473 (2018).
252. Le Guilloux, V., Schmidtke, P. & Tuffery, P. Fpocket: An Open Source Platform for Ligand Pocket Detection. *BMC Bioinformatics* **10**, 168. doi:10.1186/1471-2105-10-168 (2009).
253. Campbell, E. C. *et al.* Laboratory Evolution of Protein Conformational Dynamics. *Current Opinion in Structural Biology* **50**, 49–57. doi:10.1016/j.sbi.2017.09.005 (2018).
254. Sun, Z., Liu, Q., Qu, G., Feng, Y. & Reetz, M. T. Utility of B-Factors in Protein Science: Interpreting Rigidity, Flexibility, and Internal Motion and Engineering Thermostability. *Chemical Reviews* **119**, 1626–1665. doi:10.1021/acs.chemrev.8b00290 (2019).
255. Guerrero, F., Ciragan, A. & Iwai, H. Tandem SUMO Fusion Vectors for Improving Soluble Protein Expression and Purification. *Protein Expression and Purification* **116**, 42–49. doi:10.1016/j.pep.2015.08.019 (2015).
256. Greener, J. G., Kandathil, S. M. & Jones, D. T. Deep Learning Extends de Novo Protein Modelling Coverage of Genomes Using Iteratively Predicted Structural Constraints. *Nature Communications* **10**, 3977. doi:10.1038/s41467-019-11994-0 (2019).
257. Buchan, D. W. A. & Jones, D. T. The PSIPRED Protein Analysis Workbench: 20 Years On. *Nucleic Acids Research* **47**, W402–W407. doi:10.1093/nar/gkz297 (2019).
258. Afriat-Jurnou, L., Jackson, C. J. & Tawfik, D. S. Reconstructing a Missing Link in the Evolution of a Recently Diverged Phosphotriesterase by Active-Site Loop Remodeling. *Biochemistry* **51**, 6047–6055. doi:10.1021/bi300694t (2012).

259. Correy, G. J. *et al.* Mapping the Accessible Conformational Landscape of an Insect Carboxylesterase Using Conformational Ensemble Analysis and Kinetic Crystallography. *Structure* **24**, 977–987. doi:10.1016/j.str.2016.04.009 (2016).
260. Biel, J. T., Thompson, M. C., Cunningham, C. N., Corn, J. E. & Fraser, J. S. Flexibility and Design: Conformational Heterogeneity along the Evolutionary Trajectory of a Redesigned Ubiquitin. *Structure* **25**, 739–749.e3. doi:10.1016/j.str.2017.03.009 (2017).
261. Clifton, B. E. *et al.* Evolution of Cyclohexadienyl Dehydratase from an Ancestral Solute-Binding Protein. *Nature Chemical Biology* **14**, 542–547. doi:10.1038/s41589-018-0043-2 (2018).
262. Reetz, M. T., Carballeira, J. D. & Vogel, A. Iterative Saturation Mutagenesis on the Basis of B Factors as a Strategy for Increasing Protein Thermostability. *Angewandte Chemie International Edition* **45**, 7745–7751. doi:10.1002/anie.200602795 (2006).
263. Kamtekar, S., Schiffer, J. M., Xiong, H., Babik, J. M. & Hecht, M. H. Protein Design by Binary Patterning of Polar and Nonpolar Amino Acids. *Science* **262**, 1680–1685. doi:10.1126/science.8259512 (1993).
264. Wang, M. S. & Hecht, M. H. A Completely De Novo ATPase from Combinatorial Protein Design. *Journal of the American Chemical Society*. doi:10.1021/jacs.0c02954 (2020).
265. Goldenzweig, A. *et al.* Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability. *Molecular Cell* **63**, 337–346. doi:10.1016/j.molcel.2016.06.012 (2016).
266. Engler, C., Kandzia, R. & Marillonnet, S. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLOS ONE* **3**, e3647. doi:10.1371/journal.pone.0003647 (2008).
267. Acevedo-Rocha, C. G., Reetz, M. T. & Nov, Y. Economical Analysis of Saturation Mutagenesis Experiments. *Scientific Reports* **5**. doi:10.1038/srep10654 (2015).

9 References

268. Van Loo, B. *et al.* High-Throughput, Lysis-Free Screening for Sulfatase Activity Using *Escherichia Coli* Autodisplay in Microdroplets. *ACS Synthetic Biology* **8**, 2690–2700. doi:10.1021/acssynbio.9b00274 (2019).
269. Suter-Crazzolara, C. & Unsicker, K. Improved Expression of Toxic Proteins in *E. Coli*. *BioTechniques* **19**, 202–204 (1995).
270. Saïda, F. Overview on the Expression of Toxic Gene Products in *Escherichia Coli*. *Current Protocols in Protein Science* **50**, 5.19.1–5.19.13. doi:10.1002/0471140864.ps0519s50 (2007).
271. Gasteiger, E. *et al.* in *The Proteomics Protocols Handbook* (ed Walker, J. M.) 571–607 (Humana Press, Totowa, NJ, 2005). doi:10.1385/1-59259-890-0:571.
272. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2017).
273. Sievers, F. & Higgins, D. G. in *Multiple Sequence Alignment: Methods and Protocols* (ed Katoh, K.) 3–16 (Springer US, New York, NY, 2021). doi:10.1007/978-1-0716-1036-7_1.
274. Armougom, F. *et al.* Espresso: Automatic Incorporation of Structural Information in Multiple Sequence Alignments Using 3D-Coffee. *Nucleic Acids Research* **34**, W604–W608. doi:10.1093/nar/gkl092 (2006).
275. Guindon, S. *et al.* New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology* **59**, 307–321. doi:10.1093/sysbio/syq010 (2010).
276. Lefort, V., Longueville, J.-E. & Gascuel, O. SMS: Smart Model Selection in PhyML. *Molecular Biology and Evolution* **34**, 2422–2424. doi:10.1093/molbev/msx149 (2017).
277. Gerlt, J. A. *et al.* Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A Web Tool for Generating Protein Sequence Similarity Networks. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1854**, 1019–1037. doi:10.1016/j.bbapap.2015.04.015 (2015).

278. Zallot, R., Oberg, N. & Gerlt, J. A. The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. *Biochemistry* **58**, 4169–4182. doi:10.1021/acs.biochem.9b00735 (2019).
279. Shannon, P. *et al.* Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research* **13**, 2498–2504. doi:10.1101/gr.1239303 (2003).
280. Kabsch, W. Integration, Scaling, Space-Group Assignment and Post-Refinement. *Acta Crystallographica Section D: Biological Crystallography* **66**, 133–144. doi:10.1107/S0907444909047374 (2010).
281. Evans, P. R. & Murshudov, G. N. How Good Are My Data and What Is the Resolution? *Acta Crystallographica Section D: Biological Crystallography* **69**, 1204–1214. doi:10.1107/S0907444913000061 (2013).
282. Karplus, P. A. & Diederichs, K. Linking Crystallographic Model and Data Quality. *Science* **336**, 1030–1033. doi:10.1126/science.1218231 (2012).
283. Vagin, A. & Teplyakov, A. MOLREP: An Automated Program for Molecular Replacement. *Journal of Applied Crystallography* **30**, 1022–1025. doi:10.1107/S0021889897006766 (1997).
284. Murshudov, G. N. *et al.* REFMAC5 for the Refinement of Macromolecular Crystal Structures. *Acta Crystallographica Section D: Biological Crystallography* **67**, 355–367. doi:10.1107/S0907444911001314 (2011).
285. Adams, P. D. *et al.* PHENIX: A Comprehensive Python-Based System for Macromolecular Structure Solution. *Acta Crystallographica Section D: Biological Crystallography* **66**, 213–221. doi:10.1107/S0907444909052925 (2010).
286. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and Development of Coot. *Acta Crystallographica Section D: Biological Crystallography* **66**, 486–501. doi:10.1107/S0907444910007493 (2010).

9 References

287. Grant, B. J., Rodrigues, A. P. C., ElSawy, K. M., McCammon, J. A. & Caves, L. S. D. Bio3d: An R Package for the Comparative Analysis of Protein Structures. *Bioinformatics* **22**, 2695–2696. doi:10.1093/bioinformatics/btl461 (2006).
288. Almagro Armenteros, J. J. *et al.* SignalP 5.0 Improves Signal Peptide Predictions Using Deep Neural Networks. *Nature Biotechnology* **37**, 420–423. doi:10.1038/s41587-019-0036-z (2019).

10 Appendix

10.1 Amino acid sequences of studied proteins

For purification, proteins were cloned with an additional N-terminal affinity tag , apart from small *de novo* proteins, which were purified by HPLC (see section 8.7). Signal peptides as predicted by SignalP-5.0 [288] are shown in blue and were not included into the expression construct (*MzDLH*, *EcoDLH*).

>P91-WT

MTARKVDYTDGATRCIGEFHWDEGKSGPRPGVVVFPEAFGLNDHAKERARRLADLGFAALAADMHGDAQVF
DAASLSSTIQGYGDRAHWRRRAQAALDALTAQPEVDGSKVAAIGFCFGGATCLELARTGAPLTAIVTFHG
GLLPEMAGDAGRIQSSVLVCHGADDPLVQDETMKAVMDEFRRDKVDWQVLYLGNVHSFTDPLAGSHGIPG
LAYDATAEARSWTAMCNLFSEFG

>P91-R1

MTARKVDYTDGATRCIGEFHWDEGKSGPRPGVVVFPEAFGLNDHAKERARRLADLGFAALAADMHGDAQVF
DAASLSSTIQGYGDRAHWRRRAQAALDALTAQPEVDGSKVAAIGFCFGGATCLELARTGAPLTAIVTFHG
GLLPEMAGDAGRIQSSVLVCHGADDPLVQDETMKAVMDEFRRDKVDWQVLYLGNVHSFTDPLAGSHGWPG
VAYDATAEARSWTAMCNLFSEFG

>P91-R2

MTARKVDYTDGATRCIGEFHWDEGKSGPRPGVVVFPELFGGLNDHAKERARRLADLGFAALAADMHGDAQVF
DEASVSSTIQGYGDRAHWRRRAQAALDALTAQPEVDGSKVAAIGFCFGGATCLELARTGAPLTAIVTFHG
GLLPEMAGDAGRIQSSVLVCHGADDPLVQDETMKAVMDEFRRDKVDWQVLYLGNVHSFTDPLAGSHGWPG
VAYDATAEARSWTAMCNLFSEFG

10 Appendix

>HmDLH

METRRTITYSDGNTDYLGELYWEAGDAPRPGIIVVFPEAFGLNDHARERARRLAAIGYVAFADLHGGGAILD
SMEALGPRMQALFADRTLWRAMAGAALDTLAAQAEADAGKLGAI GFCFGGATCLELARSGAALAGIASFHG
GLKKEIDGDAGRIRASVVLVHGAEDPLL GAGTIDAVMTEFRRDQVDWQFTYYGNALHSFTDPEADTRGMDG
LAYDARTEARSWNAMA AFFDEL FN

>m3DLH

MKHREIRYTDGHTQFV GELHWDEQQGGKCPGVVVFPEAFGLNDHARERARRLAGLGYAALAADLHGDGRLI
DDMEQLRPRMEGLFGDRAAWRALARAALDTLVAQPEVDADRLAAIGFCFGGTTALELARSGASLGAI VTFH
AGLLPELPEDAGRIRGRV LVCHGAEDPLVQKEAIDAVMGEWRRDRVDWQFTFYGNAHSFTDPAADAHGMA
GLAYEPLTEARSWTAMRNLFDEVFSR

>SifDLH

MKTQSIEYASGPTRLVGH LAWDADIGARRPGIIVVFPEAFGLNAHARERAERLARLG YVALAADLLGDGRVF
DNLPSVVPSIKALYADRTAWRARA AALEVLLARPEVDRERLGAIGFCFGGSTALELARSGAPLSAVATFH
AGLLPRLPEDAGRIGSRV LICHGDDDPVVNQDALATV VDELRRDRVDWQLARYGNTVHSFTDPQADARNNP
GFAYNALADRRSWAAMRQLFDEAFALP

>m1DLH

MKTETIEYSDGGTTCIGH LAWDDTQTGPRPGIIVVFSEAYGLNDHARIRAERLAALGFVALAADLHGNGLVY
GDMASLGPAIQALYADRSARARALAAFNTLVALPQVD TNQTAAGFCFGGATCFELARTGAPLGGITV FH
AGVIPELPEDKGRISGQVLICQ GADDPVVKKEAVDAVTAELSRDKVDWQYIVYANTGHSFTDPDADARNMP
GFAYNALAEERSWMMRLQYHEIFVT

>SfDLH

MHQQPIETTENGQRHIHQFFLDETLQGPRPGVLVVFPEAFGLGDHALQRARRLAELGYAALAVDIHGEGREF
QDLAQVRPAILALFGDRAAWRARLQAAHELLRAQPQVDAARTAAIGFCFGGACSL ELARSGAPLSAIVTFH
AGLQPPEADAGKIKAKVLVCHGAEDPLMKPEPLAAILAELTRDKVDWQLLSHGNNVHSFTNPDADARGAP
GFAYNAGADRRSWAAMQGLFAEVFA

>ScDLH

MRKQKIEYGNPTQFHGWLIRDDSLDGVRPGVLVVFPEAYGLNEHAIERAERLAQLGYVALAADMHGGGVVY
SDTATLGPAIRSLFGDRAEWRARAQAALDALLAQPVDRDRVAAIGFCFGGATCLELARSGAPLSALVTFH
AGLQPPEADAGRITGKVLICHGAEDPLMKPEALNAVLAELSRDRVDWQLLSFGGVAHSFTNPDADARGAP
GFAYNANADRRSWAAMQGLFAEVFAN

10.1 Amino acid sequences of studied proteins

>SaDLH

MKSQQVVYRGAGRRFLGELYWDEAATQPAPGVLVFPDAFGLADHARERAQRLAQLGYVALAADLHGEGAVY
EDVASMVRVHLQPLFENRADWRARAQAALAALQAQTPVDAQRLAAIGFCLGGATCLELARCGAPLKAIVGFH
AGVLAPLPGDEQKIQAQVLLCQGADDPLIKKENMAAVEAEELRRDRVDWQLIVYGNVHSFTNRDAATRQSP
AMAYDAAADRRSWAAMQGLLAEVF

>MzDLH

MKPHALLHPLIVLAIAPALVQAE^{LHTEEEIN}YRVGDDQFTGYLAYDDQISGQRPGILIVHEWWGHNEFARQQ
AERLATEGFTAFALDMYGSQKADHPDNARQFMQAATENS^{DVIRERFEAMRLLQDQPTVDASKIAAQGYC}
FGGAVVLNARMGMDLAGVVS^{IHGSLASPIQAEPGRVKARVQVYTGGADQMVPADQVAALVHEMQSAGVDL}
TLTSYPGVKHSFSNPDA^{DVVAERFGMPVAYDEQAAERTWRGTLAFYQELFGR}

>HsDLH (CMBL)

MANEAYPCPCDIGHRLEYGGLGREVQVEHIKAYVTKSPVDAGKAVIVIQDIFGWQLPNTRYIADMISNGY
TTIVPDFFFVGQEPWDPSGDWSIFPEWLKTRNAQKIDREISAILKYLKQQAQKIGIVGFCWGGTAVHHLM
MKYSEFRAGVSVYGVKDSEDIYNLKNPTLFIFAENDVVIPLKDVSLLTQKLKEHCKVEYQIKTFSGQTHG
FVHRKREDCSPADKPYIDEARRNLI^{EWLNKYM}

>KpDLH

MTTTKQPGFAPAASPHAATAVHTPEEHI^{IAGETSIPSQGENMPAYHARPKNADGPLP}IVIVVQEIFGVHEH
IRDLCRRLAQEGYLAIAPELYFRQGD^{PNEYHDIPTLFKELVSKVPDAQVLADLDHVASWAARHGDAHRL}
ITGFCWGGRITWLYAAHNPQLKAAVAWY^{GKLVGEKSLNSPKHPVDIAVDLNAPVLGLYGAKDASIPQD}TV
TMRQALRAANATAEIVVYPEADHAFNADYRASYHEESA^{KDGWQRMLAWFAQYGGKKG}

>EcoDLH

MLCLKKHQ^{LRSATMPRLTAKDFPQELLDYYDYAHGKISKREFLNLA}AKYAVGGMTALALFDLLKPNYALA
TQVEFTDPEIFA^{EYITYPSPNGHGEVRGYLVKPAKMSGKTPAVVVH}ENRGLNPYIEDVARRVAKAGYIAL
APDGLNSVGGYPGNDDK^{GRELQQQVDPTKLMNDFFAAIEFMQRY}PQATGKVGITGFCYGGGVSNA^{AAVAYP}
ELACAVPFYGRQAPTADVAKIEAPLLLHFAEL^{DTRINEGWPAYEAALKANNKVYEAYI}YPGVNHGFHNDST
PRYDKSAADLAWQRTLK^{WFDKYL}S

>PkDLH

MLTEGISIQSYDGHTFGALV^{GSPAKAPAPVIVIAQEIFGVNAFMRET}VS^{WLVDQGYAAVCPDLYARQ}APGT
ALDPQDERQREQAYKLWQAFDMEAGVDLEAAIRYARHQ^{PYSNGKVGLVGYCLGGALAF}LVAAGYVDRAV
GYYGVGLEKQLNKVPEVKHPALFHMGGQ^{DHFVPAPSRQLITEGFGANPLLQV}HWE^{EAGHSFARTSSSGYV}
ASAAALANERTLD^{FLAPLQSKKP}

10 Appendix

>LbOpdB

MPVIFYIHGAGWVFGSAKTHDKLVRELAVRTHSVVVFPEYTLSPEAKYPTAIEQNYAVLQQLPEIAKDENE
DVDRLTVAGDSVGGNMTVMMLTKQRQGIKINQQLLYYPVTDANFDTPSYNQFADNYLTKEGMQWFWDQ
YTTDPKQRAEITASPLRASLDELKDLPAAMILNGEADVLRDEGEAYANKLREAGVAVTQVQFQGMIHDFVM
VNSLDDTHATRAAMDLSWSVLQRNNQ

>Smu.1393c

MAALNKEMVNTLLGPIYTCHREGNPCFVFLSGAGFFSTADNFANIIDKLPDSIGILTIDAPNSGYSPVSNQ
ANVGLRDWVNAILMIFEHFQSYLLCVHSIGGFAALQIMNQSSKACLGFIGLEPTTVMYRAGFSSDLYP
QLALRRQKLKTAADRLNYLKDLSRSHFSSQQFKQLWRGYDYCQRQLNDVQSLPDFKIRLALGEEDFKTGIS
EKIPSIVFSESFREKEYLESEYLNKHTQTKLILCGQHHLHWSETNSILEKVEQLLSNHEKL

>hBChE-7 G117H

MEDDIIITTKNGKVRGMNLTVFGGTVTAFLGIPYAQPPLGRLRFKKPQPLTKWSGIWNATKYANSCMQNID
TSFPGFHGSEMWNPNLTDLSEDCLYLNWIPAPKPKNATVMVWIYGGHFQGTSSLPVYDGKFLARVERVIV
VSMNYRVGALGFLALPGNPEAPGNMGLFDQQLALKWVQDNIAAFGGDPNRVTLFGESAGAASVSLHLLSPG
SHPLFTRAILQSGSANAPWAVMSPEEARNRTLNLAKLLGCSRENETEIIKCLRNKDPQEILDNEAFVVPYS
TPLSVNFGPTVDGDFLTDMPDTLLELGQFKKTQILVGVNKDEGTAFLVYGAPGFSKDNDSSIITRKEFQEG
KVFFPNVSEFGKESILFHYTDWEDEDRPENYRDALAEVVGDYFFICPALEFAKKYAEHGNNAYFYFEHRS
SKLPWPEWMGVMHGYEIEFVFGPLPLERRLNNTKKEEELSREIMRRWANFAKYGNPNETQNNSTQWPVFKPT
EQKYLTLNTESSRIMTKLRAQHCRFWNSFFPKV

>P91 insertion variant

MTARKVDYTDGATRCIGEFHWDEGKSGPRPGVVVFPEAFGLNDHAKERARRLADLGFALAADMHGDAQVF
TAASLSSTNMHGDAQVFQAASLSSTIQGYGDRAHWRRRAQAALDALTAQPEVDGSKVAAIGFCFGGATCL
ELARTGAPLTAIVTFHGGLLPEMAGDAGRIQSSVLVCHGADDPLVQDETMKAVMDEFRRDKVDWQVLYLGN
AVHSFTDPLAGSHGIPGLAYDATAEARSWTAMCNLFSELFG

>HsDLH-1E11

MANEAYPCPCDIGHRLEYGGLGREVQVEHIKAYVTKSPVDAGKAVIPIEAFGWQLPNTRYIADMISGNGY
TTIVPDFDFVQEPWDPSPGDSIFPEWLKTRNAQKIDREISAILKYLKQQAQKIGIVGFCWGGTAVHHL
MKYSEFRAGVSVYGIKDSEDIYNLKNPTLFIFAENDVVIPLKDVSLLTQKLKEHCKVEYQIKTFGNVHS
FTDPLAGSHGWPVAYDATYIDEARRNLIEWLNKYM

10.1 Amino acid sequences of studied proteins

10.2 Pre-steady state kinetics

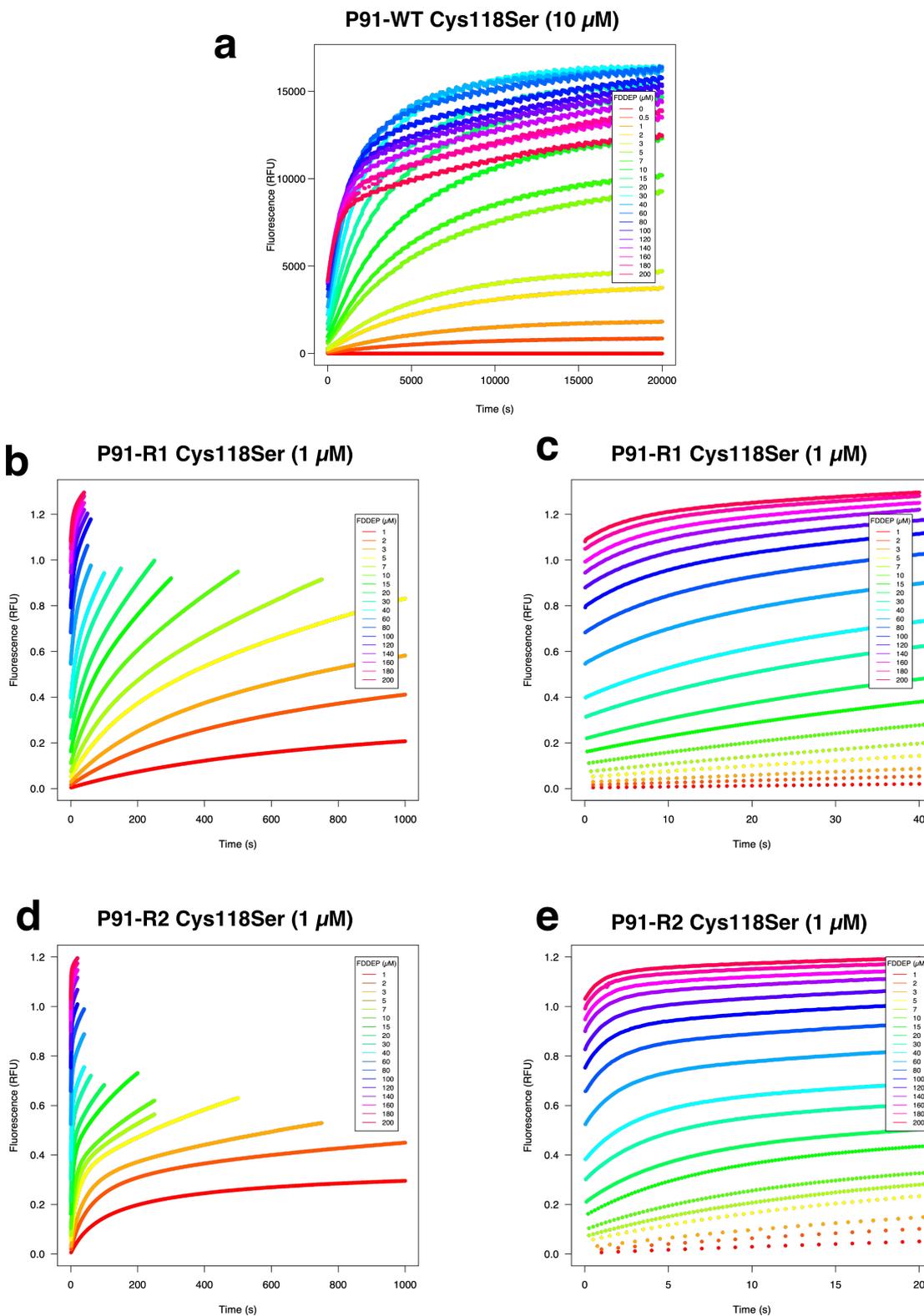


Figure 10.1: Stopped-flow reaction traces of all characterised P91 C118S variants with a concentration range of FDDEP (1–200 μM), measured in 50 mM HEPES-NaOH, 150 mM NaCl, 1 mM TCEP, pH 8.0 at 25 $^{\circ}\text{C}$. Measurement time was varied with different substrate concentrations and was increased at low substrate concentrations according to reaction rate. **(a)** P91-WT C118S (10 μM). **(b)** P91-R1 C118S (1 μM), full time range. **(c)** P91-R1 C118S (1 μM), close-up of the initial 40 s. **(d)** P91-R2 C118S (1 μM), full time range. **(e)** P91-R2 C118S (1 μM), close-up of the initial 20 s.

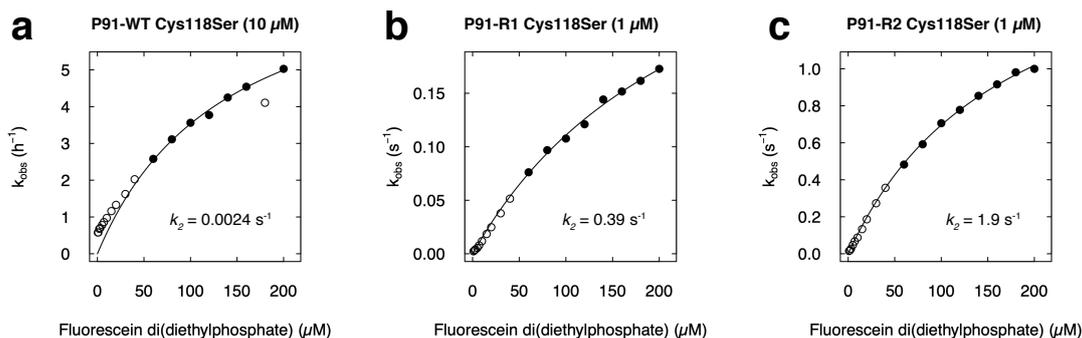


Figure 10.2: Determination of the phosphorylation rate k_2 . The parameter k_{obs} , determined from exponential fit of the kinetic burst traces measured with the substrate FDDEP, was plotted against substrate concentration and fitted to a modified Michaelis-Menten equation as described in section 8.8.2. Only substrate concentrations of $\geq 60 \mu\text{M}$ FDDEP (where substrate concentration \gg enzyme concentration) were considered for the fit (filled dots). Phosphorylation rates and standard errors of the fit were determined as **(a)** P91-WT C118S: $(2.4 \pm 0.10) \cdot 10^{-3} \text{ s}^{-1}$. **(b)** P91-R1 C118S: $(3.9 \pm 0.32) \cdot 10^{-1} \text{ s}^{-1}$. **(c)** P91-R2 C118S: $(1.9 \pm 0.067) \text{ s}^{-1}$. As the curves do not reach saturation due to low substrate solubilities, the values determined for k_2 may contain an error larger than that of the fit.

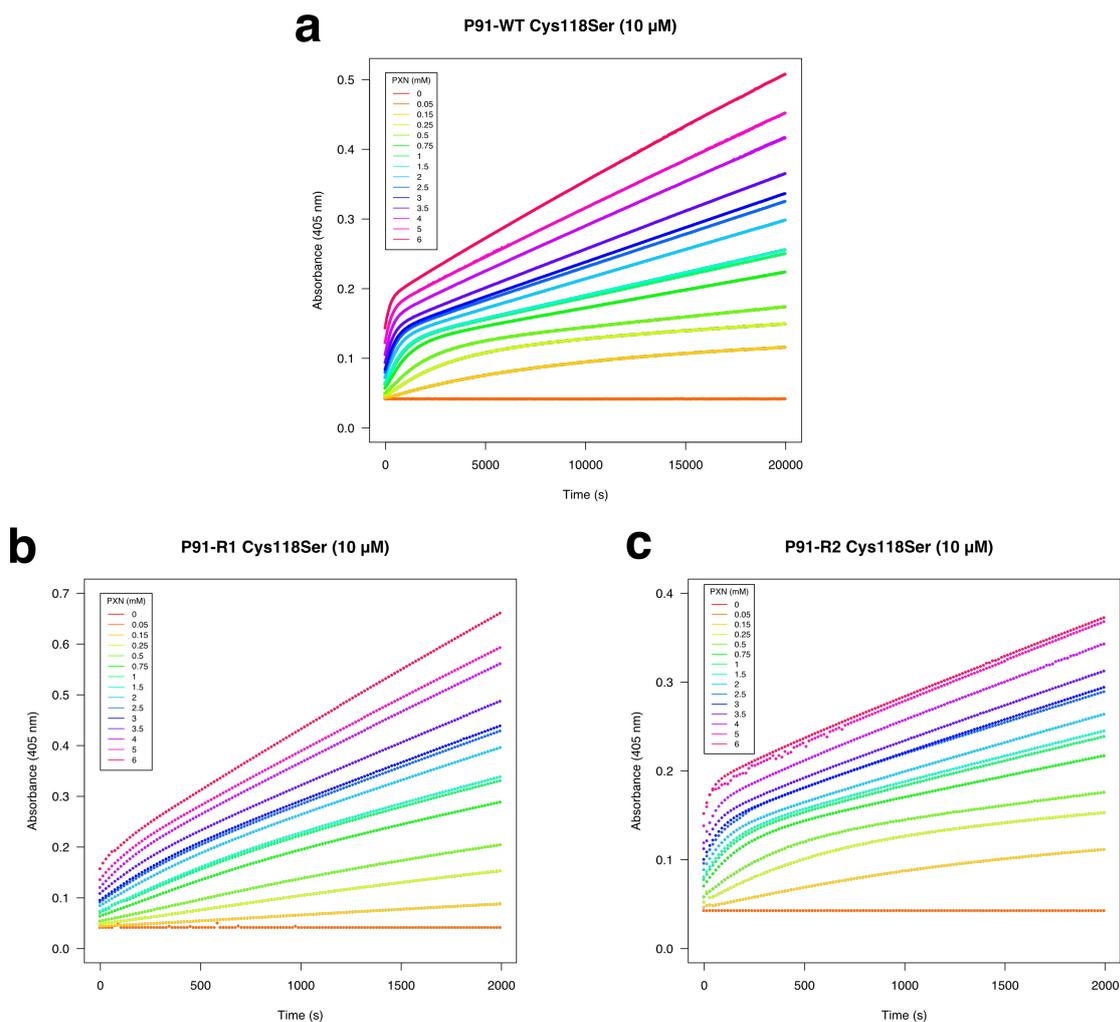


Figure 10.3: Burst traces of all characterised P91 C118S variants with a concentration range of paraoxon-ethyl (0.05–6 mM), measured in 50 mM HEPES-NaOH, 150 mM NaCl, 1 mM TCEP, pH 8.0 at 25 °C. **(a)** P91-WT C118S (10 μ M). **(b)** P91-R1 C118S (10 μ M). **(c)** P91-R2 C118S (10 μ M).

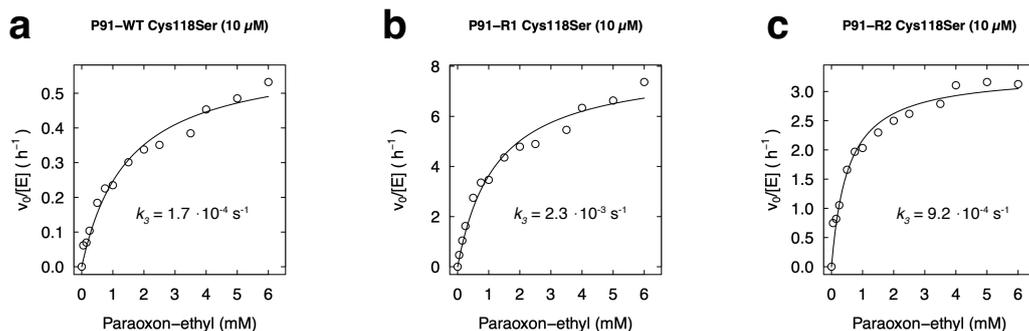


Figure 10.4: Determination of the de-phosphorylation rate k_3 . The initial rate v_0 from the second phase of the burst kinetics with the substrate paraoxon-ethyl was fitted to the Michaelis-Menten equation to determine k_{cat} and thus k_3 , as described in section 8.8.2. In the case of P91-R1 C118S, where the burst is less pronounced (probably due to a lower k_2 with the substrate paraoxon), k_{cat} gives an upper boundary and thus k_3 might be lower. De-phosphorylation rates were determined as **(a)** P91-WT C118S: $(1.7 \pm 0.10) \cdot 10^{-4} \text{ s}^{-1}$. **(b)** P91-R1 C118S: $(2.3 \pm 0.11) \cdot 10^{-3} \text{ s}^{-1}$. **(c)** P91-R2 C118S: $(9.2 \pm 0.34) \cdot 10^{-4} \text{ s}^{-1}$.

10.3 Steady-state kinetics

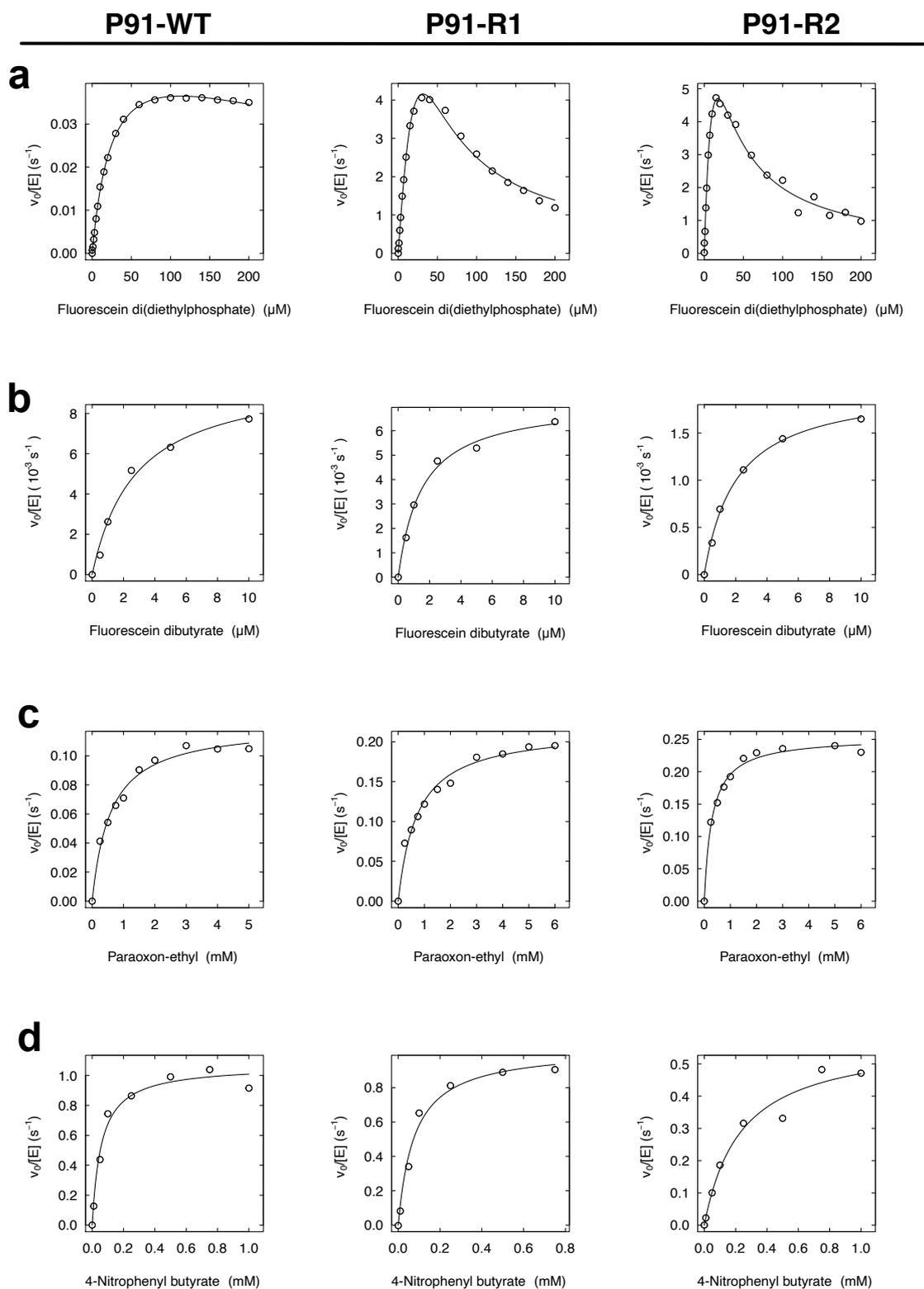


Figure 10.5: Michaelis-Menten plots of steady-state kinetics of P91-WT, P91-R1, and P91-R2 with (a) fluorescein di(diethylphosphate) 1, (b) fluorescein dibutyrate 3, (c) paraoxon-ethyl 2, and (d) *p*-nitrophenyl butyrate 4. Measured in 50 mM HEPES-NaOH, 150 mM NaCl, 1 mM TCEP, pH 8.0 at 25 °C.

Table 10.1: Steady-state kinetic parameters of P91-WT with carboxyester and phosphotriester substrates 1–4, measured in 50 mM HEPES-NaOH, 150 mM NaCl, 1 mM TCEP, pH 8.0 at 25 °C. The respective Michaelis-Menten plots are shown in Figure 10.5. Note that, due to strong substrate inhibition, estimates of k_{cat} and K_M of P91-R1 and P91-R2 for substrate 1 are extrapolations and only $\frac{k_{cat}}{K_M}$ can be regarded as precise. n.a.: not applicable.

P91 variant	Substrate	Enzyme concentration (μM)	k_{cat} (s^{-1})	K_M (μM)	K_i (μM)	$\frac{k_{cat}}{K_M}$ ($\text{M}^{-1}\text{s}^{-1}$)
P91-WT	1	0.2	0.054	27	470	2000
	2	0.2	0.01	2.8	n.a.	3600
	3	1	0.11	480	n.a.	230
	4	0.03	1.1	58	n.a.	18000
P91-R1	1	0.001	37	120	7.9	300000
	2	0.2	0.0072	1.5	n.a.	4800
	3	1	0.22	730	n.a.	300
	4	0.03	1.0	77	n.a.	13000
P91-R2	1	0.0002	15	20	14	780000
	2	0.5	0.002	0.02	n.a.	980
	3	0.5	0.27	360	57	750
	4	0.5	0.58	230	n.a.	2500

10 Appendix

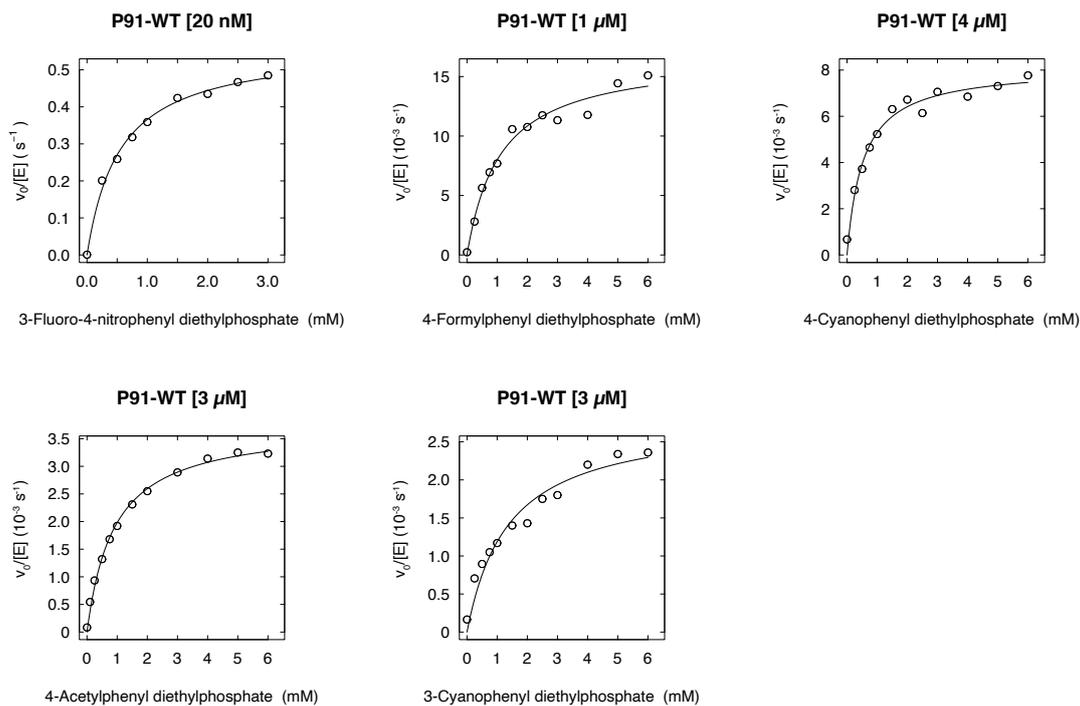


Figure 10.6: Michaelis-Menten plots for steady-state kinetics of P91-WT with linear-free energy relationship substrates 5–9 (see axis labels), measured in 50 mM HEPES-NaOH, 150 mM NaCl, pH 8.0 at 25 °C.

Table 10.2: Steady-state kinetic parameters of P91-WT with linear-free energy relationship substrates 5–9, measured in 50 mM HEPES-NaOH, 150 mM NaCl, pH 8.0 at 25 °C. The respective Michaelis-Menten plots are shown in Figure 10.6.

Substrate	Enzyme concentration (μM)	k_{cat} (s^{-1})	K_M (mM)	$\frac{k_{cat}}{K_M}$ ($\text{M}^{-1}\text{s}^{-1}$)
5	0.02	0.57	0.55	1000
6	1	0.017	1.1	15
7	4	0.0081	0.54	15
8	3	0.0038	0.9	4.2
9	3	0.0028	1.4	2.1

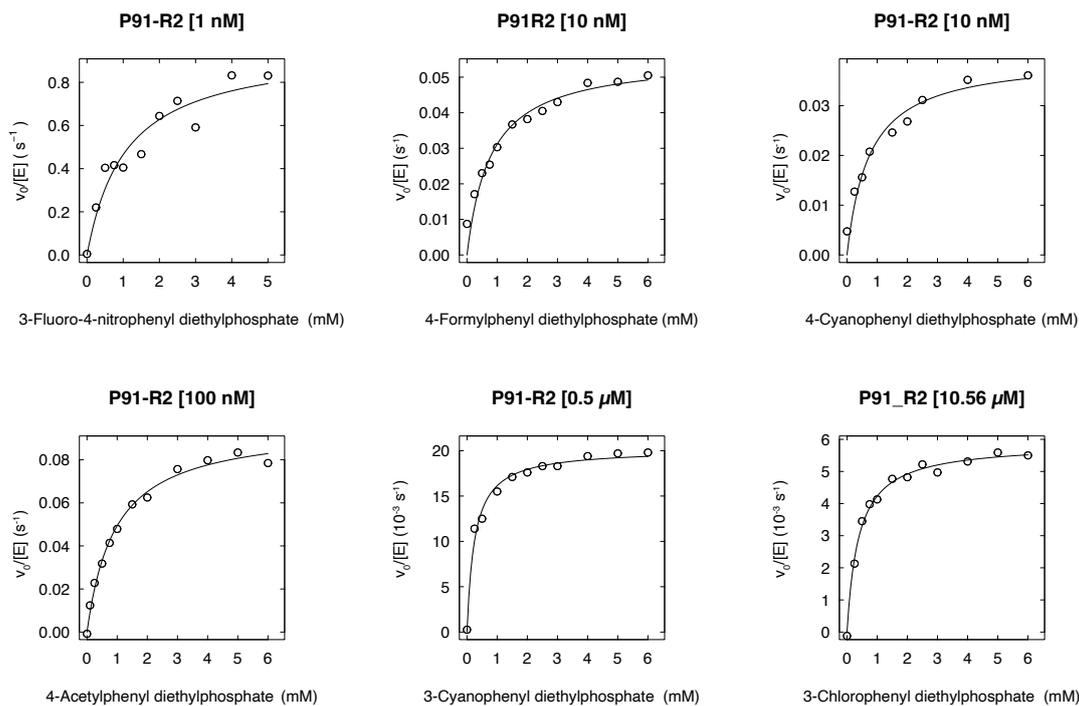


Figure 10.7: Michaelis-Menten plots for steady-state kinetics of P91-R2 with linear-free energy relationship substrates 5–10 (see axis labels), measured in 50 mM HEPES-NaOH, 150 mM NaCl, pH 8.0 at 25 °C.

Table 10.3: Steady-state kinetic parameters of P91-R2 with linear-free energy relationship substrates 5–10, measured in 50 mM HEPES-NaOH, 150 mM NaCl, pH 8.0 at 25 °C. The respective Michaelis-Menten plots are shown in Figure 10.7.

Substrate	Enzyme concentration (μM)	k_{cat} (s^{-1})	K_M (mM)	$\frac{k_{cat}}{K_M}$ ($\text{M}^{-1}\text{s}^{-1}$)
5	0.001	0.96	1.1	900
6	0.01	0.056	0.79	71
7	0.01	0.04	0.74	54
8	0.1	0.096	0.94	100
9	0.5	0.02	0.25	80
10	10.56	0.0059	0.39	15

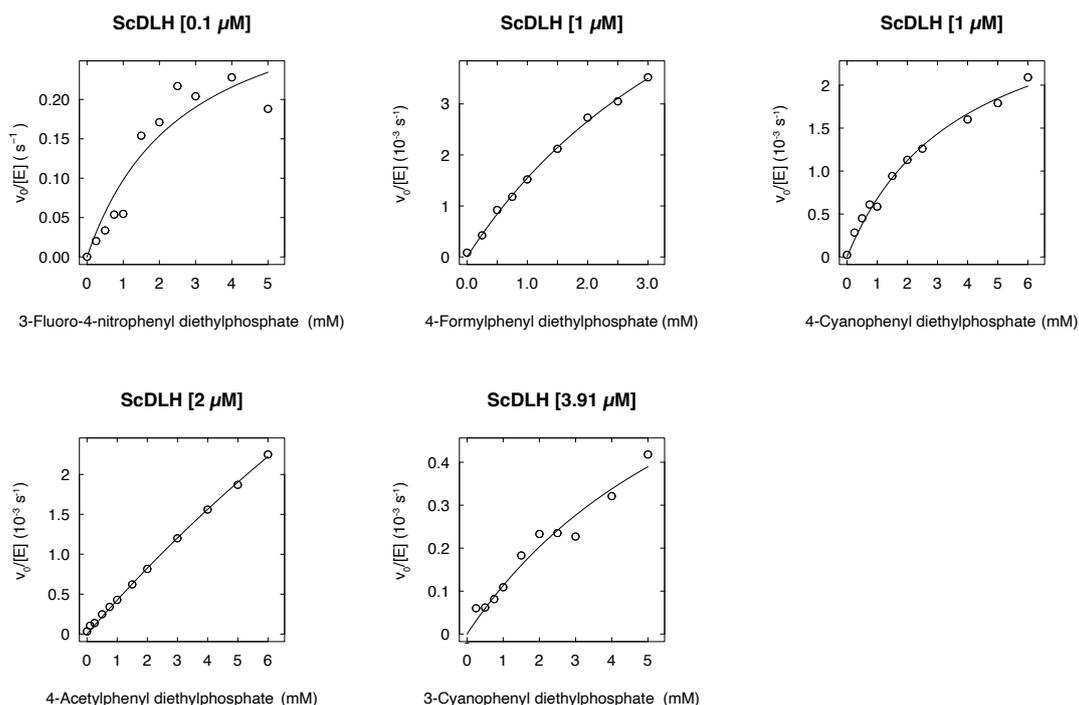


Figure 10.8: Michaelis-Menten plots for steady-state kinetics of ScDLH with linear-free energy relationship substrates 5–10 (see axis labels), measured in 50 mM HEPES-NaOH, 150 mM NaCl, pH 8.0 at 25 °C. Note that due to limited substrate solubility, the curve does not approach saturation with most substrates and, therefore, only $\frac{k_{cat}}{K_M}$ can be regarded as precise.

Table 10.4: Steady-state kinetic parameters of ScDLH with linear-free energy relationship substrates 5–9, measured in 50 mM HEPES-NaOH, 150 mM NaCl, pH 8.0 at 25 °C. The respective Michaelis-Menten plots are shown in Figure 10.8.

Substrate	Enzyme concentration (μM)	k_{cat} (s^{-1})	K_M (mM)	$\frac{k_{cat}}{K_M}$ ($\text{M}^{-1}\text{s}^{-1}$)
5	0.1	0.36	2.7	130
6	1	0.0095	5.1	1.8
7	1	0.0032	3.8	0.86
8	2	0.0015	33	0.44
9	3.91	0.001	7.9	0.13

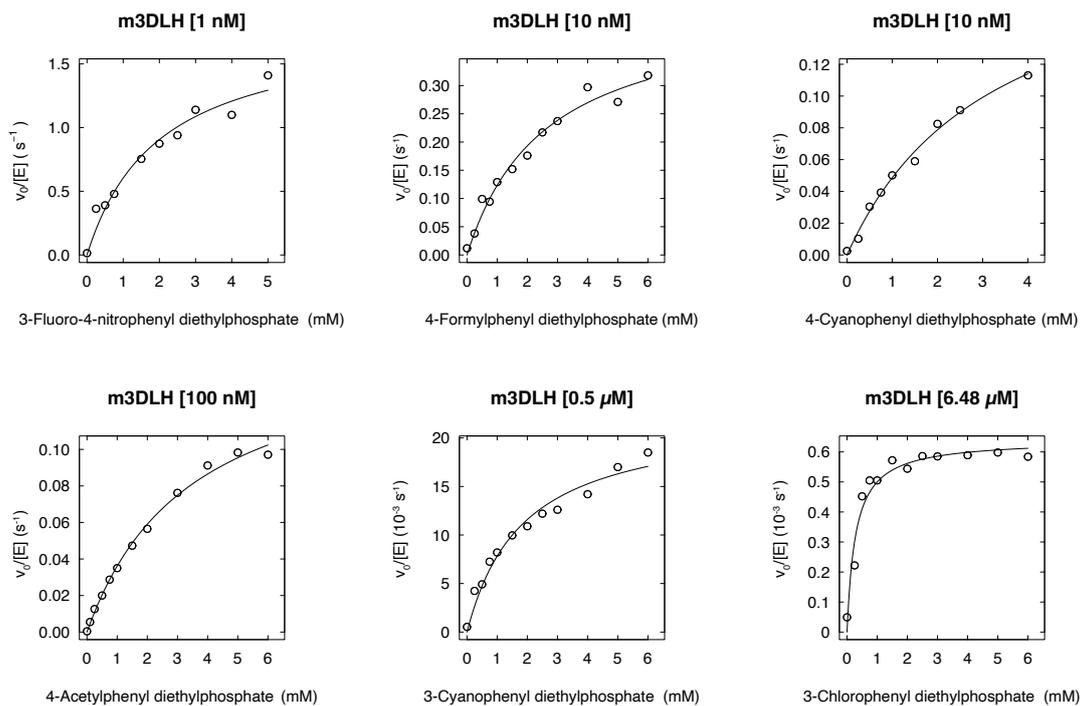


Figure 10.9: Michaelis-Menten plots for steady-state kinetics of *m3DLH* with linear-free energy relationship substrates 5–10 (see axis labels), measured in 50 mM HEPES-NaOH, 150 mM NaCl, pH 8.0 at 25 °C.

Table 10.5: Steady-state kinetic parameters of *m3DLH* with linear-free energy relationship substrates 5–10, measured in 50 mM HEPES-NaOH, 150 mM NaCl, pH 8.0 at 25 °C. The respective Michaelis-Menten plots are shown in Figure 10.9.

Substrate	Enzyme concentration (μM)	k_{cat} (s^{-1})	K_M (mM)	$\frac{k_{cat}}{K_M}$ ($\text{M}^{-1}\text{s}^{-1}$)
5	0.001	1.8	2.0	910
6	0.01	0.45	2.6	170
7	0.01	0.21	3.3	64
8	0.1	0.16	3.5	46
9	0.5	0.022	1.9	12
10	6.48	0.00064	0.28	2.3

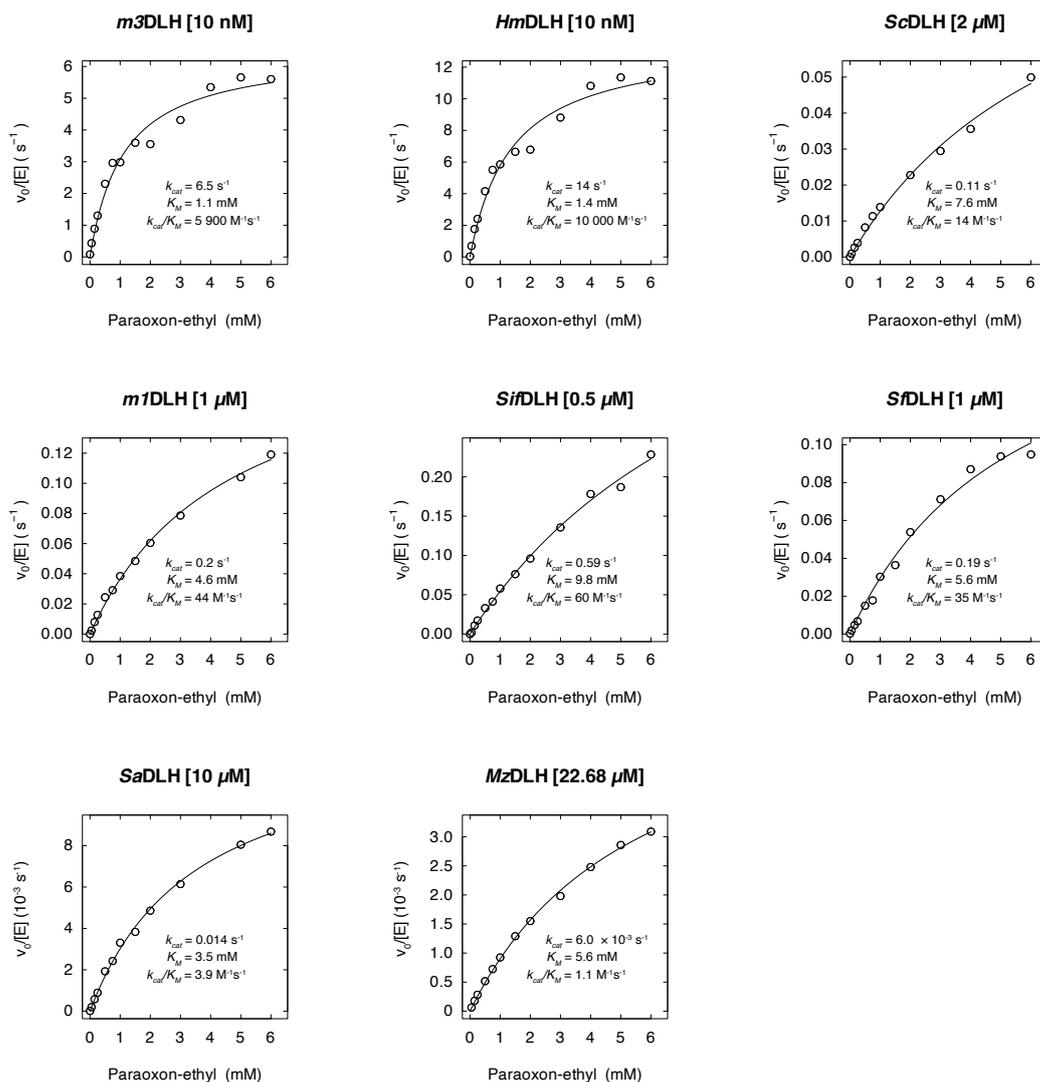


Figure 10.10: Michaelis-Menten plots for steady-state kinetics of DLH family proteins with the phosphotriester paraoxon-ethyl (2), measured in 50 mM HEPES-NaOH, 150 mM NaCl, pH 8.0 at 25 °C.

Table 10.6: Steady-state kinetic parameters of studied DLH family proteins for the hydrolysis of paraoxon-ethyl (substrate 2), measured in 50 mM HEPES-NaOH, 150 mM NaCl, pH 8.0 at 25 °C. The respective Michaelis-Menten plots are shown in Figure 10.10.

Enzyme	Enzyme concentration (μM)	k_{cat} (s^{-1})	K_M (mM)	$\frac{k_{cat}}{K_M}$ ($\text{M}^{-1}\text{s}^{-1}$)
<i>m3DLH</i>	0.01	6.5	1.1	5900
<i>HmDLH</i>	0.01	14	1.4	10000
<i>ScDLH</i>	2	0.11	7.6	14
<i>m1DLH</i>	1	0.2	4.6	44
<i>SifDLH</i>	0.5	0.59	9.8	60
<i>SfdLH</i>	1	0.19	5.6	35
<i>SaDLH</i>	10	0.014	3.5	3.9
<i>MzDLH</i>	22.68	6.0×10^{-3}	5.6	1.1

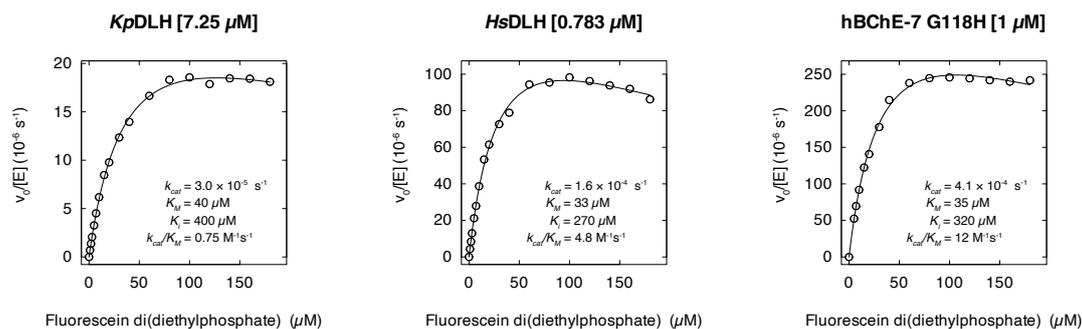


Figure 10.11: Michaelis-Menten plots for steady-state kinetics of DLH family proteins and hBChE-7 G118H with the phosphotriester fluorescein di(diethylphosphate) (FDDEP, 1), measured in 50 mM HEPES-NaOH, 150 mM NaCl, pH 8.0 at 25 °C.

Table 10.7: Steady-state kinetic parameters of studied DLH family proteins for the hydrolysis of fluorescein di(diethylphosphate)(substrate 1), measured in 50 mM HEPES-NaOH, 150 mM NaCl, pH 8.0 at 25 °C. The respective Michaelis-Menten plots are shown in Figure 10.11.

Enzyme	Enzyme concentration (μM)	k_{cat} ($10^{-4} \cdot \text{s}^{-1}$)	K_M (μM)	K_i (μM)	$\frac{k_{cat}}{K_M}$ ($\text{M}^{-1}\text{s}^{-1}$)
<i>KpDLH</i>	7.25	0.3	40	400	0.75
<i>HsDLH</i>	0.783	1.6	33	270	4.8
hBChE-7 G118H	1	4.1	35	320	12

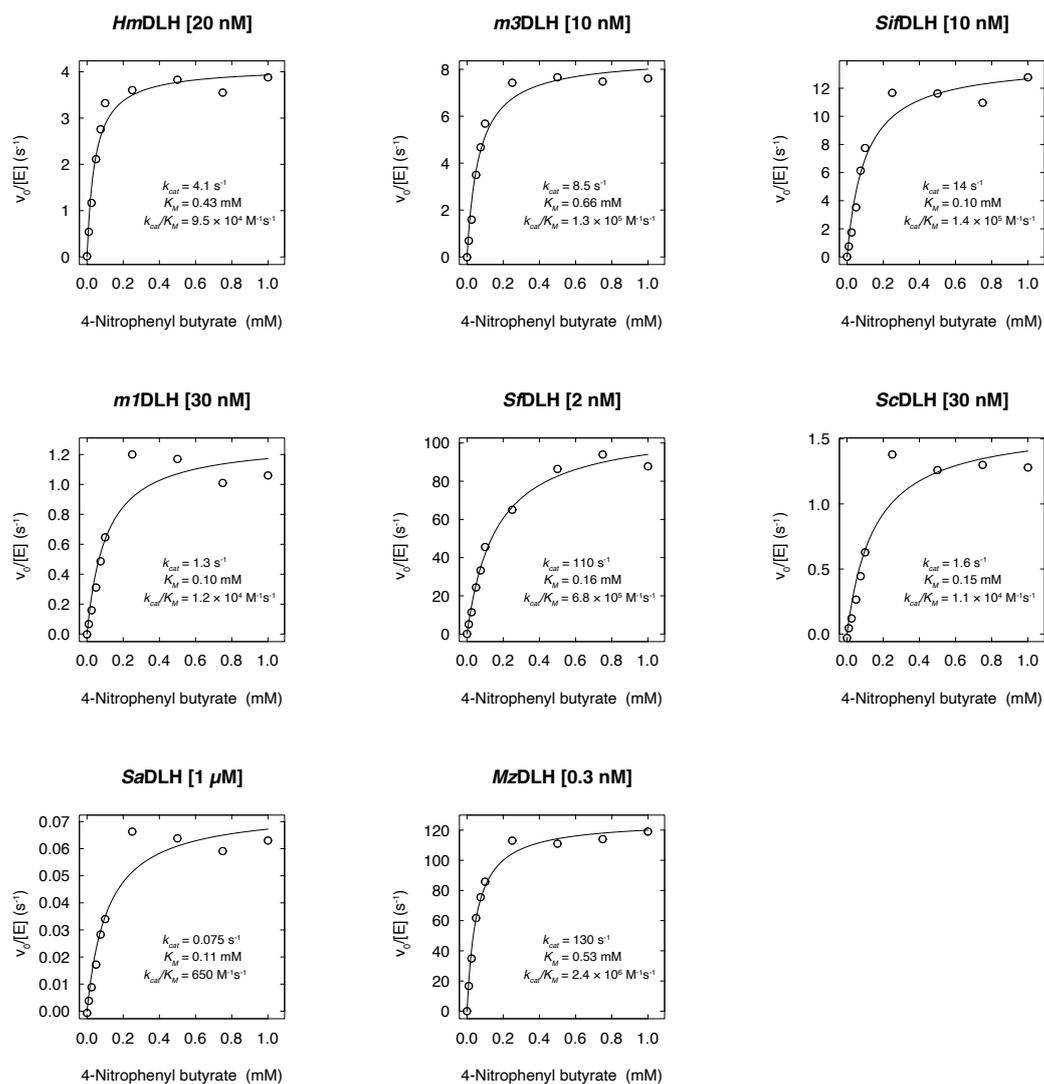


Figure 10.12: Michaelis-Menten plots for steady-state kinetics of DLH family proteins with the carboxyester *p*-nitrophenyl butyrate (4), measured in 50 mM HEPES-NaOH, 150 mM NaCl, pH 8.0 at 25 °C.

Table 10.8: Steady-state kinetic parameters of studied DLH family proteins for the hydrolysis of *p*-nitrophenyl butyrate (substrate 4), measured in 50 mM HEPES-NaOH, 150 mM NaCl, pH 8.0 at 25 °C. The respective Michaelis-Menten plots are shown in Figure 10.12.

Enzyme	Enzyme concentration (nM)	k_{cat} (s ⁻¹)	K_M (mM)	$\frac{k_{cat}}{K_M}$ (10 ³ · M ⁻¹ s ⁻¹)
P91-WT	30	1.1	0.058	18
<i>HmDLH</i>	20	4.1	0.043	95
<i>m3DLH</i>	10	8.5	0.066	130
<i>SifDLH</i>	10	14	0.1	140
<i>m1DLH</i>	30	1.3	0.1	12
<i>SfDLH</i>	2	110	0.16	680
<i>ScDLH</i>	30	1.6	0.15	11
<i>SaDLH</i>	1000	0.0075	0.11	0.65
<i>MzDLH</i>	0.3	130	0.053	2 400

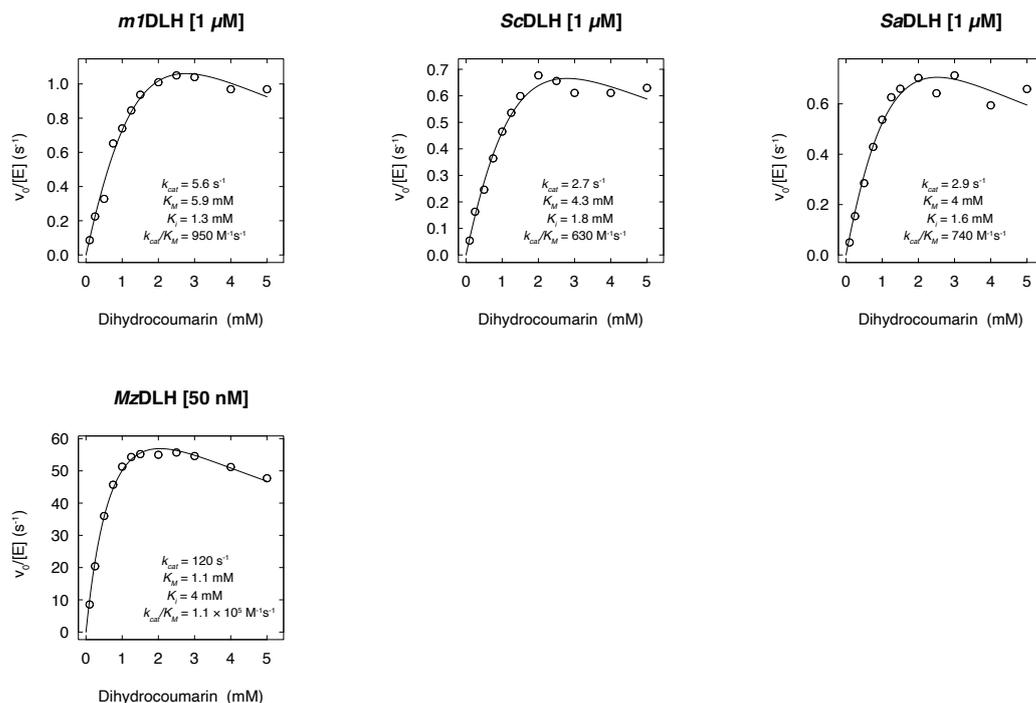


Figure 10.13: Michaelis-Menten plots for steady-state kinetics of DLH family proteins with the lactone dihydrocoumarin (11), measured in 50 mM HEPES-NaOH, 150 mM NaCl, pH 8.0 at 25 °C.

Table 10.9: Steady-state kinetic parameters of studied DLH family proteins for the hydrolysis of dihydrocoumarin (substrate 11), measured in 50 mM HEPES-NaOH, 150 mM NaCl, pH 8.0 at 25 °C. The respective Michaelis-Menten plots are shown in Figure 10.13.

Enzyme	Enzyme concentration (μM)	k_{cat} (s^{-1})	K_M (mM)	K_i (mM)	$\frac{k_{cat}}{K_M}$ ($\text{M}^{-1}\text{s}^{-1}$)
<i>mIDLH</i>	1	5.6	5.9	1.3	950
<i>ScDLH</i>	1	2.7	4.3	1.8	630
<i>SaDLH</i>	1	2.9	4	1.6	740
<i>MzDLH</i>	0.05	120	1.1	4	110 000

10.4 Crystal structures: Data collection details and refinement statistics

Table 10.10: Data collection details and refinement statistics for crystal structures of *m3DLH*, *ScDLH*, and *SfDLH*.

	<i>m3DLH</i>	<i>ScDLH</i>	<i>SfDLH</i>
Data collection			
Space group	C 1 2 1	P 1 2 ₁ 1	P 1 2 ₁ 1
Cell dimensions:			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	78.45 43.68 59.65	43.55 123.60 46.66	40.14 87.84 71.41
α , β , γ (°)	90.00 92.71 90.00	90.00 100.98 90.00	90.00 99.17 90.00
Resolution (Å)	23.75 - 1.533 (1.588 - 1.533)	40.4 - 1.85 (1.916 - 1.85)	39.63 - 1.45 (1.502 - 1.45)
<i>R</i> _{sym} or <i>R</i> _{merge}	0.1138 (0.5259)	0.08356 (0.5364)	0.0669 (0.2875)
CC _{1/2}	0.996 (0.85)	0.997 (0.909)	0.998 (0.979)
CC*	0.999 (0.958)	0.999 (0.976)	0.999 (0.995)
<i>I</i> / σ <i>I</i>	9.53 (2.13)	13.14 (3.64)	12.01 (4.43)
Completeness (%)	98.36 (95.71)	97.87 (99.76)	99.73 (99.14)
Redundancy	6.7 (5.9)	6.7 (6.8)	6.5 (6.3)
Refinement			
Resolution (Å)	23.75 - 1.533 (1.588 - 1.533)	40.4 - 1.85 (1.916 - 1.85)	39.63 - 1.45 (1.502 - 1.45)
No. unique reflections	29979 (2912)	40398 (4112)	86395 (8555)
<i>R</i> _{work} / <i>R</i> _{free}	0.1638 / 0.2017	0.1646 / 0.1943	0.1431 / 0.1797
No. atoms:			
Protein	1938	3690	3778
Ligand/ion	-	-	61
Water	205	504	317
B-factors:			
Protein	22.75	28.69	24.86
Ligand/ion	-	-	46.29
Water	30.79	38.24	32.25
R.m.s. deviations:			
Bond lengths (Å)	0.008	0.003	0.008
Bond angles (°)	0.98	0.63	1.10
PDB ID	7JKA	7JIZ	7JOP