Structural and mechanistic analysis of carbon-carbon bond hydrolases and lyases for potential industrial applications

Annika Frank

A thesis submitted for the degree of Doctor of Philosophy

The University of York

Department of Chemistry

August 2013

Abstract

The formation or cleavage of carbon-carbon bonds is often thermodynamically unfavourable and, using abiotic catalysts, only achieved under harsh reaction conditions. As an alternative, numerous enzymes are able to catalyse equivalent reactions and constitute versatile biocatalytic tools for the production of high value chemical building blocks. Three such catalysts have now been analysed with respect to their structure, function, reaction mechanism and/or substrate specificity.

The *Bacillus subtilis* phenolic acid decarboxylase (*Bs*PAD) is a carbon-carbon bond lyase. Despite the availability of an *apo*-structure, the enzyme's reaction mechanism remained speculative. Within this work, an active site mutant library was prepared and analysed and the structure of one of the variants could be determined in complex with its phenolic acid substrate. Using combined structural and kinetic data, a catalytic mechanism was proposed and confirmed previous modelling approaches.
In addition, the codon optimised genes of two carbon-carbon bond hydrolases were subcloned for recombinant protein production. Phloretin hydrolase (Phy), a *retro*-Friedel Crafts hydrolase from *Eubacterium ramulus*, was purified and crystallised. Although the crystals' diffraction quality was too poor for structure determination, a fluorescence spectrum revealed Phy to bind zinc. Together with a substrate screen and kinetic data from an active site mutant library, these findings gave novel insights into a unique metal dependent hydrolase.
As a member of the β-diketone cleaving family of enzymes, the oxidised polyvinyl alcohol hydrolase (OPH) from *Pseudomonas* sp. VM15C was analysed. After extensive solubility screening and construct optimisation, OPH could, for the first time, be purified as a soluble fusion construct with an N-terminal GST tag. The fused protein crystallised and now awaits further structural and mechanistic analysis.

It is hoped that the data obtained as part of this project will now advance the understanding of the characterised biocatalysts and create a basis for their optimisation and use in industrial applications.

LIST OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Acknowledgements

Of the many people involved in this work I would first and foremost like to thank my supervisor, Dr. Gideon Grogan. In addition to the design of the projects described in this thesis, his advice and support have been crucial to my work. He has always been extremely supportive while at the same time having an open mind towards new ideas and has created a working atmosphere that made the four years of this PhD a fantastic experience.

In addition to Gideon I would also like to thank the current and former members of his group with whom I have enjoyed working, travelling and celebrating on many occasions. I have learned a lot from them and truly enjoyed being a part of this group. The same is also true for all members of the York Structural Biology Laboratory. YSBL has been a wonderful place to work, study and write and the support and advice of PIs, staff and other students has been invaluable.

I would particularly like to thank both Sally Lewis and Simon Grist who have always kept the wet labs running perfectly and have been incredibly helpful to my practical work. I am also very grateful to Sam Hart and Johan Turkenburg for maintaining the X-ray equipment and for the many days, nights and weekends of data collection. Their support in data processing and help with understanding crystallographic problems has been extremely helpful. Here, a special mention and thanks also belong to Andrew Thompson who provided many a solution or explanation to molecular biology and crystallography problems. He has also been a brilliant housemate and his supply of tea and late night telly were greatly appreciated.

Outside of York I would like to thank all members of the Marie Curie Biotrains network. It has been a great experience to meet them, to learn about their research and especially to make so many friends on the way. Within this network, a special thanks also belongs to our close collaborators, Wolfgang Kroutil, Elina Siirola and Aashrita Rajagopalan. Working with them, learning chemistry and a bit of Austrian German, especially during my exchange visit to the Elk group's lab at the University of Graz, has been a great experience.

Last but certainly not least I would like to thank Rob and my family for their love, support and never-ending patience (and in Rob's case also invaluable lessons in macromolecular crystallography…). Danke für Alles!

Author’s declaration

I declare that all work presented in this thesis is my own unless stated otherwise. Any work and material which is not my own has been clearly referenced. Research carried out in collaboration has been acknowledged accordingly.

Annika Frank

August 2013

Chapter 1: Introduction

1. Biocatalysis

The chemical synthesis and degradation of many organic compounds can be complex and challenging. Expensive starting materials, harsh reaction conditions and multiple bottlenecks can make processes unfeasible. Furthermore, unfavourable equilibria and poor regio- and/or enantioselectivity of the catalysts may lead to low overall yields. To overcome this problem, the catalysts' biological counterparts - from isolated enzymes to whole live cells - may replace one or several components of a reaction and a purely abiotic reaction turns into what is known as "Biocatalysis". These biological components will not only be able to catalyse the desired reaction but have several advantages with respect to reaction conditions and overall productivity. Enzymes, being the active reagent in all forms of biocatalysis, are often most active at moderate temperatures and pressure as well as physiological pH values. They show high substrate specificity and can differentiate between different stereoisomers. Enzymes have a further advantage in their use as catalytic tools as they can, within limits, be optimised and shaped into desired functionalities. Alteration of the primary amino acid sequence and subsequently smaller or larger changes to secondary and tertiary structure can allow for alterations in the substrate scope, activity and selectivity. Still, no matter what reaction will be catalysed eventually, biocatalysis can conform to the Principles of Green Chemistry, which promote processes which are efficient, yet safe and sustainable (1,2).

1.2 Enzymatic C-C bond formation and cleavage in biocatalysis

Carbon-carbon bonds form the backbone of most organic compounds, many of which are important in biotechnological applications. Their formation and cleavage is essential to the synthesis and degradation of such molecules but form bottlenecks in chemical processes. Especially for the formation of C-C bonds, equilibria are often unfavourable and reaction conditions too harsh for large scale applications. Unlike carbon-heteroatom bonds, the linkage between two carbon atoms is not polar and, especially for the sp^3 hybridised C-C single bonds, very strong. As a consequence, the use of enzymes in the formation and cleavage of carbon-carbon bonds has

become an attractive alternative, allowing for the production (or degradation) of complex organic molecules on a large scale (3,4).

A number of enzymes have been characterised with respect to their ability to cleave and/or form carbon-carbon bonds. The majority are C-C lyases and C-C hydrolases, belonging to enzyme classes EC 4.1 and EC 3.7 respectively, in addition to a smaller number of oxidoreductases and transferases. Members of the two main classes commonly cleave carbon-carbon bonds in nature but differ in structure, function, reaction mechanisms and cofactor requirements. *In vitro*, a small number of these enzymes have been shown to also catalyse the reverse reaction and form C-C bonds rather than cleave them. This reaction is, however, still unexploited for many biocatalysts, especially the hydrolases, and is an area of potential interest for industrial applications (4).

The following sections will give a brief overview on the two major groups of carbon-carbon bond cleaving enzymes, C-C hydrolases and lyases, many of which are known for their biocatalytic potential. The enzymes' diversity in reactions catalysed, mechanisms, substrate scope and cofactor dependence will be outlined.

1.3 Carbon-carbon bond lyases

C-C bond lyases catalyse addition-elimination reactions across carbon-carbon double bonds. These include decarboxylations, aldol reactions, or nucleophilic substitutions. Lyases may form two products from one substrate or *vice versa*, depending on the directionality of the reaction. According to the catalysed reactions, carbon-carbon lyases form four groups, namely the carboxy-lyases or decarboxylases, aldehyde lyases, oxo-acid lyases and others (Enzyme classes 4.1.1, 4.1.2, 4.1.3 and 4.1.4 respectively).

1.3.1 Aldolases

Aldolases catalyse the formation of C-C bonds *via* an aldol addition of α-hydroxy aldehydes or ketones, to create ß-hydroxy carbonyl compounds, in some cases requiring a cofactor such as pyridoxal-5'-phosphate (PLP) for activity. A donor and acceptor molecule are combined *via* nucleophilic addition - involving formation of a

Schiff base intermediate between the enzyme's active site lysine and the donor molecule (class I aldolases) or by using a zinc ion cofactor acting as a Lewis acid to activate a donor nucleophile (class II aldolases) (4,5). Alternatively, they may cleave carbon-carbon bonds in the reverse *retro*-aldol reaction. Fully reversible aldolase reactions are well understood and widely applied, also due to their increased stereospecificity and applicability on highly substituted substrates compared to their abiotic counterparts. They accept, however, only a very limited number of donor molecules – only four carbonyl compounds for the most commonly used aldolases (3,4,6).

1.3.2 Oxo-acid lyases

Oxo-acid lyases cleave carbon-carbon bonds of 3-hydroxy acids. This can involve, for example, aldol reactions as with the widely studied, pyruvate dependent *N*-acetylneuraminic acid lyase (NeuA). NeuA catalyses the aldol condensation between *N*-acetyl D-mannosamine and pyruvate to yield *N*-acetylneuraminic acid - a signalling molecule with diverse biological functions. Other oxo-acid lyases may catalyse electrophilic substitutions such as the isocitrate lyase (ICL) in the glyoxylate cycle – a bypass to the citric acid cycle. The enzyme produces glyoxylate and succinate from isocitrate and thus supplies both a shortcut and supply of intermediates to the TCA cycle and gluconeogenesis. (3,7-9).

1.3.3 Carboxylases

Carboxylases mediate the transfer of a carboxylic acid functional group to or from their substrates. In many cases the reaction equilibrium is in favour of decarboxylation and breaking of carbon-carbon bonds (see Fig. 1.1). The reverse reaction has however been observed and can be enhanced by optimisation of experimental conditions, *i.e.* by shifting the equilibrium (10-12). A recent example for this type of reaction is the enzymatic carboxylation of a variety of phenolic and styrene substrates in carbonate buffer (13). Different decarboxylases catalyse their reactions in several ways. They may depend on organic cofactors such as thiamine diphosphate (ThDP), biotin or pyridoxal phosphate (PLP), the latter of which is especially common with amino acid decarboxylases (3,14,15). Some enzyme-family

members also make use of a small inorganic cofactor, usually a mono- or divalent metal, to catalyse their reactions. Other decarboxylases, such as certain phenolic acid decarboxylases and related enzymes, are independent of any additional cofactors (16-19).

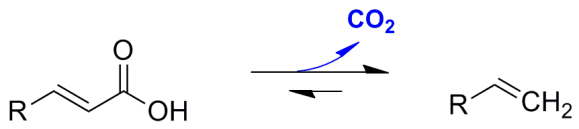


Figure 1.1: A schematic carboxylation/decarboxylation reaction. For enzymes the equilibrium is usually in favour of carbon-carbon bond cleavage.

1.3.3.1 PLP dependent decarboxylases

A large family of enzymes relies on pyridoxal phosphate (PLP) derived from Vitamin B_6 as a cofactor for catalysis, among them also a variety of decarboxylases. The majority are involved in the synthesis and breakdown of amino acids and their metabolic importance has rendered them attractive drug targets. As a consequence of their basic metabolic functions, PLP dependent enzymes are present in all forms of life but are more abundant in the less complex prokaryotes and lower eukaryotes (20). They are grouped, both structurally and functionally, into α, β and γ families, referring to the position of their substrates' main catalytic carbon. Reactions catalysed by these enzymes range from decarboxylations, transaminations, racemisations and eliminations to transfer reactions (21,22).

In PLP dependent enzymes, the cofactor forms a Schiff base with the ε-amino group of a highly conserved lysine in the active site (Fig. 1.2a). This is known as the internal aldimine. As substrate enters the active site, the imine linkage is broken and reformed between PLP and the substrate's amino group. For PLP dependent decarboxylases, this corresponds to the primary NH-group of a variety of amino acids as shown in Fig. 1.2. The external aldimine thus formed will mediate all subsequent catalysis, consistent with observations that PLP-mediated reactions can in fact take place (albeit at a reduced rate) without the enzyme being present (23).

These first steps are common to all PLP dependent enzymes and only after formation of the external aldimine do pathways split according to the final reaction outcome. In the case of decarboxylases, the carbon-carbon bond between the substrate's Cα carbon and the carboxylic acid is broken, releasing CO_2 and forming a quinonoid intermediate. The carbanion created by cleaving the C-C bond is highly unstable and relies on the conjugated system of PLP and the imine bond to relocate the negative charge and act as an electron sink, stabilising the quinonoid. For this to be most efficient, PLP and the imine bond lie in a plane orthogonal to the scissile carbon, ensuring optimal overlap of orbitals and thus movement of resonant electrons. Eventually, reprotonation then leads to release of the amine product and regeneration of the internal aldimine (21,23-25).

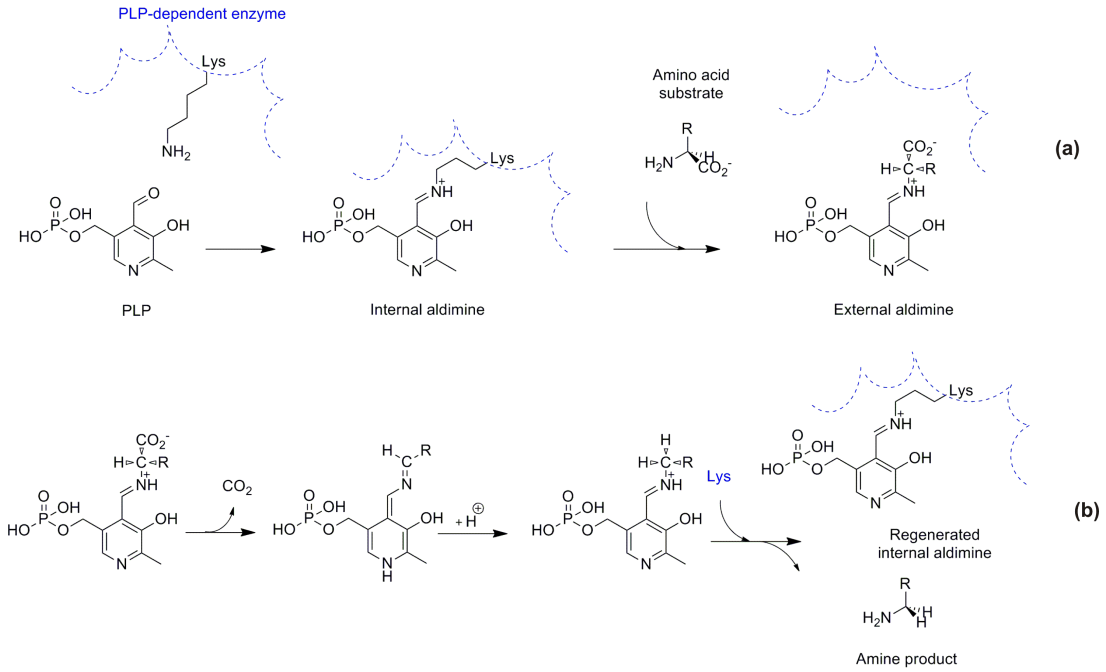


Figure 1.2: In the absence of substrate, PLP is linked to an active site lysine through an imine bond, forming the internal aldimine. As substrate enters, the imine bond is broken and reformed with the amino acids primary NH group to give the catalytic external aldimine (a). Within the latter, the carboxylate moiety is cleaved off and, following reprotonation at Cα, the imine is broken and reformed with the active site lysine, leading to release of the amine product regeneration of the internal aldimine (b) [adapted from (23)].

A well characterised enzyme that represents many of the structural and mechanistic features of PLP dependent decarboxylases is the L-3,4-dihydroxyphenylalanine

decarboxylase (DDC). DDC removes a carboxyl moiety from its amino acid substrate L-3,4-dihydroxyphenylalanine (DOPA) to give the important neurotransmitter L-3,4-dihydroxyphenylamine (dopamine). In addition to dopamine, the enzyme is involved in the biosynthesis of 5-hydroxytryptamine (serotonin) from L-5-hydroxytryptophan and is able to decarboxylate tryptophan to tryptamine. It is thus sometimes referred to as a general aromatic amino acid decarboxylase (AADC) (26). DDC proteins vary in size from around 50-100 kDa and associate into homodimers. They are members of the fold I type or α-family of PLP dependent enzymes whose structure strongly resembles that of mitochondrial aspartate aminotransferase (23,27,28). In 2001, the first structure of human DDC was solved in its PLP bound state as well as in complex with the inhibitor (2*S*)-3-(3,4-dihydroxyphenyl)-2-hydrazinyl-2-methyl-propanoic acid (carbidopa) (29,30).

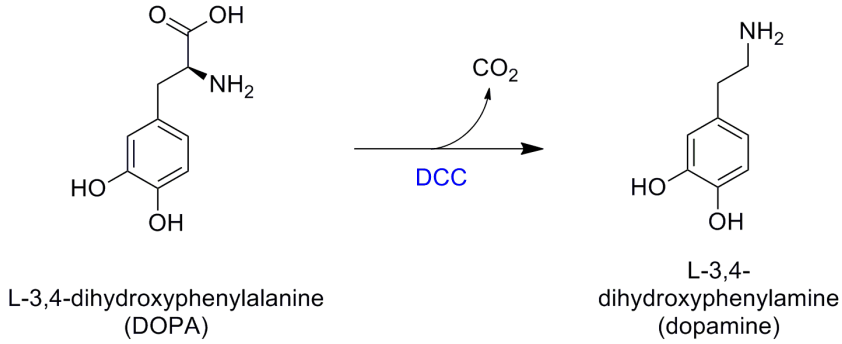


Figure 1.3: DDC converts L-3,4-dihydroxyphenylalanine (DOPA) into the neurotransmitter L-3,4-dihydroxyphenylamine (dopamine). (29).

The active site is located in the centre of DDC, within the large subunit and protected from solvent while the enzyme is in a closed state. The latter is obtained by significant conformational changes of the asymmetric monomers on binding of the cofactor, making dimerisation essential to DDC activity (30). Analysis of substrate binding as well as sequence conservation and site-directed mutagenesis experiments, have given an important insight into decarboxylation by PLP dependent enzymes of the α–family. The catalytic lysine is highly conserved, even in more distantly related animal and plant amino acid decarboxylases (31). The recognition of PLP and transaldimination induce a concerted loop movement and closure of the active site to

the surrounding solvent. The newly positioned loops bring amino acid residues involved in substrate binding into direct proximity of the active site (30). In the complex structure of DDC and carbidopa, the bound inhibitor is found in a plane orthogonal to the hydrazone bond and the cofactor once the external aldimine has fully formed. This corresponds to the proposed arrangement of PLP and its imine bond with the native DOPA substrate. In addition, the PLP pyridine ring's nitrogen forms a salt bridge with the carboxylate moiety of an aspartate side chain (Asp 271) (29,31). Both features are thought to enhance the aldimine's function as an electron sink and stabilise the decarboxylated quinonoid intermediate through extended resonance (23,29). For catalysis to be completed, the latter then has to be protonated. In the structure of human DDC, Tyr 332 was found to be in a suitable position for this reaction and was suggested to be a potential proton donor. Eventually, Lys 303 is thought to mediate release of the amine product through nucleophilic attack on the external aldimine and transaldimination with PLP (31).

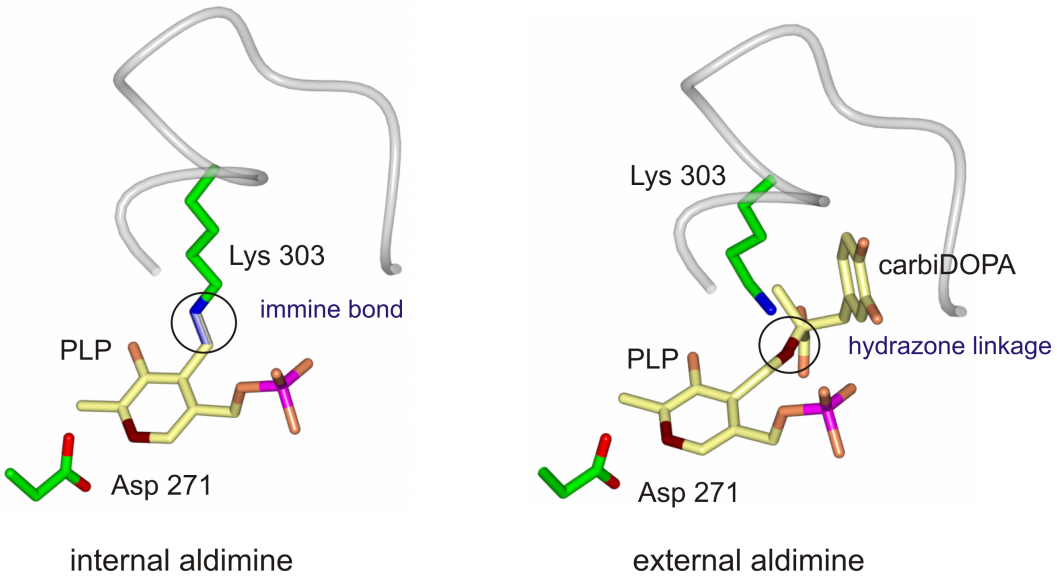


Figure 1.4: The internal and external aldimines formed in DOPA decarboxylase between Lys 303 and PLP or Lys 303 and the DOPA mimic carbidopa respectively. The catalytic lysine and its imine bond with PLP lie in a plane orthogonal to the cofactor's pyridine ring. The same arrangement is observed between PLP and the bound substrate, with the imine bond being represented by a hydrazone linkage in the carbidopa inhibitor.

The structural and mechanistic knowledge about PLP dependent decarboxylases is essential in the design of drugs against a variety of diseases. Abundance and activity

of DDC is strongly correlated to Parkinson's disease (PD) (32). Another example is African sleeping sickness which is caused by the parasite *Trypanosomas brucei*. It is widely treated using α–difluoromethylornithine, an inhibitor of the parasite's PLP dependent ornithine decarboxylase, which disrupts cellular polyamine synthesis (33).

1.3.3.2 ThDP dependent decarboxylases

The use of thiamine diphosphate (ThDP), formerly known as thiamine pyrophosphate (TPP), as a cofactor, has been established for a large number of enzymes. These include dehydrogenases, transketolases and oxidoreductases as well as a variety of decarboxylases. The latter make up the largest family of ThDP dependent enzymes and accept a variety of 2-keto acid substrates which lack an inherent electron sink to facilitate decarboxylation (34,35). The cofactor is derived from Vitamin B1 by phosphorylation and available to prokaryotic and eukaryotic enzyme classes. It is made up of two heterocycles, a thiazolium and a 4'- aminopyrimidine ring, connected through a methyl "bridge" (see Fig. 1.5) (35,36).

Analogous to the PLP dependent enzymes, the initial steps in ThDP binding are identical for the different enzyme classes. Unlike PLP however, ThDP is not covalently attached to the enzyme. The $GDGX_{26}N(C)N$ motif is highly conserved and suggested to be essential for binding of a divalent metal such as Mg^{2+}. This in turn coordinates the cofactor, with the assistance of a variety of active site residues (36). For catalysis to start, the cofactor needs to be bound within the active site in a V-shaped conformation. The methylene group is held in place through hydrophobic interactions while the pyrimidine ring's nitrogens form stabilising hydrogen bonds to the surrounding active site residues (36,37). Whereas the cofactor usually exists in a tautomeric equilibrium between its 4'-aminopyrimidine and the 1,4-iminopyrimidine forms, the latter is thought to be favoured by hydrogen bonding with a basic active site residue. The V-shaped conformation positions the pyrimidine ring's 4' NH-group close enough to abstract a proton from the thiazolium ring's C2 carbon. This leads to formation of the C2 carbanion and the so-called ylide, which is the dipolar active form of the cofactor (see Fig. 1.6) (36-38).

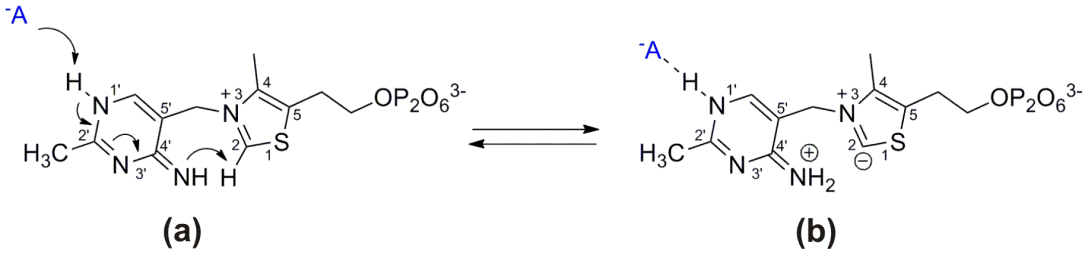


Figure 1.5: If ThDP is stabilised in its 1,4-iminopyrimidine form (a), activation of the cofactor can occur in response to deprotonation of the thiazole ring's C2 carbon by the pyrimidine ring's 4'-NH group. This leads to formation of a carbanion and the dipolar cofactor ylide (b) [adapted from (37,39)].

As substrate enters the active site, its carbonyl group will be attacked by the carbanion (acting as a nucleophile). A covalent bond is formed between them, which results in a first tetrahedral intermediate. The cofactor's thiazolium ring, in particular the carbon-nitrogen double bond, can now act as an electron sink during subsequent cleavage of the scissile bond. For ThDP dependent decarboxylases, the carboxylate moiety is oriented perpendicular to the cofactor, which ensures optimal overlap of orbitals. This overlap accommodates free electrons and facilitates their movement by closer proximity to neighbouring atoms, ensuring an optimal distribution of charges. The resulting high resonance stabilises the intermediate as the carbon-carbon bond is cleaved, analogous to the PLP dependent decarboxylases (35,37,40). In addition, the substrate binding pocket and cofactor are thought to be positioned at a sufficient distance in the protein to create a steric strain on the carboxylic acid and favour decarboxylation (38). The 2α carbanion-enamine intermediate formed after bond cleavage is also known as the activated aldehyde. It is common to all ThDP dependent enzymes, with reaction pathways splitting subsequent to this stage. Its protonation will form the third and last intermediate, 2α-hydroxyalkyl-ThDP. Deprotonation of this intermediate will lead to separation of the aldehyde leaving group in decarboxylases and regeneration of the cofactor ylide (see Fig. 1.6) (35,37). Alternatively, the carbanion intermediate can serve as a donor molecule in a condensation reaction, leading to formation of a new carbon-carbon bond. The latter option is particularly interesting for the commercial use of ThDP dependent enzymes. The electrophilic acceptor was found to bind predominantly in one of two active site pockets. These pockets are positioned in a distinct spatial orientation with

respect to the aldehyde donor and determine the stereochemical outcome of the reactions. As the recently discovered S-pocket is largely inaccessible in most ThDP dependent enzymes, the majority of reactions are R-selective (41). The understanding of structure and mechanisms of ThDP dependent decarboxylases now allows for the synthesis of complex, chiral organic compounds (36,41). The most prominent example is the synthesis of (*R*)-phenyl acetylcarbinol (*R*-PAC) using biotransformations. The molecule is a precursor of L-ephedrine which is used, among other applications, in the treatment of asthma (37,42,43).

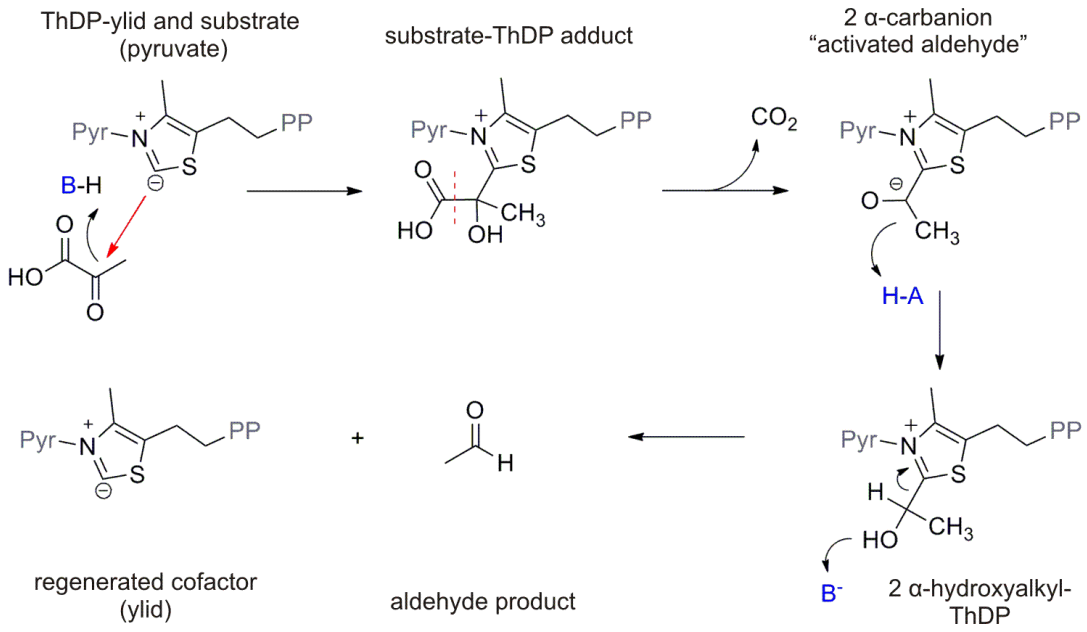


Figure 1.6: The conversion of pyruvate into acetaldehyde, at the expense of CO_2, is a model reaction for ThDP dependent decarboxylation. It involves three intermediate adducts, the sequential decarboxylation, protonation and deprotonation of which lead to release of the aldehyde product and regeneration of the cofactor. The second intermediate, the activated aldehyde, exists in a tautomeric equilibrium with its enamine form (not shown) (37-39).

The best characterised and most representative of the ThDP dependent decarboxylases are pyruvate decarboxylases (PDCs). PDCs are abundant in bacteria, yeast and plants and involved in alcohol fermentation processes. The enzyme catalyses the conversion of pyruvate to acetaldehyde at the expense of carbon dioxide (CO_2) (see Fig. 1.6). The aldehyde is then reduced to ethanol or serves as the electrophilic acceptor in a condensation side reaction with accumulated aldehyde product, leading to the production of acetoin. High resolution structures of apo-, holo

and ligand bound protein as well as kinetic studies on mutant PDC enzymes have provided detailed information about the general reaction mechanisms of ThDP dependent decarboxylases (see above) (35-37).

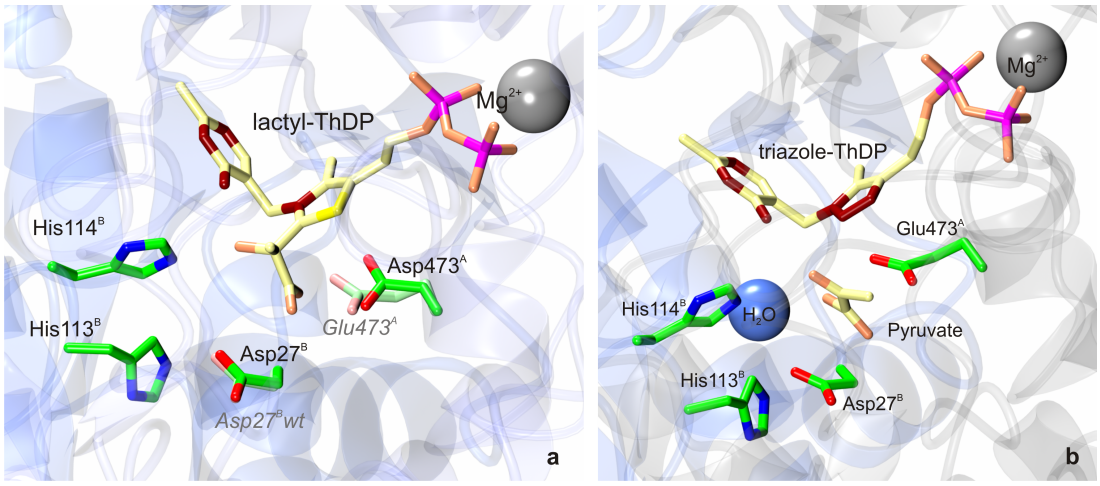


Figure 1.7: Various reaction intermediates have been trapped in the active site of a *Zymomonas mobilis* PDC using a weakly active mutant (Glu473Asp, lactyl adduct, a) and a triazole ThDP analogue, free pyruvate, b). They show ThDP in its V-shaped conformation, with the phosphate moiety coordinated by Mg^{2+}. The proposed catalytic acid-base residues $Asp27^{B}$ and $Glu473^{A}$ (A and B denoting different PDC monomers) are within suitable distance to act as proton donors and acceptors to the reaction intermediates. Image a shows a superposition of the wild type residues (pastel coloured) onto the mutant PDC structure (38,40,44).

Most PDC enzymes are made up of homodimers as their basic catalytic unit but associate into tetramers in solution. Each monomer is divided into three domains, all showing an open α/β fold. Upon entry into the active site, a flexible loop and adjacent alpha-helix move to bring catalytic residues into close proximity of ThDP. A glutamate residue is thought to deprotonate the pyrimidine ring's 4'NH group which in turn abstracts a proton from the thiazolium ring to create the catalytically active ylide (38). As pyruvate enters the active site, its carboxylate group is coordinated by two histidine residues, the so-called HH motif which is common to a number of ThDP dependent decarboxylases. It is furthermore held in place by several hydrogen bonds, involving the aforementioned glutamate (from both monomers), the histidines as well as an aspartate residue (37,38). As the ylide attacks

the pyruvate's keto group, a lactyl adduct forms and the ketone's oxygen is protonated (35). This was proposed to be catalysed by the thiazolium rings 4'NH_2 group.

If the previously protonated glutamate "returns" a proton to the cofactor's 4'NH group, the negatively charged residue could then destabilise the pyruvate's CO_2 moiety and enhance subsequent decarboxylation (38). The penultimate step, protonation of the C2α-carbanion/enamine intermediate, has been suggested to be catalysed by the same glutamate (or a glutamate-histidine-aspartate triad) - presumably *via* a catalytic water that has replaced the carboxylate leaving group. (37,38,40). Finally, the catalytic glutamate or an aspartate have been proposed to accept a proton from the cofactor's 4'NH_2 or the aforementioned water molecule. These could then act as bases and deprotonate the 2α-hydroxyethyl-ThDP, leading to release of the acetaldehyde substrate (38).

1.3.3.3 Metal dependent decarboxylases

Whereas many decarboxylases make use of complex organic molecules to fix their substrates and act as an electron sink during catalysis, others rely on small, inorganic cofactors such as metals. A number of enzymes use both for a joint structural and catalytic support. Decarboxylases relying on metal ions as their only cofactor may be non-oxidative, analogous to the PLP and ThDP dependent enzymes. Others however incorporate molecular oxygen into their substrates and thus catalyse oxidative decarboxylation reactions (45). Organic cofactor-dependent enzymes often make use of imine bonds and conjugated or thiazolium ring systems to relocate electrons and stabilise the carbanion intermediate. In contrast to that, metal dependent decarboxylases rely on polarizing ions to delocalize the charges within their substrates (46).

For decarboxylases employing metal ions and dioxygen as cofactors, the catalysed reactions can be oxidative or non-oxidative and always involve the formation of radical intermediate species. Oxidative reactions have been observed for a number of non-heme, iron-dependent enzymes catalysing the decarboxylation of α-ketoglutarate (45). The best studied example of a decarboxylase using molecular oxygen in

catalysis, however without inserting any oxygen into the product, is oxalate decarboxylase. The enzyme catalyses the conversion of oxalate into pyruvate and requires both Manganese and O_2 as cofactors for catalysis. The highly coordinated Manganese ion in the enzyme's active site may bind the incoming dioxygen as well as oxalate. An electron transfer from metal cofactor to dioxygen followed by a second transfer from oxalate to cofactor leads to formation of a superoxide metal species and a radical substrate intermediate respectively. Both are thought to stabilise the carbanion formed during subsequent decarboxylation. The product anion radical then receives an electron from the metal ion as well as being protonated by an active site residue, allowing it to dissociate from the cofactor (45,47-49) (see Fig. 1.8). Oxalate decarboxylase has a range of potential biotechnological applications in the breakdown of harmful oxalate accumulation, particularly in food plants. It can furthermore be used in diagnostics of hyperoxaluria, the increased levels of oxalate in urine samples (50).

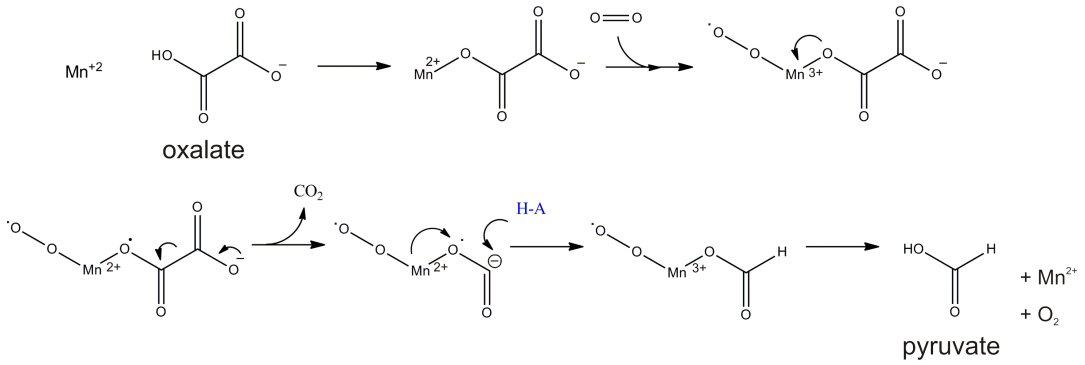


Figure 1.8: The decarboxylation of oxalate relies on the presence of a dioxygen molecule to create a metal superoxide. The latter can withdraw electrons from the bound substrate, creating a radical intermediate that is amenable to decarboxylation. As the single electron is returned from metal to product, the latter is released and the cofactor is regenerated [adapted from (45,47-49)].

A number of decarboxylases were found to use divalent metal ions as their only cofactor, without the need for dioxygen to relocate electrons during the reaction. The most abundant catalytic metals are Mn^{2+} and Mg^{2+}, although several enzymes were found to employ Zn^{2+} as a cofactor (45,51). The ions coordinate or even ligate to one or two of the substrate's oxygen atoms, especially of the carboxylate leaving group.

The proposed reaction mechanisms vary significantly but generally rely on the metals to act as a Lewis acid and accept free electrons during the decarboxylation reaction. They subsequently stabilise the negatively charged decarboxylation intermediates which are often enolates or dienolates. Protonation by an enzymatic acid base residue or an activated catalytic water then allows for product dissociation (45,51-55).

A detailed decarboxylation mechanism has been proposed for the *Rhizobium* 2,6-dihydroxybenzoate (γ-resorcylate) decarboxylase which relies on a Zn^{2+} atom as it's only cofactor. The enzyme decarboxylates 2,6- dihydroxybenzoate to 1,3-dihydroxybenzene (see Fig. 1.9) and its structure has been solved in complex with the substrate as well as with the substrate analogue and inhibitor 2,6-dihydroxybenzaldehyde.

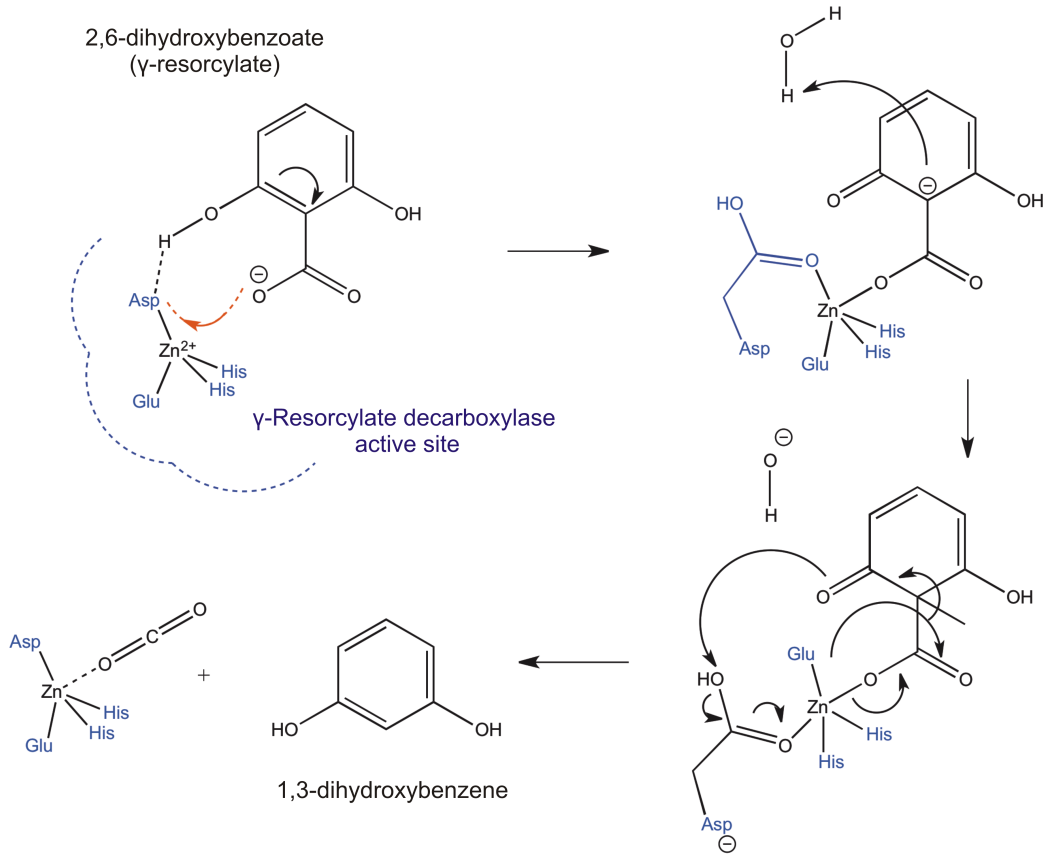


Figure 1.9: The decarboxylation of 2,6-dihydroxybenzoate proceeds *via* the formation of a carbanion intermediate and protonation by a catalytic water. The subsequent electron relay, mediated by the catalytic metal, leads to carbon-carbon bond cleavage and release of the carboxylate moiety (ligated to the metal cofactor) (adapted from (51)).

It was proposed that, upon substrate entry into the active site, the enzyme's Zn^{2+} cofactor comes into close proximity of the carboxylate's negatively charged oxygen atom whereas the benzene ring's 2-OH group forms a hydrogen bond with an aspartate residue. A charge relay (Fig. 1.9, red line) from the carboxyl oxygen to the aspartate, *via* the positively charged cofactor, ligates the substrate to Zn^{2+} and lowers the amino acid's pKa (an effect which is enhanced by the polarising effects of a His-Glu–Asp catalytic triad). The latter can subsequently deprotonate the dihydroxybenzoate's 2-OH group. A second charge relay leads to formation of a C1-carbanion intermediate, the negative charge of which is stabilised by the metal cofactor. An activated water molecule donates a proton to the carbanion after which coordinated protonation at C2 and decarboxylation release 1,3 dihydroxybenze and CO_2 from the active site (51).

1.3.3.4 Cofactor independent decarboxylases

A common feature of all enzymatic decarboxylation reactions is the need to destabilise the substrate in a way that favours removal of the carboxylate moiety. At the same time, an electron sink has to be provided which can accommodate the free electrons. Often negatively charged intermediates are formed in the process and have to be stabilised by the enzyme. Cofactor dependent enzymes can make use of complex organic molecules providing both electron sink and stabilising positive charges. Alternatively, metals are used to provide the stabilising charge, with delocalisation of electrons taking place within the substrate. The effects of both types of cofactors are enhanced by enzymes' surrounding active site residues. Cofactor independent decarboxylases on the other hand are unique in that they rely entirely on their primary and secondary structure, as well as the inherent properties of their substrates, to afford decarboxylation (45). Among other aspects, the absence of cofactors and the need for respective recycling systems makes this enzyme family particularly interesting for biocatalytic applications.

A limited number of cofactor independent decarboxylases have been characterised to date. These include the α–aryl malonate decarboxylase (AMDase), the malonate semialdehyde decarboxylase (MSAD), the 2-Oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline (OHCU), the orothidine 5'-monophosphate decarboxylase (OMPDC) decarboxylase and a variety of phenolic acid decarboxylases (PADs). The

first three enzymes listed catalyse decarboxylation reactions involving the formation of an enediolate, enolate or enol intermediates respectively. Decarboxylation of OMPDC leads to formation of a carbanion intermediate whereas phenolic acid substrates are thought to be protonated prior to decarboxylation to give quinone methide reaction intermediates (45,56-59). Most cofactor independent decarboxylases are thought to bind the scissile carboxylate within a predominantly hydrophobic pocket. The unfavourable interaction between non-polar amino acid residues and the deprotonated (negatively charged) moiety was proposed to afford substrate destabilization. The intermediates formed after decarboxylation are then stabilised by positively charged amino acid residues and/or delocalization of electrons, either within the substrate or with the help of an enzymatic oxyanion hole. Protonation or rearrangement of the stabilised intermediate eventually leads to product release (45,56-60).

A well characterised cofactor independent decarboxylase is the α–aryl malonate decarboxylase (AMDase) from *Bordetella bronchiseptica*. The enzyme converts α-aryl-malonates into (R)-α-aryl-carboxylates with high enantioselectivity (see Fig. 1.11). Its structure has been solved in the *apo*-state as well as in complex with an analogue of the proposed enediolate intermediate. AMDase has a molecular weight of approximately 24 kDa and exists as a monomer in solution (61). Two β-sheets enclose the enzyme's active site and are surrounded on all four sides by α-helices and flexible loops. When the enzyme's structure was solved, two potential oxyanion holes were detected in the active site (the "dioxyanion hole"). Their residues, in particular the backbone amides, form an extensive hydrogen bonding network around the two oxygen atoms of a benzylphosphate molecule which resembles the enediolate. This way the negatively charged intermediate is stabilised, the reaction's activation energy is lowered and decarboxylation facilitated. The substrate's aromatic ring is coordinated through stacking interactions with a glycine pair's amide bond and a proline residue, bringing it into a coplanar conformation with the enediolate moiety. The carboxylate leaving group is found within a small hydrophobic pocket (see Fig. 1.10) (56).
From the structures solved, Okrasa *et al.* have proposed a mechanism for AMDase catalysed decarboxylation. As substrate enters the active site, its two carboxylate moieties and the aromatic ring are bound and coordinated as described above. The

carboxylate group positioned in the small hydrophobic pocket is in its deprotonated state and hence negatively charged. The hydrophobic surrounding destabilises the group, leading to cleavage of the carbon-carbon bond. Once cleaved, the carbon dioxide is neutral and stable within the active site pocket. The free electron pair "released" upon decarboxylation relocates and leads to formation of an enediolate intermediate. Its negative charge is stabilised by resonance within the molecule's coplanar aromatic ring as well as by delocalisation into the adjacent enzymatic oxyanion hole.

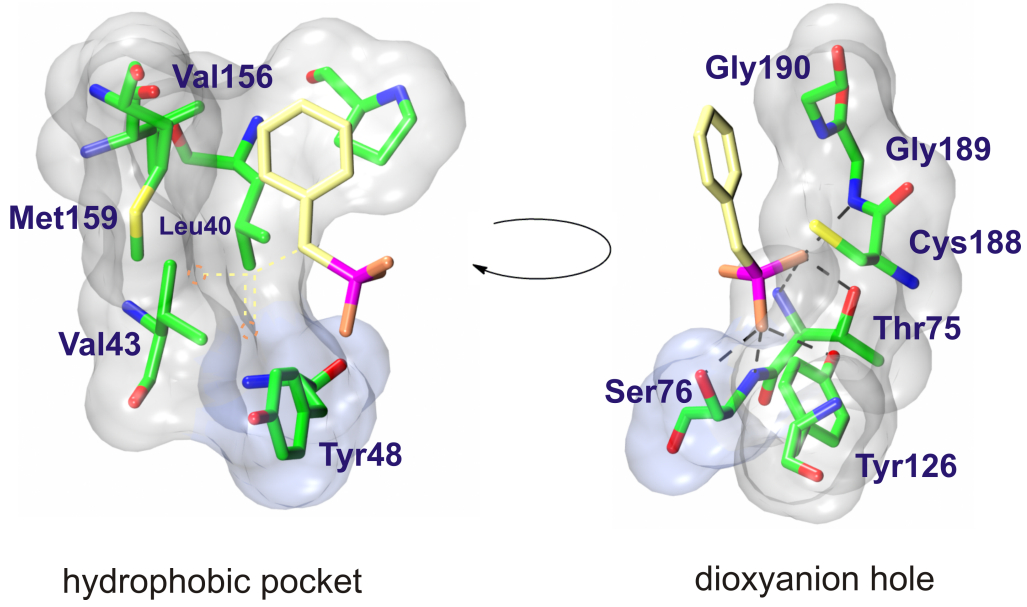


Figure 1.10: The AMDase active site possesses a hydrophobic pocket and a dioxyanion hole. Whereas the former loosely coordinates and destabilises the carboxylate leaving group (dotted line, absent in complex with benzylphosphate analogue), the latter stabilises the enediolate intermediate (61) [adapted from (56,62)].

Eventually, for the product to be released, the Cα carbon can be protonated by a catalytic cysteine in the active site. Its position favours protonation on the enediolate's *si*-face and yields exclusively the (*R*)-enantiomer of α-aryl-carboxylate product. Mutagenesis experiments in which the catalytic cysteine has been introduced or repositioned to the opposite site of the large binding pocket led to formation of racemic product or complete inversion of the enzyme's stereospecificity respectively (56,62-64). The enzymatic decarboxylation of α–arylmalonates can thus lead to the synthesis of a variety of α–arylcarboxylate products. It involves the introduction of a stereocenter and yields products of particularly high enantiomeric

excess compared to abiotic synthesis, making the process highly desirable for downstream synthetic applications (62).

AMDase serves as a good model for mechanisms in cofactor independent decarboxylases. Although exact modes of leaving group destabilisation and delocalisation of electrons in reaction intermediates vary, the enzyme's structural features and proposed mechanism have much in common with other members of the family. The key reaction stages are shared and highlight how the elaborate interplay between substrates and catalysts circumvent the need for external cofactors.

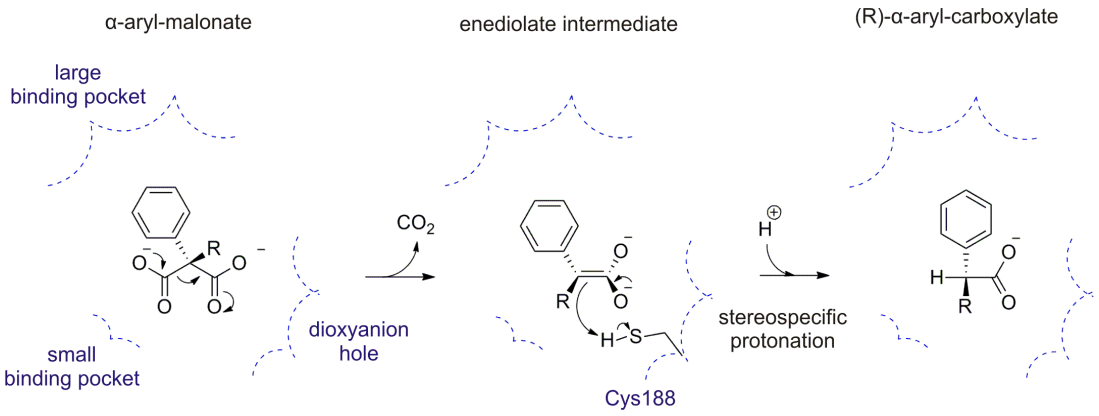


Figure 1.11: The decarboxylation of α–aryl malonate by AMDase proceeds *via* the formation of an enediolate intermediate and stereospecific protonation yields the carboxylate product at high enantiomeric excess. Substrate and reaction specificity are defined by the enzyme's binding pockets, formation of the unstable intermediate is enabled by delocalisation of electrons into the oxyanion hole (adapted from (56,62)).

Outliers in this family are the phenolic acid decarboxylases (PADs). Their catalytic mechanisms do not appear to involve charged intermediates and the leaving carboxylate is fixed in a polar pocket rather than destabilised (60,65). Yet, they catalyse comparable decarboxylation reactions, without the need for an exogenous metal or organic cofactor. Their unique structure and mechanisms will be discussed in chapters three and four.

Cofactor	Stabilisation of intermediate (electron sink)
Organic	Resonance within cofactor, especially via imine bonds
Metal/ O_2	Radical species, metal superoxide
Metal	Positive charge of metal, resonance within substrate
None	Resonance within substrate, neutral intermediates (PADs)

Table 1.1: All decarboxylases face the formation of unstable reaction intermediates as catalytic bottlenecks and have developed various mechanisms to lower the free energy of such transition states. The cofactor independent decarboxylases are unique in not requiring a second, exogenous catalyst in this process, which makes them particularly interesting targets for a wide range of biocatalytic applications.

1.4 Carbon – carbon bond hydrolases

Hydrolases acting on C-C bonds belong to enzyme commission (EC) subgroup 3.7.1 and accept a variety of diketones as substrates. They mediate reactions using a catalytic water molecule and are, in most cases, independent of complex organic cofactors. The cleaved substrate's leaving group is added onto the water hydroxyl group to give a carboxylic acid in addition to a methylketone product (see Fig. 1.12).

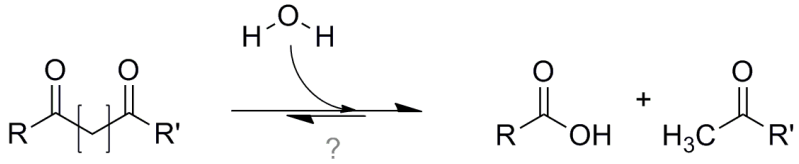


Figure 1.12: Carbon-carbon hydrolases may cleave diketone substrates to give a carboxylic acid as well as a methylketone product. As for the corresponding lyases, reaction equilibria strongly favour C-C bond cleavage over formation [adapted from (66)].

Various hydrolytic mechanisms exist that have been suggested to share a number of common features (67). In many mechanistic aspects, carbon-carbon hydrolases resemble the previously described lyases. The substrates may be subjected to a nucleophilic attack, either by an enzymatic residue or an activated water molecule. This requires the molecule to provide an electron sink, in other words an

electrophilic functional group which readily takes up a free electron pair. This sink is either inherent or can, in some cases, be created during the reaction by an enol-keto tautomerisation (68,69). After the nucleophilic attack, an oxyanion is thought to form which is a tetrahedral or enolate intermediate, and has to be stabilised, often but not always, by way of an enzymatic oxyanion hole (66,70). As described in section 1.3.3.4, this "hole" structure provides hydrogen bonding between backbone amides or side chains and the reaction intermediate – an interaction which would be exclusive for this stage and not formed with the final product, thus lowering the overall activation energy for the reaction (67). Protonation of the leaving group often leads to cleavage of the carbon-carbon bond. For certain groups of hydrolases this is then thought to involve the formation of an acyl-enzyme intermediate and a second tetrahedral intermediate, which is released as the enzyme partner is re-protonated – analogous to the peptide cleavage by serine proteases (67). In whichever way the reaction is catalysed however, this C-C bond cleavage differs from those of hydrolytic reactions on heteroatomic carbon bonds, which are more polar and thus more reactive.

Equilibria for the reversal of carbon-carbon hydrolysis are extremely unfavourable. The reaction can only be achieved by depriving the enzyme of catalytic water molecules, as is possible in organic solvents (71). A number of enzymes of biocatalytic importance are known to be active in a variety of organic solvents, among these also carbon-carbon bond hydrolases. Of the latter, two β-diketone hydrolases, for example, have recently been shown to retain their full or partial activity, even at high solvent concentrations (72). A reversal of the C-C cleavage reaction into carbon-carbon bond formation has been observed for numerous lyases but only for a single hydrolase and with very low yields to date – it remains to be determined now if other hydrolases can catalyse the formation of carbon-carbon bonds in a similar manner (3,71). The biocatalytic potential for such reactions would be considerable as C-C bond hydrolases can act on a variety of substrates, including ß-diketo acids and neutral ß-diketones, esters and thioesters.

1.4.1 Serine hydrolases

Structurally, many C-C bond hydrolases belong to the superfamily of α/ß hydrolases that consist of a central, multistranded ß-sheet surrounded by α-helices. Three loops,

at C-terminal ends of these helices, each carry a residue of an active site catalytic triad (73). Typically these residues are a nucleophile, a histidine and an acid. Identity, function and requirement, especially for the nucleophile, vary within the α/ß hydrolase family (74). This is consistent with the large variety of reactions catalysed by members of this superfamily, despite the strong similarities in the overall fold. They range from hydrolysis of carbon-carbon and carbon-heteroatom bonds to lipase, dehalogenase and esterase activities (73).

All α/ß hydrolases known to cleave carbon-carbon bonds possess a highly conserved serine, histidine and aspartate as part of their catalytic triad. They share a common catalytic motif (Gly-X-Ser-X-Gly), centred around the nucleophile which is also known as a "lipase box" (73). The particular composition of the catalytic triad is reminiscent of the well characterised serine proteases which are also members of this structural superfamily. Their reaction mechanism has long served as a model for serine dependent carbon-carbon hydrolysis (see Fig. 1.13).

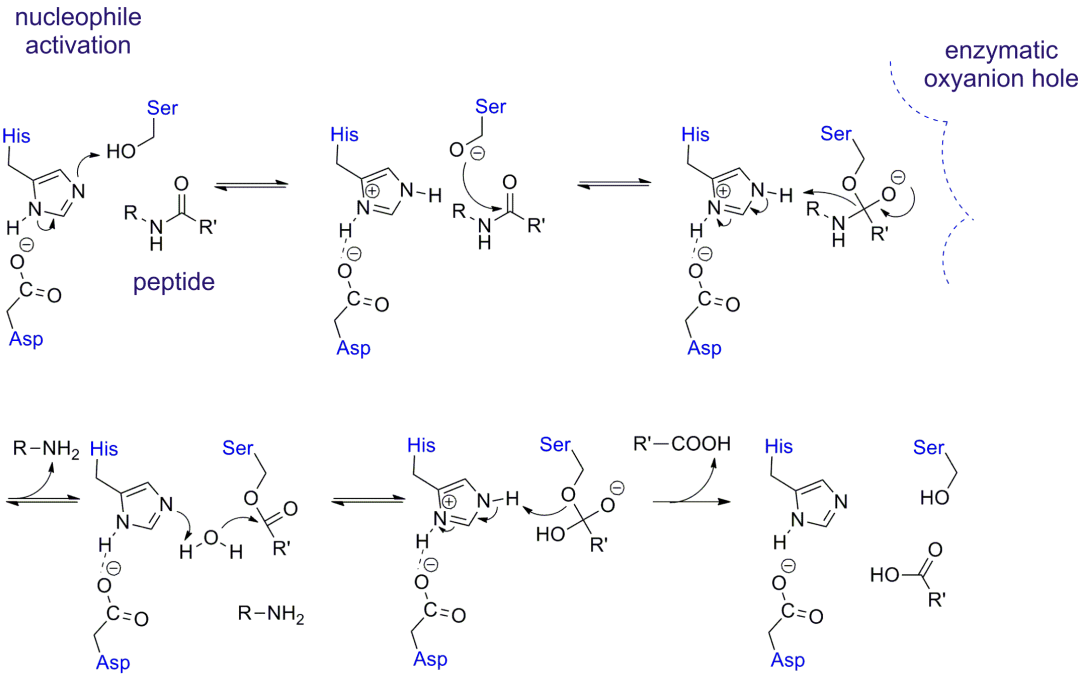


Figure 1.13: The cleavage of peptide bonds by trypsin proteases serves as a model reaction for serine dependent hydrolases [adapted from (75,76)].

The nucleophile in these enzymes is deprotonated by histidine, creating an imidazolium cation side chain which can in turn be stabilised by the triad's aspartate. As a peptide is recognised and bound by the enzyme, the deprotonated serine carries

out a nucleophilic attack on the substrate's carbonyl, leading to formation of a first tetrahedral intermediate which is stabilised within the enzyme's oxyanion hole. The catalytic histidine (the most conserved residue in all characterised α/ß hydrolases) then fulfils a triple function: protonation of the negatively charged intermediate allows for bond cleavage and separation of the leaving group. The residue then activates a water molecule which attacks the acyl-enzyme intermediate to give a second tetrahedral intermediate. For the product to be released, this then needs to be protonated, catalysed again by the same histidine side chain. This last reaction regenerates the triad's amino acid side chains for a new round of catalysis (76).

1.4.1.1 Serine MCP hydrolases

A large number of bacterial serine dependent carbon-carbon hydrolases are involved in the breakdown of aromatic compounds. This process constitutes part of the so-called *meta*-cleavage pathway and the enzymes are known as MCP hydrolases accordingly. Natural and non-natural aromatic compounds (often substituted with a variety of functional groups) are abundant but often harmful to the environment. Due to the ring structure's resonance the molecules are very stable and their degradation challenging to microorganisms (77). To be accessible substrates for hydrolysis, the aromatic substrates have to be hydroxylated and cleaved by the upstream action of dioxygenases or monooxygenases (in the case of phenols). This cleavage occurs either between (intradiol) or adjacent (extradiol) to the two hydroxyl groups and gives *ortho* and *meta* cleavage products respectively. These can then be further degraded by serine dependent hydrolysis (77-79).

The best characterised MCP hydrolases are the 2-hydroxy-6-keto-nona-1,9-dienoic acid 5,6 hydrolase (MhpC) and the 2-hydroxy-6-keto-6-phenylhexa-2,4-dienoic acid hydrolase (BphD). Both act on 2-hydroxy-6-oxo- dienoic acid (HODA) substrates to give two carboxylic acid products: 2-hydroxypenta-2,4-dienoic acid (both MhpC and BphD) as well as succinic and benzoic acid (MhpC and BphD respectively) (see Fig. 1.14).

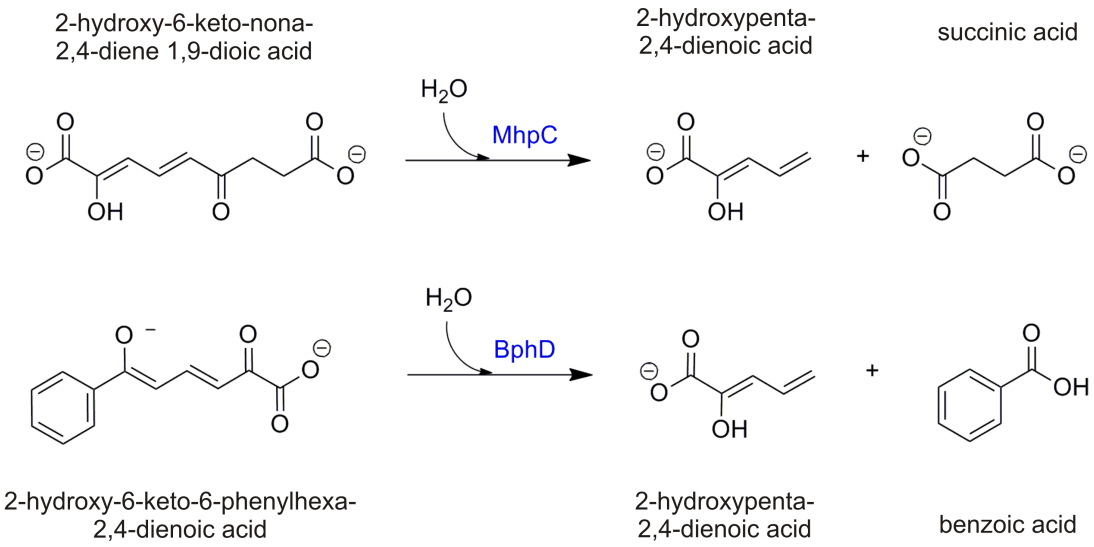


Figure 1.14: MhpC and BphD catalyse the hydrolytic cleavage of HODA ring fission products into two carboxyalic acid products.

The structures of both enzymes have been solved and resemble each other closely, with the exception of their quaternary structure (MhpC is a dimer, BphD a tetramer in solution). They show the typical α/β–hydrolase fold of twisted beta sheets and a distinct α–helical lid domain. Position and conformation of the catalytic triad is nearly identical in both enzymes, with each amino acid typically positioned on a flexible loop between β–sheets and interconnecting α-helices (see Fig. 1.15). The serine residue is positioned centrally on a so-called "nucleophile elbow" which is the most highly conserved structural feature of the α/β–hydrolases (see Fig. 1.15) (73).
In 1997, Henderson *et al.* and Lam *et al.* identified the main catalytic steps in MCP hydrolysis using MhpC as a model system. In a set of experiments carried out at low (sub-optimal) pH to slow down catalysis and observe pre steady-state kinetics, MhpC was found to catalyse hydrolysis across the substrate's C5-C6 bond. This reaction is thought to be a two step process: the substrate could enter the active site in its thermodynamically more stable dienol form and be protonated by an active site acid/base residue to give the keto tautomer. The keto group would then be able to act as an electron sink in the subsequent carbon–carbon bond cleavage reaction. The presence of a keto intermediate is supported by the observation that no additional UV signals seem to be observed, indicating a reduction in conjugation within the molecule (69,80). The authors proposed that the originally unfavourable ketonisation is made possible by binding of the (very stable) dienol substrate with its

carbonyl in a non-planar and thus destabilising position with respect to the conjugated dienol (69). This potential substrate destabilisation was later also observed for BphD (see Fig. 1.16) (68).

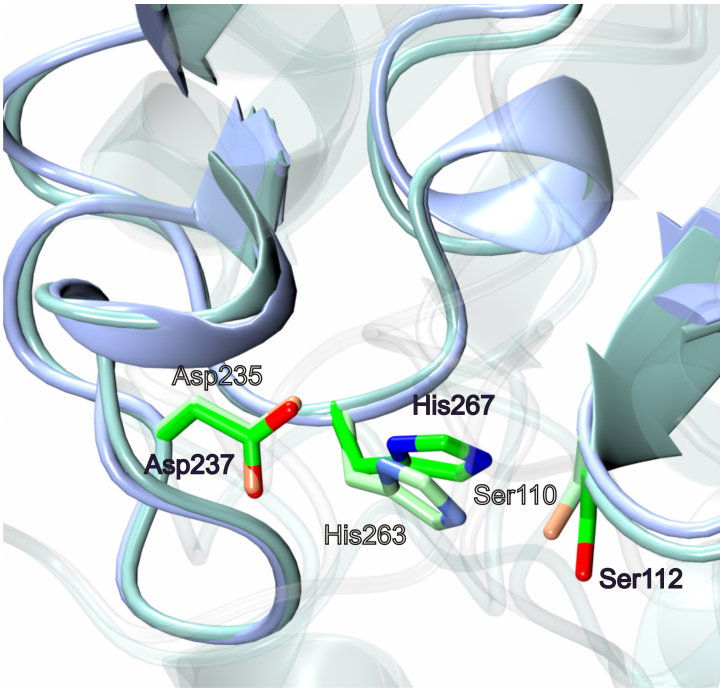


Figure 1.15: As Serine dependent α/β hydrolases, MhpC and BphD show a very similar and highly conserved fold and are thought to make use of the equally conserved catalytic triad Ser-Asp-His (bold and pale sidechains for MhpC and BphD respectively).

Following determination of these intermediate steps, it remained controversial if carbon-carbon bond cleavage is catalysed by a general base mechanism or by way of an acyl-enzyme intermediate. The former would involve activation of a water molecule (by the catalytic triad's serine) and its nucleophilic attack on the keto-form of the substrate to create a geminal diol intermediate. In the alternative proposal, the water is replaced as a nucleophile by the serine which would lead to formation of a covalent acyl-enzyme intermediate (81,82). The latter was found to be the case for BphD in 2012, when the first complex structure was solved which showed substrate covalently attached to the catalytic serine. Parallel quenching experiments supported the presence of the intermediate in solution and led to the authors to propose a detailed mechanism for covalent catalysis (see Fig. 1.16) (83).

As BphD substrate 2-hydroxy-6-keto-6-phenylhexa-2,4-dienoic acid (HOPDA) enters the active site, it is thought to be in its thermodynamically more favourable dienolate form. The molecule is bound in two distinct areas, the polar (P) and the non-polar (NP) subsite, recognising the dienolate moiety and aromatic ring respectively. HOPDA is thought to be bound by BphD in a non-planar and destabilising conformation, favouring the formation of a C5 carbanion diketo species. This carbanion is subsequently protonated by serine, leading to completion of both the keto-tautomerisation of the substrate and activation of the nucleophile. It also differentiates the proposed mechanism from the well characterised serine proteases which require activation of serine through the histidine – aspartate dyad. The activated serine can now carry out a nucleophilic attack on the C6 carbonyl to give the first tetrahedral, covalent acyl-enzyme intermediate. An electron relay through this intermediate leads to separation of the first product (2-hydroxypenta-2,4-dienoic acid, HPD) during which the previously formed keto group acts as an electron sink for carbon-carbon bond cleavage. The aforementioned aspartate-histidine dyad is now thought to be required for deprotonation and activation of a catalytic water. The latter can then attack the acyl-enzyme adduct to give a second tetrahedral intermediate. As the previously protonated histidine donates the hydrogen to the serine, hydrolysis is complete, the covalent bond is broken and benzoate released as the second reaction product. At the same time, the serine nucleophile is reprotonated and thus regenerated for the next round of catalysis. As for many carbon-carbon bond cleaving enzymes, all of the BphD reaction intermediates are stabilised within an enzymatic oxyanion hole, in this case the backbone amides of a methionine and a glycine residue (83).

BphD is the first MCP hydrolase for which a covalent acyl-enzyme intermediate has been observed in a crystal structure. It is unclear to date if catalysis by the closely related MhpC is identical as solvent exchange experiments still favour a geminal-diol intermediate and general base mechanism (81,83).

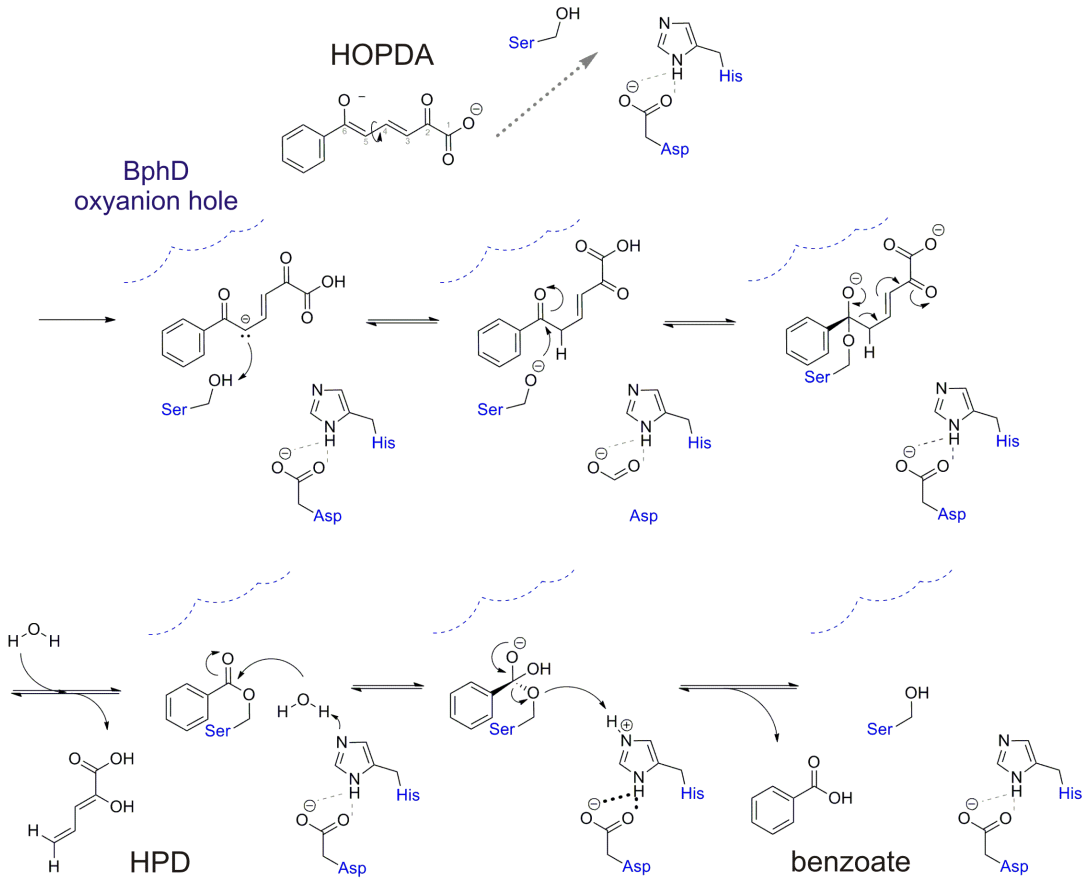


Figure 1.16: The first crystal structure of an MCP hydrolase (BphD) covalently bound to its substrate led to the detailed proposal of a mechanism involving an acyl-enzyme intermediate. Nucleophile activation occurs in a concerted reaction with the substrate's enol to keto tautomerisation. Two covalent intermediates are formed before both substrates, HPD and benzoate, are released and the active site is regenerated for the next round of catalysis.

1.4.1.2 Serine β-diketone/Friedel-Crafts hydrolases

Outside the bacterial breakdown of aromatic compounds described above, a large proportion of carbon-carbon hydrolases accept exclusively 1,3-β-diketones as substrates. They are accordingly known as the β-diketone hydrolase subfamily and include serine-dependent catalysts of the α/β–hydrolase family as well as serine-independent enzymes with alternative folds and mechanisms.

β–diketones are essential building blocks for a number of applications. In the chemical industry they are used as synthetic starting materials or ligands, in the pharmaceutical industry they are of interest for their potential antibacterial, antiviral and anti-carcinogenic properties (84,85). Aside from the presence or absence of an

enzymatic nucleophile in catalysis, β-diketone hydrolases are grouped according to their substrate preferences. β-diketo-acid hydrolases require a carboxylate moiety in their substrates whereas β-diketohydrolases accept "neutral" substituents (aliphatic or alicyclic) (66).

Within the family of β-diketone hydrolases, only few members depend on a catalytic serine. These include the Ayg1p hydrolase from the mould *Aspergillus fumigatus* as well as the 2,6-dihydroxy-*pseudo*-oxynicontine hydrolase (DHPONH) from *Arthrobacter nicotinovorans*. Both are members of the α/β-hydrolase family and posess a serine – histidine – aspartate catalytic triad. Whereas Ayg1p has been proposed to hydrolyse its substrate 1,3,6,8 - tetrahydroxynaphtalene by way of a covalent acyl-enzyme intermediate, analogous to BphD described above, DHPONH is thought to mediate the activation of a catalytic water (86,87). Both Ayg1p as well as DHPONH cleave carbon-carbon bonds to separate an acyl leaving group from an aromatic ring, analogous to the abiotic Friedel-Crafts reaction. They can thus also be grouped with the Friedel-Crafts hydrolases (see Fig. 1.17).
The DHPONH enzyme was first purified from a recombinant expression host and characterised in 2005. The dimeric enzyme was shown to cleave its substrate 2,6-dihydroxy-*pseudo*-oxynicontine into 2,6-dihydroxypyridine (DHP) and γ-N-methylaminobutyrate (MAB) and was accordingly thought to be the first hydrolase acting on a heteroaromatic substrate. The primary sequence, in particular the GXSXG motif, indicated the presence of a catalytic triad centred around a serine nucleophile in the enzyme. An *apo*-structure of DHPONH was subsequently determined and allowed for the proposal of two potential reaction mechanisms (see Fig. 1.17) (87,88).

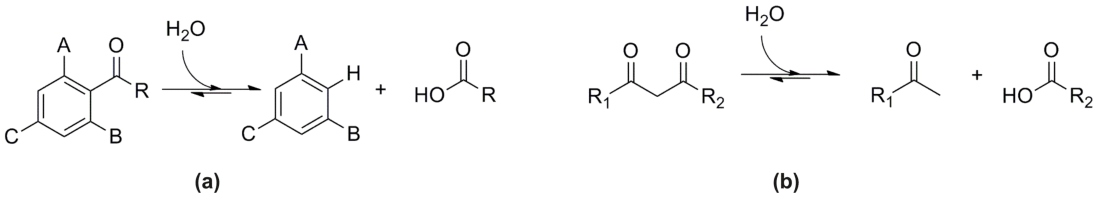


Figure 1.17: A number of hydrolases have been shown to catalyse the *retro*-Friedel-Crafts reaction and cleave acyl groups from aromatic rings (a). Depending on the ring's substituents, the substrate can be turned into a β-diketone and the enzyme can cleave a carbon-carbon bond to give both methylketone and carboxylic acid products (b).

Upon substrate entry into the active site, the O1 carbonyl is thought to be fixed within an enzymatic oxyanion hole, made up of glutamine backbone amides. To give the 1,3 diketone moiety required for nucleophilic attack, the O2 hydroxyl then has to be deprotonated which was proposed to be catalysed by a glutamate side chain, acting as a catalytic base. The same glutamate is thought to transfer the abstracted proton onto the C3 carbon, completing diketone formation. The protonation at C3 is facilitated by "force" applied onto the atom by the residues of the flexible N-terminal domain, favouring its sp^3 hybridisation. Pathways then split subsequent to this point. In the mechanism favoured by the authors, a water molecule is activated through deprotonation by the catalytic glutamate. Its hydroxyl can carry out a nucleophilic attack on the C1 keto group, leading to carbon-carbon hydrolysis and separation of the MAB product. The water's second hydrogen is required for reprotonation of O2, generating DHP with the aromatic ring as a good leaving group. The Ser-His-Asp triad is not directly involved in hydrolysis but essential for supporting the catalytic water in this hypothesis. In the alternative mechanism, the C1 carbon can be attacked by the serine nucleophile which is in turn activated by the His-Asp dyad. An acyl-enzyme intermediate is formed and subsequently hydrolysed by the catalytic water (activated as described above). The O2 oxygen in this mechanism may be reprotonated by the triad's histidine (87).

Confirmation as to which mechanism applies in serine dependent hydrolysis of DHPON now requires future experiments and ideally a complex structure with potential reaction intermediates.

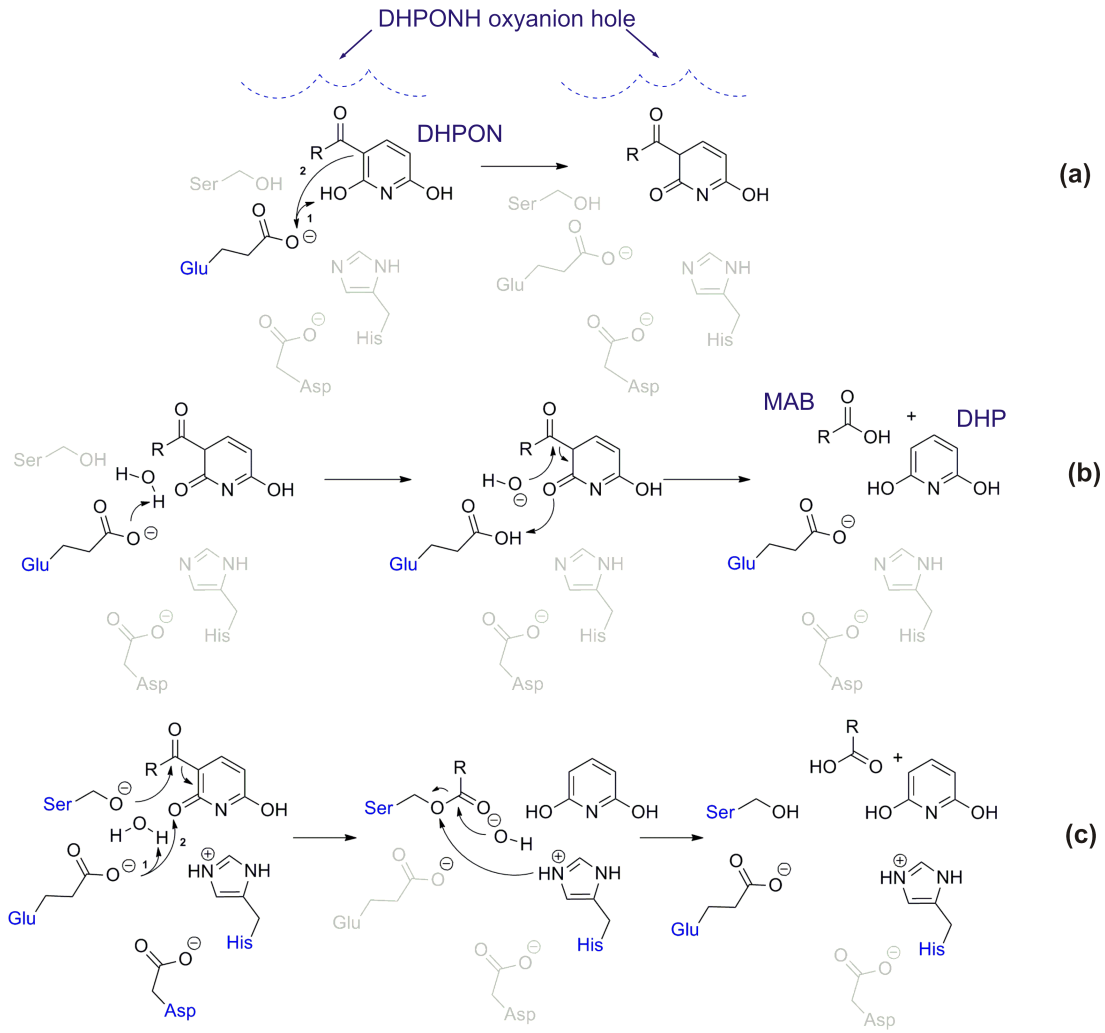


Figure 1.18: A mechanism has been proposed for the hydrolysis of DHPON into DHP and MAB. After the substrate is converted into a 1,3 diketone (a), the authors suggested hydrolysis to proceed through a general base catalysed mechanism and nucleophilic attack by an activated water (b). Alternatively, an acyl-enzyme intermediate could be formed and subsequently hydrolysed (c) [adapted from (87)].

Another important example of a serine dependent β-diketone hydrolase is the oxidised polyvinyl alcohol hydrolase (OPH) from *Pseudomonas sp.* VM15C which cleaves its β-diketone substrates into carboxylic acids and monoketones (89). Analysis of its structure, function and mechanism were part of this project and will be discussed in greater detail in chapter 6.

1.4.2 Non-serine hydrolases

A number of carbon-carbon hydrolases have developed mechanisms which are independent of a catalytic serine to act as a nucleophile or as a stabiliser thereof. Accordingly, these enzymes show alternative folds to the previously described α/β hydrolases and their reaction mechanisms differ. Whereas some of the non-serine hydrolases depend on a catalytic metal, others are completely independent of any exogenous cofactors.

1.4.2.1 Non-serine hydrolases (metal dependent)

Within the current EC classification for carbon-carbon hydrolases acting on ketonic substances, seven enzymes have been shown to bind divalent metals, predominantly Mg^{2+} but also Mn^{2+}, Ca^{2+} and Zn^{2+}. The best characterised members of this group are the fumarylacetoacetate hydrolase (FAH) and the Friedel-Crafts hydrolases diacetylphloroglucinol hydrolase (PhlG) and phloretin hydrolase (Phy).

The apo-structure of mouse FAH was solved in 1999 and constituted the first ever structure of a carbon-carbon bond cleaving hydrolase. It showed a novel α/β fold, different to the previously characterised α/β hydrolases by lacking a lipase box and flexible loops carrying the classical catalytic triad. The structure allowed for the proposition of a detailed mechanism for serine independent carbon-carbon hydrolysis relying on a catalytic metal (90). While still possessing a catalytic triad, including a glutamate and histidine acid/base pair, the serine nucleophile in FAH is thought to be replaced by a water molecule. The latter is positioned and coordinated by a glutamate and the catalytic metal - Mn^{2+}. As it enters the active site, the β–diketone substrate fumarylacetoacetate (and all subsequent negatively charged intermediates and products) are stabilised by the metal cation and coordinated within an unusual oxyanion hole. The latter is made up of side chains (rather than backbone amides) of a lysine, arginine and glutamine residue. After deprotonation and activation of the catalytic water by histidine, the hydroxide nucleophile attacks the C4 carbonyl and forms a tetrahedral oxyanion (alkoxide) intermediate. Rearrangement of free electrons leads to cleavage of the carbon-carbon bond, formation of the first product, fumarate, and an acetoacetyl carbanion which is stabilised by resonance. Proton

transfer to the latter from the fumarate's carboxylate or by the oxyanion hole's lysine gives the second product, acetoacetate (see Fig. 1.19). If mediated by the catalytic histidine, this transfer would also regenerate the active site's base for subsequent rounds of hydrolysis (90,91).

In addition to Mn^{2+}, which is essential for catalysis, a second metal (probably Na^{2+}) is thought to be required for FAH structural stability and has been proposed to be involved in intersubunit communication (92,93).

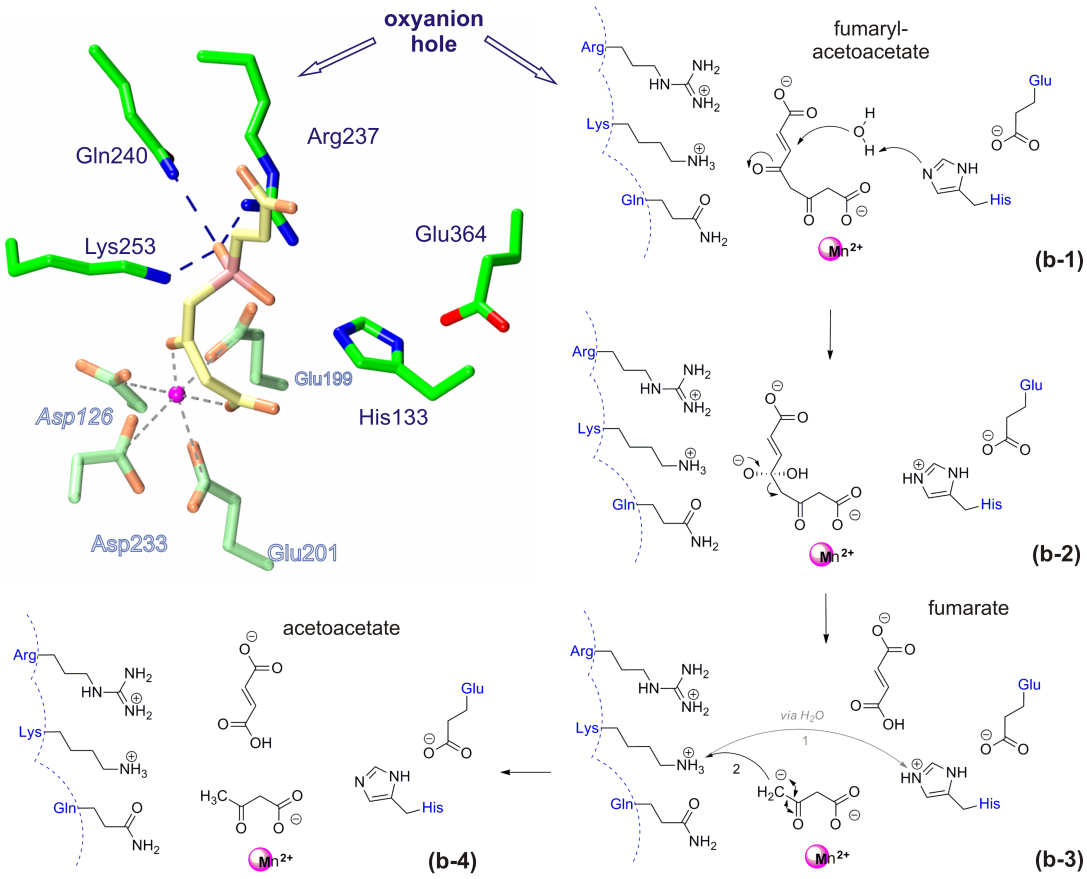


Figure 1.19: The structure of FAH was solved in complex with the substrate mimic 4-[(2-carboxyethyl)hydroxyphosphynil]-3-oxobutyric acid (CEHPOBA). The catalytic metal is highly coordinated within the active site and positions the catalytic water and substrate (grey dashed lines, image **a,** water not shown). During the reaction (**b1 – b4**), a lysine is shown to act as the final proton donor to the carbanion/enolate intermediate. Alternatively, the hydrogen could originate from protonated fumarate product [adapted from (90,91)].

Other examples of non-serine, metal dependent carbon-carbon hydrolysis can be found with the Friedel-Crafts hydrolases (which also include the serine dependent, metal independent Ayg1p and DHPONH, see above). The best characterised reaction

of this kind is the hydrolysis of diacetylphloroglucinol (DAPG) into monoacetylphloroglucinol (MAPG) and acetate as catalysed by the diacetylphloroglucinol hydrolase (PhlG). PhlG is part of a large operon in *Pseudomonas* species whose genes and respective protein products produce and transport DAPG as an important antimicrobial compound. The enzyme was first characterised by Bottiglieri *et al*. and found to be a 35 kDa carbon-carbon hydrolase which associates into a homodimer in solution. The purified enzyme showed very narrow substrate specificity and would not hydrolyse any natural DAPG derivatives such as mono- or triacetylphloroglucinol (TAPG) (94). In 2010, the 3-dimensional structure of PhlG was solved to 2 Å resolution and, despite the absence of complexed substrate or product, provided information about the strict substrate specificity and a potential reaction mechanism.

In accordance with the lack of sequence motifs common to α/β hydrolases, such as the lipase box, PhlG showed a novel Bet-v1-like fold, making it the first carbon-carbon hydrolase member of this structural superfamily. The structure revealed two features important to substrate specificity and reaction mechanism. Buried within the hypothesised C-terminal active site, the authors detected electron density into which only a transition metal could be fitted. The latter was identified as a divalent zinc atom which is highly coordinated by two histidine and two glutamate residues. The authors have proposed this Zn^{2+}, together with a single water molecule, to be essential in catalysis. In the suggested reaction mechanism, the metal is thought to lower the water's pk_a sufficiently to allow for the loss of a proton, turning it into a hydroxyl nucleophile. The latter could then attack the scissile carbon-carbon bond between the DAPG's aromatic ring and the acetyl moiety to form a tetrahedral intermediate. Rearrangement of free electrons and protonation of the formed acetic acid would complete the reaction (see Fig. 1.20). The metal-coordinating glutamate residues have been suggested to shuttle the proton abstracted from the catalytic water to the MAPG product (95).

Modelling of the entrance channel from protein surface to the proposed active site presented a dumbbell shaped cavity which could, allowing for minor conformational changes on binding, accommodate the DAPG substrate. The distribution of polar and aromatic residues within this channel was shown to reflect interactions with the substrate. These include π-stacking interactions between phenylalanine and tryptophan residues and the aromatic ring as well as hydrogen bond formation of

tyrosine residues and an asparagine with the DAPG hydroxyl groups and the scissile acetyl moiety respectively (see Fig. 1.21). Higher substituted phloroglucinol derivatives such as TAPG would not fit into this channel whereas others may not be positioned appropriately, explaining the narrow substrate spectrum of PhlG. Both substrate specificity and reaction mechanism have also been addressed by site-directed mutagenesis (SDM) experiments which supported the structural findings (95).

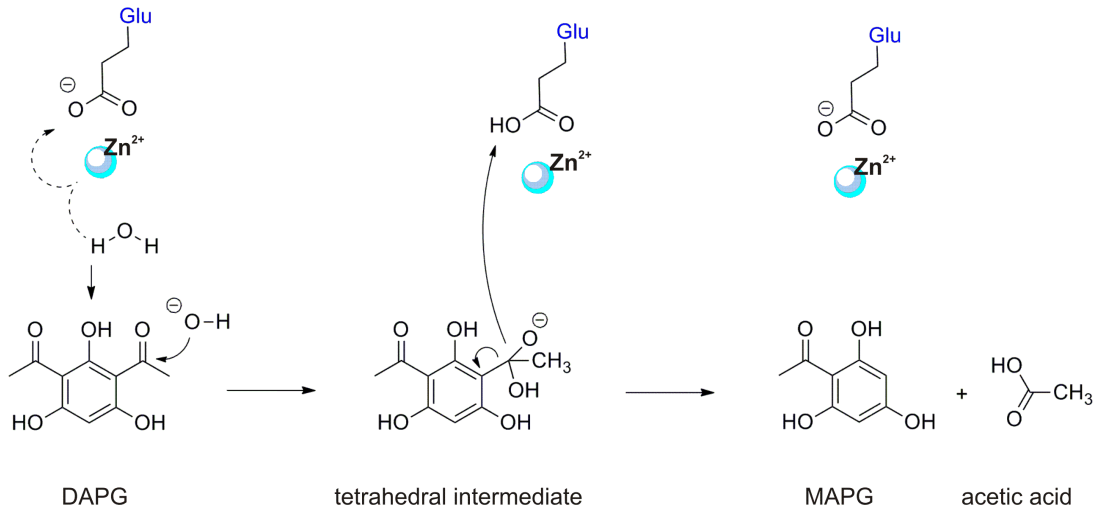


Figure 1.20: PhlG catalyses the hydrolysis of DAPG to give MAPG and acetic acid. The side chain shown could represent either of the two metal coordinating glutamates which are thought to help shuttle the proton separated from the activated water molecule (dashed line) to the acetate product [adapted from (95)].

PhlG represents a structurally novel carbon-carbon hydrolase and the first known to catalyse a serine-independent *retro*-Friedel-Crafts acylation using a catalytic transition metal. The enzyme shows low sequence similarity to a number of hypothetical and uncharacterised proteins as well as to a second metal-dependent Friedel-Crafts hydrolase. The latter, phloretin hydrolase (Phy), cleaves the plant derived flavonoid phloretin into phloroglucinol and phloretic acid (96). Its structure and mechanism were analysed as part of this project and will be discussed in greater detail in chapter 5.

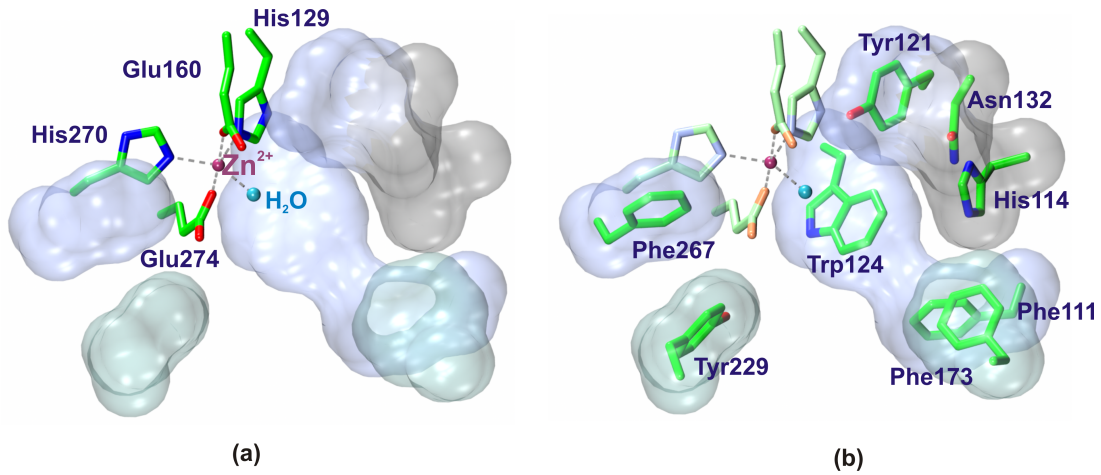


Figure 1.21: The PhlG active site is defined by the divalent zinc atom, penta-coordinated through glutamate and histidine sidechains as well as the catalytic water (a). The cavity is restrictive in size, contributing to the narrow substrate specificity. It is made up of aromatic residues thought to be involved in π-stacking interactions with the DAPG benzene ring (phenylalanines and tryptophan) as well as the hydrogen bond donors proposed to fix the substrate's hydroxyl and acetyl moieties (tyrosines, histidine and asparagine respectively) [adapted from (95)].

1.4.2.2 Non-serine hydrolases (cofactor independent)

In addition to the examples described above, a number of hydrolases have developed mechanisms of carbon-carbon bond cleavage completely independent of either serine nucleophile and/or exogenous cofactors like divalent metals.

The best characterised member of this family is the oxocamphor hydrolase (OCH) acting on β–diketone substrates. The enzyme hydrolyses its substrate 6-oxocamphor into (2*R*,4*S*)-α-campholinic acid and (2*S*,4*S*)-α-campholinic acid with a 6:1 excess of the former diastereoisomer. It thus catalyses the important but rare biocatalytic carbon-carbon bond cleavage of a prochiral substrate to give a diastereomeric product at high enantiomeric excess (*ee*) (see Fig 1.22) (97,98). Structurally, OCH is a member of the crotonase superfamily(99), named after the first characterised model enzyme Enoyl-CoA hydratase (crotonase). Members of this family show a common fold of two perpendicular β–sheets surrounded by α-helices, known as the ββα unit. Despite significant structural similarities, members of the crotonase superfamily

catalyse a diverse range of reactions on carbon-heteroatom as well as carbon-carbon bonds. The latter include decarboxylation, C-C bond formation and hydrolysis. Crotonases have thus attracted much attention as potential biocatalysts (70).

When OCH was isolated and characterised, the enzyme was found to be the first member of the crotonase family not to rely on its substrates' forming a thioester bond to a coenzyme A (CoA) molecule. Without the addition of exogenous cofactors, OCH hydrolysed a number of non-enolisable α,α disubstituted mono- and bicyclic β–diketone substrates into keto acids (97,98). The condensation of two esters to form a diketone is known as a Claisen condensation (or Dieckmann condensation if it occurs within a single molecule to give a cyclic product). Accordingly, hydrolysis by OCH is known as a *retro*-Claisen reaction.

OCH has a monomeric weight of 30 kD and associates into a dimer of trimers in solution which is a typical feature of many crotonase type enzymes. Each monomer has an individual active site, situated within a central hydrophobic pocket. Structures of wild type and mutant OCH as well as the *Anabaena β*–diketone hydrolase (ABDH) orthologue in their *apo*- and complexed states have provided an insight into hydrolysis of cyclic *β*–diketones by crotonase like enzymes (99-101).

Before hydrolysis can take place, the camphor has to be converted into the 6-oxocamphor through hydroxylation followed by oxidation at the 6-*endo* position. As this diketone enters the OCH active site, a histidine-aspartate dyad is thought to activate a catalytic water for nucleophilic attack at the (pro-*S*) side. This leads to formation of a tetrahedral oxyanion and defines the strict stereochemistry at the product's C4 carbon. The tetrahedral oxyanion rearranges into an enolate intermediate with simultaneous breaking of the C-C bond and ring opening. Protonation of the enolate intermediate at the product's C2 carbon, subsequent tautomerisation and relaxation of the ring structure lead to formation of the (2*R*,4*S*) and (2*S*,4*S*)-α-campholinic acid products at a 6:1 ratio. As an equivalent ratio is found in the abiotic *retro*-Claisen reaction of 6-oxocamphor, it has been proposed that protonation and tautomerisation are non-enzymatic and under thermodynamic control (98,99,101).

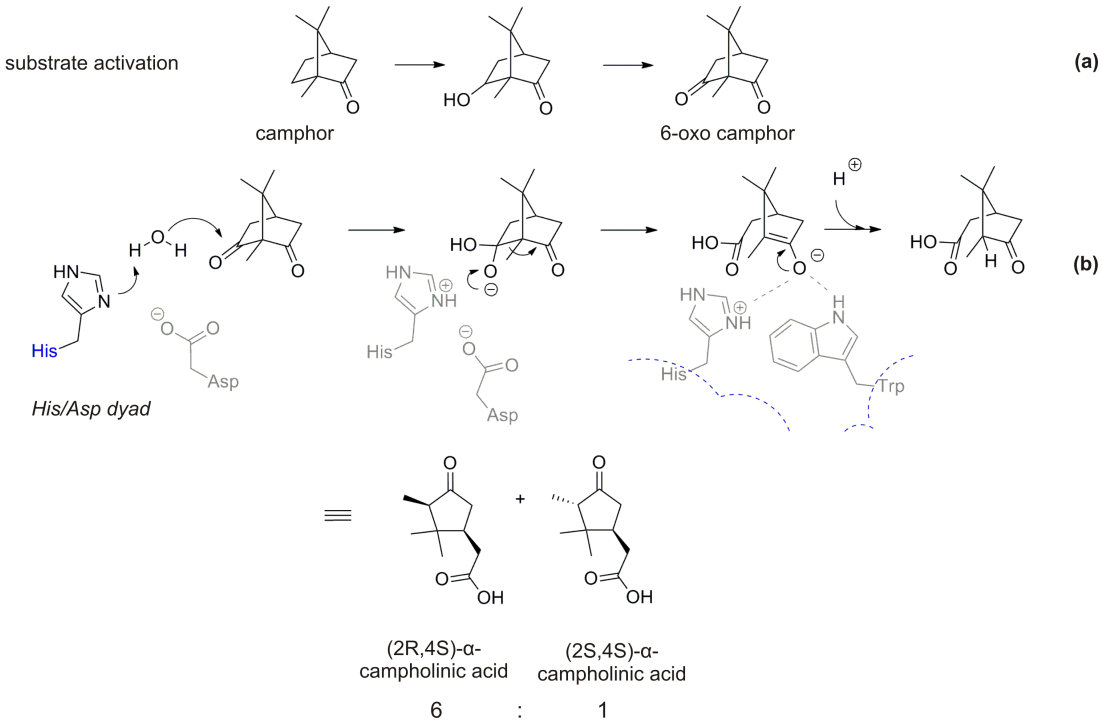


Figure 1.22: Hydrolysis of 6-oxocamphor (derived from camphor by hydroxylation and subsequent oxidation (a)) proceeds through attack of a nucleophilic water at the substrate's (pro-*S*) site. Non-enzymatic protonation at the C2 carbon gives the product with a 6:1 excess of the (2*R*,4*S*) over the (2*S*,4*S*) diastereomer (b) [adapted from (101)].

A number of features distinguish the described structures and mechanism from previously characterised crotonase-like enzymes, leading to the suggestion that coenzyme A independent members may form their own subfamily (101). In contrast to crotonase family members with coenzyme A-bound substrates, OCH and its ABDH orthologue have a strict requirement for non-enolizable substrates (97). Furthermore, neither enzyme possesses a classical oxyanion hole motif for stabilisation of negatively charged enolate intermediates. In the product complex structure of an OCH mutant enzyme, the enolate-derived carbonyl was shown to be in close proximity to a tryptophan as well as the catalytic histidine residue. The intermediate's negative charge may thus be stabilised by hydrogen bonding interactions with side chains of both residues rather than through the backbone amides of standard oxyanion holes (see Fig. 1.22 and Fig. 1.23). The unusual presence of two spatially separated intermediates may even require a second oxyanion hole equivalent. Asp154 has been proposed to fulfil this role by stabilising the tetrahedral intermediate prior to bond cleavage (101).

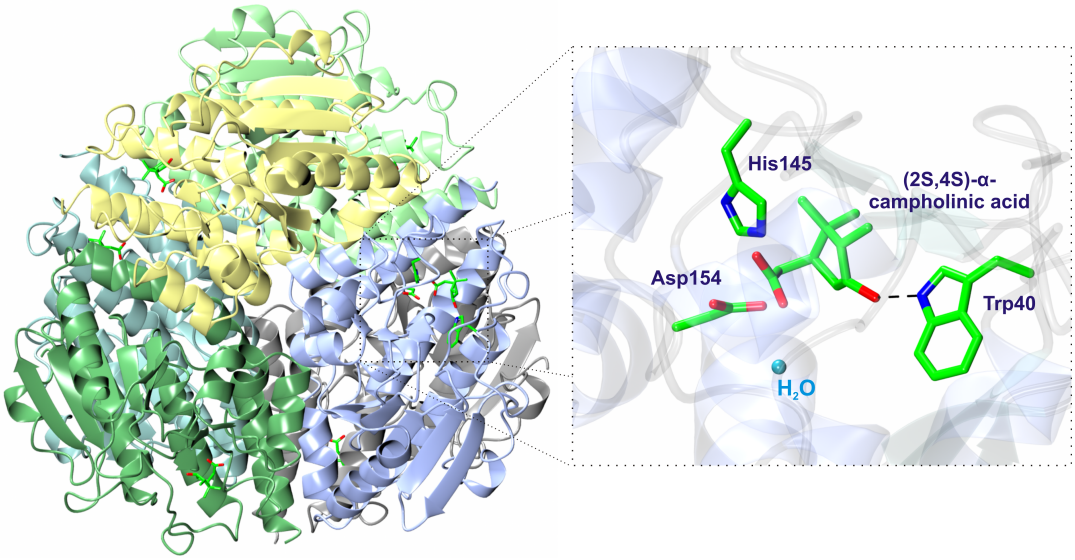


Figure 1.23: OCH, like many enzymes of the crotonase fold, associates into a dimer of trimers in solution. The active site (close up) is unusual in its lack of a defined oxyanion hole. In its place, the enolate intermediate is thought to be stabilised by hydrogen bonding to side chains of a histidine (not shown) and a tryptophan. Histidine 145, which is thought to act as a catalytic base, is positioned and stabilised by aspartate 154 [adapted from (101)].

Their ability to catalyse desymmetrisation reactions makes crotonase type hydrolases like OCH highly valuable for biocatalytic applications. They are furthermore interesting engineering targets, especially with a view to improve or invert their stereochemistry (97,100).

1.4.3 Carbon-carbon bond forming hydrolases

To date, only one C-C hydrolase is thought to have catalysed the reverse reaction, *i.e.* the formation rather than cleavage of carbon-carbon bonds. In experiments analysing the catalytic promiscuity of *E.coli* MhpC, Li *et al.* have observed the formation of a monoethyl ester product resulting from γ–acylation of a dienol substrate in organic solvents. Yields for this reaction were however very low (1.6 %) and insufficient for NMR analysis of the product (102).

1.5 Project aims

The enzymes described above are examples of the growing family of lyases and hydrolases acting on carbon-carbon bonds. They show a large variety in structures, mechanisms and cofactor requirement but share the ability to overcome unfavourable reaction energies and cleave inherently non-reactive bonds. Within this project and in collaboration with Professor Wolfgang Kroutil's group from the University of Graz, three carbon-carbon bond cleaving enzymes, a lyase and two hydrolases, were analysed with respect to their structure, mechanisms and substrate scope. None of these enzymes depend on complex organic cofactors and/or recycling systems thereof, making them attractive targets for biocatalytic applications. As a consequence, all work could be carried out as part of the Marie Curie Biotrains network and involved, in addition to the lyase and hydrolase work described, the structural analyses of biocatalytic enzymes from other network partners.

The first enzyme analysed was a phenolic acid decarboxylase from *Bacillus subtilis* (*Bs*PAD). This lyase is the member of a well-characterised family of enzymes which decarboxylate their substrates to give industrially useful styrene derivatives. Although *apo*-structures of *Bs*PAD and other related enzymes were available, their catalytic mechanism remained speculative in the absence of a ligand complex. The aim of this work was thus to prepare a mutant library, collect kinetic data and trap the enzyme's substrate in the active site of a weakly active mutant. In collaboration with the University of Graz, *Bs*PAD would furthermore be the candidate for a cascade catalysis with a fungal oxygenase from *Trametes hirsuta*. If used in a one pot reaction, this cascade is hoped to allow for the production of high value natural vanillin from plant derived renewable resources.

To analyse the enzyme catalysed hydrolysis of carbon-carbon bonds, two targets were chosen which were thought to differ in structure, function and mechanism. Phloretin hydrolase from *Eubacterium ramulus* (Phy), is involved in plant derived flavonoid hydrolysis in the human gut and catalyses a rare enzymatic *retro*-Friedel Crafts reaction. At the start of this work, the structure of Phy was unknown but thought to differ from previously characterised C-C hydrolases. Its mechanism was, accordingly, speculative. The aim of the project was thus to obtain information

on structure, mechanism and substrate scope of this unusual member of the C-C hydrolase family.

Oxidised polyvinyl alcohol hydrolase (OPH) from *Pseudomonas* sp. VM15C was previously shown to cleave β–diketone substrates and can be used for degradation of the latter in waste-water treatment. As previous attempts to purify the protein from a recombinant host were unsuccessful, primary aim of this work was to create a soluble expression construct, followed by determination of the enzyme's structure and mechanism.

Ultimately, if the reaction of either hydrolase could be reversed, this would create an important foundation for the enzymatic synthesis of carbon-carbon bonds in high value chemical building blocks.

Chapter 2: General Methods

The following chapter will give a brief theoretical background on selected materials and methods which are particular to this work. Detailed experimental setups and protocols will be given separately with the individual projects.

2. 1 Gene cloning, expression and protein purification

2.1.1 Ligation-independent cloning

All enzymes characterised as part of this project were purified to high homogeneity. Their respective genes thus required a selective expression system which would supply an affinity-tagged construct for inducible protein production and purification from a recombinant host. For this purpose, the YSBLIC3C vector was chosen into which all genes could be introduced by way of ligation independent cloning (LIC).
The LIC method was first developed in 1990 by Aslanidis and de Jong (103). Within the York Structural Biology Laboratory it was later optimised and adapted for its use with a derivatised Novagen pET-28a vector. This created the YSBLIC and YSBLIC-3C vectors which have since allowed for the standard and high throughput cloning and expression of a large number of hexahistidine tagged protein constructs (103,104). The method is based on the creation of complementary nucleotide overhangs in both vector and insert DNA which anneal spontaneously. This circumvents the need for a ligase reaction and, as it significantly reduces the tendency of vector DNA to self-anneal, no phosphatase treatment is required.
To prepare the respective overhangs, the pET-28a vector has been modified to introduce a LIC site between the multiple cloning site's NcoI and NdeI restriction sequences. This site consists of two 3'-5' nucleotide stretches deficient in deoxyadenosine triphosphate (dATP), a BseRI recognition sequence and a restriction site for the human rhinovirus (HRV) 3C protease (see Fig. 2.1). Upon BseRI digest, the vector is linearised such that the dATP deficient nucleotide stretch is exposed at each strand's 5' end. Complementary stretches, deficient in deoxythymidine triphosphate (dTTP) nucleotides, are attached to the 5' ends of insert DNA by a polymerase chain reaction (PCR) using specifically designed LIC primers.

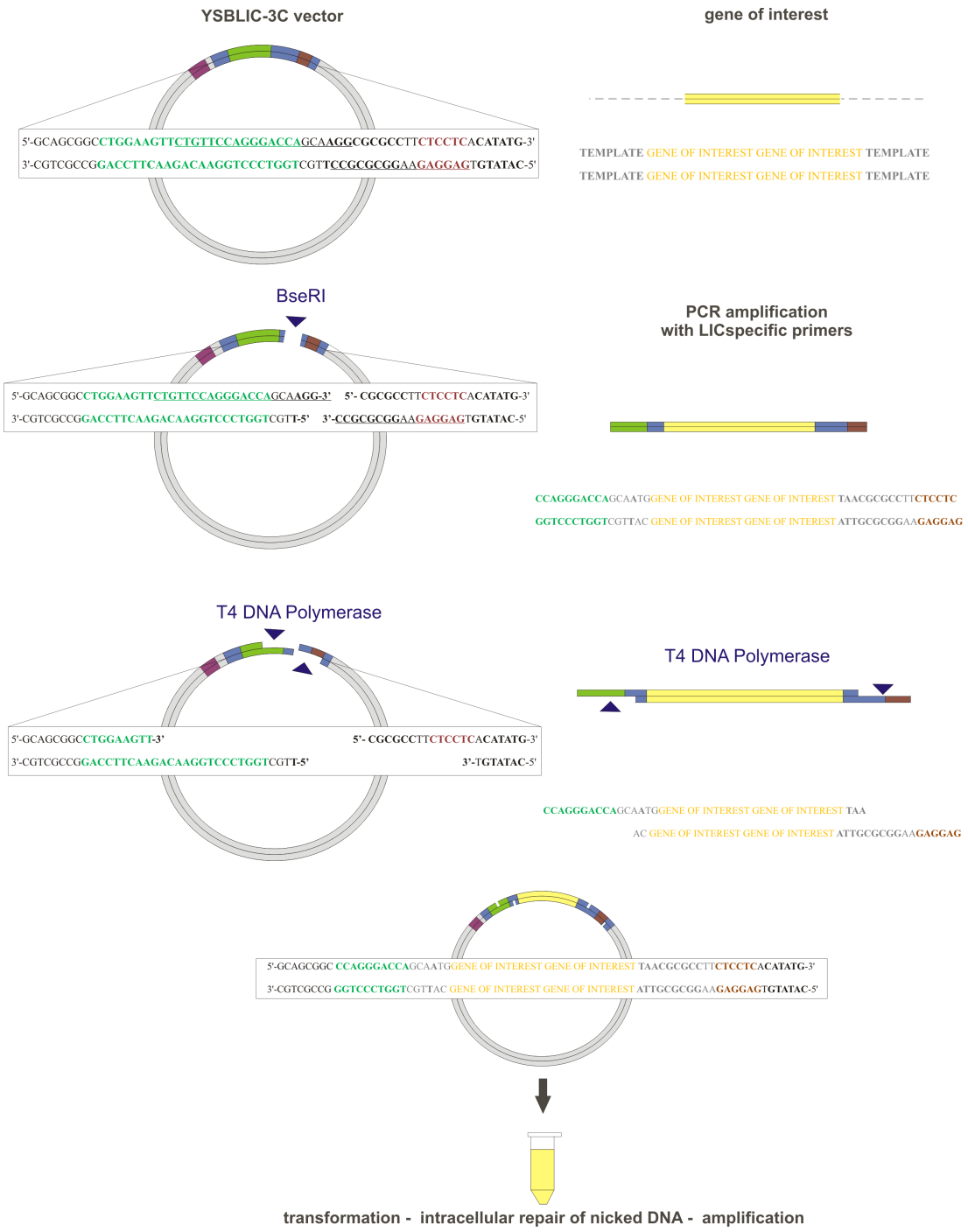


Figure 2.1: Ligation independent cloning into the YSBLIC-3C plasmid allows for high throughput preparation of affinity tagged gene constructs. The vector provides a BseRI recognition site (brown) for linearization and a sequence encoding the HRV-3C protease cleavage site (green) between hexahistidine tag (purple) and gene of interest (yellow) [adapted from presentation slides prepared by Mark Fogg and (104)].

The 5'-3' exonuclease activity of T4 polymerase is subsequently used to synthesise 5', single stranded, 14 nucleotide overhangs on both vector and insert. The addition

of excess dTTP and dATP to the vector and insert reactions respectively, ensures that the presence of a complementary base inhibits the T4 enzyme's exonuclease activity as the nucleotide deficient sequence ends. The highly complementary 5' single stranded overhangs thus created anneal spontaneously if vector and insert DNA are combined.

This process creates circular plasmid DNA that is stable enough for the transformation of a highly competent *E. coli* cloning strain. The microorganism's repair and amplification system will close the nicks between vector and insert and allow for production of high copy numbers of recombinant plasmid (see Fig. 2.1) (103,104). The presence of a kanamycin resistance marker then allows for selection of recombinant clones on antibiotic-containing growth medium.

The YSBLIC3C vector allows for the expression of a gene which, when transcribed and translated in the cell, will give rise to a hexa-histidine tagged protein construct ideal for affinity purification. The affinity tag can be cleaved specifically using the HRV3C protease if required (104).

2.1.2 In-Fusion cloning

In addition to the LIC cloning described above, the In-Fusion™ technique was used by the University of York's Technology Facility to prepare fusion constructs between the oxidised polyvinyl alcohol hydrolase and a variety of solubility enhancing protein tags. The technique allows for the insertion of genes (amplified with gene and vector specific primers) into any vector of choice. Analogous to the LIC method, 15 nucleotide overhangs are created on the insert. These are then specifically recognised by the proprietary In-Fusion enzyme™ and fused to their complementary vector sequences. The method is suitable for both prokaryotic and eukaryotic expression systems (105).

2.1.3 Gene expression and protein production in E.coli recombinant hosts

The Gram-negative bacterium *Escherichia coli* was chosen as a host system for all projects described in this work. Its fast growth rates and high levels of gene expression and protein production make it an ideal candidate for small to large scale purification processes [pET manual].

The most commonly used expression strain is *E.coli* BL21 (DE3). The cells possess a T7 RNA polymerase gene (provided by the lysogen DE3) which is under the control of a *lac*-promoter system. It is ideal for protein production from pET vectors, such as YSBLIC-3C, the multiple cloning sites of which are, in return, under the control of a T7 promoter. By intracellular inhibition of the *lac*-repressor through allolactose or its analogue isopropyl β-D-1-thiogalactopyranoside (IPTG), gene expression and subsequent protein production can be initiated. The control of gene expression levels can thus improve yields and/or enhance solubility of recombinantly produced protein.

The BL21 strain was originally derived from B834, a methionine auxotroph (Met^-) variant of *E.coli* which, like all the "B" strains, lacks the *lon* and outer membrane (*omp*) protease genes. The B834 strains are commercially available as (DE3) variants and are commonly used for the expression of selenomethionine labelled proteins required in macromolecular crystallography. They have also been shown to help with the controlled production of proteins which are particularly toxic to their recombinant host (106,107).

As an alternative to *E.coli* BL21 or B834 (DE3) cells, the Rosetta (DE3) variant is frequently used for the expression of genes with unusual codon usage, particularly of eukaryotic or viral origin. The strain provides an additional plasmid encoding tRNAs complementary to the rarely used codons of arginine, glycine, isoleucine, leucine and proline. Their presence can prevent mismatching of amino acids and premature termination of the translation process and thus lead to translation of heterologous gene targets in *E.coli* (107,108).

All three strains described above can be obtained as pLysS variants. These contain, on an additional plasmid or the rare-tRNA encoding plasmid in case of Rosetta strains, the T7 lysozyme gene. This gene's protein product specifically binds and inhibits T7 RNA polymerase. The pLysS variants thus limit background expression of the gene of interest in the uninduced cells. After induction and inactivation of the *lac*-repressor, sufficient T7 RNA polymerase is produced to outweigh the inhibitory effect of lysozyme and induce expression of the gene of interest. This control of basal or background gene expression is thought to help cell viability and stable production of protein products which are toxic to the host cells.

In addition to the regulation of expression levels of the gene of interest, a variety of *E.coli* strains have been engineered to enhance folding and solubility of recombinantly produced proteins. The Lemo21 (DE3) strain allows for fine-tuning of T7 RNA polymerase expression and thus indirectly controls the *lac* promoter in pET28-based YSBLIC3C plasmids (109,110). Without addition of L-rhamnose to cultures, Lemo cells are comparable to regular BL21(DE3) pLysS strains, expressing T7 lysozyme and thus inhibiting background expression from the *lac*-promoter as described above. Addition of an L-rhamnose gradient to cultures allows for titration of T7 lysozyme expression and activity of the *lac*-promoter accordingly. For toxic proteins this can prevent or lower the formation of inclusion bodies. The Arctic Express™ strain makes use of the co-expressed cold shock chaperonins cpn60 and cpn10. Expression at low temperatures induces production of these chaperonins which may then, in turn, support the folding of a protein of interest (111,112).

2.2 Purification of recombinant proteins by affinity and size-exclusion chromatography

2.2.1 Metal-affinity chromatography

A significant advantage of gene expression and protein production from the YSBLIC-3C vector is the presence of a cleavable N-terminal hexahistidine tag. This allows for purification of the protein of interest by immobilised metal affinity chromatography (IMAC). The method was first introduced for general protein purification in 1975 and is based on the interaction between the basic side chain, in particular the imidazole ring, of histidine and a divalent metal ion acting as a Lewis acid (113). This metal can be Zn^{2+} or, less commonly, Co^{2+} and is immobilised onto a carbohydrate resin to which it shows a higher binding affinity than to the protein ligand. Of the metal's six coordination sites, four interact with the resin whereas two are free to interact with the protein (see Fig. 2.2) (114). The nature of the interaction between metal and histidine side chains allows for comparatively high specificity binding of the tagged protein. When eluted, purities exceeding 90% can be observed accordingly. This specific elution is either based on a variation in pH or, as applied in this work, by replacement of the histidine side chains with increasing concentrations of free imidazole in an elution buffer (113,114).

2.2.2 Size-exclusion chromatography

As IMAC yields protein of high, yet insufficient, purity for macromolecular crystallography, a second chromatography step is commonly added to each purification. This process, known as gel filtration or size exclusion chromatography (SEC), is based on separation by molecular weight and frequently used as a final step in the purification of macromolecules. The SEC columns used for this project are made up of Sephadex™, an inert and stable packed bed mixture of cross linked dextran and agarose. The spherical particles form a porous sieve through which larger molecules can pass without considerable interaction with the stationary phase while smaller molecules diffuse into the pores. A heterogeneous protein sample will be separated accordingly: the higher molecular weight its components have, the earlier they elute from the column. SEC thus allows for purification of proteins from contaminants of different molecular weight. When calibrated with an appropriate marker, SEC columns can also be used to determine the molecular weight as well as the mono- or multimeric state of a previously uncharacterised protein (115).

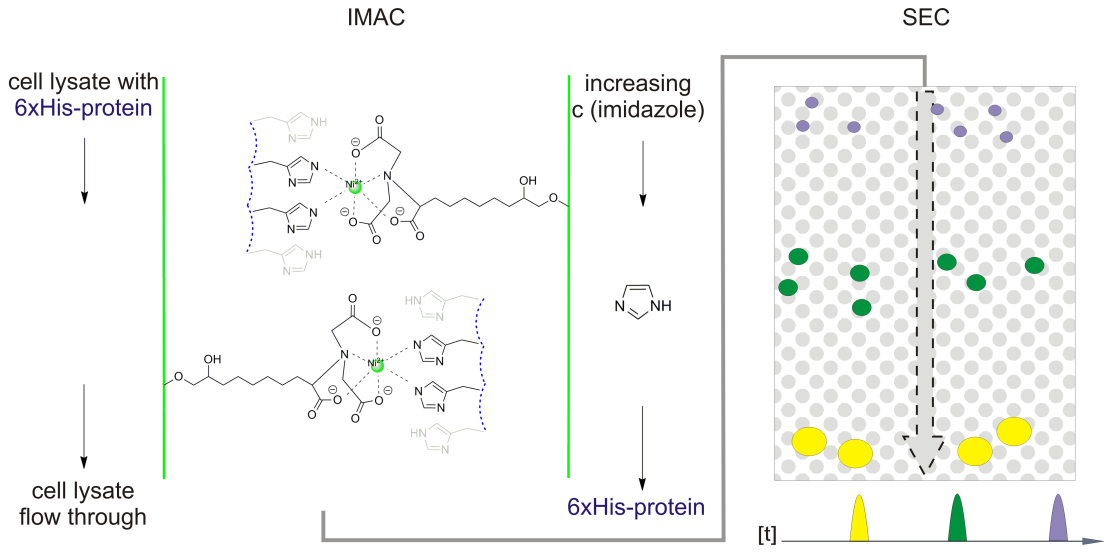


Figure 2.2: The metal ion (green) in IMAC purification columns is highly coordinated by the carbohydrate resin, in this case to give Ni-nitriloacetic acid (Ni-NTA). Two of its coordination sites are free to interact with the histidine side chains. A gradient of imidazole can replace these interactions and specifically elute the tagged protein [adapted from (114)]. In the subsequent SEC step, the protein of interest is separated from lower or higher molecular weight contaminants as a consequence of their longer and shorter retention times on a porous stationary phase respectively.

2.3 *Improvement of protein solubility*

A considerable number of proteins are insoluble when produced in heterologous expression systems. Reasons for this are diverse and cannot always be derived from their sequence or experimentally. They include post-translational modifications (in particular for eukaryotic proteins), the presence of signal sequences, requirements for chaperones, unfavourable expression conditions and others. Whereas some of these problems can be overcome by the use of genetically engineered expression systems (see above), others may require optimisation of genes, protein production and purification methods.

2.3.1 Removal of Signal sequences

The presence of signal sequences on the gene of interest may target the translated protein to the membrane/cell wall or certain cellular compartments *in vivo*. If it is not essential to protein folding or function, removal of the signal sequence may retain the protein in the cytosol and help with its purification from soluble cell fractions.

2.3.2 Buffer screening -The Sparse Matrix approach

In 2000, Lindwall and co-workers designed and published a buffer-based screen aimed at improving the solubility of overexpressed proteins. This screen was based on the observation that transfer of gene expression and protein production into a heterologous host often leads to the formation of inclusion bodies. The latter contain aggregated, mis- or partially folded proteins which are hoped to be solubilised (refolded) in the appropriate conditions. The authors proposed three variables to influence aggregation or allow for retrieval of protein from inclusion bodies already formed (116). By altering a buffer's pH, the protein is moved closer or further from its isoelectric point (pI). The charge, especially of surface exposed residues, determines the interaction between a protein and its surrounding solvent (the solvation state) and is thus thought to affect the formation of aggregates. Similarly, the salt content of a protein's environment can alter its solubility as it is described in the so-called Hofmeister series. Their influence on electrostatic repulsive forces or hydrophobic interactions is thought to affect correct protein folding, either stabilising

or preventing it (117). A third set of conditions, the presence of additives such as sugars, alcohols and detergents, is thought to further influence folding and stability, either directly or by preventing the protein from interacting with other components of the host cell. The detergents in particular could interact with surface-exposed hydrophobic residues and thus prevent or disrupt aggregation between macromolecules.

With a randomized collection of thirty buffers, the authors hope to cover conditions optimal for recovery of a soluble target protein. They propose that the combination of pH, salts and other additives may resemble the polypeptide's natural microenvironment better than the mild standard buffer preparations commonly used today (116).

2.3.3 Preparation of fusion constructs

In addition to altering the protein's expression and purification environment as described above, its solubility can be enhanced by the attachment of a solubility tag. The method has been used for several decades and with a large variety of constructs, from simple polypeptides to full-length protein tags. The rationale for increased solubility of the fusion construct is often unclear (118).

The most commonly used protein solubility tags are the *E. coli* maltose binding protein (MBP) and the *Schistosoma japonicum* glutathione S-transferase (GST). Others include the ubiquitous green fluorescent protein GFP, the hydrophilic *E. coli* thioredoxin A (*trxA*) and N-utilization substance A (NusA) as well as the immunity protein 9 (Im9). MBP in particular has been shown to improve the solubility of proteins which would, without the fusion, be expressed but largely remain unfolded or partially folded in cellular inclusion bodies. The exact mechanisms by which the fusion supports solubility are however still unknown. It has been suggested to stabilise some of its "passenger proteins" to favour their native folding over aggregation whereas others may be positioned suitably for molecular chaperones to assist in folding (118-120). There are no structural or functional similarities between the commonly used protein solubility tags (118) and mechanisms of solubilisation may have to be determined individually in the future.

2.4 Alteration of single amino acids by site-directed mutagenesis

An essential requirement for the characterisation and optimisation of proteins, in particular those involved in biocatalytic processes, is the ability to change single amino acid residues. These changes are usually (but not essentially) focused on the enzyme's active site. They can be used to knock out or reduce activity to highlight the importance of selected catalytic residues or to trap a substrate, reaction intermediate or product in the active site. Alternatively, residues can be altered to change substrate scope, regio- and/or enantioselectivity or to simply increase a catalyst's known activity. The most common method to change single or multiple selected residues is by amplification of template DNA using mutagenic primers.

Early DNA modifications were chemical- or radiation induced. The first site-directed mutagenesis (SDM) experiments were carried out in 1978 by Hutchison and co-workers who changed wild type and mutant single stranded phage DNA into their respective counterparts. Amplification of template DNA using mismatched primers and subsequent transfection of *E.coli* hosts then allowed them to select for the desired phenotypes (121). The method has since been optimised and is now routinely carried out on double-stranded DNA vectors using a set of two primers complementary to the gene of interest. These can be fully or partially overlapping, the latter being preferred for templates with a high ratio of guanine and cytosine (GC-rich DNA).

The SDM reaction is in many aspects analogous to regular PCR-based gene amplifications (see Fig. 2.3). The mutagenic primer pair anneals to the single strands of a heat-denatured DNA plasmid template and is extended in opposite directions (5'-3'). The primers' overall complementarity to the gene of interest allows for tight binding to the template, despite the presence of one to three mismatched bases per mutation and one or several mutations per primer. Increased extension times as compared to single gene amplifications then lead to replication of the full template. The resulting single stranded circular DNA associates into a double stranded, nicked plasmid suitable for direct use in transformation reactions, without the need for ligation. The nicks will subsequently be repaired in the recombinant host. In early experiments, the ratio of template to amplified DNA (cycle number) would determine the relative amount of mutant clones after transformation of a host system by the template/mutant DNA mixture. The use of DpnI restriction endonuclease has

resolved this problem by specifically degrading the methylated parental DNA while leaving the non-methylated amplification products intact (2,122,123).

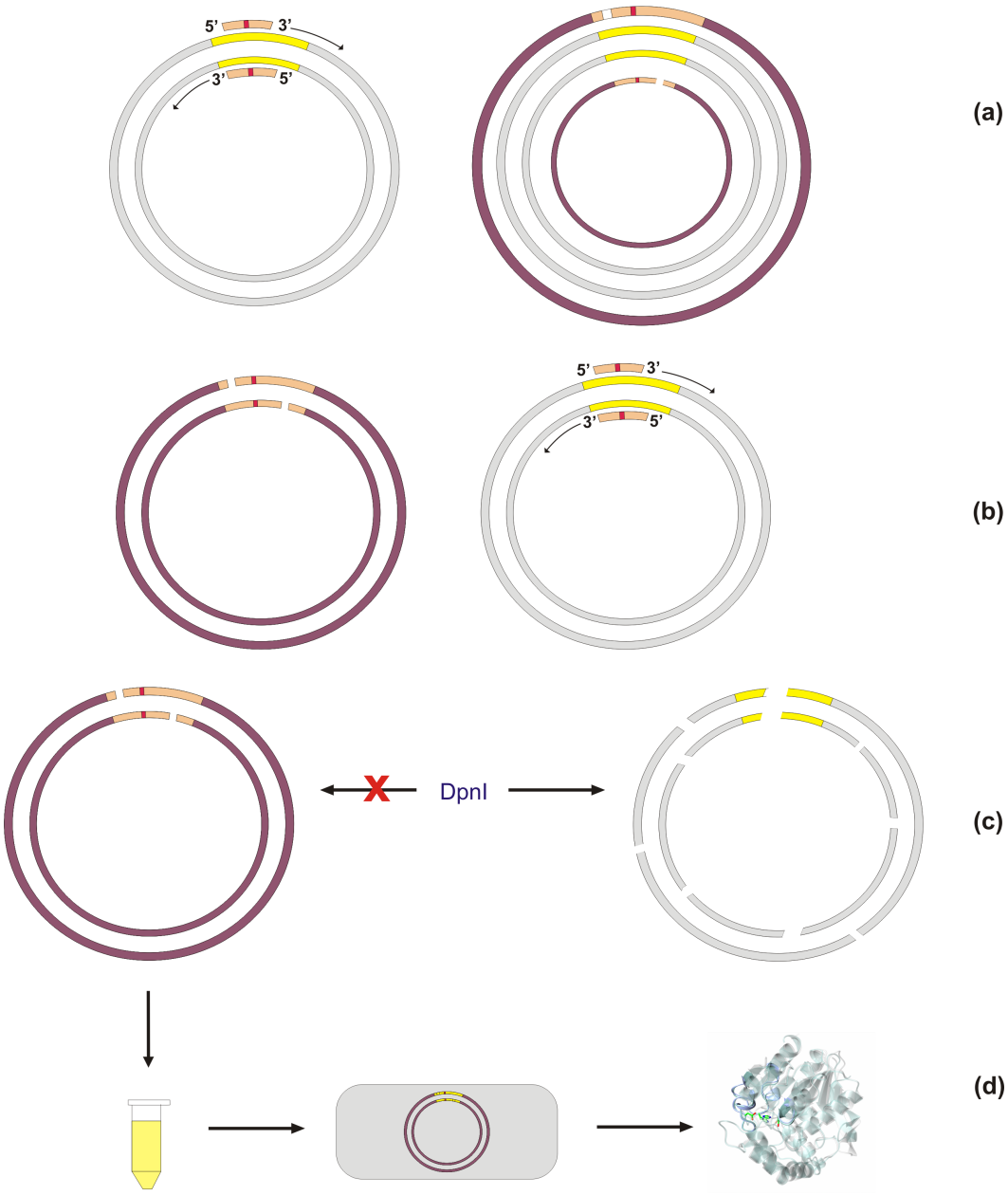


Figure 2.3: SDM using a pair of mismatch primers is an efficient method of altering single or multiple residues on a gene of interest (yellow, mutation shown in red). After amplification (a), the newly synthesised strands (purple) associate into a double stranded, nicked plasmid; primers only anneal and extend from the original template (grey) (b). Methylated "parental" DNA is specifically recognised and digested by the DpnI restriction endonuclease whereas the mutant plasmid remains intact (c). A recombinant host transformed by this plasmid can repair the nicked strands and amplify the mutant vector. It can then be used for sequencing, gene expression and production of the mutant protein [adapted from (123)].

2.5 Determination of protein kinetics

Within this project, all enzymes that could be purified with sufficient yields were subjected to experiments allowing the observation of substrate turnover and the determination of kinetic rate constants. This involved native proteins as well as a number of active site mutants. The detailed analysis of enzyme kinetics can give important information on the mechanisms of substrate binding and turnover. It defines both efficiency and specificity of the catalyst. Kinetic data thus complement structural data or provide important information in their absence.

2.5.1 Michaelis Menten kinetics

The most basic model for enzymatic conversion of a substrate into a product by way of a single transition state (the enzyme-substrate complex) was proposed in 1913 by Leonor Michaelis and Maud Menten (124). It was based on the observation that an enzymatic reaction initially follows a linear relationship between substrate concentration and reaction velocity before levelling off and reaching a plateau. Once the plateau is reached, a further increase in substrate concentration has no significant effect on the rate of the reaction (see Fig. 2.4). Michaelis and Menten derived an equation describing the catalytic rate at a specific substrate concentration, taking into consideration the protein's maximum catalytic potential and the strength of substrate binding:

$$V_0 = \frac{\mathrm{V_{max}}\,[S]}{[S] + K_M}$$

Eq. 2.1

$$[V_0] = moles * s^{-1}$$

Their equation holds true as long as an enzyme reaction is in its "steady state", i.e. a constant equilibrium between the formation and dissolution of the enzyme bound transition state.

The Michaelis-Menten equation provides all constants required to describe simple catalysis, i.e. the conversion of a substrate into a product by way of an energetically less favourable, enzyme bound transition state. The starting point for the equation is the assumption that catalysis is composed of four different sub-reactions, each

defined by a rate constant k. The formation of an enzyme substrate complex [ES] from free enzyme [E] and free substrate [S] is followed by the release of free product [P] from regenerated enzyme [E]. These two reactions are described by the reaction constants k_1 and k_3. k_3 is known as the enzyme's catalytic efficiency and thus also called the rate constant k_{cat}. Each can, in theory, occur in reverse. As the re-association of product into the transition state (k_4, red) is negligible however, only the dissociation of enzyme-substrate complex into free molecules is considered in the equation and given the rate constant k_2.

$$[\mathrm{E}][\mathrm{S}] \underset{k_2}{\overset{k_1}{\rightleftarrows}} [\mathrm{ES}] \underset{k_4}{\overset{k_3}{\rightleftarrows}} [\mathrm{E}][\mathrm{P}]$$

Eq. 2.2

$$[\mathrm{k}] = (\mu\mathrm{M}[\mathrm{S}] * \mu\mathrm{M}[\mathrm{E}]^{-1}) * \mathrm{s}^{-1} = \mathrm{s}^{-1}$$

The first parameter described within the course of the reaction is K_M, the Michaelis constant. It is both an indicator at which concentration the enzyme reaction reaches half its maximal velocity and can be a sign of how strongly a substrate is bound. When the substrate concentration corresponds to the Michaelis constant, the initial velocity of a reaction is equal to half the maximum velocity:

$$V_0 = \frac{[\mathrm{S}] * V_{\max}}{[\mathrm{S}] + K_M} \quad \text{if } [\mathrm{S}] = K_M \text{ then } \quad V_0 = \frac{1}{2} V_{\max}$$

Eq.2.3

It is obtained by setting the rate of enzyme substrate complex formation and breakdown equal (as in the definition of steady state kinetics) after which the equation can be limited to the rate constants alone:

$$k_1 * [\mathrm{E}][\mathrm{S}] = k_3 * [\mathrm{ES}] + k_2 * [\mathrm{ES}] = (k_3 + k_2) * [\mathrm{ES}] = (k_3 + k_2) * [\mathrm{ES}]$$

$$\frac{k_1}{(k_3 + k_2)} = \frac{[\mathrm{ES}]}{[\mathrm{E}][\mathrm{S}]} \quad \frac{(k_3 + k_2)}{k_1} = \frac{[\mathrm{E}][\mathrm{S}]}{[\mathrm{ES}]} = K_M$$

Eqs. 2.4, 2.5

$$[K_M] = \text{moles} * \mathrm{l}^{-1} = \mathrm{M}$$

For the Michaelis constant to be high, the concentration of [ES] and the rate constant $k1$ for its formation accordingly need to be low and *vice versa* for low K_M values. K_M can thus be an indicator of how tightly a substrate binds to an enzyme (the enzyme's specificity). As the equation however also includes k_3 (k_{cat}), its contribution to K_M has to be taken into consideration prior to evaluating substrate binding affinities.

Catalytic activity is well described by K_M, V_{max} and k_{cat} under saturating substrate conditions, i.e. when the total amount of enzyme corresponds approximately to the amount of enzyme in complex with substrate ($[E_{tot}] = [ES]$, $V_{max} = k_{cat}[E_{tot}]$). The enzymatic reaction has to be defined by the k_{cat}/K_M value however when substrate concentrations remain significantly below K_M. At this point it is assumed that most of the active sites are unoccupied (i.e. $[E_{tot}] = [E]$) and knowledge about the turnover number k_{cat} and the Michaelis constant can be combined to give a new determinant of the initial rate V_0:

$$V_0 = k_{cat}[ES] \quad K_M = \frac{[E_{tot}][S]}{[ES]} \quad V_0 = \frac{k_{cat}}{K_M} * [E_{tot}][S]$$

Eq. 2.6

$$\frac{k_{cat}}{K_M} = \frac{V_0}{[E_{tot}][S]} = s^{-1} * M^{-1}$$

Accordingly, the k_{cat}/K_M ratio defines the rate of catalysis for a fixed concentration of substrate below the K_M and (predominantly) free enzyme and is stated as the number of substrate molecules turned over per second by one mole of enzyme. The k_{cat}/K_M rate could be particularly important for analysing "natural" protein kinetics as it has been proposed that the K_M resembles the substrate concentrations a protein would encounter *in vivo* (125,126) .

2.5.2 The Lineweaver-Burk plot

For enzymes following simple Michaelis-Menten kinetics, rate constants can be derived from a graph plotting substrate concentration versus reaction rate V_0 (see Fig. 2.4). As the resulting curve can only provide approximate values however (especially for V_{max}, which is approached but never reached), their reciprocals are often plotted

in addition. These produce what is known as the Lineweaver-Burk plot. It produces a straight line, the x- and y-axis intercepts of which correspond to -1 x K_M^{-1} and -1 x V_{max}^{-1} respectively. The line's gradient corresponds to K_M/V_{max} (see Fig. 2.4, inlet) (125,126).

$$V_0 = \frac{V_{max} * [S]}{[S] + K_M}$$

$$\frac{1}{V_0} = \frac{[S] + K_M}{V_{max} * [S]} = \frac{[S]}{V_{max} * [S]} + \frac{K_M}{V_{max} * [S]} \quad = \quad \frac{1}{V_{max}} + \frac{K_M}{V_{max}} * \frac{1}{[S]}$$

Eq. 2.7

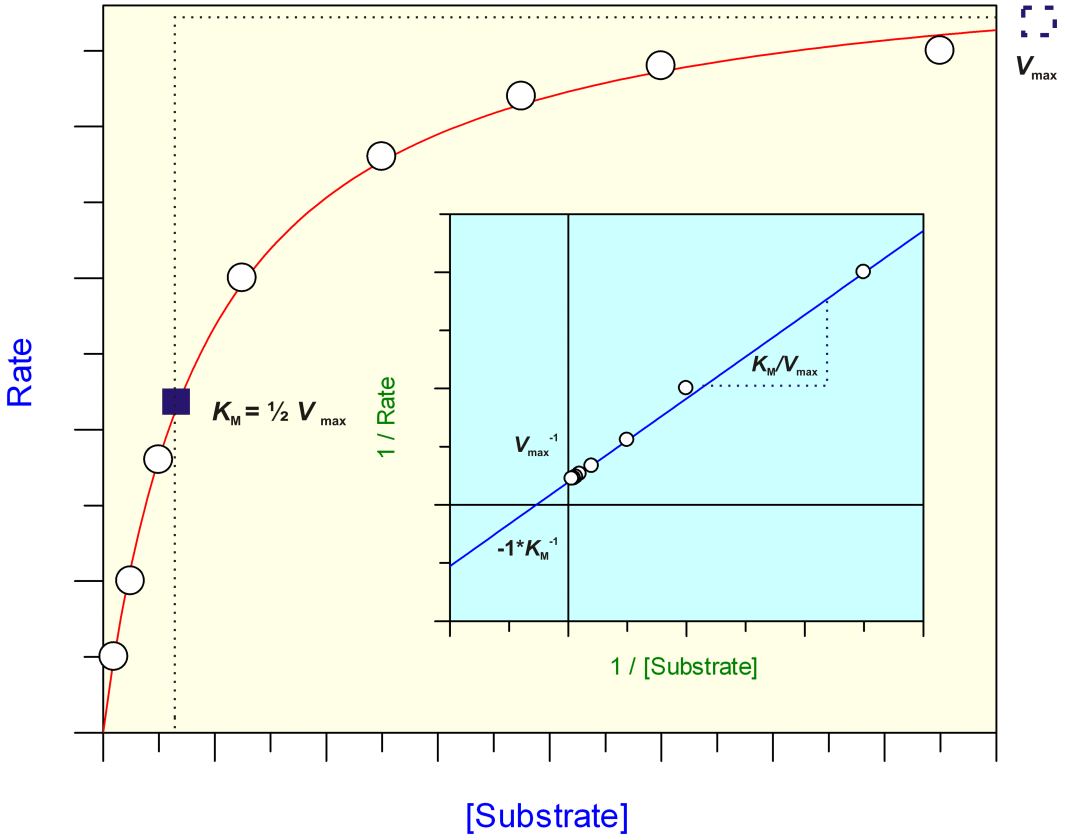


Figure 2.4: The reaction kinetics for enzymatic conversion of a substrate into its product, described by the Michaelis Menten equation, can be visualised by plotting the substrate concentration dependent reaction rate (K_M = blue square, V_{max} = dashed square). Alternatively, the reciprocal values of both can be a represented in the Lineweaver-Burk plot (inlet window). The latter facilitates determination of the rate constants and the k_{cat}/K_M ratio which correspond to intercepts of a linear trend line with the graph's axes and its gradient respectively.

The Lineweaver Burke plot presents disadvantages in the determination of protein kinetics as the fitting of a straight line can be unreliable, particularly for the densely clustered data points collected at higher substrate concentrations. As a consequence, substrate dependent initial reaction velocities ([S]/V_0) were plotted exclusively in the course of this work. From this, kinetic parameters could be determined by computational fitting of a Michaelis Menten curve or manual estimation of K_M and V_{max} values from the resulting graph (see Fig. 2.4)

The knowledge of an enzyme's catalytic efficiency and its tendency to bind a certain substrate are essential for all biocatalytic processes. Michaelis-Menten kinetics thus provide the basis for protein engineering, ligand screening, process scale-up and optimisation. The method of determining kinetic constants for a reaction depend on the protein, its substrate and, if applicable, cofactors involved in catalysis. The measurements required can be based on the consumption of substrate, the appearance of product or the turnover of cofactor molecules accordingly.

2.6 Analytical techniques

2.6.1 UV spectrophotometry

UV spectrophotometry can be used for real-time observation of substrate turnover by an enzyme. It is based on absorbance differences between substrate and product or the reaction states of a cofactor (for example if it is oxidised or reduced).
As monochromatic light is passed through a solution of substrate, electrons (particularly within π-bonds) are transferred into a higher energy level when "hit" by a photon and thus absorb part of the incoming energy. This absorbance A (or the difference in transmitted light T) can be measured. It corresponds to:

$$T = \frac{I}{I_0} \quad A = -\log T$$

I_0 corresponds to the incoming radiation and I to the transmitted radiation. Eq.2.10

Atoms and molecules have an absorbance maximum at a defined wavelength which corresponds to the energy difference between the excited and ground state. It can lie within the visible electromagnetic spectrum (between 400 and 700 nm) or in the range of ultraviolet radiation (10 – 400 nm).

Ultraviolet-visible (UV-VIS) spectroscopy makes use of this property. A substrate solution is exposed to monochromatic light at a wavelength corresponding to the previously determined absorbance peak of either substrate, product or cofactor. The transmitted light intensities are measured by a detector and used to determine the change in absorbance (see equation 2.10). As protein is added, catalysis changes the solution's absorbance properties. In other words, substrate breakdown, product formation or turnover of cofactor can be observed in real time as a consequence of the molecules' changes in absorbance spectra. Using Beer's law, the change in absorbance can then be transformed into a change in concentration, *i.e.* the molecular units of catalytic turnover and reaction rates:

$$\Delta A = \Delta(\varepsilon * l * c) \qquad \Delta c = \Delta(\frac{A}{\varepsilon * l})$$

Eq. 2.9

The change in concentration c of product, substrate, or a specific cofactor state can thus be calculated from the change in absorbance if the molecule's extinction coefficient ε and the pathlength l are known (127).

2.6.2 High performance liquid chromatography

Whereas spectrophotometric methods are used to observe real-time enzymatic catalysis, other methods rely on the quantitative and qualitative analysis of reaction samples taken at various time points. High-performance liquid chromatography (HPLC) is a highly sensitive method enabling the user to detect even trace amounts of product and follow turnover of substrate and/or cofactor. The molecules of interest can be extracted into an organic or aqueous phase and are subsequently separated on a chromatography column. As a consequence, these methods can be used for both purification as well as analysis.

Separation of a reaction sample's components on a HPLC column is achieved through their differential interaction with the stationary phase (the packing material) and a mobile phase (the solvent). These interactions are usually based on polarity. For normal-phase HPLC, the columns packing material is a polar substance which will interact with polar functional groups in the sample. The organic solvent is non-polar and will preferentially enhance transport of less or non-polar molecules accordingly. In reverse-phase HPLC, which is the most commonly used method, these conditions are inverted – the stationary phase is non-polar (it contains long hydrocarbon chains) whereas the water-based solvent is polar. As for UV-VIS spectroscopy, HPLC can be used to determine protein kinetics. It is the method of choice when high sensitivity and a large detection range (with respect to a sample's chemical composition and concentration) are required. Once the molecules of interest are eluted from the column, a detector measures their absorbance. In combination with the retention times unique to substrate, product or cofactor, this allows for both qualitative and quantitative assessment of each reaction and can be used to visualise a catalytic process, as described above.
The stationary phases of HPLC columns can also be composed of chiral materials. As stereoisomers of the same molecule interact differently with such surfaces, chiral columns allow for their separation and identification (128).

2.6.3 Gas chromatography

Components in an enzyme-catalysed reaction can also be detected and separated by gas chromatography (GC). In GC based analysis, all components to be measured have to be extracted into an organic solvent after the enzymatic reaction – analogous to HPLC based methods. The resulting solution is then heated until it reaches its gas phase. GC analysis is thus limited to molecules that are sufficiently volatile, yet stable, at the required temperatures. The heated sample is mixed with an inert carrier gas (the moving phase) and transported through a small diameter (<1 mm), spiral column. This column is lined with the stationary phase, a material that specifically interacts and retains the carbon-rich analytes over the carrier gas. The interactions between stationary phase and analytes can be based on polarity (as described above), solubility/volatility or other chemical interactions. As for HPLC, GC columns with

chiral stationary phases allow for the differentiation between stereoisomers of the same molecule.

Unlike UV-VIS spectroscopy and HPLC, the molecules of interest are not identified or quantified based on their absorbance properties. As each reaction component leaves the column, it passes through a flame. Decomposition of the organic analytes leads to formation of ions (particularly carbocations in organic compounds) which are attracted to an electrode. The resulting current is measured: its strength is proportional to the analyte's concentration, the retention time between injection and signal serves as an identifier as it is unique for its molecular composition. The experimental output (voltage) does not allow for direct calculation of the analyte's concentration, as with Beer's law for UV based methods. As a consequence, GC methods rely on reference samples and calibration curves for quantitative results (129,130).

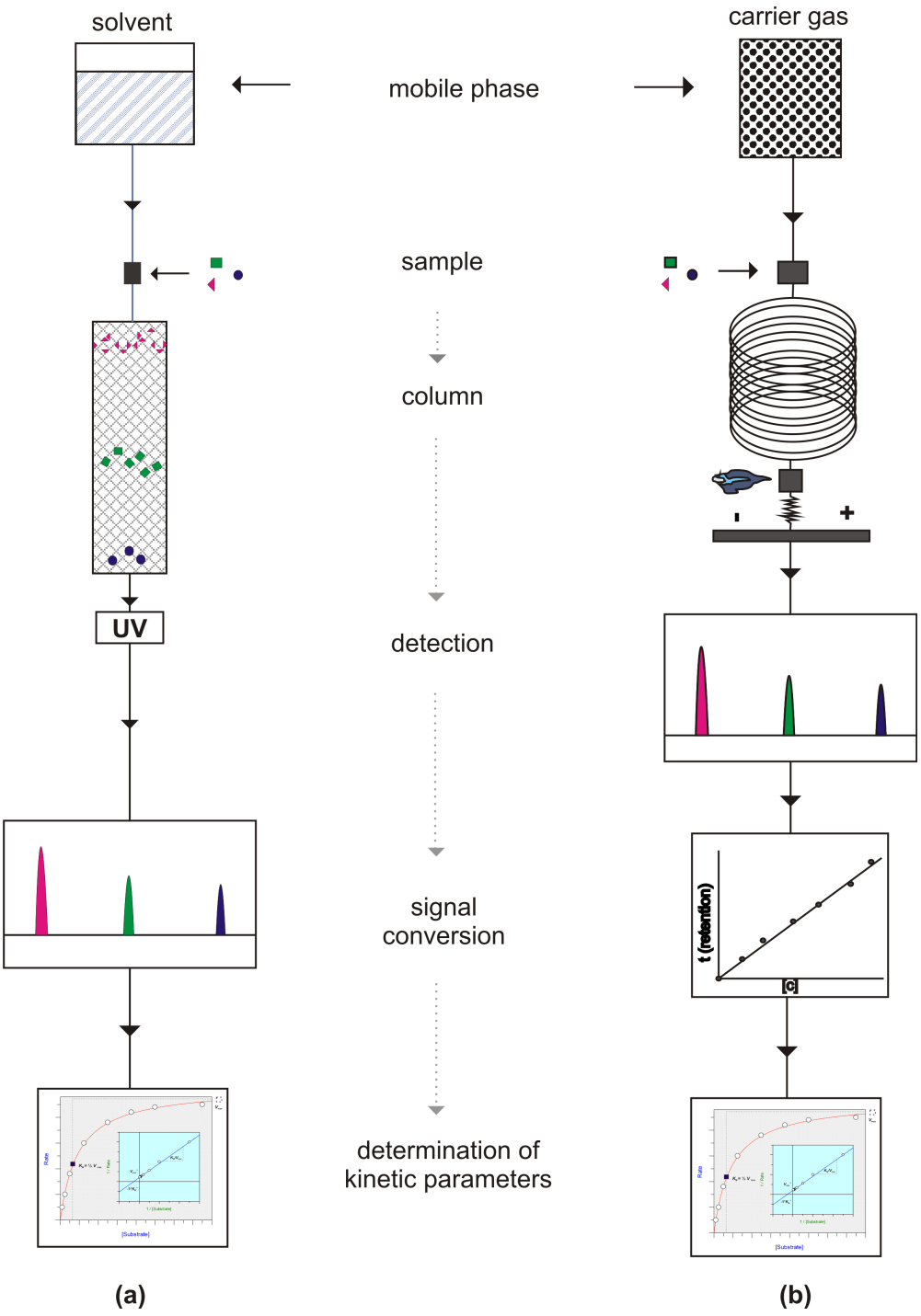


Figure 2.5: HPLC (a) and GC (b) analysis are used for qualitative and quantitative analysis of enzymatic reaction samples. Both methods rely on the transport of analytes by a carrier substance. The molecules of interest are retained to various degrees by a stationary phase in the analytic column. Their relative amount is then determined by UV absorbance (HPLC) or, after combustion, through the specific current the analytes' ions create on a measuring electrode (GC). Whereas UV signals can be converted into concentrations directly, the electric signal has to be evaluated with the help of a previously prepared calibration curve (129,130).

Chapter 3: Preparation of a *Bs*PAD mutant library

3.1 Introduction

3.1.1 Phenolic acid decarboxylases

Phenolic acid decarboxylases (PADs) belong to the family of cofactor-independent carbon-carbon bond lyases. They catalyse the removal of a carboxylate functional group from a number of natural phenolic acids and show varying substrate scopes and specificities between the different family members. PADs are predominantly found in bacteria (the Gram positive *Bacillus*, *Lactobacillus, Pseudomonas, Aerobacter* and *Pediococcus* species and the Gram negative *Klebsiella oxytoca*) as well as in lower eukaryotes like yeast. They constitute an important defence mechanism for their hosts as they come into contact with plant derived phenolic acids. The latter are abundant in nature and can be found in plants where they cross-link the cell wall polysaccharides and hence become a breakdown product during lignin degradation (131). They share a common aromatic ring substituted by at least one hydroxyl moiety in *para*-position to a crotonic ((*E*)-2-butenoic) acid side chain, making them derivatives of cinnamic acid (see Fig. 3.1). Some of the most abundant phenolic acids are ferulic-, *p*-coumaric- and caffeic acid (132).

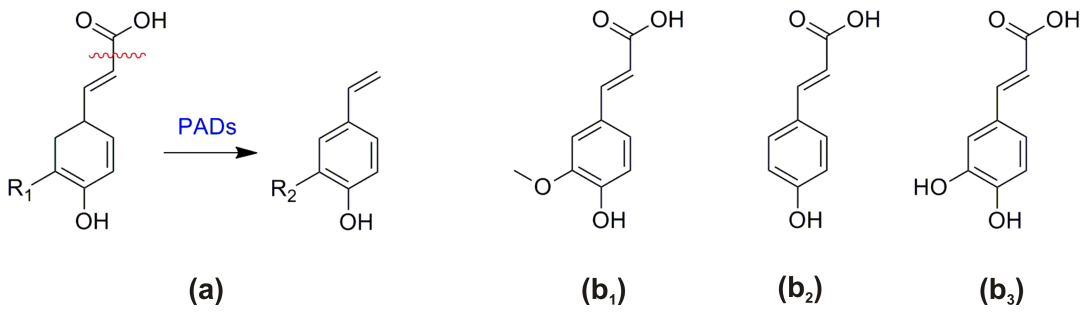


Figure 3.1: Phenolic acids are identified by their phenol ring, substituted with at least one hydroxyl, and their carboxylic acid functional group. Their decarboxylation leads to formation of a *p*-hydroxystyrene (a).The phenolic acid substituents (R) may vary in number, character and position. The most abundant phenolic acids are: ferulic acid (R = (CH_3-O, b_1), *p*-coumaric acid (no R, b_2) and caffeic acid (R = OH, b_3).

In the food industry, phenolic acids contribute to off-flavours, especially in their volatile form, and may negatively influence aroma and taste, not least because of their astringent properties. Depending on their application however, they also have desired properties, especially in the production of fermented foods and alcoholic beverages and in their function as antioxidants (131-133). Natural phenolic acids have been shown to exert inhibitory growth effects on microorganisms exposed to them. For the extensively studied *Lactobacillus* species for example, this has been suggested to be due to disruptions to the cytoplasmic membrane (131). As a consequence, many microorganisms have evolved PAD enzymes for their specific decarboxylation in a process known as the phenolic acid stress response (133). As the carboxylate moiety is removed, the acids are converted into volatile styrenes (see Fig. 3.1), which are considerably less toxic to the organisms.

The crystal structures of four prokaryotic PADs have been solved to date. All associate into dimers and show the conserved overall protein fold of a central, flattened β-sandwich flanked by an α-helical region on one end and flexible loops on the other, more open (solvent exposed) side (see Fig. 3.2). The structural conservation is most pronounced for the sandwich motif, which is thought to accommodate the proteins' active sites. It also reflects the high primary sequence conservation between PAD proteins, of both known and unknown 3-dimensional structure, in this area (60,133,134).

Several mechanisms could be proposed for phenolic acid decarboxylation. The most comprehensive PAD mechanism was published by Hashidoko *et al.* in 1998 (135) for the enzyme from *Klebsiella oxytoca* (*Ko*PAD). It was based on experiments in which the reaction's stereochemistry could be observed using deuterium labelled substrates and nuclear magnetic resonance (NMR). This mechanism was later supported by the availability of PAD crystal structures (including one ligand complex) and modelling approaches (60,134,135). Though differing in the assignment of catalytic residues, all hypotheses include specific substrate recognition of the carboxylate and the *para*-hydroxyl group, deprotonation of the latter and an electron relay through the protein. Protonation at the C8-carbon is then thought to lead to the formation of a *para*-quinone methide intermediate. This will eventually be subject to decarboxylation, followed by a second electron relay and formation of the phenolic acid's respective styrene derivative (60,134,135) (see Fig. 3.3).

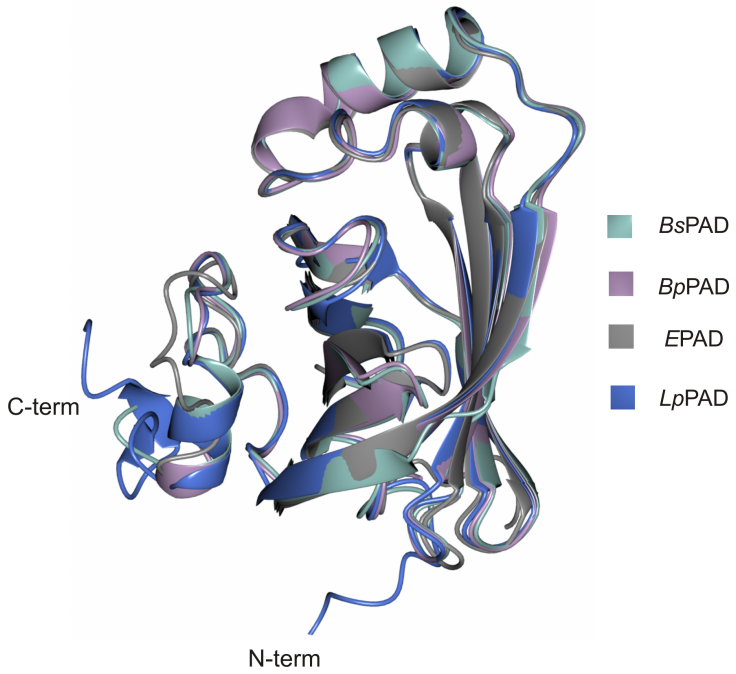


Figure 3.2: The superposition of all four PAD crystal structures determined to date shows a clear conservation of the overall fold, particularly around the core of the protein. The native enzyme sources were *Bacillus subtilis* (*Bs*PAD, green, 2P8G), *Bacillus pumilus* (*Bp*PAD, purple, 3NAD), *Enterobacter* sp. Px6-4 (*E*PAD, grey, 3NX1) and *Lactobacillus plantarum* (*Lp*PAD, blue, 2GC9).

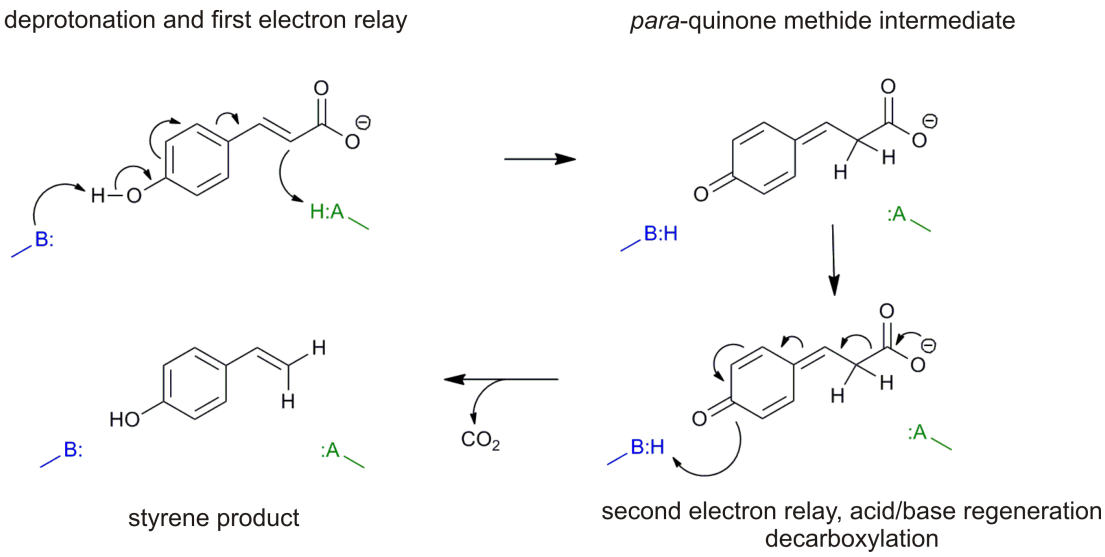


Figure 3.3: Detailed reaction mechanisms have been suggested for the enzymatic decarboxylation of phenolic acids. The proposals differ in their assignment of enzymatic residues as catalytic acids (A, green) or catalytic bases (B, blue). Also the nature of the proton donor for the substrate's scissile C8 carbon (A, green) is unclear [adapted from (135)].

Other, unrelated enzymes involved in the turnover of identical or similar substrates also provide information towards possible mechanisms and catalytic residues. Tyrosines forming a 'pincer' - like arrangement for recognition of the substrate's phenolic hydroxyl and arginine residues stabilising a quinone-methide intermediate have been observed in the Hydroxycinnamoyl-CoA hydratase lyase (HCHL) of *Pseudomonas fluorescens* as well as the Vanillyl-alcohol oxidase (VAO) from *Penicillium simplicissimum* (136,137).

3.1.2 PAD enzymes in Biotechnology

The interest in PAD enzymes for biotechnological applications stems from both their use in substrate degradation and in product formation. The release of phenolic acids and their decarboxylated phenol derivatives can have desired as well as unwanted effects in the food industry. Both contribute (off-)flavours and the latter gives astringent properties, especially to fermented foods and beverages (131). A thorough understanding of the fermentation process and the potential use of genetically modified organisms could thus be of advantage.

Phenolic acid decarboxylation into styrene derivatives is, in addition, an important starting point for various potential downstream processes. Polymerisation of the styrene products as well as the selective transformation of the ferulic acid decarboxylation product, 4-vinylguaiacol, to give vanillin, would allow for the formation of high value chemicals from natural, renewable resources (138). In addition, the PAD reaction was shown to be reversible *in vitro*, leading to carbon-carbon bond formation. This would allow for the fixation of gaseous carbon dioxide waste and its use in the production of valuable chemicals (13).

3.1.3 Bacillus subtilis Phenolic acid decarboxylase (BsPAD)

As described above, decarboxylase activity has been observed for a number of microorganisms, in particular the Gram positive *Bacilli*. Of these, the soil dwelling *Bacillus subtilis* was found to be the first microorganism to decarboxylate all three of the most abundant phenolic acids. The *padC* gene responsible for this ability, a 528 bp open reading frame, was first isolated and cloned into an *E. coli* host by Cavin and co-workers (18). Within the recombinant host, it could be expressed and

translated into a 161 amino acid polypeptide, the *Bacillus subtilis* phenolic acid decarboxylase (*Bs*PAD). The purified protein had a molecular weight of approximately 20 kDa and was shown to assemble into a dimer in solution. It specifically decarboxylated ferulic-, *p*-coumaric- and caffeic acid (in descending order of catalytic efficiency) into 2-methoxy-4-vinyl phenol, 4-vinyl phenol and 4-vinyl-2-methyl phenol respectively (see Fig. 3.4). Other substrates lacking either the *para*-hydroxyl group or a double bond in the crotonic acid side chain (*i.e.* a conjugated system) were not turned over by the enzyme (18). Prior to the availability of structural data, sequence alignments and the preparation of chimeric libraries gave a first insight into the relationship between *Bs*PAD and other previously characterised PAD enzymes. Differences in substrate specificity appeared to be the result of a few amino acid alterations between otherwise highly conserved sequences. (18,133).

The first crystal structure of the *Bs*PAD apo enzyme (PDB code 2P8G) was determined by the Joint Center for Structural Genomics initiative (JCSG) in the United States in 2007. It shows the typical PAD features previously described (see section 3.1.1), including the two prominent beta sheets surrounding and enclosing the enzyme's active site as a β-sandwich. One end of this tunnel-like structure is covered by α-helices, the other by flexible loops connecting the β-strands. (see Fig. 3.4). A tail like structure, comprising a long flexible loop and another α-helix, lies adjacent to the β-sheets and forms the C-terminal end of the polypeptide. In its dimeric state, these termini face in opposite directions.

When superposed with homologous enzymes, *Bs*PAD shows a strong conservation in three-dimensional structure, particularly around the central region (see above and Fig. 3.2). In this area, a number of residues can be identified which have been implied in catalysis for structurally and/or functionally related proteins such as the aforementioned *Lp*PAD, *Ko*PAD, VAO and HCHL (60,135,136,139) (see Fig. 3.4). These residues include a potential catalytic acid for removal of the substrate's phenolic hydroxyl (Glu64), a step essential for initiation of the first electron relay in PAD enzymes. For this part of the reaction to occur, the acid is required to be in its deprotonated state prior to catalysis. The resulting negative charge, in addition to those developing during substrate deprotonation and formation of the quinone-methide intermediate, may be stabilised by a positively charged arginine side chain (Arg41). A threonine residue in position 68 (as well as a second threonine in position

98, identified in the course of this project) may coordinate and/or activate a catalytic water which has been proposed to serve as a potential proton donor to the intermediate (60). For substrate recognition, either of the phenolic hydroxyl or the carboxylate moiety, two closely positioned tyrosine residues (Tyr11 and Tyr13) could be of importance (136,139). Both may also play a role in the observed substrate specificity of the enzymes in question. A third tyrosine in position 19 was previously found to constitute part of a hydrophobic pocket and may further contribute to substrate recognition and binding, the latter involving its rearrangement on binding of a phenolic acid molecule (60,136,139).

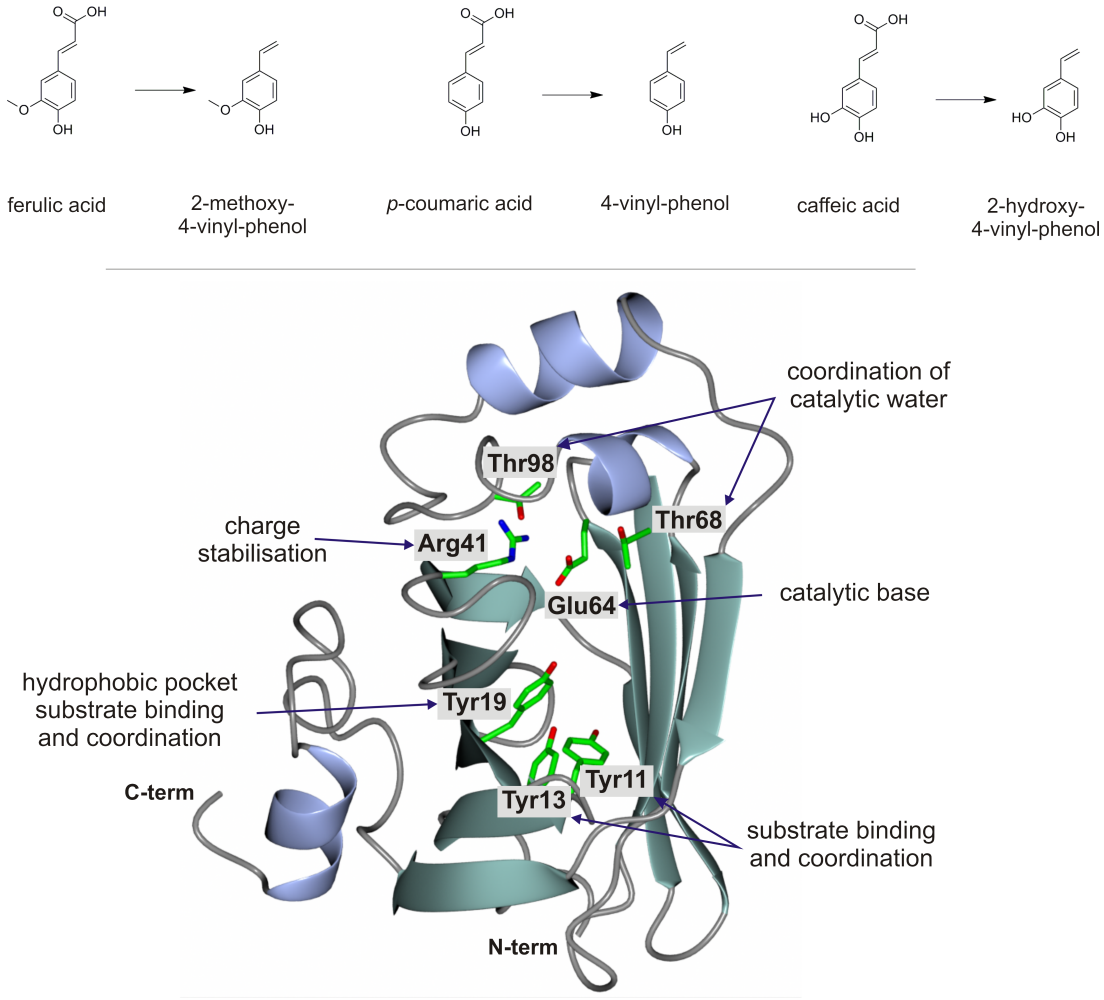


Figure 3.4: The crystal structure of *Bs*PAD resembles that of other, related prokaryotic PAD enzymes, particularly around the conserved central region (see Fig. 3.2). The enzyme was the first example of a PAD able to convert three phenolic acid substrates, ferulic-, *p*-coumaric- and caffeic acid (in decreasing order of catalytic efficiency). Putative catalytic residues, inferred from homologous and/or mechanistically related enzymes, are shown.

3.1.4 Project aim

At the start of this project, a structure of the *Bs*PAD enzyme had been solved in the absence of a ligand. In addition, partial information regarding the overall reaction cycle was available from homologous enzymes and mechanistically related catalysts. To understand the complete and detailed mechanisms of phenolic acid decarboxylation however, it is necessary to identify and understand the interactions between catalytic active site residues and the substrate, reaction intermediate(s) and product. For this reason, a mutant library encompassing the potential catalytic residues outlined above was prepared. From this it was hoped to gain an insight into their importance for the reaction. This could ultimately allow for confirmation of previously suggested reaction mechanisms or the proposal of a new type of decarboxylation reaction. In addition, it was hoped that a weakly active mutant protein could be used to fix a ligand molecule in crystallisation experiments, eventually leading to the determination of a complex structure.

3.2 Materials and Methods

3.2.1 Production and purification of wild type BsPAD

3.2.1.1 Cell culture and protein production

The *padC* gene encoding the wild type phenolic acid decarboxylase from *Bacillus subtilis* (18) ($BsPAD^{wt}$) was amplified and cloned into the YSBLIC3C vector by William Eborall, a Biochemistry B.Sc. project student at York in 2008. This construct, hereafter named YSBLIC3C:*Bs*PAD was then kindly provided for further experiments.

For gene expression and protein production, competent *E. coli* BL21 cells were transformed by the YSBLIC3C:*Bs*PAD vector. The cells were then plated onto a solid lysogeny broth medium (LB agar plates) containing 30 μg/ml kanamycin and incubated at 37°C for 12-18 h. From this plate, single colonies were picked and used to inoculate overnight cultures of 5 mL liquid lysogeny broth (LB medium) at the

same concentration of the selective antibiotic. These were, in return, used for the inoculation of 0.5-4 L of LB protein production cultures, with a 1:100 ratio of starter culture to fresh medium. These cultures were again grown in the presence of 30 µg/ml kanamycin and induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 1 mM as soon as their optical density at 600 nm (OD_{600}) had reached a value of 0.6-1.0. Cultures were then incubated for 16 h at 18 °C while shaking at 180 revolutions per minute (rpm). Harvested cells were resuspended in 30 mL of 50 mM Tris/HCl buffer with 300 mM NaCl at pH 7.5 (hereafter named "resuspension buffer") per litre of culture and disrupted by sonication at 12-14 amplitude microns for 3 x 45 s. The soluble fraction was obtained by centrifugation in a Sorvall RC5B Plus centrifuge at 15,000 rpm for 30 min.

3.2.1.2 Protein purification by Ni^{2+}-affinity chromatography

As a consequence of the cloning procedure, the expressed protein carried an N-terminal hexahistidine tag and could be purified in two subsequent steps. A 5 mL HiTrap™ Chelating HP column was charged using 0.1 M Ni_2SO_4 and washed with three column volumes of resuspension buffer prior to loading of the soluble supernatant obtained from crude lysate centrifugation (see above). After loading, the column was washed with resuspension buffer containing 30 mM imidazole (hereafter named "low imidazole buffer") and attached to an ÄKTA™ fast protein liquid chromatography (FPLC) system. The latter allowed for an automated wash step of ten column volumes, using low imidazole buffer, prior to elution of the protein in a 100 mL gradient from low to "high imidazole buffer" (50 mM Tris/HCl buffer with 300 mM NaCl and 500 mM imidazole at pH 7.5). Elution of the target protein was detected by measuring the column eluent's absorbance at 260 and 280 nm. Peak fractions were then collected accordingly and concentrated for subsequent purification by size-exclusion chromatography.

3.2.1.3 Protein purification by size-exclusion chromatography

Following Ni^{2+}-affinity chromatography, peak fractions were collected as described above and concentrated to a final volume of 2 mL. This sample could be purified by

size-exclusion chromatography using an S75 Superdex™ gel filtration column attached to the ÄKTA™ FPLC system. After loading of the protein sample it could be separated from impurities according to molecular weight and was eluted in resuspension buffer at an average flow rate of 0.5-1.0 mL/min. Peak fractions were detected as described above and pooled. The combined sample's absorbance was then measured at a wavelength of 280 nm (A_{280}) and, using Beer's law, the protein concentration was calculated from A_{280}, the known path length and the protein's unique extinction coefficient. The latter was predicted from its primary sequence using the ProtParam server (140).

Purified protein was frozen in liquid nitrogen and stored at -18 °C for later use in kinetic experiments. For crystallisation experiments, a part of the purified batch was retained and stored at 4 °C.

3.2.1.4 Hexahistidine-tag cleavage

A portion of the purified *Bs*PAD was set aside to remove the hexahistidine tag. For this, the protein concentration was adjusted to 1 mg/mL and incubated with HRV-3C protease at a ratio of 1:10 (w/w) for 16 h. To separate the protein from its cleaved tag and the protease, a second Ni^{2+}-affinity purification was carried out as described above. As the low imidazole buffer was replaced by resuspension buffer, imidazole free, cleaved *Bs*PAD was eluted whereas the separated histidine tag and the 3C protease were retained on the column until a 0-500 mM imidazole gradient was applied.

3.2.1.5 Determination of protein purity

During each purification, 10 µL samples were taken from sonicated whole cells or insoluble lysate, soluble lysate, affinity column flow through and selected peak protein fractions. Each sample was mixed with a denaturing loading buffer and heated to 94°C for approximately 5 min. Its protein contents and purity were then visualised by electrophoresis on a 12% sodium dodecyl sulphate polyacrylamide gel (SDS PAGE). For estimation of protein size, a Biorad™ low molecular weight marker was run in parallel on each gel.

3.2.2 Site-directed mutation of active site residues

For analysis of the proposed *Bs*PAD active site's key catalytic residues, a mutant library including seven single and one double mutant was prepared by site-directed mutagenesis. Residues mutated included two tyrosines thought to be involved in recognising and binding the substrate's *para*-hydroxyl group or carboxylate (Tyr11 and Tyr13), a potential constituent of a flexible loop restricting access to the active site (Tyr19), two residues acting as stabilisers and catalytic acid during catalysis (Arg41 and Glu64 respectively) as well as two nucleophiles implied in the coordination of a potential catalytic water (Thr68 and Thr98) (60). This process involved amplification of the wild type template plasmid using mutagenic primers, digest of parental DNA, transformation of a host organism by the mutant plasmid, amplification and isolation of the plasmid and subsequent determination of the gene sequence to confirm the desired mutation (see chapter 2, section 2.4.).

3.2.2.1 Primer design

A set of two complementary primers were designed for each single and for the double mutation and are shown in table 3.1. These primers contained one to three mismatches with respect to the wild type *Bs*PAD gene in the centre of their sequence.

Pos.	Mut.		Primer sequences
11	Tyr ↓ Phe	Fwd:	ggaagccacatgatt **ttt** acgtatgaaaacgga
		Rev:	tccgttttcatacgt **aaa** aatcatgtggcttcc
13	Tyr ↓ Phe	Fwd:	cacatgatttatacg **ttt** gaaaacggatgggaa
		Rev:	ttcccatccgttttc **aaa** cgtataaatcatgtg
11, 13	Tyr ↓ Phe	Fwd:	ctttatcggaagccacatgatt **ttt** acg **ttt** gaaaacggatgggaataccgag
		Rev:	tattcccatccgttttc **aaa** cgt **aaa** aatcatgtggcttccgataaagttttccattgc
19	Tyr ↓ Ala	Fwd:	gaaaacggatgggaa **gcc** gagatttatattaaa
		Rev:	tttaatataaatctc **ggc** ttcccatccgttttc
41	Arg ↓ Ala	Fwd:	ggaatggttgccgga **gcc** tgggttcgagatcag
		Rev:	ctgatctcgaacccaggctcc **ggc** aaccattcc
64	Glu ↓ Ala	Fwd:	aaagtgtcttggaca **gcg** ccgactggcacggat
		Rev:	atccgtgccagtcgg **cgc** tgtccaagacacttt
68	Thr ↓ Val	Fwd:	acagagccgactggc **gtg** gatgtttcattaaac
		Rev:	gtttaatgaaacatc **cac** gccagtcggctctgt
98	Thr ↓ Ala	Fwd:	gcatgaacatcctgaaatt**gcg**gtttgctaccaaaatg 3'
		Rev:	cattttggtagcaaac**cgc**aatttcaggatgttcatgc 3'

Table 3.1: Primers used for the preparation of a *Bs*PAD active site mutant library by site-directed mutagenesis. Nucleotide triplets encoding the altered amino acid are highlighted in red. Position (Pos.) and nature of the mutation (Mut.) are shown on the left.

3.2.2.2 Site-directed mutagenesis

The mutagenic primers were used to amplify the YSBLIC:*Bs*PAD vector as part of the reaction mixture shown in table 3.2 (top). The reaction was carried out using a Fisher Scientific TC-312™ thermocycler and the conditions outlined in table 3.2 (bottom).

Reaction component	Concentration	Volume [ul]	final concentration
forward primer	20 pmol/µL	1	0.4 µM
reverse primer	20 pmol/µL	1	0.4 µM
template DNA (YSBLIC:*Bs*PAD)	50 ng/uL	1	1 ng/µL
deoxyribonucleotides (dNTPS)	25 mM	1	0.2 mM
pfu turbo® hot start DNA polymerase buffer	5x	5	1x
pfu turbo® hot start DNA polymerase	2.5U/ µL	1	0.05 U/ µL
deionised water	-	40	-

Cycles	Temperature [°C]	Time [min:sec]	Reaction
1	95	0:30	initial denaturation
	95	0:30	denaturation
25	50	1:00	annealing
	68	6:00	extension
-	4	∞	hold/storage

Table 3.2: Reaction components (top) and amplification conditions (bottom) for the preparation of a *Bs*PAD active site mutant library by site-directed mutagenesis.

3.2.2.3 Protein production and purification of BsPAD mutants

Following the amplification, a restriction digest with DpnI endonuclease allowed for selective removal of non-mutated (hemimethylated) template DNA. For this, 10 units of New England Biolabs™ DpnI restriction endonuclease were added to each reaction. The samples were centrifuged at 13,000 rpm for one minute prior to incubation at 37°C for 3-5 h. An *E. coli* cloning strain (*E. coli* XL-1 blue or *E. coli* XL10 gold) was subsequently transformed by 5 µL of digested plasmid DNA. Colonies resulting from the transformation were used for the preparation of 5 mL overnight cultures in LB medium containing 30 µg kanamycin per mL of culture.

Subsequent isolation of plasmid DNA was achieved using a Sigma GenElute™ plasmid miniprep kit according to the supplier's specifications. After extraction, 300-500 ng of isolated plasmid DNA were supplied to the University of York's Technology Facility for determination of the gene of interest's nucleotide sequence and confirmation of the desired mutation. Sequencing chromatograms were evaluated using the ChromasLite™ software and the resulting nucleotide sequence was translated into amino acids using the Swiss Institutes of Bioinformatic's Expasy server (141). The final output could then be compared to the *Bs*PAD wild type sequence by pairwise sequence alignment using the EMBOSS Needle web server.

After sequence verification, an *E.coli* expression strain (*E.coli* BL21 (DE3)) was transformed by the mutant plasmids. The genes of interest were expressed and the proteins produced and purified as described for the wild type *Bs*PAD enzyme. To avoid contaminations, particularly for parallel purifications of several proteins, 5 mL HiTrap™ Chelating HP columns were stripped with three column volumes of 0.1 M EDTA and 0.5M NaCl followed by three column volumes of 0.5 M NaCl prior to recharging with $NiSO_4$ and equilibration with three column volumes of resuspension buffer as described above.

3.2.3 Analysis of enzyme kinetics

3.2.3.1 Spectrophotometric analysis

Kinetic data for the wild type protein and the active site mutants were initially collected by spectrophotometric analysis of substrate conversion. Comparison of UV-spectra for ferulic and *p*-coumaric acid and the products of their decarboxylation, 2-methoxy-4-vinyl-phenol and 4-vinyl-phenol respectively (all obtained from Sigma, Poole, Dorset, UK), show a significant decrease in absorbance, exclusive to the substrates, at a wavelength of 310 nm (data kindly provided by William Eborall). According to this, a kinetic assay was previously developed (142). Within this assay, substrate concentrations ranging from 5-160 μM (ferulic acid) or 5-450 μM *p*-coumaric and caffeic acid), were prepared to a total volume of 1 mL in a solution of resuspension buffer. A GBC Cintra™ spectrophotometer was then used to measure the drop in absorbance the addition of 6 μg (corresponding to $3.2x10^{-10}$

moles) of purified *Bs*PAD enzyme causes to this series of substrate concentrations. Absorbance changes where measured over a timeframe of 50-80 s, within a linear region of the observed spectra. Each sample was prepared and measured in triplicate. Michaelis Menten kinetics (V_{max}, K_M and k_{cat} as well as k_{cat}/K_M) were approximated for the wild type protein and three substrates (*p*-coumaric-, ferulic- and caffeic acids) using Beer's law and the equations described in section chapter 2, section 2.6.1. Phloretic acid, which lacks carbon chain structures essential for its conversion by *Bs*PAD, served as a negative control and was tested at a concentration of 50-400 µM. The same assay was carried out for each mutant and ferulic acid as a substrate. Michaelis Menten kinetics determined for the wild type protein varied by several orders of magnitude from literature reference values. This was found to be a consequence of signal saturation at higher substrate concentrations and led to abandonment of any further spectrophotometric assays.

Phloretic acid was additionally tested as a potential inhibitor of *Bs*PAD. For this, the effect of increasing concentrations of phloretic acid on a decarboxylation reaction was shown. In the first set of measurements, decarboxylation in a 125 µM *p*-coumaric acid solution was compared to reactions containing 125 µM *p*-coumaric and 0, 125 or 250 µM phloretic acid. Measurements were performed as described above. For the second set, decarboxylation in a 200 µM *p*-coumaric acid solution was compared to the same reaction in solutions containing 200 µM *p*-coumaric acid and 0, 200 or 400 µM phloretic acid.

3.2.3.2 HPLC analysis

As an alternative to UV spectroscopy, a HPLC assay was developed which allowed for qualitative and semi-quantitative observation of substrate degradation and product formation by the *Bs*PAD^{wt} and the active site mutant proteins. Enzyme reaction samples were prepared by incubating 6 µg ($3.5x10^{-10}$ moles) of purified protein in 1 mL of buffer containing 1 mM ferulic acid. The same assay was also carried out using *Bs*PAD^{wt} as well as the weakly active mutants Tyr11Phe, Tyr13Phe, Tyr19Ala and Thr98Ala and a substrate concentration range of 0.25 - 2.5 mM ferulic acid. Enzyme-free buffer solution with ferulic acid was used as a negative control and furthermore served in the preparation of a substrate calibration curve.

After 6 h of incubation at 30°C in a waterbath, 500 µL of ethyl acetate were added to 500 µL of each enzyme reaction and centrifugation at 14,000 rpm for 2 min separated aqueous and organic phases. The latter was retrieved and evaporated overnight to obtain decarboxylated product and non-converted substrate. Prior to running, samples were re-dissolved in 50% methanol. HPLC analysis was carried out on a Waters Alliance® high performance liquid chromatograph, using a SunFire™ C18 column (internal diameter 4.6 x 250 mm, 5 µm particle size) at a temperature of 25°C and a flow rate of 1 mL/min. Running conditions were optimized using substrate and product standards to achieve optimal separation of peaks. Calibration and protein reaction samples were then eluted in an isocratic flow of 50% solvent A (water) and 50% solvent B (methanol + 0.1% acetic acid). Samples were injected in 50 µL aliquots, each prepared and run in duplicate. Peaks were then detected at a wavelength range from 210-400 nm and evaluated using the Empower™ 3 chromatography data software. Peak areas of ferulic acid and vinyl guaiacol were determined at 280 nm and compared to the standard curves for quantitative analysis.

3.2.3.3 GC analysis

The HPLC analysis described above provided data on comparative activity between the wild type and mutant *Bs*PAD enzymes. Due to solubility problems at substrate concentrations exceeding 2.5 mM however, no Michaelis Menten kinetics could be determined. This problem was overcome by optimising the ferulic acid stock preparation with the kind help of Dr Aashrita Rajagopalan from the University of Graz. Solubility was maintained into concentrations exceeding 10 mM by adjusting the buffer solution's pH while solubilising the acidic substrate. In addition, repeated sonication of the solution in a waterbath was required during the solubilisation process.

After this optimisation of substrate solubility, a GC based assay was developed to determine full Michaelis Menten kinetics for the wild type and a subset of weakly active mutant proteins (Tyr11Phe, Tyr13Phe and Tyr19Ala). For preparation of product standards, a 0.5-10 mM concentration range of 3-methoxy-4-vinyl-phenol was prepared in ethyl acetate and injected onto a HP5 column using a HP-6890 gas chromatograph from Agilent. Helium was used as a carrier gas, the detector temperature was set to 320 °C. For the column, a temperature gradient from 90-

180 °C was applied at a rate of 3 °C per minute and peak areas were measured for 8 different substrate concentration standards. As previous GC experiments had shown the presence of product peaks in ferulic acid standards, the injector temperature was lowered to 165 °C. When testing a 10 mM ferulic acid control sample under the conditions described above, this did not yield a significant product peak. For determination of enzyme kinetics, purified protein was added to 1.0 mL of buffer containing 0.5 – 10.0 mM ferulic acid and the samples were incubated without shaking at 30 °C. The amount of protein added to the reaction samples had to be optimised for detection and varied between wild type and mutant *Bs*PAD, ranging from 0.3 µg (1.6×10^{-11} moles, *Bs*PAD^{wt}) to 3 µg (1.6×10^{-10} moles, *Bs*PAD^{Tyr19Ala}) and 6 µg (3.2×10^{-11} moles, *Bs*PAD^{Tyr11Phe} and *Bs*PAD^{Tyr13Phe}). The reaction was stopped after 20 min after which substrate and product could be extracted twice in 500 µL of ethyl acetate. The combined samples could be injected onto a HP5 GC column using the method outlined above. Each reaction was performed in duplicate to obtain statistically significant results. Using the previously prepared standards, initial reaction velocities (V_0) could eventually be plotted against substrate concentrations. The Michaelis constant and maximum reaction velocity V_{max} were approximated and k_{cat} values could be determined subsequently as described in section chapter 2, section 2.5.1.

3.3 Results

3.3.1 Production and purification of wild type BsPAD

The wild type *Bs*PAD protein could be produced recombinantly in an *E. coli* BL21(DE3) expression strain and was purified to high homogeneity. A litre of culture routinely yielded approximately 35 mg of protein after two purification steps. The sharp elution peak resulting from increased concentrations of imidazole during Ni^{2+}-affinity chromatography indicates that the protein carried the desired hexa-histidine tag. Elution of *Bs*PAD in low imidazole buffer after his-tag cleavage also indicated that the N-terminal HRV-3C protease recognition site was intact and allowed for removal of the small peptide tag with high precision. Visualisation of purified *Bs*PAD protein on a denaturing SDS PAGE gel confirmed the purity

previously suggested by elution of a single peak during size exclusion chromatography. The elution volume during gel filtration as well as the protein band's position on a gel corresponded to a molecular weight of approximately 40 kDa (dimer in solution) and 20 kDa (monomer) respectively (see Fig. 3.5). This is in accordance with the expected literature values (18). The purified protein preparations were stable at 4 °C for several days, protein frozen in liquid nitrogen retained activity over many months. From the latter, *Bs*PAD kinetic properties could thus be analysed with respect to substrate degradation and product formation.

3.3.2 Preparation of a BsPAD mutant library

Using site-directed mutagenesis, eight active site *Bs*PAD mutants were prepared. These included seven single mutants (Tyr11Phe, Tyr13Phe, Tyr19Ala, Arg41Ala, Glu64Ala, Thr68Val and Thr98Val) as well as one double mutant (Tyr11Phe/Tyr13Phe). Sequencing of the mutant YSBLIC:*Bs*PAD plasmids confirmed the successful introduction of each of the listed amino acid alterations without the appearance of other, undesired, mutations (see Fig. 3.6).

3.3.3 Production and purification of a BsPAD mutant library

After confirmation of the desired mutation by sequencing, all mutant genes were expressed and translated into soluble protein in a similar way to the wild type protein. No significant *Bs*PAD inclusion bodies were observed on SDS PAGE gels (see Fig. 3.5). Within the library, protein production levels varied but were generally slightly reduced compared to the wild type PAD enzyme.

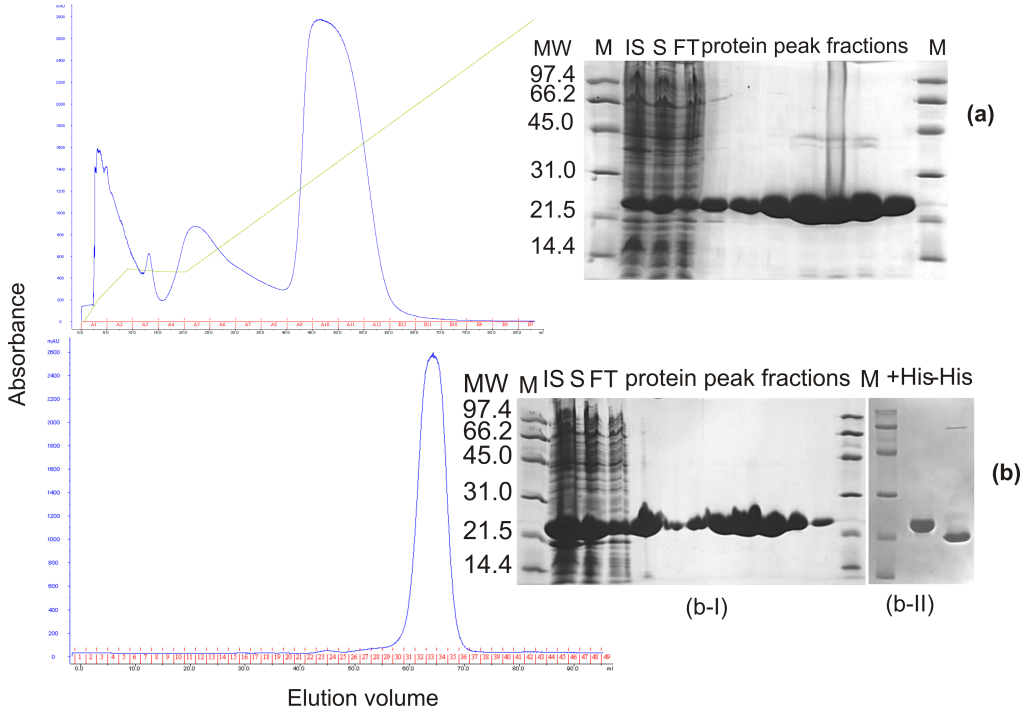


Figure 3.5: *Bs*PAD^{wt} protein was purified by Ni^{2+}-affinity (a) and size-exclusion chromatography (b). The protein elutes from a size-exclusion column in a single peak which does not show any major impurities on an SDS PAGE gel. The protein's molecular weight of approximately 22 kDa (b-I) is reduced to 19 kDa after removal of the hexa-histidine tag (b-II).

```
BsPADwt            1 MENFIGSHMIYIYENGWEYEIYIKNDHTIDYRIHSGMVAGRWVRDQEVNI     50
                     ||||||||||:||:||||||||||||||||||||||||||||||||||||
BsPADTyr11,13      1 MENFIGSHMIFIFENGWEYEIYIKNDHTIDYRIHSGMVAGRWVRDQEVNI     50

BsPADwt           51 VKLTEGVYKVSWTEPTGTDVSLNFMPNEKRMHGIIFFPKWVHEHPEITVC    100
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
BsPADTyr11,13     51 VKLTEGVYKVSWTEPTGTDVSLNFMPNEKRMHGIIFFPKWVHEHPEITVC    100

BsPADwt          101 YQNDHIDLMKESREKYETYPKYVVPEFAEITFLKNEGVDNEEVISKAPYE    150
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
BsPADTyr11,13    101 YQNDHIDLMKESREKYETYPKYVVPEFAEITFLKNEGVDNEEVISKAPYE    150

BsPADwt          151 GMTDDIRAGRL    161
                     ||||||||||
BsPADTyr11,13    151 GMTDDIRAGRL    161
```

Figure 3.6: A sequence alignment between the Tyr11Phe/Tyr13Phe mutant (top line) and the *Bs*PAD^{wt} protein (bottom line) shows the site-specific introduction of mutations into the active site.

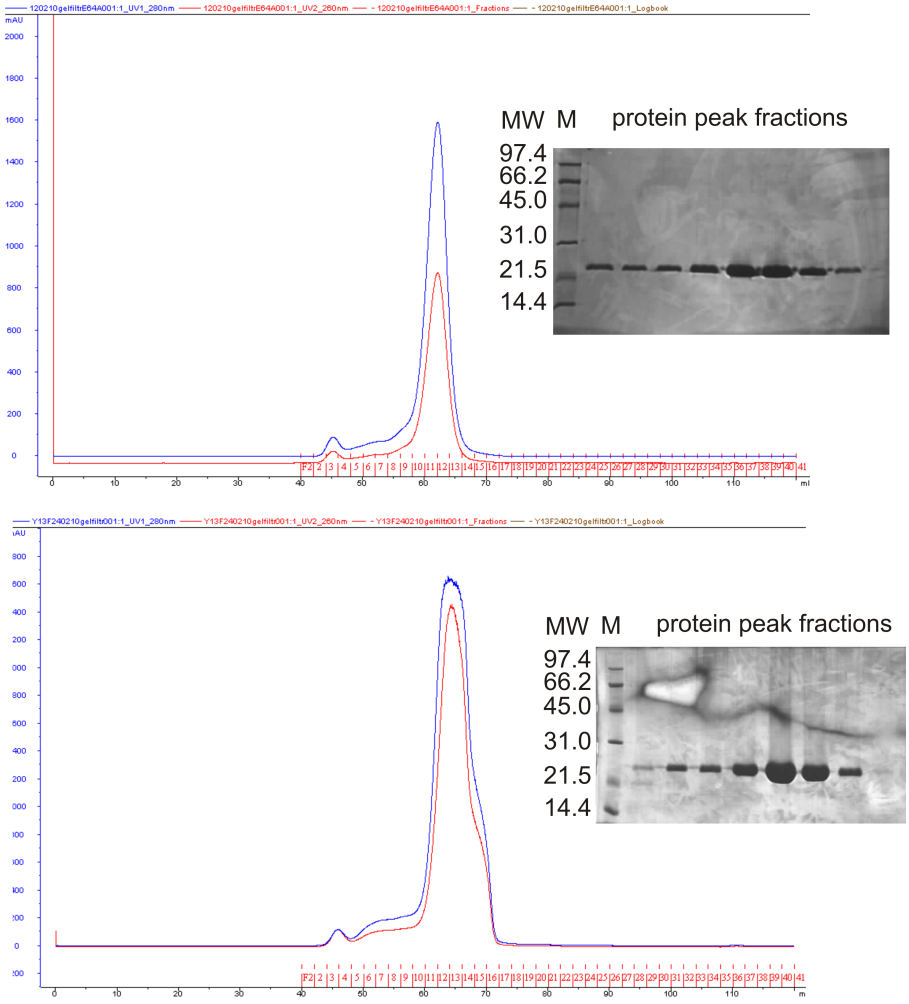


Figure 3.7: All mutant *Bs*PAD genes were expressed and translated into soluble protein. Size-exclusion chromatograms and SDS PAGE gels of BsPAD Thr68Val and Tyr19Ala mutants are shown to represent proteins with conservative and non-conservative mutations respectively.

3.3.4 Analysis of enzyme kinetics

3.3.4.1 Spectrophotometric analysis of BsPAD wild type and active site mutants

To observe decarboxylation of ferulic, *p*-coumaric and caffeic acid by *Bs*PAD, a spectrophotometric assay was developed previously (142). The assay makes use of the drop in absorbance at 310 nm between the phenolic acid substrate and its decarboxylated styrene derivative products. When a concentration range of 5-180 or 5-450 μM of substrate was measured for ferulic as well as *p*-coumaric and caffeic

acid respectively, apparent Michaelis-Menten kinetics were observed for the wild type protein. In addition, decarboxylation could be observed for the Tyr19Ala mutant (see Fig. 3.8 d).

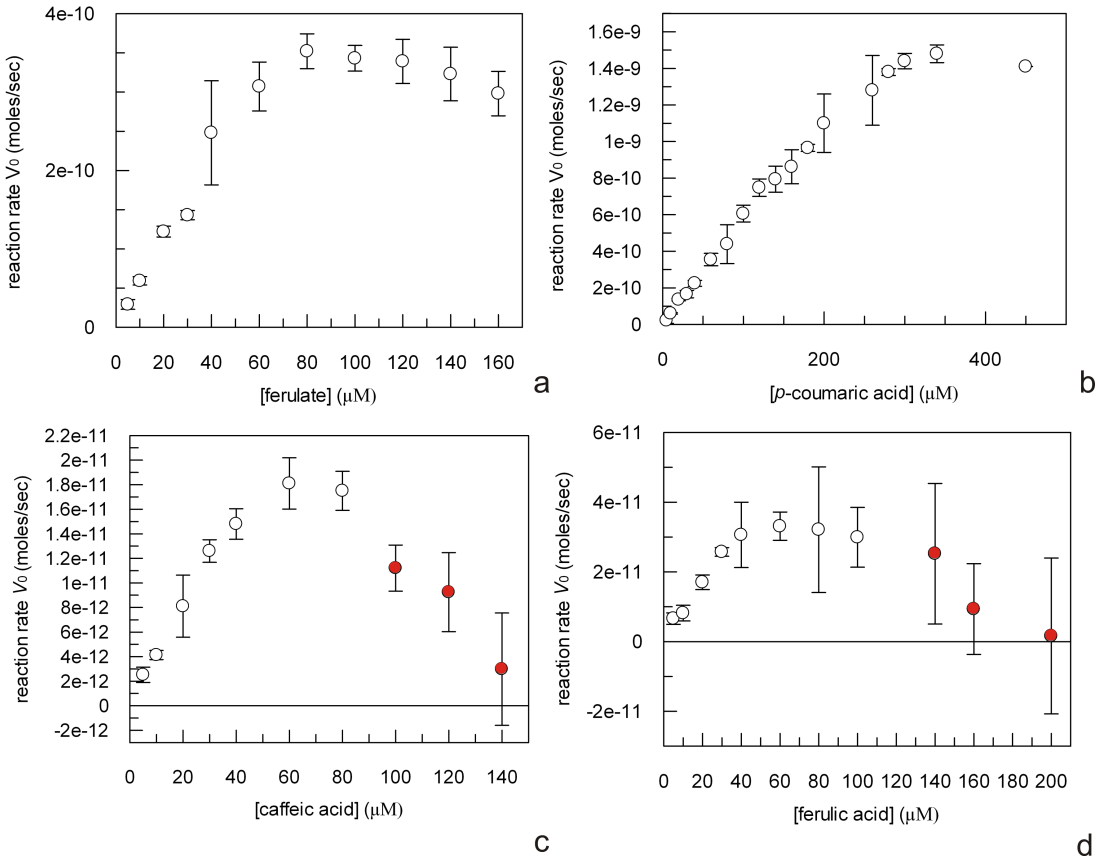


Figure 3.8: Concentration dependent decarboxylation could be observed for *Bs*PAD^{wt} protein and ferulic acid, *p*-coumaric acid and caffeic acid substrates (a-c). Of the mutant library, only the Tyr19Ala mutant showed activity, albeit at a significantly reduced level compared to the wild type protein (d). The *Bs*PAD^{wt} decarboxylation of caffeic acid and the Tyr19Ala mutant's reaction with ferulic acid appear to be subject to substrate inhibition (red data points).

On plotting concentration dependent reaction velocities for the *Bs*PAD^{wt} reaction, V_{max}, K_M and k_{cat} values of 3.5 x 10^{-10} moles/s, 35 µM and 1.0 s^{-1} were estimated for the ferulic acid substrate, 1.5 x 10^{-9} moles/s, 180 µM, 4.3 s^{-1} for *p*-coumaric acid and 1.8 x 10^{-11} moles/s, 25 µM and 0.05 s^{-1} for caffeic acid. For the latter, a steep drop in reaction rates was observed at concentrations above 60 µM, indicating the presence of substrate inhibition. Kinetic constants estimated for the active site mutant,

BsPAD^{Tyr19Ala}, showed a significantly decreased maximum velocity and turnover numbers as well as an increased Michaelis constant of 3.3 x 10^{-11} moles/s, 0.10 s^{-1} and 20 µM respectively. No activity could be observed for any other mutant within the course of the assay. The reaction constants for the *Bs*PAD wild type protein differed significantly from previously published literature values (18). On closer inspection of the absorbance values, especially for ferulic acid, it was observed however that concentrations > 80-100 µg/mL led to signal saturation at the given wavelength of 310 nm. This makes measurements in this area unreliable (as can be seen from the high standard deviations in figure 3.8 c and d).

Phloretic acid is very similar in structure to *p*-coumaric acid but lacks the carbon chain structures suggested to be essential for its conversion by *Bs*PAD, namely a conjugated carbon chain essential for the catalytic electron relay (see Fig. 3.9). The molecule is thus not expected to be decarboxylated by the enzyme, could however act as a possible inhibitor that might block the *Bs*PAD active site. However, no clear competitive inhibition could be observed when the molecule was assayed against ferulic or *p*-coumaric acid (see Fig. 3.9).

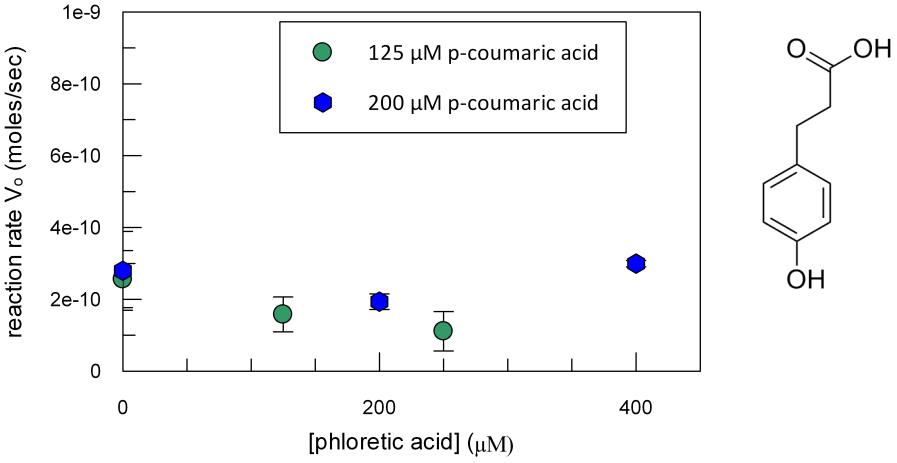


Figure 3.9: When tested as a potential inhibitor, phloretic acid (right) did not appear to compete for the enzyme's active site with the *p*-coumaric acid substrate. The substrate : inhibitor ratios are shown above each data point.

3.3.4.2 HPLC analysis

To overcome the limitations of signal saturation observed in spectrophotometric assays, a HPLC-based assay was developed. While the overall concentration range

was still limited, this assay allowed for the qualitative and semi-quantitative comparison of decarboxylation by wild type and mutant enzymes.

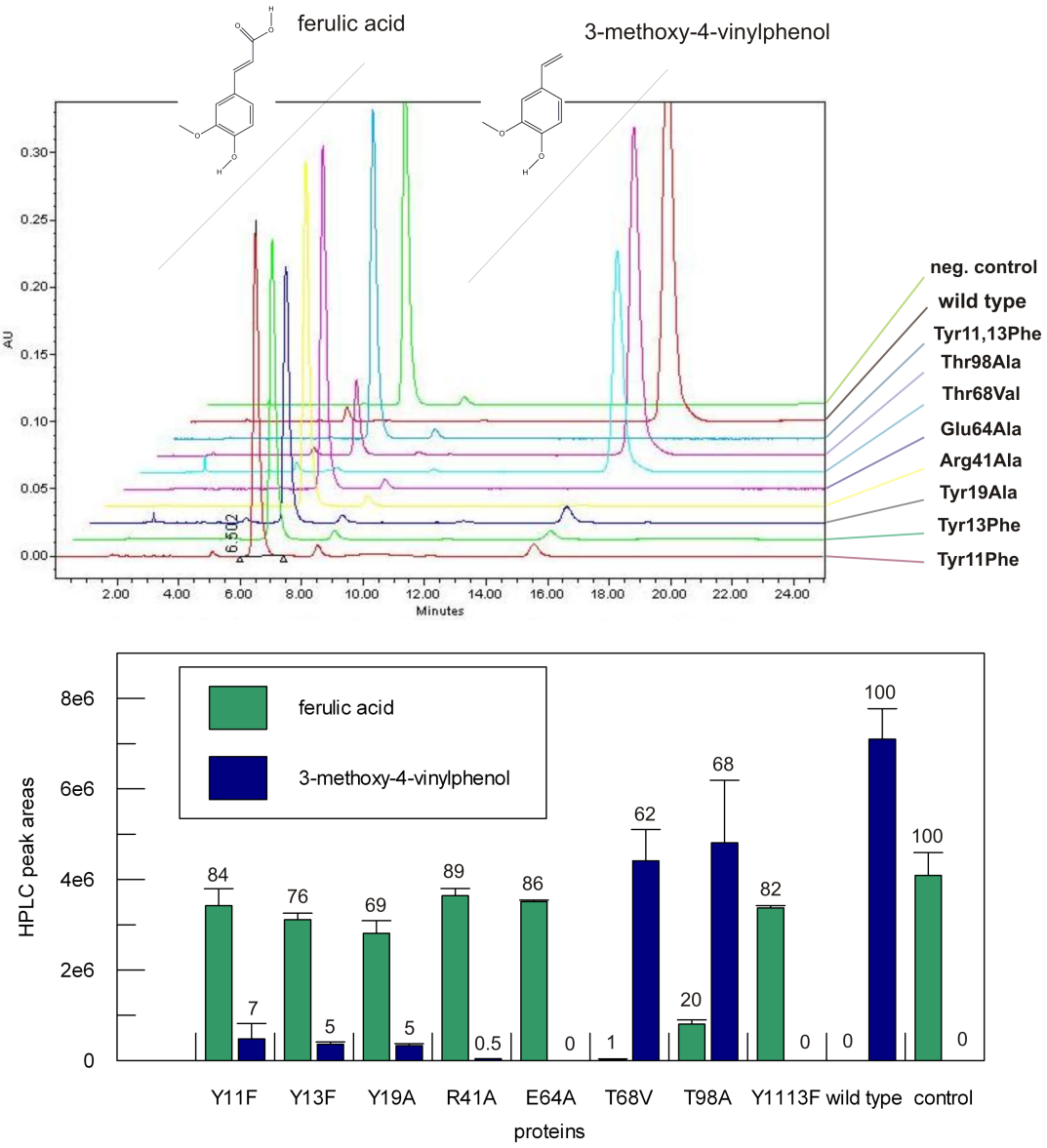


Figure 3.10: A HPLC peak chromatogram shows ferulic acid degradation and vinyl guaiacol formation by *Bs*PAD^{wt}, proteins of the active site mutant library and an enzyme free reaction serving as a negative control (top). The experiment allowed for qualitative and semi-quantitative evaluation of decarboxylation of ferulic acid into vinyl guaiacol (bottom image, green and blue bars respectively). The numbers on top of each column represent substrate turnover/product formation relative to the wild type protein in per cent (%).

When comparing catalysis at a fixed concentration of ferulic acid (1 mM) and after a reaction time of six hours, all substrate was converted by *Bs*PAD^{wt}. The 3-methoxy-

4-vinylphenol product peak was observed after an expected retention time of 15.5 min. The negative control (enzyme free reaction) presented a peak for the non-converted substrate after a retention time of 5.2-6.5 min, depending on the experimental setup. Both correspond to standards run in advance of the reaction samples (results not shown). The Thr68Val and Thr98Ala mutants also retained high activity levels, with an apparent product formation above 50% compared to the wild type protein. Alterations of the three tyrosine residues in positions 11, 13 and 19 led to a significant reduction in enzymatic activity with 16, 24 and 31% respectively of substrate being converted within the course of the assay. The Arg41Ala and Glu64Ala single mutants as well as the Tyr11Phe,Tyr13Phe double mutant did not yield any detectable formation of the vinyl guaiacol product.

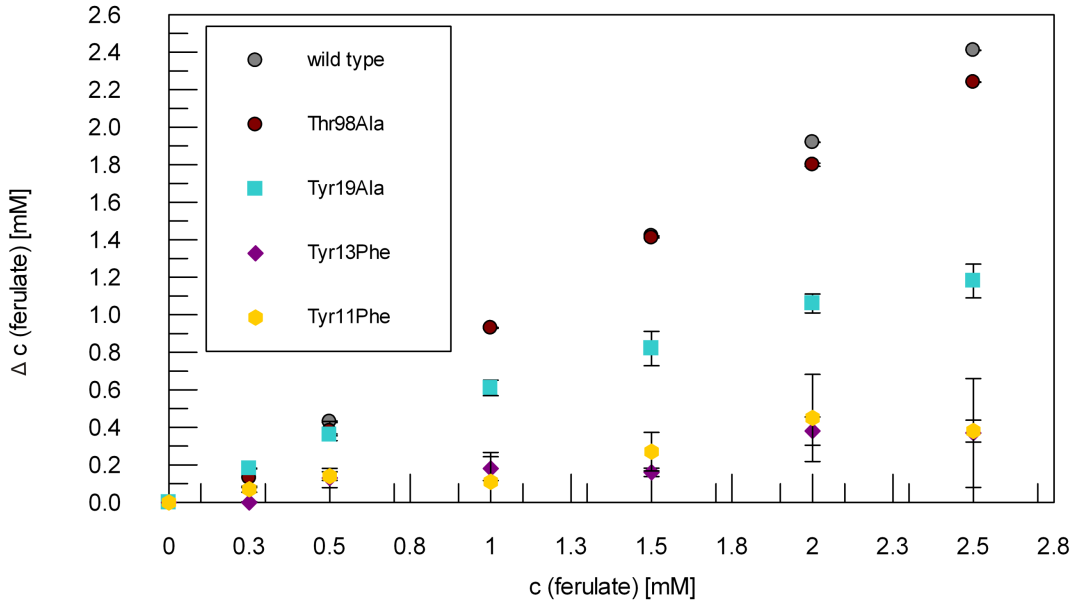


Figure 3.11: Initial attempts were made to determine kinetic constants for the decarboxylation of ferulic acid by *Bs*PAD^{wt} and mutant enzymes. Only the Tyr11Phe, Tyr13Phe and Tyr19Ala mutants approach maximum velocities within the concentration range assayed. The Thr98Ala variant and wild type protein show a linear relationship between substrate concentration and substrate consumption (Δ c ferulic acid).

On analysing ferulic acid concentration dependent decarboxylation by *Bs*PAD^{wt} and the weakly active mutants Tyr11Phe, Tyr13Phe, Tyr19Ala and Thr98Ala, only the first two mutants approach saturating concentrations. The wild type protein and other active site mutants however show a linear range of substrate turnover/product formation for the concentrations assayed. Their data did thus not allow for the

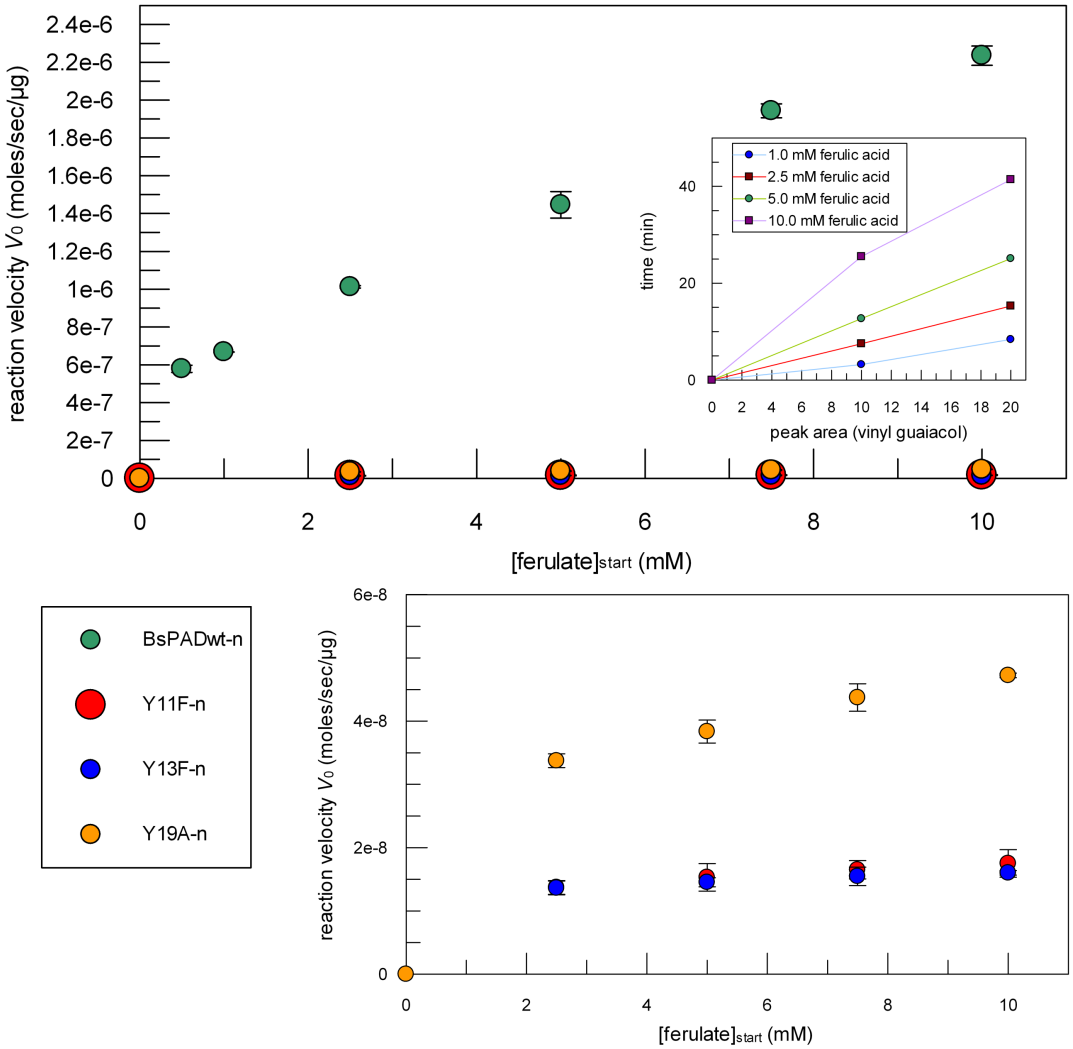
determination of Michaelis constants. Catalytic rate and efficiency could not be determined with the available data as the assay time of six hours exceeds the linear range of product formation.

3.3.4.3 GC analysis

To obtain quantitative data and, eventually, rate constants for the decarboxylation of ferulic acid by *Bs*PAD^{wt} and three of the weakly active mutant enzymes, the experimental set-up was optimised and samples adapted for analysis by gas chromatography. The initial limitations of substrate concentration were overcome by adapting the solution's pH and repeated sonication during the solubilisation process. With this method, solutions exceeding concentrations of 10 mM ferulic acid were easily prepared. In addition, the initial reaction pH remained equal for all samples analysed.

The GC method developed allowed for detection of the vinyl guaiacol product with a retention time of approximately 9.6 min, corresponding to a column temperature of approximately 120 °C. When a concentration range from 0.5 – 10 mM of the styrene product was assayed, a linear calibration curve could be prepared from the resulting peak areas (see Fig. 3.12, inlet). In initial experiments, inlet temperatures of 250 °C were shown to degrade ferulic acid into vinyl guaiacol on entering the chromatography column. A control reaction with 10 mM ferulic acid was thus run to exclude such signal interference from residual substrate in reaction samples. With a reduced inlet temperature of 165°C, no product peak could be detected at the corresponding retention time.

For enzyme reaction samples, product formation was analysed after 20 min. This allowed for the determination of reaction rates V_0 within a time frame corresponding to linear product formation, see Fig. 3.12. Reaction rates of ferulic acid-concentration dependent vinyl guaiacol formation were then plotted, allowing for the estimation of reaction constants V_{max} and K_M. From these, the turnover number k_{cat} and catalytic efficiency values (k_{cat}/K_M) could be determined for the wild type as well as the assayed mutant enzymes and are listed in table 3.3. Due to the lack of data points at substrate concentrations below 0.5 and 2.5 mM (for wild type and mutant enzymes respectively) the increase in initial reaction velocity may in fact be steeper and the Michaelis constants lower than the approximated values.



*Bs*PAD variant	V_{max} [moles/s/µg]	K_M [mM]	k_{cat} [s^{-1}]	k_{cat} / K_M [s^{-1}/M]
Wild type	6.8E-10	< 4.5	4.3	9.6 x 10^3
Tyr11Phe	1.1E-10	< 1.0	0.35	3.5 x 10^2
Tyr13Phe	9.7E-11	< 1.0	0.31	3.1 x 10^2
Tyr19Phe	1.5E-10	< 0.75	0.95	1.3 x 10^3

Figure 3.12/Table 3.3: : Kinetic constants of ferulic acid decarboxylation were estimated for the BsPAD^{wt} protein and three of its active site mutants within a period of linear activity (20 min, top graph, inlet). After extraction of the substrate and product into ethyl acetate, measurements were taken using an optimised gas chromatography assay. Vinyl guaiacol peaks observed at a column temperature of 120 °C were quantitated using a previously prepared standard curve. The mutant activities are also depicted individually (bottom graph).

The BsPAD^{wt} protein shows the highest reaction velocity with an estimated V_{max} of 2.3 x 10^{-6} moles/s/µg (where V_{max} units were normalised for protein concentration as the amount of enzyme added to the reactions varied between wild type and mutant assays). With 4.5 mM, the Michaelis constant is equally the highest of all the variants tested. The turnover number of wild type protein amounts to 4.3 x 10^4 s^{-1} while the k_{cat}/K_M value is 9.6 x 10^6 s^{-1}/M. Replacing a tyrosine in position 19 with an alanine led to a 4-fold reduction in the enzyme's k_{cat} value when compared to the wild type protein, as well as a 9-fold reduced Michaelis constant. The mutation of tyrosine in position 11 and 13 to phenylalanine had a comparable effect on both mutants and led to the strongest reduction in activity of all enzymes assayed. The catalytic turnover number was reduced by a factor of 10 when compared to BsPAD^{wt} and was accompanied by a 4-fold decrease in K_M.

3.4 Discussion

The microbial decarboxylation of phenolic acids has long been of interest for biotechnological applications. Many of the enzymes known to catalyse this reaction have thus been identified, isolated and characterised from prokaryotic and eukaryotic sources. Previous kinetic studies on mutant PAD enzymes and deuterium exchange experiments, as well as structural information from unrelated proteins with similar substrate spectra, have allowed for the proposal of several mechanisms. Except for one substrate analogue complex however, a ligand-bound structure of PAD enzymes remained elusive and the suggested mechanisms awaited structural conformation. Despite knowledge of the enzyme's *apo*-structure, the phenolic acid decarboxylase from *Bacillus subtilis*, *Bs*PAD, previously lacked kinetic as well as structural information on ligand binding. The aim of this project was thus to provide both. A mutant library of previously proposed catalytic residues was prepared and the mutations' effects on catalysis assayed. It was thus hoped to create a weakly active PAD enzyme and solve its structure with a substrate, intermediate or product molecule "trapped" in the active site.

Each of the mutations introduced to the *Bs*PAD active site affected catalysis, however to varying degrees. The first two alterations analysed were the mutation of a

Tyr11 and Tyr13 to phenylalanine. In enzymes such as the vanillyl alcohol oxidase (VAO) from *Penicillium simplicissium* hydroxycinnamate and the hydratase-lyase (HCHL) from *Pseudomonas fluorescens*, ligand bound structures have shown their substrates' phenolic hydroxyls to be coordinated by tyrosine pairs during enzymatic catalysis (136,139). In contrast to this, models of ligand binding in a high-resolution *apo*-structure as well as mutagenic studies of the *Lactobacillus plantarum* PAD enzyme (*Lp*PAD) suggested the two tyrosines to form hydrogen bonds to the phenolic acid's carboxylate. They could thus stabilise the CO_2 leaving group. In either position, both residues could be important determinants of substrate binding and specificity (see Fig. 3.13). The change from tyrosine to phenylalanine is conservative and replaces the wild type side chain with a side chain of similar size. The effect on protein folding was thus expected to be minimal and the mutant gene sequences were translated into soluble proteins. The mutation does however remove the side-chain's hydroxyl group and thus its ability to form potential hydrogen bonds with a substrate. When analysed by UV-VIS spectroscopy, no substrate conversion could be observed within the course of the assay, indicating that the mutants would be either inactive or catalysis was reduced to a level too low to be observed with this method. The latter was shown to be the case using HPLC based analysis of ferulic acid decarboxylation. After a period of six hours, only small amounts of substrate could be converted by either of the two mutants. When kinetic constants were determined, turnover numbers were significantly reduced compared to the wild type enzyme which corresponds to previous findings for the *Lp*PAD protein. All three experiments thus indicate an importance for both residues, Tyr11 and Tyr13, in catalysis. Although, when compared to the wild type, the reduced Michaelis constant may be an indicator of tighter substrate binding, this value has to be seen in context of the overall reduced activity and catalytic efficiency (k_{cat}/ K_M). The hydrogen bonding functions of both tyrosine side chains are thus likely to be involved in substrate binding and important for overall catalysis.

Each individual residue is however not absolutely essential to catalysis as the lack of one hydrogen bond donor only reduces the overall reaction rate. Removal of both, as it was shown with the Tyr11Phe/Tyr13Phe double mutant, completely abolishes catalytic activity. In neither UV-spectrophotometric nor HPLC based assays, i.e. short or long term analysis of substrate turnover, could any ferulic acid

decarboxylation be observed. It is thus likely that, upon substrate binding, the phenolic acid is recognised and bound to the active site by way of hydrogen bond formation to Tyr11 and Tyr13. The lack of one side chain may, to some degree, be compensated for by the binding abilities of the other. Lack of both renders the protein completely inactive. Which part of the phenolic acid substrate is bound by the two tyrosine residues cannot be defined by kinetic data alone and requires the presence of structural data which will be discussed in chapter 4.

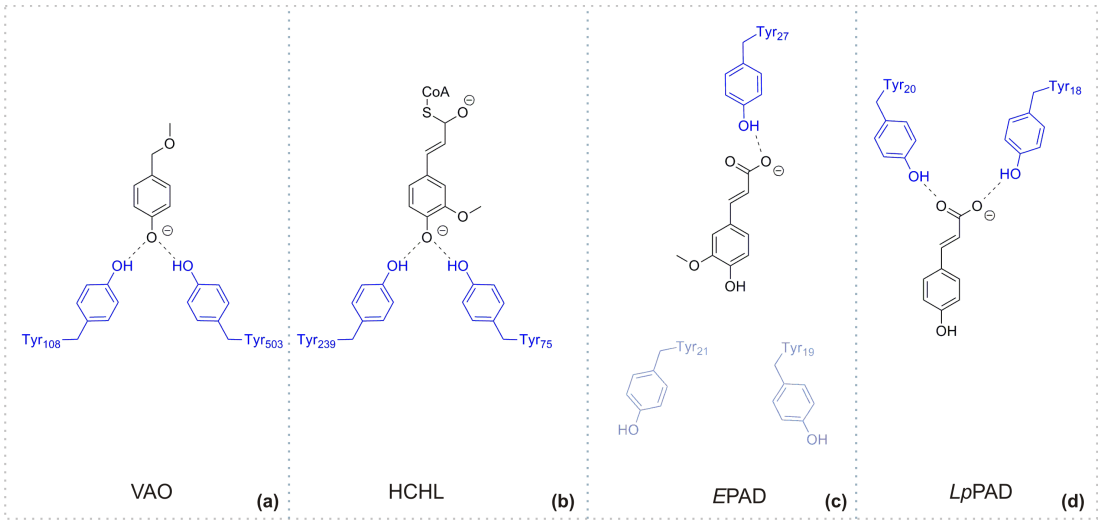


Figure 3.13: Whereas the unrelated but mechanistically similar enzymes VAO and HCHL (a and b) suggest a role for a role Tyr11 and Tyr13 in substrate binding at the *para*-hydroxyl group, modelling of the phenolic acid ligand into an *Lp*PAD structure has implied an important function in recognition and binding of the carboxylate moiety (d). In contrast to both proposals, a mechanism independent of either tyrosine residue has been suggested based on the complex structure of an *Enterobacter* PAD enzyme (c) [adapted from (60,134,136,139)].

A further tyrosine side chain in close proximity to positions 11 and 13 is Tyr19. When the *apo*-structure of a phenolic acid decarboxylase from *Lactobacillus plantarum* (*Lp*PAD) was solved, the authors modelled a *p*-coumaric acid molecule into the enzyme's active site (60). The *Lp*PAD Tyr26 side chain, which is equivalent to *Bs*PAD Tyr19, was found to be part of a hydrophobic pocket surrounding the central, non-polar core of the phenolic acid substrate. This pocket is thought to contribute to substrate binding and specificity by forming a confined space for the aromatic ring and aliphatic side-chain. When Tyr19 is replaced by an alanine in *Bs*PAD, catalysis is reduced, albeit at lower levels than for the Tyr11 and Tyr13

mutants described above. Turnover can still be observed in real-time experiments such as UV-VIS spectroscopy. The Tyr19 equivalent in *Lp*PAD has previously been suggested to be part of a flexible loop involved in restricting access to the enzyme's active site (60). The reduced turnover and catalytic efficiency of the mutant could thus be explained by the enzyme's altered efficiency in switching between open and closed active site conformations. Both substrate access and release would be impaired. Alternatively, as Tyr19 only constitutes a small part of this loop and the effect of its mutation on catalysis is considerable, the residue may interact with substrate or solvent. This possibility has to be analysed in the context of a ligand complex structure and is discussed in chapter 4. A third function of this residue, proposed for the conserved equivalent in *E*PAD, could be the fixation of the substrate's carboxylate moiety by hydrogen bonding. In view of the significant residual activity observed however, this role can more likely be assigned to alternative residues, such as Tyr11 and or Tyr13, as described above.

For decarboxylation to be initiated, the *para*-hydroxyl group's proton has to be removed by a catalytic acid. A side-chain able to fulfil this function would have to be positioned at one end of the active site pocket (which end cannot be defined for certain in the absence of a ligand complex). Based again on both *apo*-structure and modelling of the ligand complex (60) however, the residue most likely to act as the catalytic acid was proposed to be a glutamate in position 64 (see Fig. 3.14). Mutation of this amino acid to alanine renders the mutant protein completely inactive, supporting its function as one of the main catalytic residues. A similar mechanism was proposed for the *E*PAD protein. The catalytic residue assigned to this step however (Glu134), differs significantly from the findings for *Bs*PAD described here and for *Lp*PAD (60).

For catalysis to be initiated by deprotonation, the Glu64 side chain has to be deprotonated and hence negatively charged in the absence of a ligand. Furthermore, the phenolic acid's *para*-quinone methide intermediate will be negatively charged after its deprotonation. This surplus of free electrons (negative charge) is energetically unfavourable and requires stabilisation by a positively charged active site residue (see Fig. 3.14). Based on the *Bs*PAD *apo*-structure, the substrate-binding model for the *Lp*PAD homologue and, again, a similar function in the non-related

VAO and HCHL enzymes (60,136,139), this stabiliser is most likely an arginine in position 41. In accordance with this proposed function as well as kinetic data from a comparable *Lp*PAD mutant, replacement of Arg41 with an alanine residue abolishes all catalytic activity in the enzyme. The stabilisation of high-energy intermediates and thus reduction of activation energy for a chemical reaction is the key feature of enzyme-based catalysis (143). Loss of a stabilising side chain such as arginine should thus return reaction rates close to those of a non-catalysed decarboxylation reaction. The latter would not be detectable within the course of the assays described. Again, the assignment of catalytic residues described here differs compared to the previously proposed EPAD mechanism in which an asparagine (Asn23) and a water molecule have been suggested to create an appropriate polar environment for deprotonation (134).

Both Arg41Ala and Glu64Ala are non-conservative mutations. Replacement of the charged, extended side chains with the short, non-polar alanine could lead to disruptions in protein folding. Both proteins were however predominantly produced in their soluble form by the *E. coli* host, the Glu64Ala mutant also crystallised under varying conditions (see chapter 4). It was thus assumed that the observed changes in catalytic activity are more likely the consequence of a lack in catalytic residues rather than partially or misfolded mutant proteins.

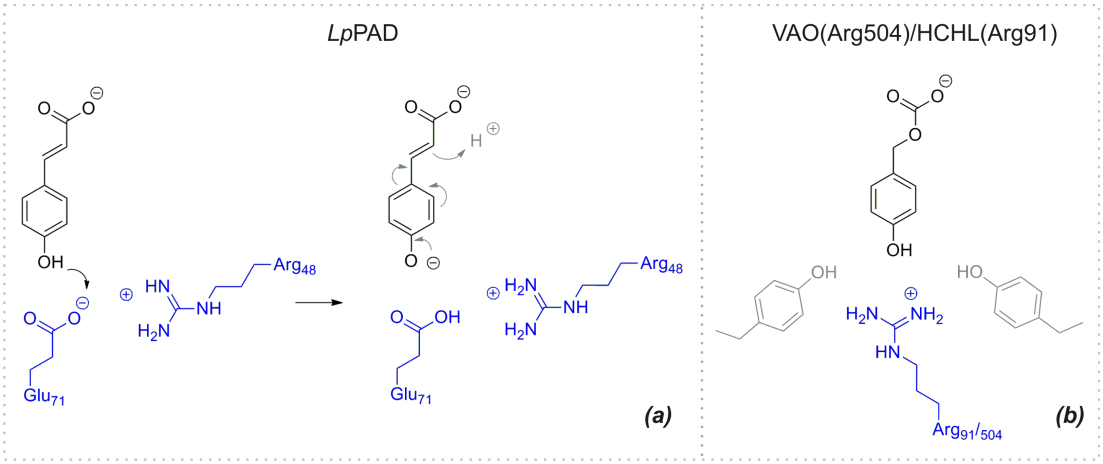


Figure 3.14: Modelling of a PA ligand into the *Lp*PAD active site as well as previous mutagenesis experiments have suggested an importance for the highly conserved *Bs*PAD Glu64 and Arg41 side chains in deprotonation of the phenolic hydroxyl and stabilisation of negative charges respectively (a). For the latter residue, similar observations of charge stabilisation have been made in the unrelated VAO and HCHL enzymes (b) [adapted from (60,136,139)].

When preparing the active site mutant library, the last two residues altered (from N- to C-terminus) where two threonines in position 68 and 98. Whereas the former underwent a conservative substitution to valine, the latter was changed to an alanine. In previous literature on *Lp*PAD, Thr68 (or rather its respective counterpart in the homologous enzyme) had been suggested to coordinate a water molecule and constitute part of a large polar network surrounding the substrate's hydroxyl group (60). The water molecule could serve as a proton donor during formation of the *para*-quinone methide intermediate and would thus be essential for catalysis to progress from the kinetically unfavourable intermediate to the decarboxylated styrene product. A similar step, without the assignment of an activating or coordinating amino acid residue, has been proposed for catalysis by the *E*PAD homologue. The change of Thr68 to a valine should replace the former with a sterically similar side-chain. As it lacks a hydroxyl group at the C8 carbon however, the valine side chain is non-polar. This property change reduces the mutant's catalytic activity and makes the residue a potential candidate for coordination of the catalytic water through polar interactions, as described above. As considerable residual activity is observed however, Thr68 could not be the only coordinating residue for this water. Its replacement by a non-polar side chain must thus be compensated for by another coordinating residue. Alternatively, activity of the valine mutant could be an indicator that another amino acid, rather than the catalytic water, acts as a proton donor to the C8 carbon. During preparation of the mutant library, preliminary, low resolution structural data became available for one of the mutant proteins, *Bs*PAD Tyr19Ala (see chapter 4). A highly conserved threonine in position 98 was found to be positioned in close proximity to an area of unmodelled electron density. As the latter was thought to represent a molecule of *p*-coumaric acid, Thr98 was suggested to be of potential catalytic importance. When the residue was replaced non-conservatively with an alanine however, activity levels remained close to those of the wild type protein, as could be seen from the long-term, HPLC based decarboxylation assay. It is thus likely that, despite its proximity to the substrate, Thr98 does not fulfil a significant catalytic function.

In summary, the mutant library prepared gave a first insight into the functional importance of various *Bs*PAD active site residues. Two tyrosines in position 11 and 13 are likely to be important for substrate recognition and binding. Whereas the lack

of one can be compensated for by the other, a double mutant is catalytically inactive. A third tyrosine in position 19 is less essential to catalysis itself, may however constitute part of a flexible loop which is involved in restricting access to the enzyme's active site. Alternatively, the residue may have other functions which are less obvious in the absence of structural data. Mutations in two of the amino acid residues completely abolish all enzymatic activity, supporting their proposed functions in catalysis. Arg41 is thought to be essential in stabilising negative charges arising from formation of the *para*-quinone methide intermediate. A glutamate in position 64 is the most likely catalytic acid/base and thought to remove a proton from the *para*-hydroxyl group – thus initiating catalysis. Thr68 may be important, albeit not essential, in the coordination of a catalytic water. The latter could in turn protonate the kinetically unfavourable intermediate and allow for formation of the styrene product. As for Tyr19, confirmation of this function requires structural data in addition to the kinetic information presented above.

The findings presented herein coincide and/or agree with kinetic and structural data presented previously for the *Lactobacillus plantarum* PAD protein (60). In contrast to this, the residues assigned to key functions in substrate binding and catalysis, differ signicantly from those proposed for the *Enterobacter* homologue EPAD (134).

Chapter 4: Structure determination of a *Bs*PAD mutant in complex with *p*-coumaric acid

4.1 Introduction

4.1.1 Structures of phenolic acid decarboxylases

To date, the three-dimensional structures of four bacterial phenolic acid decarboxylases have been determined. These include the PAD enzymes from *Lactobacillus plantarum* (*Lp*PAD), *Enterobacter* species Px6-4 (*E*PAD), *Bacillus subtilis* (*Bs*PAD) and *Bacillus pumilus* (*Bp*PAD) (60,134,144). Their coordinates have been deposited in the protein data bank (PDB) with the accession codes 2GC9, 3NX1, 2P8G and 3NAD respectively. All proteins show high degrees of sequence conservation and are equally well conserved in their secondary, tertiary and quarternary structures (see Figures 4.1 – 4.3).

The most significant feature of all phenolic acid decarboxylases is their common core of two sheets of five anti-parallel β-strands which assemble into an orthogonal β-sandwich. The latter is a common tertiary structure in enzymes and characterised by the spatial arrangement of the β-sheets into a flattened "sandwich-like" structure. As with the classical β-barrel, the last β-sheet (from N- to C-terminus) forms a network of backbone hydrogen bonds to the first, creating a stable and rigid pore like structure. β-barrels and sandwiches are often characterised by alternating polar and non-polar amino acids in the primary protein sequence. As a consequence of the β-sheet architecture, in which the amino acid side chains alternately face opposite directions, this allows for an overall structure of a hydrophilic core and a non-polar frame or *vice versa*. The former is particularly common in membrane proteins integrated into the cellular lipid bilayer. For cytosolic β-barrel proteins and PAD enzymes acting on small, hydrophobic substrates, a polar "wall" protecting a non-polar core and active site is commonly found (60,134,145).

In their three-dimensional structure, PAD enzymes show significant similarities to members of the lipocalin family (which are made up of four-stranded, anti-parallel β-sheets (144,146)) and the fatty acid binding proteins or FABPs (which show the PAD like arrangement of five-stranded anti-parallel β-sheets) (60,147). Both enzyme

families are known to bind small hydrophobic ligands and show extended flexible loops between the first two β-sheets of their core structure. These loops are involved in restricting access to the active site and are also common (in position and function) to the characterised PAD enzymes (60,144,146).

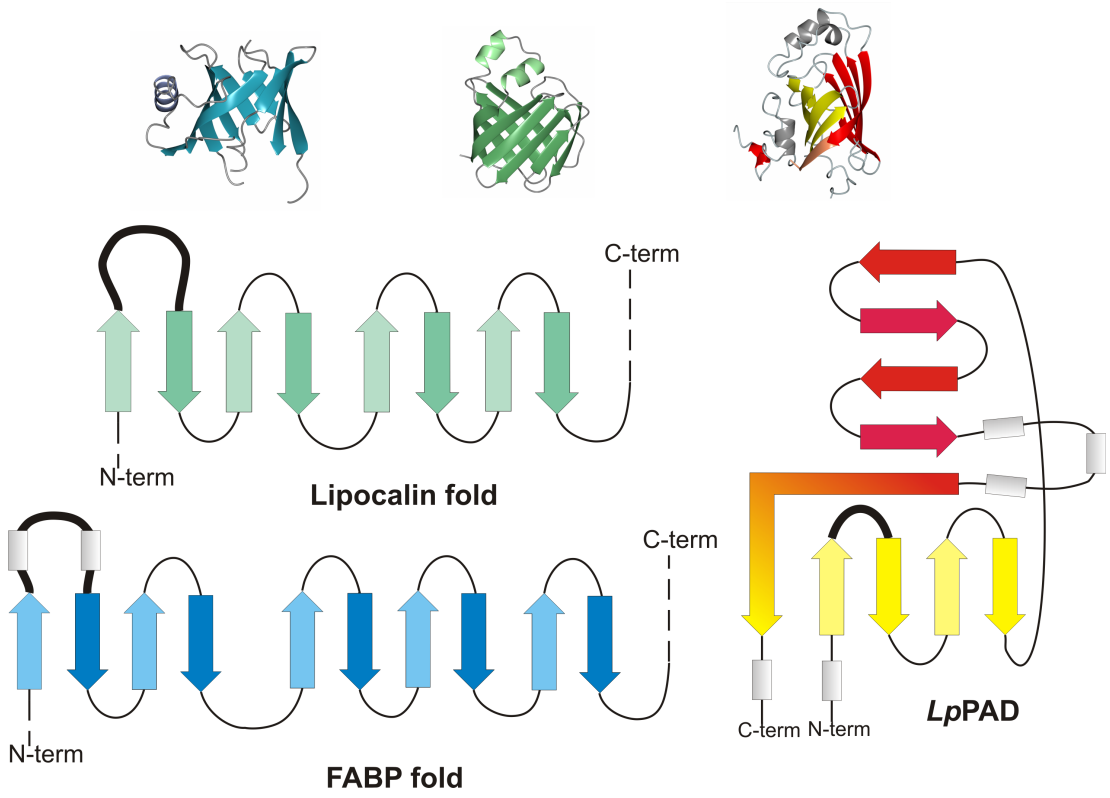


Figure 4.1: PAD enzymes show strong similarities in secondary and tertiary structure to members of the FABP and lipocalin protein families. Their β-sheet core structure arranges into an orthogonal β-sandwich, as shown for *Lp*PAD, in which the most N-terminal and C-terminal β-sheets can hydrogen bond to form a continuous, flat barrel. For the lipocalin and FABP fold only the core structures are shown, missing N- and C-terminal elements are represented as dashed lines. The flexible loop in control of active site accessibility is highlighted in bold for each protein [adapted from (60,146)].

The PAD active sites vary in size. Whereas the *Lp*PAD protein shows a narrow binding cavity with solvent accessible volume only slightly above the estimated volume of the ligand, the *E*PAD enzyme presents a significantly larger pocket. A feature all proteins share however is a reduction in the active site's accessible volume when the proteins' change from their "open" into a "closed" and (most likely) ligand bound state (60,134).

As described above, the active site of PAD proteins comprises a hydrophobic pocket and is thus lined with non-polar residues. The exceptions are a few selected polar residues thought to be key players in catalysis. They are stabilised inherently through the formation of salt bridges and hydrogen bonds. They include, among others, glutamate, arginine and threonine side chains (60,134,144).

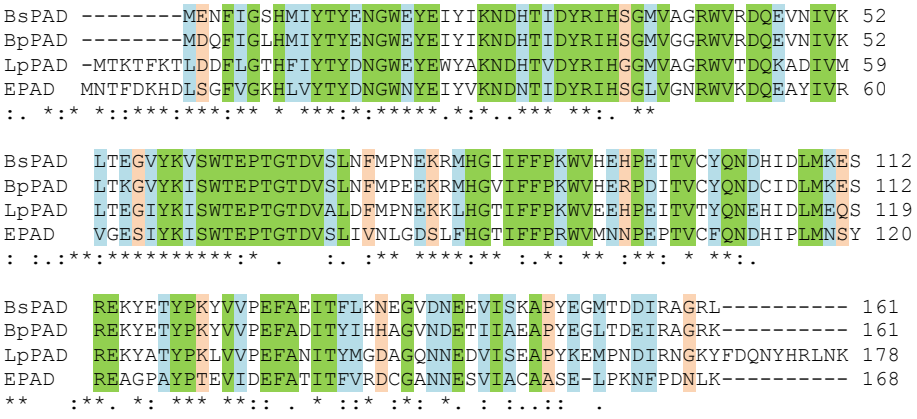


Figure 4.2: All PADs show a high degree of sequence conservation, particularly throughout their N-terminal amino acid residues. An alignment using the CLUSTALΩ algorithm is shown, highlighting fully conserved residues in green (*), conservation of strongly similar properties in blue [:] and conservation of weakly similar properties in coral (.) (148)

All bacterial phenolic acid decarboxylases characterised to date associate into homodimers in solution. When determining and analysing the *Lp*PAD *apo*-structure, Rodriguez *et al.* discovered a so-called “hydrophobic zipper” at the interface between the two dimers. This zipper is characterised by polar residues at the core and hydrophobic residues at the edges of the subunit interface. Based on the high sequence conservation of the residues involved (60), it was proposed to be highly conserved between the PAD proteins. For *E*PAD, dimerisation appears to have additional functional consequences. In contrast to other PADs, the active site of which is enclosed within individual monomers, one of the *E*PAD active sites is buried within the dimer interface and thought to be catalytically inactive(134).

4.1.2 The BsPAD apo-structure

In 2007, the *apo* (ligand free) structure of *Bs*PAD was solved by the Joint Centre of Structural Genomics (149) and deposited in the protein data bank (PDB) under the accession code 2P8G. It shows the two prominent β-sheets described above, arranged in an orthogonal sandwich fold. One end of this tunnel-like structure is covered by α-helices, the other by flexible loops connecting the β-strands (see Fig. 4.3).

The side chains lining the inside of the β-sandwich are, with few exceptions, predominantly hydrophobic. A tail like structure, comprising a long flexible loop and an α-helix, lies adjacent to the β-sheets and forms the C-terminal end of the polypeptide. In its dimeric state, these termini face opposite each other. The same features can be seen in several other members of the PAD family as well as in the lipocalin proteins.

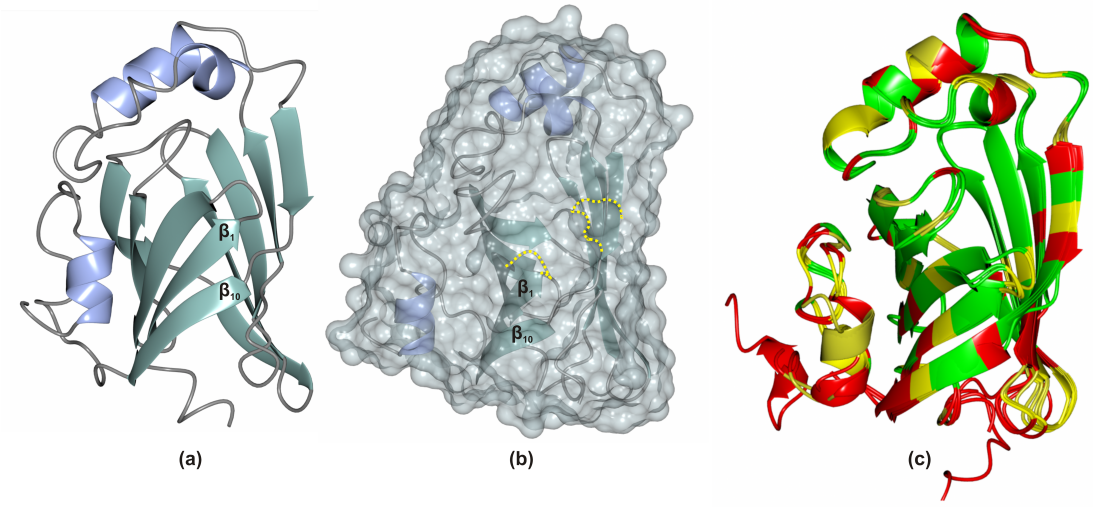


Figure 4.3: The three-dimensional structure of *Bs*PAD shows the overall β-sandwich fold common to all PAD proteins. The two most N- and C-terminal β-strands (β_1 and β_{10}), which interact through hydrogen bonding, are highlighted (a). A surface view superposed onto the enzyme's tertiary structure highlights the open entry channel to the putative active site and the flexible loop thought to be involved in conformational changes (yellow dashed line) (b). A structural superposition of *Bs*PAD with *Lp*PAD, *E*PAD and *Bp*PAD highlights the enzyme family's similarities. Conserved residues are shown in green, moderately conserved residues in yellow and non-conserved amino acids in red.

4.1.3 Project aim

The project work described in the following sections was aimed at obtaining the three-dimensional structure of a *Bs*PAD protein complexed with either ligand, reaction intermediate or product. Although a ligand complex structure has previously been published for the *E*PAD protein (134), its implications on substrate binding and catalysis differed from kinetic data and modelling studies of PAD homologues such as *Lp*PAD (60). This structural work paralleled and followed the preparation of a mutant library, as described in chapter 3. It is based on the assumption that an inactive or weakly active mutant, with reduced reaction rates and possibly increased binding affinities, could trap sufficient amounts of the molecule of interest in the protein's active site. This could allow for successful co-crystallisation of a complex or soaking of the ligand into an *apo*-crystal. Structural data on ligand binding was then hoped to complement kinetic data for the mutant library and thus give an insight into the reaction mechanism and, most importantly, the key catalytic residues. This knowledge could, eventually, inspire protein engineering experiments to improve or alter *Bs*PAD activity and substrate selectivity.

4.2 Materials and Methods

4.2.1 Crystallisation of BsPAD^{wt} and mutant proteins

For crystallisation and structure determination by X-ray crystallography, each *Bs*PAD protein preparation was purified to high homogeneity as described in chapter 3, section 3.2. For the wild type, hexa-histidine-tagged protein, three crystallisation screens were then set up in 96 well Swissci MRC™ crystallisation plates using the sitting-drop vapour diffusion method. The procedure was repeated for non-tagged *Bs*PAD^{wt} protein. A Mosquito® automated liquid handling robot allowed for mixing of 300 nL of protein solution at a concentration of approximately 20 mg/mL with an equal amount of crystallisation solution (mother liquor). The latter included the Hampton PACT screen as well as the INDEX and Clear Strategy Screens (CSS I+II) from Molecular Dimensions™ (150,151). The 24 well CSS I+II screen was assayed in duplicate. The screens were differentiated only by the addition

of 0.1 M MES buffer at pH 5.6 or 0.1 M Tris buffer at pH 8.0, allowed for individual adjustment of the solutions' pH values. In addition to the *apo*-protein setup, crystallisation screens were carried out with protein in a ferulic acid solution. For this, a 500 mM ligand stock was prepared in dimethylsulfoxide (DMSO) and diluted to approximately 5 mM final concentration in the protein solution. After incubation on ice for 60 min, the solution was centrifuged for 2 min at 13,000 rpm to separate precipitated protein. Crystallisation screens were then set up as described above.

As the CSS screen proved most successful in producing protein crystals for the wild type, both in the presence and absence of the ferulic acid, it was used as a standard screen for all mutant library proteins. Initial 96 well screens were set up as described above for the *Bs*PAD^{wt} protein, however without the addition of ligands. Potential crystal hits were then scaled up and optimised using the hanging drop vapour diffusion method. In 24 well Linbro type crystallisation plates, 1-2 µL of protein solution were mixed with 1-2 µL of mother liquor at varying ratios and incubated over 1 mL of the mother liquor. For the latter, both pH and precipitant concentration were varied in optimisation screens. For the Tyr11Phe mutant, various detergents were added to the crystallisation conditions to slow down growth and separate the fast forming and clustered crystals. These included ethylene glycol, butane-1,4-diol, hexane triol and glycerol and were added to final concentrations of 0-10 % (volume/volume). Showing the most promising results, crystal growth of the Tyr19Ala mutant was eventually optimised from the CSS C11 condition to the following solution: 0.2 M KSCN, 17 % PEG 1000, 17 % PEG 8000 and 0.1 M Tris/HCl pH 9.0. At this point, protein at a concentration of 20 mg/mL was incubated in the presence of 3.3 mM *p*-coumaric acid ligand, as described above for the wild type, before preparing crystallisation trays. This produced well-diffracting crystals for which however no significant (refinable) ligand density could be observed in the active site after data collection and processing. As a result, the substrate concentration was increased to 15 mM *p*-coumaric acid in further experiments. Obtained crystals were cryoprotected in their crystallisation conditions, with the addition of 10 % glycerol as well as 15 mM *p*-coumaric acid, and vitrified in liquid nitrogen for storage and data collection.

4.2.2 Data collection and processing

Crystals of wild type and Tyr19Ala mutant proteins were tested for diffraction using an in-house X-ray radiation source and a MARRESEARCH MAR345 imaging plate detector. Only the Tyr19Ala mutant crystals diffracted perceptibly and were sent off for data collection at the Diamond Light Source synchrotron in Didcot. For a crystal grown in the presence of 3.3 mM *p*-coumaric acid, 180° of data were collected on beamline I-03, at an oscillation angle of 0.5° and to a resolution of 2.1 Ångstroms (Å) (hereafter called dataset 3.3).

A second crystal of the same mutant, grown in the presence of 15 mM *p*-coumaric acid, was used for the collection of a 180° dataset with identical oscillation angles, at beamline I24. It is hereafter called dataset 15.0. The second dataset's resolution extended to 2.5 Å, however processing was limited to 3.0 Å as the inclusion of higher resolution data lead to a significant loss in overall completeness.

After collection, all data were auto-processed by the Xia2 data reduction software (at Diamond), making use of the *XDS, XSCALE* and *CCP4* program suite (152). For dataset 15.0, data was additionally re-processed with *XDS* (153) and two additional datasets of 80 and 180 images were created in attempts to lower the R_{merge} values and compare qualities of the resulting electron density maps.

4.2.3 Structure solution and refinement

The structure of the ligand bound Tyr19Ala mutant was determined by molecular replacement. This was performed using *MOLREP* of the *CCP4* suite of programs (154) and a single monomer of wild type protein (PDB accession code 2P8G), excluding waters, as a model. The dataset was subjected to 10 cycles of rigid body refinement using *REFMAC5* (155). 5% of the data were retained to calculate R_{free} values from an independent test set of randomly chosen reflections. These reflections are excluded from subsequent rounds of refinement and the difference between *R* values for working and test set thus serve as an indicator for over-fitting of the model to the experimental data. The Tyr19Ala model was then built and adjusted to electron density maps using *Coot* (156) and subsequent cycles of restrained refinement by *REFMAC5*. Library files for the ligand were created using PRODRG

(157), modelled into the appropriate electron density and refined at an occupancy of 1.0. Structure validation was carried out as part of the *Coot* program suite.

Coordinates for the BsPAD^{Tyr19Ala} complex structure, obtained from dataset 15.0, were deposited with the Protein Data Bank (PDB) and were assigned the accession code 4ALB.

4.3 Results

4.3.1 Crystallisation of BsPAD^{wt} and mutant proteins

After purification of the *Bs*PAD^{wt} protein, three crystallisation screens were prepared for both hexa-histidine tagged and cleaved protein. Whereas the latter did not produce any detectable crystals, all three screens containing the tagged protein showed formation of a variety of crystal forms. When crystallisation experiments were repeated for the CSS screen and in the presence of ferulic acid substrate however, only small and densely clustered, broomstick-shaped crystals were observed. As scale-up experiments for this condition did not yield any further crystal growth, structural work with the wild type protein was abandoned.

Having produced the most abundant hits for the wild type protein, the CSS (I+II) screen was used exclusively for crystallisation experiments with the *Bs*PAD mutant library. With the latter it was hoped to obtain crystals with substrate, intermediate or product trapped in the active site of a weakly active or inactive mutant. Three of these mutants (Tyr11Phe, Tyr13Phe and Glu64Ala) crystallised within 24 hours, forming dense clusters of needles. Scale-up experiments in the presence of ferulic or *p*-coumaric acid and a variety of detergents, added to slow down growth and lead to the formation of single crystals, were however unsuccessful. A fourth mutant, *Bs*PAD^{Tyr19Ala}, crystallised significantly slower under 96-well, sitting drop conditions, within a time frame of several days to weeks. Eventually however, a variety of crystal forms appeared in conditions containing varying concentrations of potassium thiocyanate (KSCN) and polyethylene glycol (PEG) of different molecular weights. Crystal shapes ranged from needle like structures to plates and elipses. Crystal conditions from the CSS screen could then be optimised in 24-well scale up

trays to yield reproducible crystals growth of the *Bs*PAD^{Tyr19Ala} mutant protein within 24 hours. In the final conditions, 1.5 µL of protein, at a concentration of approximately 20 mg/mL, crystallised when mixed with 1.0 µL of mother liquor containing 0.2 M KSCN, 17 % PEG 8K, 17% PEG 1K and 0.1 M Tris buffer at pH 9.0, in the presence of either ferulic acid or *p*-coumaric acid (3-15 mM). The final crystals have a clothes peg-like appearance of two closely attached needles forming a hollow space between them (see Fig. 4.4).

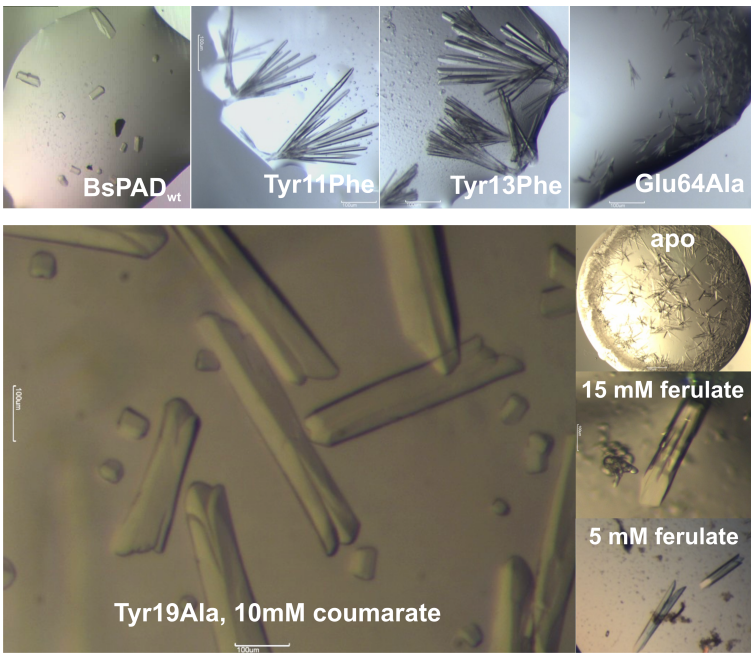


Figure 4.4: Crystals of the *Bs*PAD^{wt} protein and its mutant library (top) formed in the presence and/or absence of the ferulic and *p*-coumaric acid substrate. The Tyr19Ala variant crystallised in the presence and absence of two of its substrates, *p*-coumaric and ferulic acid (bottom).

4.3.2 Diffraction properties of BsPAD Tyr19Ala crystals

Of all mutant proteins, only the Tyr19Ala variant formed reproducible, single crystals in the presence of its phenolic acid substrates. Prior to data collection at the synchrotron, crystals grown in the presence of 3.3 mM ferulic acid were tested using an in-house X-ray source and the MAR3 imaging plate detector and yielded good diffraction of up to 2.5 Å resolution. This resolution increased to 2.0 Å when crystals were exposed to synchrotron radiation at the Diamond Light Source in Oxford (see Fig. 4.5). When a dataset was collected and processed however, the

electron density appearing in the difference map of the enzyme's active site was not sufficient to allow for modelling and refinement of a substrate complex (see below). As a result, co-crystallisation was repeated with ligand concentrations of 15 mM *p*-coumaric acid and produced crystals of similar appearance. Their diffraction quality was however inferior to those grown at lower substrate concentrations and did not exceed 3.5 Å when tested in-house. At the synchrotron, this resolution was improved however and a second dataset was collected to a resolution of 2.5 Å (see Fig. 4.5 and table 4.1).

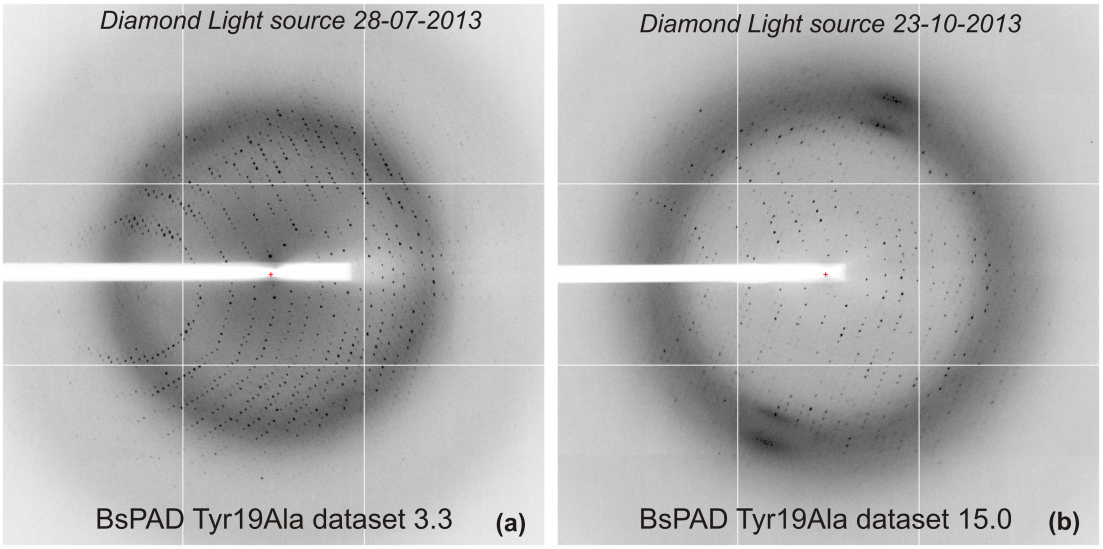


Figure 4.5: X-ray diffraction data was collected for the *Bs*PAD^{Tyr19Ala} mutant crystallised in the presence of 3.3 mM (a) and 15.0 mM (b) at the Diamond Light Source synchrotron in Didcot.

4.3.3 Data collection, processing and refinement

At the Diamond Light Source synchrotron, 180° of data (360 images at an oscillation angle of 0.5°) were collected for dataset 3.3 and 15.0. Both crystals indexed and processed in space group $P3_221$, with dataset 3.3 showing signs of twinning in its cumulative intensity distribution. With the chosen data collection strategy the resulting multiplicity was high (11, see table 4.1). This was intentional and aimed at obtaining the most optimal information on ligand binding.

Data processing was carried out using the Xia2 automated data reduction system (152) and repeated manually using XDS for dataset 15.0 (153). For the latter, a 3.0 Å

resolution cut-off was applied during all processing procedures as higher resolution data was of significantly inferior quality. The manually processed 15.0 data set was then used in particular for the preparation of datasets with reduced multiplicity. These were aimed at lowering the high R_{merge} values observed after Xia2 processing. Data reduction led to a decrease in R_{merge} from 83% to 77% and 52% in the highest resolution shell for datasets of 360, 180 and 80 images respectively. However, the signal to noise ratios ($I/\sigma I$) decreased (see table 4.2) and the maps deteriorated significantly.

	*Bs*PAD Tyr19Ala dataset 3.3	*Bs*PAD Tyr19Ala dataset 15.0
Beamline	Diamond IO3	Diamond IO4
Wavelength (Å)	0.9763 Å	0.9763 Å
Resolution (Å)	36.19-2.08 (2.19-2.08)	53.60-3.03 (3.10-3.03)
Space Group	$P3_221$	$P3_221$
Cell dimensions (Å)	a = b = 110.6; c = 91.6	a = b = 107.21; c = 92.80
Cell dimensions (°)	90, 90, 120	90, 90, 120
Unique reflections	37157 (2637)	11708 (778)
Completeness (%)	100	99.2 (98.2)
R_{merge}	0.12 (0.97)	0.11 (0.86)
$R_{p.i.m.}$	n/a	0.051 (0.39)
Multiplicity	10.5 (10.4)	10.8 (11.0)
$<I/\sigma(I)>$	14.0 (2.6)	21.4 (3.3)
R_{work}/ R_{free} (%)	16.5/21.9 (38.2/49.0)	20.1/ 26.3 (32.9/33.7)
Protein atoms	3950	3852
Ligand atoms	-	36
Solvent atoms	57	13
B-factors Protein	36.9	52.5
B-factors Ligand	-	54.7
B-factors Solvent	34.1	28.1
Average B ($Å^2$)	36.9	44.6
r.m.s.d bonds (Å)	0.026	0.011
r.m.s.d angles (^{o})	2.15	1.44

Table 4.1: Data collection and refinement values are shown for the Tyr19Ala mutant protein co-crystallised with 15 mM *p*-coumaric acid (3). Values in parentheses are for the highest-resolution shell.

Number of images	360	180	80
R_{merge}	0.023 (0.83)	0.020 (0.77)	0.013 (0.517)
$I / \sigma I$	40.4 (3.3)	29.9 (2.5)	22.2 (2.0)
Completeness (%)	98.4 (98.9)	97.0 (97.0)	91.0 (99.9)
Redundancy	9.8 (10.6)	5.1 (5.4)	2.4 (2.4)
Unique reflections	11624 (774)	11483 (747)	11198 (744)
R_{work} / R_{free} (%)	21.6 / 30.7 (28.8 / 46.1)	20.3 / 32.4 (30.1 / 56.2)	21.1 / 30.7 (31.5 / 49.8)
Protein atoms	3908	3915	3871
B-factors Protein (Å^2)	41.3	39.6	51.1
R.m.s. deviations Bond lengths (Å)	0.014	0.030	0.014
R.m.s. deviations Bond angles (°)	1.50	2.57	1.56

Table 4.2: Data collection and refinement values are shown for the Tyr19Ala mutant protein co-crystallised with 15 mM *p*-coumaric acid, processed with XDS and truncated from 360 images (left) to 180 (middle) and 80 images (right) prior to processing (3). Values in parentheses are for the highest-resolution shell.

Scaled and merged data could be used to determine the structure of the *Bs*PAD^{Tyr19Ala} mutant protein. Calculation of the Matthew's coefficient (158,159) indicated the presence of three monomers in the asymmetric unit (with a solvent content of 55 %) for which strong contrasts could be detected during molecular replacement. The molecular replacement solution was refined against the original (experimental) reflection file, leading to a significant reduction in R_{work} and R_{free} values and thus showing a reduction in bias towards the molecular replacement model.

Using *Coot*, the protein model could be fit to the resulting electron density map and was improved through iterative rounds of building and refinement using *REFMAC5*. For most datasets the difference between R_{work} and R_{free}, being an indicator of the agreement between a current model and the experimental data, remained below 10 % (12 % for the XDS 1-180 truncated dataset).

4.3.4 Data validation

For the final model of the 15.0 dataset, the Ramachandran plot calculated with Coot showed 93.4 % of the *Bs*PAD^{Tyr19Ala} residues to lie in the most favoured regions while 5.1 % and 1.5 % lie in additionally and generously allowed regions respectively.

4.3.5 Analysis of the BsPAD ^{Tyr19Ala} mutant structure

After the final rounds of model building and refinement of dataset 3.3 and 15.0, the *Bs*PAD^{Tyr19Ala} mutant's structures could be visualised and analysed using both *Coot* and the molecular graphics program *CCP4mg* (160). The mutant crystallised in the presence of 3.3 mM ferulic acid showed clear electron density for all residues, with the exception of the first two N-terminal amino acids. The crystals growing in the presence of 15.0 mM *p*-coumaric acid showed inferior diffraction quality and resulting electron density maps are equally less well defined than those of dataset 3.3. Most residues could however be fitted well into the density, with the exception of a loop region encompassing residues 13 to 16.
The final structures strongly resemble the *Bs*PAD wild type *apo*-structure as it was deposited in the protein database under accession number 2P8G and can be superposed as shown in figure 4.6. The protein associates into a homodimer, with 3 monomers present in the asymmetric unit. The central orthogonal β-sandwich forms a hollow cavity enclosing the putative active site in the centre of each monomer. This cavity is, with few exceptions, lined by predominantly hydrophobic residues whereas the *Bs*PAD solvent exposed surface is made up of polar side chains. This is a feature generally described for PAD proteins (see section 4.1). One opening of the β-sandwich is covered by α-helices, connected to the β-strands by way of flexible loops. The opposing end of the sandwich is more solvent exposed but also partially covered, in this case exclusively by flexible loops. As for the wild type protein, a C-terminal helix extends away from the β-sheets, in opposing directions for each dimer.

The only significant structural difference between the wild type and both mutant protein structures is seen in the flexible loop region which, for the latter, is displaced

by 3-4 Å compared to the 2P8G structure (see Fig. 4.6). As described above, electron density is not well defined for this area in dataset 15.0, indicating a region of high flexibility in the protein. Processing and refinement of the 3.3 dataset showed small areas of positive electron density in the mFo-DFc difference map, *i.e.* electron density which cannot be accounted for by the molecular replacement model, in the protein's active site. The areas were small and insufficient for modelling and refinement of a ligand molecule into any of the three monomers. An averaged non-crystallographic symmetry (NCS) map, however, outlined the potential shape of a molecule of ferulic acid, with its carboxylate moiety and methoxy side chain protruding from the phenol ring (see Fig. 4.7).

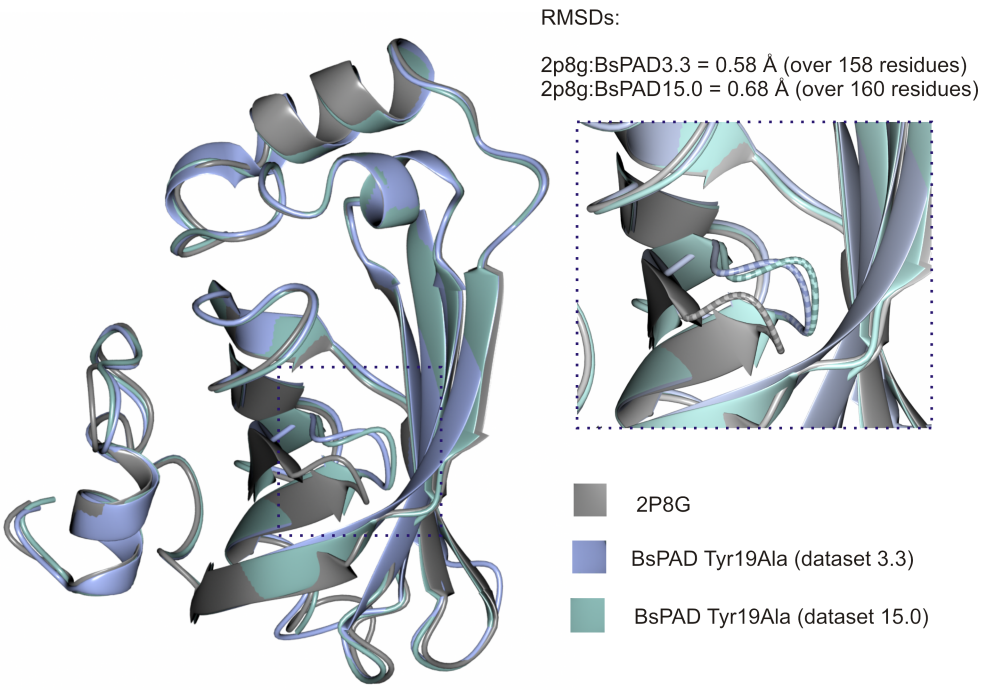


Figure 4.6: The superposition of both *Bs*PAD^{Tyr19Ala} mutant structures with the wild type enzyme as it was deposited in the Protein Data Bank (2P8G) shows little difference in the overall secondary and tertiary structure. The exception is a flexible loop (magnified, right image), encompassing residues 13 to 16, which restricts access to the mutant enzyme's active site.

For the 15.0 dataset, difference density appeared in the same area of the protein's active site and was sufficient for modelling and refinement of a molecule of *p*-coumaric acid with an occupancy of 1.0.

The best fit was obtained if the substrate's phenolic hydroxyl group was positioned at the more C-terminal pole of the active site cavity, in hydrogen bonding distance to residues Arg41, Glu64 and Thr68. The substrate's carboxylate group faces in the opposite direction and is positioned close to the tyrosine pair Tyr11 and Tyr13 as well as the alanine sidechain which replaces the wild type's Tyr19 (see Fig. 4.7).

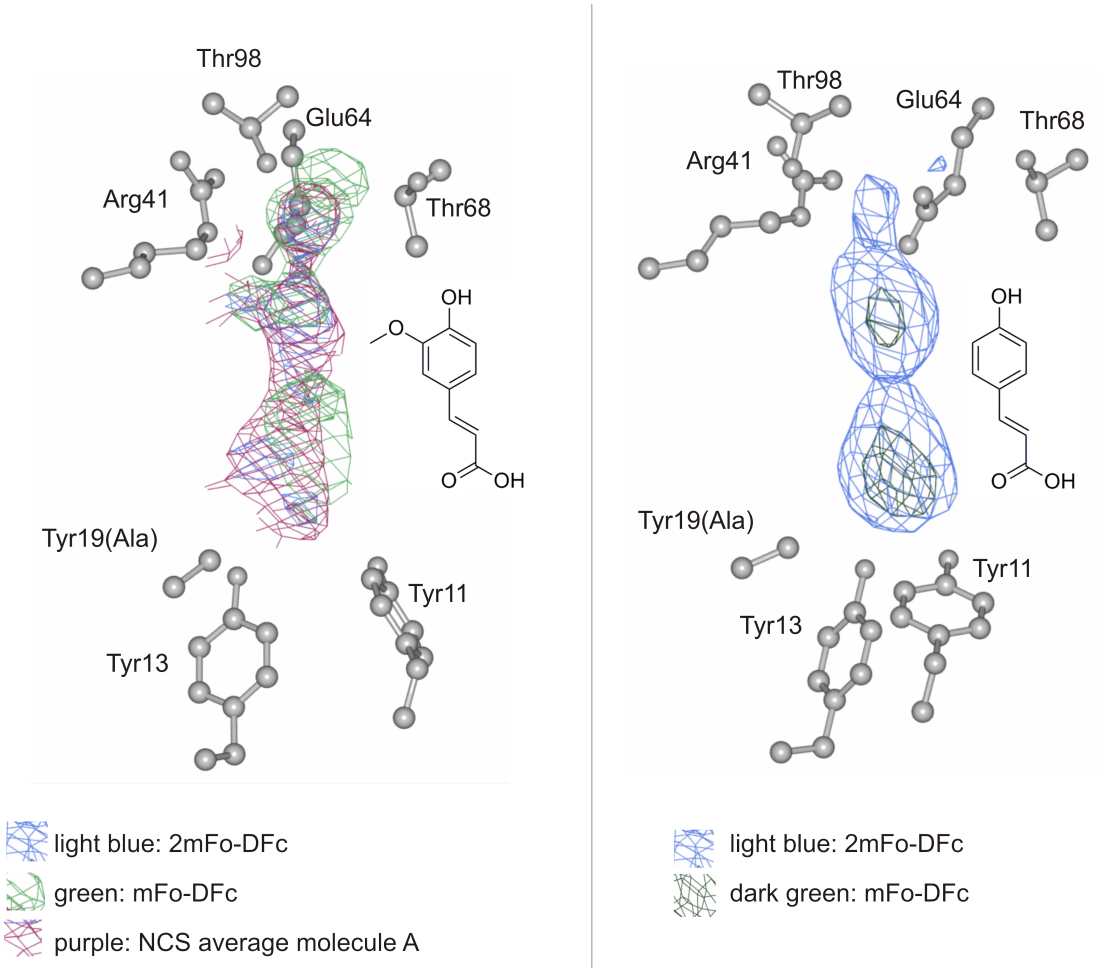


Figure 4.7: The 2mFo-DFc and mFo-DFc electron density maps of the *Bs*PAD^{Tyr19Ala} mutant's active sites show density for both substrate molecules, ferulic acid (dataset 3.0, left) and *p*-coumaric acid (dataset 15.0, right). The former is only recognisable if the non-crystallographic symmetry (NCS) averaged 2mFo-DFc density of all three monomers in the asymmetric unit is displayed onto a single chain (purple, left). All 2mFo-DFc maps are shown at a contour level of 1.0 sigma (σ), the mFo-DFc positive difference density is contoured at 3.0 σ.

4.4 Discussion

Since the first solution of a three dimensional PAD enzyme structure in 2006 by the Joint Centre for Structural Genomics, the coordinates of three more bacterial homologues haven been deposited in the Protein Data Bank.

In 2010, Rodriguez *et al.* were able to propose a detailed mechanism based on the computational modelling of a ferulic acid molecule into their high resolution *Lp*PAD *apo*-structure (60). One year later, Gu *et al.* published a the first report of a ligand complex between the *Enterobacter* PAD and the substrate analogue Na-ferulate (134). As findings with respect to ligand orientation and functionality of the various active site residues in catalysis differed significantly between the two proposals, however, further evidence was required to elucidate the key players in PAD catalysed decarboxylation.

After the preparation of a *Bs*PAD active site library, as described in chapter 3, all mutant proteins were thus subjected to crystallisation experiments. It was hoped to find an inactive or weakly active variant in the active site of which a substrate, reaction intermediate or product could be "trapped" and visualised in a three-dimensional mutant structure. After screening of all mutant proteins, the Tyr19Ala variant proved to be the only enzyme to produce single, diffracting and reproducible crystals. Initial attempts of co-crystallisation in the presence of 3.3 mM of ferulic acid yielded high quality diffraction to 2 Å and the structure could be determined easily by molecular replacement using the *Bs*PAD^{wt} structure (2P8G) as a template. The difference density observed in the enzyme's active site was however poor and only after combination of information from all three symmetry related active sites in the asymmetric unit could the potential outline of a phenolic acid be distinguished. Refinement against a molecule of ferulic acid was not possible. This finding was in contrast to the structure obtained after co-crystallisation in 15 mM *p*-coumaric acid. The diffraction's resolution was reduced to 3.0 Å. In the active site however, clear difference electron density could be observed and enabled modelling and refinement of the ligand phenolic acid with an occupancy of 1.0 (see Figures 4.7 and 4.8)

The difference in diffraction quality between the two datasets 3.3 and 15.0 could be the consequence of the latter enzyme's remaining activity in the presence of excess substrate. Despite a significant reduction in catalytic efficiency, the Tyr19Ala mutant

still turns over substrate at a rate easily detectable using UV-VIS spectroscopy (see chapter 3). In addition, although enzyme and substrate were incubated on ice prior to the preparation of crystallisation trays, product formation was discernible through the typical smell of the styrene. It is thus likely that enzyme activity (mobility) and the presence of a non-homogenous population of substrate, possibly intermediate and product could have had a negative impact on crystal order and diffraction quality. This is further supported by the observation that the electron density of a flexible loop, which is thought to cover the active site during catalysis (see below), is considerably less defined in the 15.0 structure. This difference is pronounced even after taking the generally inferior map quality into consideration.

On closer inspection of the data collection statistics, a high merging *R*-factor (R_{merge}) of 83% in the highest resolution shell could be observed. This R_{merge} value originates from the comparison of reflection intensities with the average of their symmetry related counterparts and has long served as a quality indicator for crystallographic data. As the number of reflections however increases with multiplicity, so does the distribution of their intensities (161). As a consequence, the overall R_{merge} values increase for highly redundant data such that collected for the *Bs*PAD^{Tyr19Ala} crystal. As an alternative to R_{merge}, the precision-indicating merging *R* factor (R_{pim}) can be stated which takes the data's redundancy into consideration and is accordingly significantly lower (162,163). To highlight the importance of including redundant data, the Tyr19Ala 15.0 dataset was re-processed with the omission of 0, 180 and 280 images to give datasets of 360, 180 and 80 images respectively. This should give an insight into the effects of a loss of multiplicity on the overall quality of electron density maps and, in particular, presence of the ligand's difference density. The high symmetry of the $P3_221$ space group means that only 30° of data (60 images at an oscillation angle of 0.5°) should be required to collect all (hkl) reflections at least once and acquire a complete dataset. As expected, the R_{merge} values of the newly processed datasets decrease with the reduction in data (and multiplicity). Equally however, the I/σI values decrease. *I/σI* values signify the average intensity of reflections divided by their respective average standard deviation and thus correspond to the data's signal to noise ratio (161). Low values indicate an increase in the average standard deviation and thus a strong reduction in data quality and reliability after loss of the redundant reflections. The general loss in amount and

quality of information obtained from the measured reflections thus explains the significant reduction in quality of the resulting electron density maps and the ligand's difference density. This holds true for each truncated dataset, even after several iterative rounds of model building and refinement.

The *Bs*PAD^{wt} and both Tyr19Ala mutant structures are highly similar and superpose with low root mean square deviations (RMSDs) of 0.58, 0.68 and 0.37 Å (between wild type and Tyr19Ala 3.3, wild type and Tyr19Ala 15.0 and between the two mutant structures respectively). The only significant difference can be observed in the loop region between residues Tyr13 and Gly16. For the mutant proteins this loop is shifted towards the active site's entry and occludes access for solvent molecules. This is most likely the consequence (and at the same time an indicator) of ligand binding. For catalysis to occur, the ligand has to be fixed in the active site and its mobility would be increased in the presence of more solvent molecules in a fully accessible active site. Compared to some of the maximal enzymatic rates of catalysis (143), *Bs*PAD is a fast and efficient enzyme, with a catalytic efficiency approaching 1 x 10^4 s^{-1}/M (see chapter 3, table 3.3). An altered side chain close to the loop region, as it is present in the Tyr19Ala mutant, may negatively influence flexibility and thus hamper catalysis.

At a contour level of 1σ, the 2mFo-DFc map of the 3.3 dataset shows a small area of electron density in the centre of the Tyr19Ala mutant's active site after molecular replacement and refinement. This density is significantly more pronounced in the 15.0 structure and surrounds additional positive difference density of the mFo-DFc map at a contour level of 3 σ. The phenolic acid concentration was increased by a factor of five during co-crystallisation experiments while all other conditions remained unchanged. The density observed is thus likely to correspond to the bound ligand rather than a trapped solvent molecule. The best fit to a molecule of *p*-coumaric acid could be obtained if the latter was oriented with its carboxy-terminus facing the more N-terminal part of the active site and forming hydrogen bonds with the tyrosine pair Tyr11 and Tyr13. The *para*-hydroxy group at the opposite end of the molecule is, in this conformation, within hydrogen bonding distance to the glutamate in position 64 and the ε-nitrogen of an arginine in position 41.

In addition, a small area of density could be observed in the 2mFo-DFc electron density map which is thought to correspond to a water molecule coordinated by the Glu64 and Thr68 side chains. This water has been observed in the homologous *Lp*PAD *apo*-structure and was proposed, once activated, to act as the proton donor to the *para*-methide quinone intermediate (60).

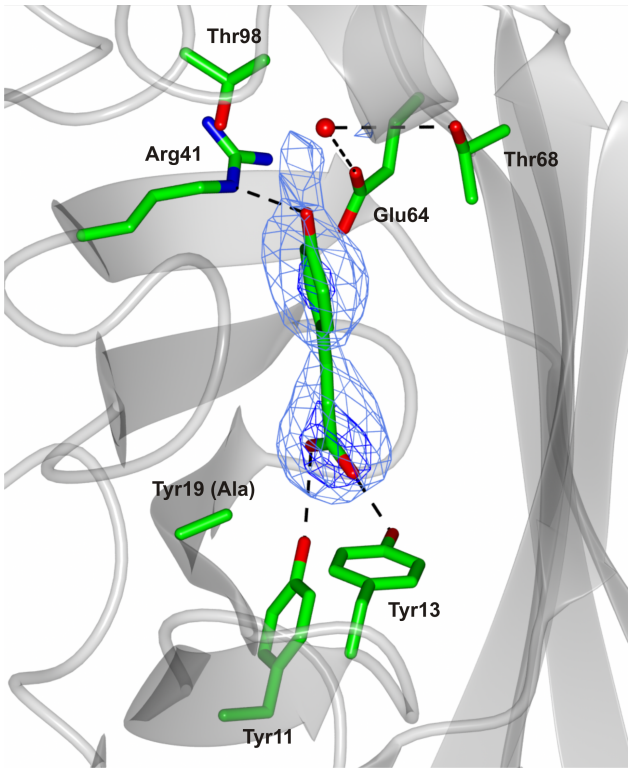


Figure 4.8: At a contour level of 1σ, an area of unmodelled density was detected in the 2mFo-DFc electron density map (light blue) of the *Bs*PAD^{Tyr19Ala} active site. Into this area, the *p*-coumaric acid ligand could be fitted and was shown to form hydrogen bonds with a number of active site residues. Electron density for the ligand was clearest at intermediate stages of refinement (shown above). The water molecule was built and refined at later stages and superposed from the final structure (4ALB). The mFo-DFc electron density map (dark blue) is shown at a contour level of 3 σ.

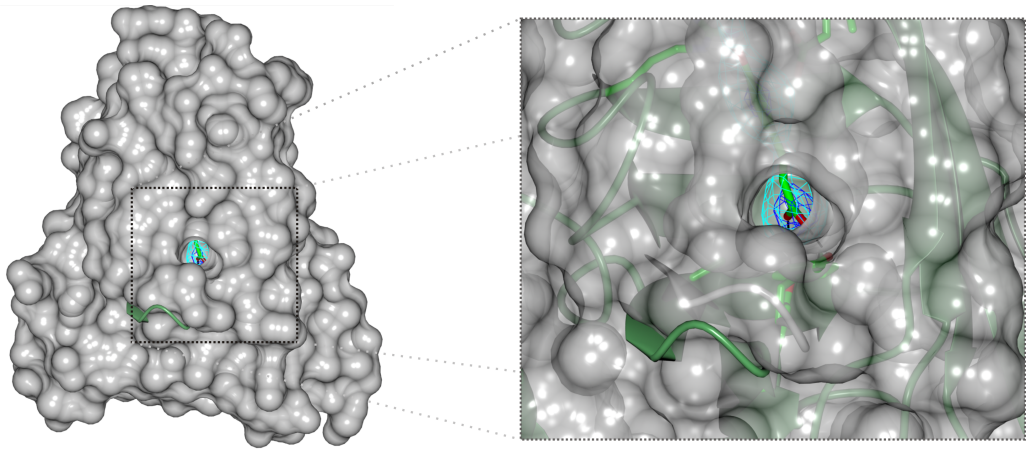


Figure 4.9: A superposition of the mutant enzyme (surface and ribbons, grey) with the wild type structure (2P8G, green ribbons) highlights the loop movement and subsequent closure of the active site upon ligand binding.

The ligand conformation observed corresponds entirely to the model previously proposed for the *Lp*PAD structure (60) and is in very good agreement with the kinetic data obtained for the *Bs*PAD active site mutant library (see chapter 3). Based on this information, a detailed reaction mechanism for phenolic acid decarboxylation in *Bs*PAD can be proposed:

As the phenolic acid substrate enters the active site, it is recognised specifically by interactions of the active site residues with both of its polar "ends". The carboxy-terminus is secured through hydrogen bonding interactions with Tyr11 and Tyr13 whereas the *para*-hydroxy group is positioned within hydrogen bonding distance of Arg41. A movement of the flexible loop encompassing residues 13 to 16 occludes solvent access to the active site (see Fig. 4.9). This locks in the substrate and intermediate in close proximity of the catalytic residues and enhances the proximity effect responsible for the high efficiency of enzymatic reactions (143).

To initiate catalysis, the Glu64 sidechain can then act as a catalytic base and abstracts a proton from the *para*-hydroxy group. This leads to an electron relay and formation of the *para*-methide quinone intermediate as described previously. Parallel protonation of the intermediate could be accomplished from two different positions (see Fig. 4.10, I and II). In the *Lp*PAD *apo*-structure a highly coordinated water molecule was observed in the active site and proposed the act as a proton donor. In the same position, electron density which could correspond to a catalytic water, was observed in the Tyr19Ala structure. It is found within hydrogen bonding distance to

Glu64 and Thr68 and could be coordinated and activated by the latter residue's side-chain. This would also explain the significant reduction in activity for the Thr68Val mutant. Alternatively, movement of the flexible loop brings the Tyr19 side chain (mutated to alanine in this structure) into close enough proximity for the phenolic hydroxyl to donate its hydrogen to the intermediate. This would be an alternative explanation for the mutant's reduced activity but equally imply that the hydrogen bond donor function can be replaced by an alternate residue (or a water molecule) in the absence of a tyrosine side chain. Eventually, the *para*-methide quinone intermediate collapses and a second electron relay in the opposite direction leads to breaking of the C8-C9 bond and release of CO_2 (see Fig. 4.10, III and IV).

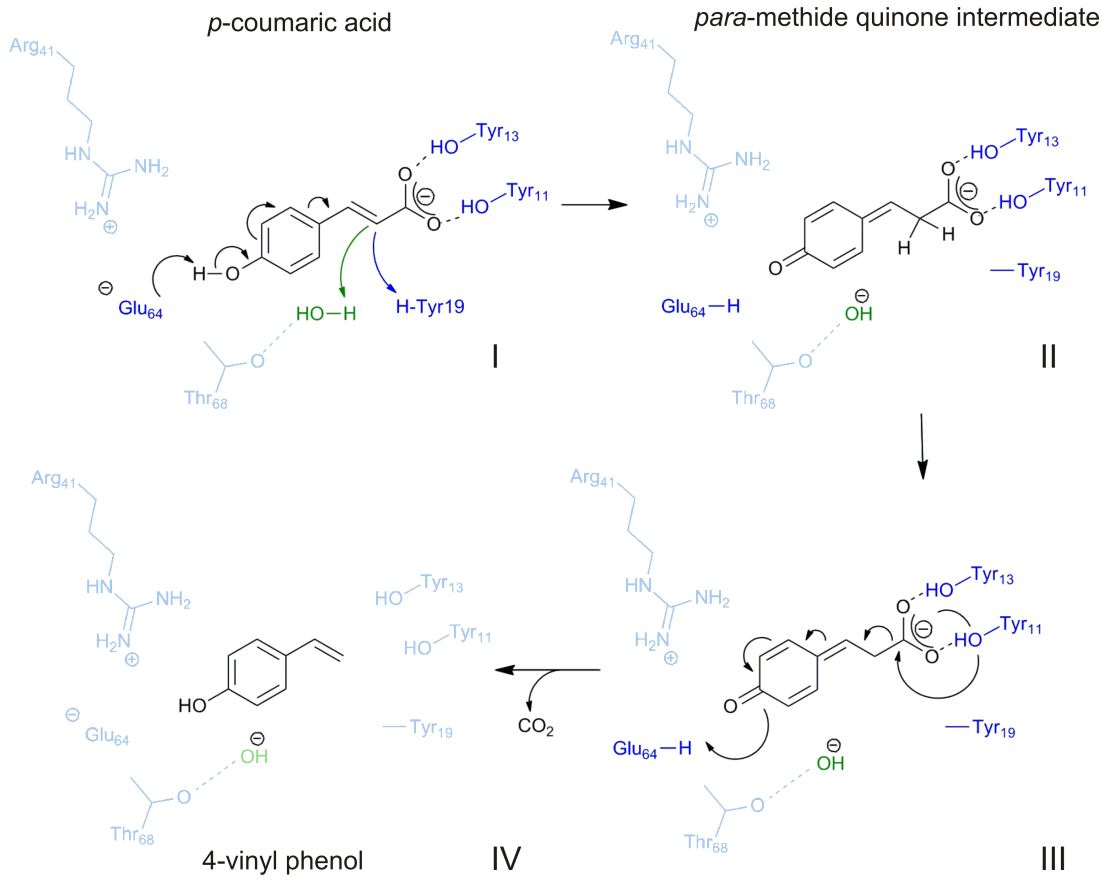


Figure 4.10: A catalytic mechanism was proposed for phenolic acid decarboxylation by *Bs*PAD. In this proposal, Glu64 acts as the catalytic base and initiates the first electron relay by deprotonation of the substrate's *para*-hydroxy group. Two catalysts could act as proton donors to the formed *para*-methide quinone intermediate – Tyr19 (blue pathway) or an activated catalytic water (green pathway) coordinated by Thr68. A second electron relay in the opposite direction then leads to decarboxylation and release of the styrene product. The positively charged Arg41 side chain acts as a stabiliser, in particular for the Glu64 catalyst.

With re-protonation of the *para*-hydroxyl group by Glu64, the catalytic base is regenerated and formation of the styrene product is completed. Movement of the flexible loop is likely to be required for exit of the product from the *Bs*PAD active site and preparation for a new round of catalysis. Catalytic solvent molecules could thus be also be regenerated or exchanged.

In conclusion, the first structure of a *Bs*PAD mutant protein in complex with its *p*-coumaric acid ligand is presented. In conjunction with the kinetic data from an active site mutant library, the findings lend strong support to the previously proposed and widely accepted decarboxylation mechanism from *Klebsiella oxytoca*. They furthermore identify the key catalytic residues in the *Bacillus subtilis* enzyme and thus explain how the reaction is made possible by the protein's active site structure and composition.

Chapter 5: Structural and mechanistic studies into serine-independent carbon-carbon hydrolysis by phloretin hydrolase

5.1 Introduction

5.1.1 Flavonoids

With the growing commercial and pharmaceutical interest in natural compounds, many members of the large class of plant secondary metabolites, the flavonoids, have been discovered, analysed and chemically synthesised over the last decades.

The molecules are ubiquitous in plants where they function in a wide variety of processes, including pigmentation, photosynthesis and electron transport as well as enzyme regulation, hormone activities and anti-infective defence mechanisms. They are synthesised from the aromatic amino acids tyrosine and phenylalanine by way of cinnamic acid derivatives and their subsequent condensation with several units of acetate (164). Most members of the flavonoid family are characterised by a three ring structure, two aromatic rings (A and B) interconnected by a pyran or pyrone ring (C). (164,165). They can be subdivided, based on their structure, into anthocyanidins, flavanones, flavones, and chalcones. The latter three subclasses are all characterised by the presence of a ketone moiety in their C-ring or the open, 3-carbon chain in case of the chalcones (see Fig. 5.1) (164). As flavonoids are taken up daily, on a milligram scale, by humans and other mammals through their diet, their multiple functionalities are also exploited by other physiological systems than plants. The various subclasses are thought to be beneficial in the defence against microbial and viral infections, aid in the protection from many chronic disorders and cancer as well as being anti-inflammatory agents. Flavonoids are able to chelate metals and may interact directly with proteins, for example through nucleophilic addition (165), but they are foremost known for their pronounced antioxidant properties. This feature is a consequence of their hydroxylation pattern, especially on the B ring, which allows them to donate protons to radical molecules and delocalise charges over their conjugated aromatic backbone (166,167).

5.1.2 Chalcones and the dihydrochalcone phloretin

Within the flavonoid subclasses, the so-called chalcones differ in their lack of the C-ring and connect ring A and B by way of a three-carbon, α-β unsaturated carbonyl chain. They are the product of a condensation reaction between hydroxy-acetophenones and benzaldehyde derivatives, both naturally and in chemical (abiotic) synthesis (164,168). In addition to their proposed benefits to human health, chalcones serve as intermediates in the synthesis of heterocyclic compounds (169).
One of the most abundant flavonoids metabolised as part of the human diet is the dihydrochalcone phloretin. The phloretin A ring constitutes a tri-hydroxylated phloroglucinol moiety and the α-β bond is, in contrast to other chalcones, saturated. In its glycosylated form, phloridizin (2-β-D-glucose), phloretin is abundant in fruits such as apples and, as all members of the flavonoid family, thought to contribute to human health by acting as an efficient radical scavenger. This feature is attributed in particular to the 2,6-dihydroxyacetophenone moiety which allows for free electron delocalisation and stabilisation after donation of a hydrogen to the radical (170).
Phloretin is an intermediate of flavonoid metabolism in the human gut, where it can originate directly from cleavage of the phloridizin glucose moiety. Alternatively, the sugar moiety of the flavone apigenin may be removed by hydrolysis, followed by reduction of a double bond in the 2-3 position, isomerisation to give a chalcone and eventually reduction of the open carbon chain's C=C double bond to give the dihydrochalcone phloretin (see Fig. 5.1) (171,172).

5.1.3 Flavonoid metabolism by Eubacterium ramulus

Within the human gut, flavonoid metabolism was found to be dominated by the Gram positive, rod-shaped, strictly anaerobic *Eubacterium ramulus*. The organism was first discovered in 1975 (173) and subsequently re-isolated and characterised by Blaut and co-workers (172). *E. ramulus* was found to grow on the flavonoid quercetin-3-glucoside and catalyse its breakdown from hydrolysis of the sugar moiety down to reduction of the hydrolytic cleavage product 1,3,5-trihydroxybenzene (phloroglucinol). From these findings, the first detailed metabolic pathway for flavonoid metabolism by this organism could be proposed (see above and Fig. 5.1) (171,172).

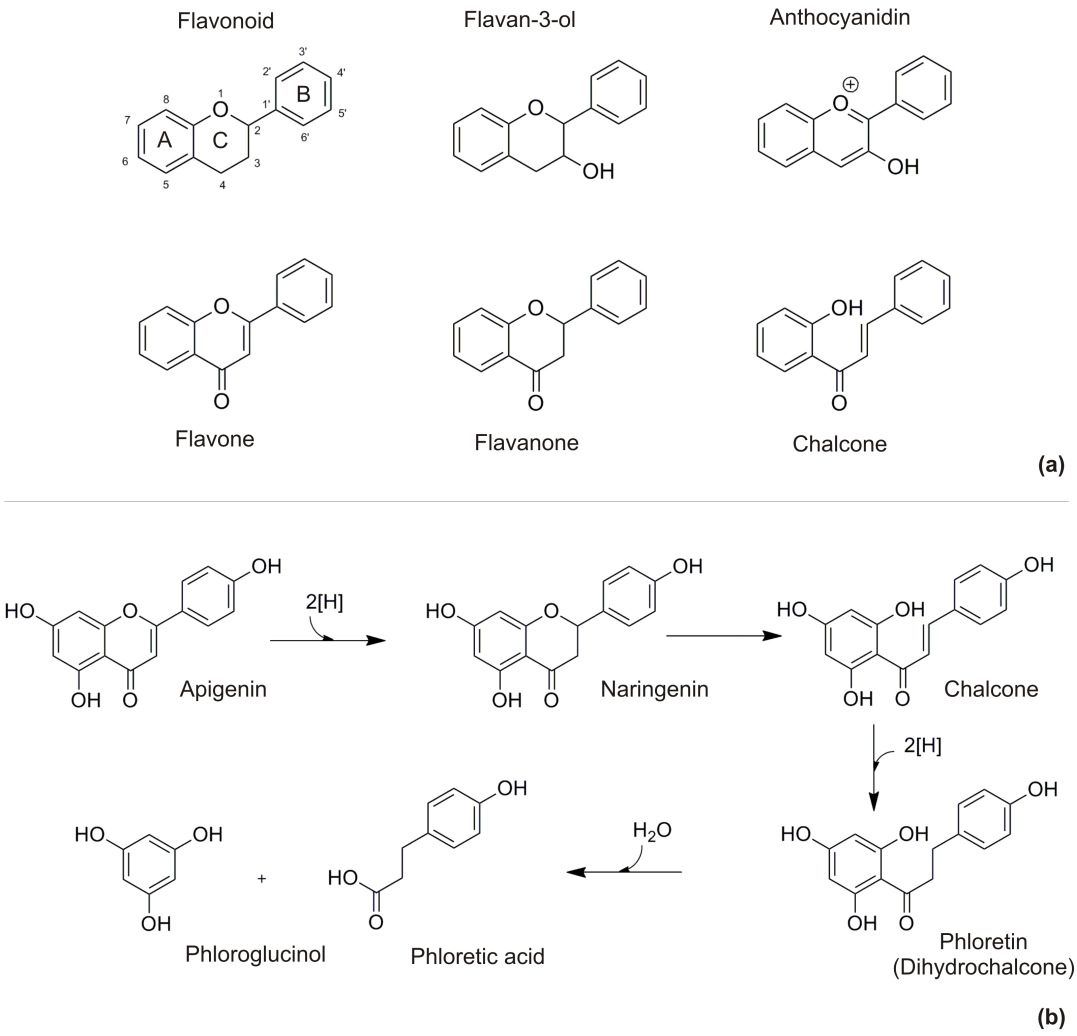


Figure 5.1: Flavonoids are plant secondary metabolites and classified according to the structure of their C-ring (or the absence thereof in the case of chalcones) (a). After hydrolytic removal of the sugar moiety, flavonoids can be reduced and isomerised into chalcones. In the proposed *E. ramulus* pathway, downstream metabolism then involves further reduction to give the dihydrochalcone phloretin, which is hydrolysed into phloroglucinol and phloretic acid (b).

5.1.4 A Friedel-Crafts hydrolase acting on phloretin

A step of particular interest for biocatalytic applications is the hydrolytic cleavage of phloretin into phloroglucinol and phloretic acid (see Fig. 5.1). This reaction constitutes a rare enzyme-catalysed hydrolytic *retro*-Friedel Crafts acylation in which a water molecule is introduced into a substrate containing an aryl and adjacent keto-group. The disruption of a carbon-carbon bond between the two groups leads to

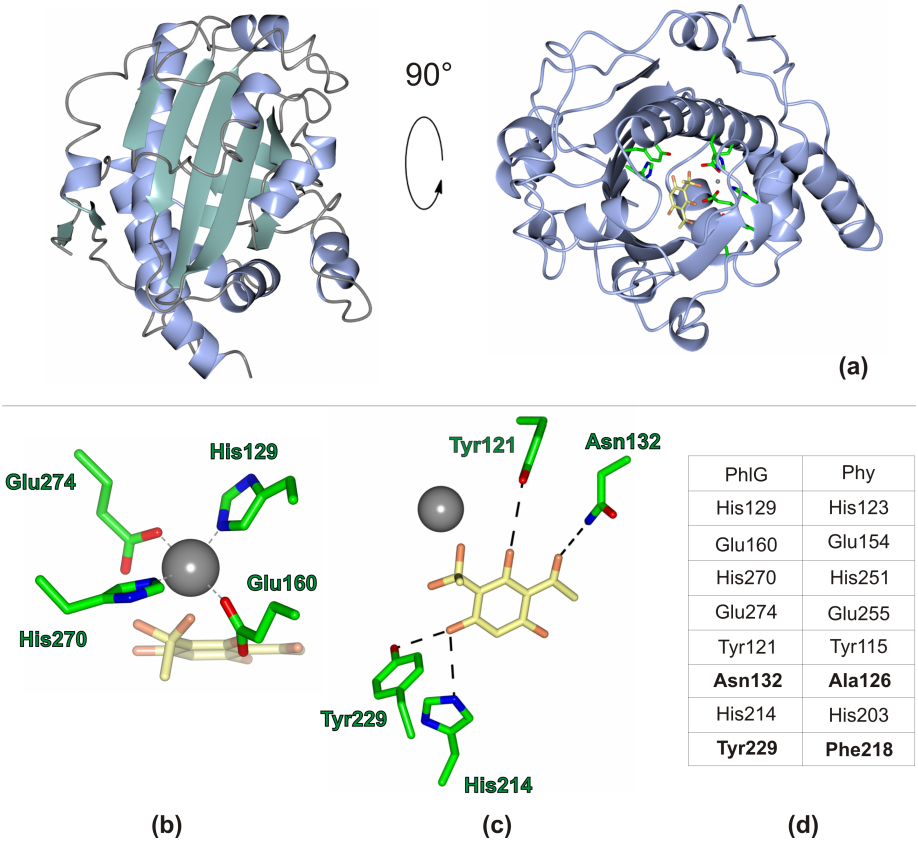
formation of a methylketone and a carboxylic acid. This reaction has been described for four different enzymes to date, the 2,6-dihydroxy-pseudo-oxynicotine hydrolase from *Arthrobacter nicotinovorans* (DHPONH) (87), the fungal heptaketide cleaving hydrolase Ayg1p (86) (for both enzymes see chapter 1, section 1.4.1.2), the diacetylphloroglucinol hydrolase from *Pseudomonas fluorescence* (PhlG) and a phloretin hydrolase from and *Eubacterium ramulus* (Phy). The enzymes have been found in both bacterial and fungal species (95,96,174,175).

The phloretin hydrolase gene from *E. ramulus* was the first to be sequenced in 2004. It could be recombinantly expressed in an *E. coli* host from which the translated protein was purified and characterised (96). Phy was shown to be a 30 kDa enzyme which associates into a homodimer in solution. Initial substrate screens showed the protein to convert phloretin into phloroglucinol and phloretic acid whereas structurally related substrates such as the phloridizin glycoside, 1,3-diphenyl-1,3-propandione and *trans*-1,3-diphenyl-2,3-epoxy-propan-1-one were not converted. The enzyme was shown to be inhibited by the chelating *o*-phenanthroline and accordingly proposed to bind a catalytic metal. Both features distinguish the *E. ramulus* Phy protein from its homologue in *Aspergillus niger* which displays a large substrate range and is only inhibited to a low level by the metal chelating agent (96,175).

5.1.5 Phy and PhlG

On closer inspection of the primary sequence, only few enzymes were shown to resemble the Phy amino acid sequence, most of them being hypothetical or non-characterised proteins. A 24% sequence identity however was found between Phy and the above mentioned PhlG hydrolase. 41% of the enzymes' amino acids are similar to each other. PhlG catalyses the hydrolytic conversion of diacetylphloroglucinol (DAPG) into monoacetylphloroglucinol (MAPG) and acetic acid and thus acts on a structurally similar substrate to Phy. Both enzymes cleave an acetyl group off their substrate's phloroglucinol moiety. Neither Phy nor PhlG contain the GXSXG motif common to many other carbon-carbon hydrolases (73) and both have thus been proposed to catalyse bond cleavage by a mechanism other than the classic catalytic serine triad. When the 3-dimensional structure of PhlG was solved in 2010, this notion was supported by the observation that the protein adapts

the Bet v1-like fold novel to carbon-carbon hydrolases (95). Furthermore, a metal atom was bound in the putative active site, coordinated by four highly conserved residues (two histidines and two glutamates). Based on these observations, computational modelling of substrate binding and kinetic data on an active site mutant library, a reaction mechanism was proposed for PhlG (95). This mechanism involves the binding and coordination of DAPG as shown in figure 5.2. A hydrogen bonding network was proposed to exist between active site amino acid side chains and the substrate's hydroxyl and acetyl moieties. Activation of a catalytic water by the divalent metal, through reduction of its pK_a, was suggested to initiate catalysis. The resulting hydroxide ion would be able to carry out a nucleophilic attack on the substrate's carbon-carbon bond, between the aryl and ketone moiety, leading to formation of a tetrahedral transition state. Rearrangement of the latter would result in carbon-carbon bond cleavage and product formation.



PhlG	Phy
His129	His123
Glu160	Glu154
His270	His251
Glu274	Glu255
Tyr121	Tyr115
Asn132	**Ala126**
His214	His203
Tyr229	**Phe218**

(d)

Figure 5.2: The P. fluorescence PhlG protein is the closest known sequence homologue to Phy and, as the latter, cleaves an acetyl moiety off its substrate's phlotoglucinol moiety. When the structure of a PhlG protein from *P. fluorescens* was determined, the authors were able to model a molecule of DAPG, in its tetrahedral transition state, into the active site (a).

The bound metal was shown to be coordinated by two histidines and two glutamates (b) and the model also revealed potential interactions of the transition state with active site residues through hydrogen bonding (c). Putative active site residues in Phy from *E. ramulus* are listed, with less or non-conserved sidechains highlighted in bold (d).

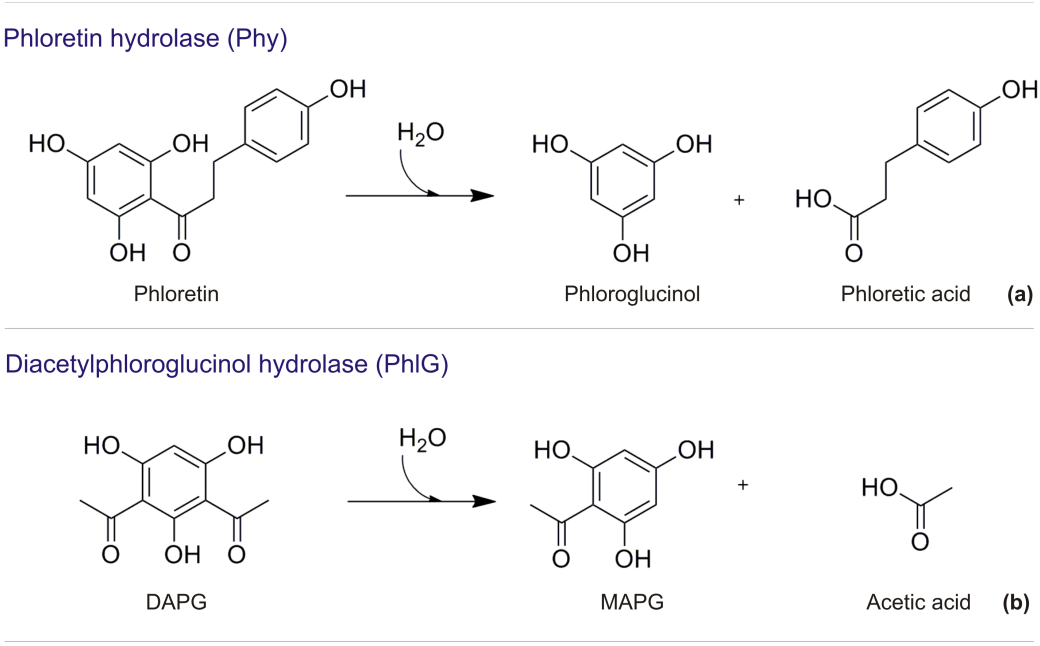


Figure 5.3: Both Phy (a) and PhlG (b) catalyse the hydrolytic cleavage between an aryl and a ketone moiety on their structurally related substrates in a formal *retro*-Friedel Crafts acylation reaction

5.1.6 Friedel-Crafts hydrolases in biocatalysis

In recent years, Friedel-Crafts hydrolases have become increasingly interesting targets from a biocatalytic perspective (176). The abiotic reaction equivalent requires a Lewis acid to activate an acid chloride by removal of the Cl^- group. The resulting carbocation can then carry out a nucleophilic attack on a benzene ring to give the acetophenone product. A reversal of the enzymatic *retro*-Friedel Crafts reactions described above would allow for the production of acylated aromatic intermediates for chemical synthesis under milder and considerably greener conditions and circumvent the need for a metal catalyst.

5.1.7 Project aims

In view of the possible applications of enzymes catalysing a formal *retro*-Friedel Crafts reaction and the potential of reversing this reaction into carbon-carbon bond formation, a more detailed understanding of these catalysts is required. For this, phloretin hydrolase from *Eubacterium ramulus* was chosen as a target enzyme, with little information being available about either structure or mechanism. It was hoped to obtain X-ray crystallographic data on the wild type enzyme to gain information on protein folding, metal coordination and substrate binding. Phy shares a number of highly conserved, putative catalytic residues with PhlG (see Fig. 5.2). Based on computational modelling, previous kinetic analyses of PhlG (95) and using a high resolution structure of the latter, a Phy active site mutant library was designed in an attempt to shed light on similarities and/or differences to the proposed PhlG mechanism. Furthermore, the introduction of two additional hydrogen bond donors present in PhlG (Asn132 and Tyr229) but absent in the Phy wild type (Ala126 and Phe218) was aimed at expanding the enzyme's substrate scope. Eventually, a screen of Phy and PhlG with previously untested phloroglucinol and hydroxyacetophenone derivatives should provide information on both enzymes' substrate requirements.

5.2 Materials and Methods

5.2.1 Ligation independent cloning of the E. ramulus phy synthetic gene

The synthetic, codon optimised *phy* gene, cloned into a pET21a(+) vector, was kindly provided by Professor Wolfgang Kroutil's group from the University of Graz. Initial expression tests showed bands for soluble protein at the correct molecular weight on 12% SDS PAGE gels. As the vector sequence however indicated the presence of a stop codon between protein and histidine tag, the *phy* gene was subcloned into the YSBLIC3C vector, using the ligation independent cloning protocol.

The LIC3C vector was prepared by amplification of the purified template in the *E. coli* XL-1 blue cloning strain, isolated using a Sigma GenElute™ plasmid Miniprep kit and purified from a 1% agarose gel using a Sigma GenElute™ gel

extraction kit. The pure plasmid was linearised by incubation with 5 units of BseRI restriction endonuclease (NEB™) per µg of DNA at 37 °C for 90 min and subsequently gel purified as described above. The hydrolase encoding synthetic gene was amplified from the pET21a(+) vector using the YSBLIC-3C specific primers sequences shown in table 5.4 (section 5.2.14) and amplification conditions shown in table 5.1.

Reaction component	Concentration	Volume [ul]	final concentration
forward primer	20 µM	1	0.4 µM
reverse primer	20 µM	1	0.4 µM
template DNA	50 mM	1	1 ng/µL
deoxyribonucleotides (dNTPS)	25 mM	1	0.2 mM
KOD® hot start DNA polymerase buffer	5x	5	1x
$MgSO_4$	25 mM	3	1.5 mM
DMSO	100 %	1	2 % (v/v)
KOD® hot start DNA polymerase	1.0U/ µL	1	0.02 U/ µL
deionised water	-	36	-

Cycles	Temperature [°C]	Time [min:s]	Reaction
1	94	2:00	initial denaturation
25	94	0:30	denaturation
	50	0:30	annealing
	72	0:30	extension
1	72	3:00	Final extension
-	4	∞	hold/storage

Table 5.1: The *phy* gene was amplified from the pET21a(+) vector in a PCR reaction using LIC specific primers and optimised cycling conditions. Reaction components (top) and amplification conditions (bottom) are shown and correspond to LIC cloning reactions for both full length protein and the truncated construct described below.

The gene was then gel purified as described above for the LIC3C vector. For preparation of the single stranded overhangs, both plasmid and insert were then individually incubated at 22 °C for 30 min under conditions listed in table 5.2.

Vector		Insert	
Linearised vector	0.5-1.0 pmol	purified gene of interest	0.2 pmol
T4 polymerase (Novagen™)	5 U	T4 polymerase (Novagen™)	1.25 U
T4 polymerase buffer (10 x)	10 µL	T4 polymerase buffer (10 x)	2 µL
dTTP (25 mM)	10 µL	dATP (25 mM)	2 µL
Dithiothreitol (DTT) (100 mM)	5 µL	Dithiothreitol (DTT) (100 mM)	0.5 µL
H_20	To 100 µL	H_20	To 20 µL

Table 5.2: Reaction components are shown for the T4 polymerase reactions which led to the addition of complementary single standed overhangs on linearised vector and insert DNA.

The reaction was stopped by a further incubation at 75 °C for 20 min and (for the vector reaction only) purified using a Sigma GenElute PCR clean-up kit. Using 1.5 µL of vector and 3 µL of insert DNA, the created single stranded overhangs were eventually annealed at room temperature for 10 min prior to the addition of 1.5 µL of 25 mM EDTA and a further incubation for 10 min. For amplification, an *E. coli* XL-1 Blue cloning strain was transformed by the annealing products. The resulting colonies were used for the preparation of 5 mL overnight cultures from which amplified plasmids were purified with a Sigma GenElute™ plasmid Miniprep kit. To check for successful insertion of the gene into its vector, a double digest using 1 unit each of *NcoI* and *NdeI* restriction endonucleases (NEB™) per µL of DNA was carried out at 37 °C for 90 min. The result was visualised on a 1% agarose gel. Eventually, the vector constructs were sent for sequencing with the University of York's Technology Facility to verify the sequence and completeness of the *phy* gene in its vector.

5.2.2 Purification of Phy^{wt}

For purification of the Phy protein, competent *E. coli* BL21 (DE3) expression cells were transformed by the YSBLIC3C:*phy* plasmid and colonies were picked for preparation of 5 mL LB starter cultures. The latter were used for inoculation of 2 to 4 L of autoinduction medium at a ratio of 1:100 (starter culture : growth medium) (177). In contrast to the Studier reference, the amount of metals added was reduced by a factor of five, with 0.1 mL of 1000 x metal solution being added to each 500 mL culture. After inoculation, all cultures were incubated at 37° C for 18 h while shaking at 180 rpm. Cells were then harvested by centrifugation at 5000 rpm for 15 min using a Sorvall RC5B centrifuge and the resulting pellet was resuspended in 30 mL of 50 mM Tris buffer (pH 7.5), containing 300 mM NaCl and 30 mM imidazole, per litre of culture. Ultrasonication at an amplitude of 14,000 microns for 3 x 45 s, with 30 s cooling intervals, was used to disrupt cells. Soluble lysate was subsequently separated from the crude cell extract by centrifugation at 15,000 rpm for 30 min. The resulting supernatant was then loaded onto a 5 mL HiTrap™ Chelating HP Nickel affinity column. After recycling of the flow-through for 30 min and a final loading step, the column was washed with ten column volumes of the resuspension buffer. The hexahistidine-tagged Phy protein was eluted using a 30-500 mM imidazole gradient over 20 column volumes. The pooled peak fractions were concentrated and loaded onto a Superdex™ S75 gel filtration column and equilibrated with 50 mM Tris buffer (pH 7.5) containing 300 mM NaCl. During all stages of purification, 10 µL samples were taken to evaluate protein purity on a 12% SDS PAGE gel. After the final purification step, all proteins were stored at 4 °C to be used in crystallisation experiments or flash-cooled in liquid nitrogen and stored at -18 °C for kinetic experiments. For a subset of the purified proteins, the N-terminal hexahistidine tag was removed by protease digest. For this, protein samples were brought to a concentration of 1 mg/mL and incubated with the human rhinovirus (HRV) 3C protease at a ratio of 1:10 (w/w) (protease : protein of interest) for 18 h at 4 °C. Repeated Ni^{2+}-affinity purification allowed for separation and elution of the cleaved protein in gel filtration buffer. The cleaved polypeptide and histidine tagged protease were eluted from the column in the same buffer containing a 0-500 mM imidazole gradient.

5.2.3 Purification of Phy^{SeMet}

As only one structural model at comparatively low sequence identity was available for determination of a Phy protein structure by molecular replacement (see section 5.1), a Phy derivative was purified in which methionine residues were replaced by selenomethionine. It was hoped that protein crystals of this heavier derivative would support phase calculation during structure determination. For purification of a Phy selenomethionine derivative, 50-100 µL of methionine auxotrophic *E.coli* B834 (DE3) competent cells were transformed by the YSBLIC3C:Phy wild type plasmid. The resulting colonies were picked and used for inoculation of 5 mL of LB overnight cultures containing 30 µg/mL of kanamycin as the selective antibiotic. These cultures were in turn used to inoculate 2 L of PASM-5052 selenomethionine autoinduction medium (177) at a ratio of 1:100 (starter culture : expression medium). In contrast to the Studier reference, the amount of metals added was reduced by a factor of five, with 0.1 mL of 1000 x metal solution being added to each 500 mL culture. Selenomethionine cultures were incubated at 37 °C for 22 hours while shaking at 180 rpm, before cells were harvested and the protein purified as described above for the native Phy hydrolase.

5.2.4 Prediction of protein disorder and preparation of a truncated Phy construct

Initial crystallisation screens showed only few conditions in which microcrystals or crystalline precipitate formed and the observed hits were difficult to reproduce. In view of this result, the protein sequence was analysed for disordered regions. The latter may impair the crystallisation process and their removal may help promote protein order and crystal formation. For disorder prediction, several programs were used and compared, including the RONN (178) and DISOPRED (179) servers. Once disordered regions were determined, a LIC3C-specific primer was designed for subcloning of the truncated construct. As a full length YSBLIC3C:*phy* gene had been prepared previously, only a forward primer for removal of 10 N-terminal residues had to be designed and is shown in table 5.4. The combination of truncated forward and full-length reverse gene primers was then used to carry out LIC cloning as described above for the full length gene. Reaction components and amplification

conditions are outlined in table 5.1. These procedures resulted in preparation of the YSBLIC3C:Phy^{-10N} plasmid.

5.2.5 Purification of Phy^{-10N}

The truncated protein, hereafter named Phy^{-10N}, was produced, purified and used for the preparation of crystallisation screens as described for the native, full length hydrolase. Cells of the *E. coli* BL21(DE3) expression strain were transformed by the YSBLIC:Phy^{-10N} plasmid and cultured for protein production. In contrast to the methods described above however, resuspension and purification buffers used were based on 50 mM HEPES buffer (pH 7.5) containing 100 mM KCl.

5.2.6 Computational modelling of the Phy 3-dimensional structure

In the absence of structural data, kinetic experiments involving the mutation of putative active site residues were guided by computational modelling of the Phy tertiary structure. For this, the protein's amino acid sequence was submitted to the PHYRE^2 protein fold recognition server. In subsequent steps, this server creates a sequence alignment, predicts secondary structural elements such as α-helices, β-sheets and coils. These elements are then compared to a fold library, allowing for prediction of a complete tertiary structure for the protein of interest. Confidence scores are assigned to each prediction (180).

5.2.7 Preparation of a Phy mutant library

A total of nine Phy active site mutants were designed based on previous kinetic data, a modeled ligand complex of the PhlG homologue [He] and a computational model of Phy (see sections 5.1.5, 5.2.6 and figure 5.2). All mutant genes/plasmids were subsequently prepared by site directed mutagenesis. For this, primers containing the desired single point mutation were designed using the *PrimerX* software and are listed in table 5.4 (section 5.2.14). For mutagenesis, each primer pair was used for amplification of the wild type YSBLIC3C:*phy* template under the conditions outlined in table 5.3.

Subsequent to this amplification, the parental DNA was removed by the addition of 20 units of DpnI restriction endonuclease to the reaction mixture. After centrifugation at 13,000 rpm for 60 s, this digest was incubated at 37 °C for 3-4 h after which 50 – 100 µL of supercompetent XL-10 Gold cells were transformed by 2-5 µL of the mutant plasmid DNA.

Reaction component	Concentration	Volume [ul]	final concentration
forward primer	20 pmol/µL	0.5	0.4 µM
reverse primer	20 pmol/µL	0.5	0.4 µM
template DNA (YSBLIC:*phy*)	100 ng/ul	1	2 ng/µL
deoxyribonucleotides (dNTPS)	2 mM	5	0.2 mM
KOD (*) or Pfu turbo™ (**) hot start DNA polymerase buffer	10x	5	1x
$MgSO_4$(*)	25 mM	3	1.5 mM
DMSO	100 %	1	2 % (v/v)
KOD (*) or Pfu turbo™ (**) hot start DNA polymerase	1.0 U/µL (*) 2.5U/ µL(**)	1	0.02 U/ µL (*) 0.05 U/ µL (**)
deionised water	-	33* (36**)	-

Cycles	Temperature [°C]	Time [min:s]	Reaction
1	95	5:00	initial denaturation
30	95	1:00	denaturation
	50	1:00	annealing
	72	10:00	extension
1	95	1:12	final denaturation
1	72	5:00	final extension
-	4	∞	hold/storage

Table 5.3: The reaction components and amplification conditions shown were used for the preparation of a *Phy* active site mutant library by site-directed mutagenesis. Polymerases required varied between mutants: whereas PfuTurbo™ polymerase was used to amplify the Phe218Tyr and His203Ala mutants (**), KOD polymerase was used to prepare all other variants (*).

Colonies resulting from the transformation reaction were picked and used for the preparation of 5 mL overnight cultures in LB medium containing 30 μg/mL kanamycin as the selective antibiotic. Plasmid DNA was isolated from these cultures using a Sigma GenElute™ plasmid Miniprep kit and submitted to the University of York's Technology Facility for sequencing. Using the Emboss Needle webserver (181), the obtained results were aligned with the *phy* wild type gene to verify presence and identity of the desired mutation.

5.2.8 Protein production and purification of the Phy mutant library

For protein purification, competent *E. coli* BL21 (DE3) and B834 (DE3) cells were transformed by the mutant plasmid DNA, cultured and soluble proteins purified as described for the wild type Phy. The lack of a distinct chromatography peak during metal affinity purification was considered to be the consequence of either insoluble proteins or a complete lack of gene expression and/or protein production. This was observed for the $Phy^{His251Ala}$, $Phy^{Glu255Gln}$ and $Phy^{Ala126Asn}$ variants. For analysis of these proteins, three different expressions strains, *E. coli* BL21 (DE3), Rosetta (DE3) and B834 (DE3), were transformed by the respective plasmid DNA and tested for protein production and solubility in small-scale expression tests. Picked colonies were used to prepare 5 mL overnight cultures from which 100 mL expression cultures could be inoculated at a ratio of 1:100 (starter culture : expression medium). For the Glu255Gln mutant, only *E.coli* Rosetta colonies were obtained and could be cultured. At an optical density of 0.6 – 1.0, protein production was induced by the addition of IPTG to a final concentration of 0.1 mM and cultures were incubated for 18 h at 37, 30 or 18 °C. For visualisation of protein production, 10 mL were aliquoted from each culture and cells were harvested in a Sorvall RT6000 D centrifuge at 5,000 rpm for 15 min. After resuspension in 1 mL of 50 mM Tris buffer containing 300 mM NaCl, cells were disrupted by sonication at 14,000 amplitude microns for 3 x 3-4 s with 5 s cooling intervals. The crude cell extracts were subsequently centrifuged at 13,000 rpm for 30 min. 10 μL samples were taken from the supernatant and from insoluble pellets resuspended in 1 mL of buffer, representing soluble and insoluble cell fractions respectively. All samples were eventually separated and visualised on a 12 % gel using SDS PAGE.

As no soluble protein production could be observed from these experiments, further purification of the three respective mutants was abandoned.

5.2.9 Crystallisation of Phy^{wt}, Phy^{SeMet},Phy^{-10N} and mutant proteins

The purified wild type Phy (native), selenomethionine, truncated and mutant proteins were subjected to crystallisation trials in 96 well Swissci MRC™ trays, using 300 nL drops in the sitting-drop vapour diffusion method. Crystallisation screens assayed included the CSS I+II, PACT and INDEX as well as Newcastle, Hampton, PGA, PEG and additive screens for the native, full length, Phy^{wt} protein. Selenomethionine, truncated and mutant proteins were screened using the CSS I+II, PACT and INDEX conditions only. Protein and mother liquor were incubated at a ratio of 1:1 (v:v).

Whereas several of the variants yielded small crystals or crystalline precipitate, only the wild type protein (full length, native) gave reproducible, single, diffracting crystals using the INDEX screen (Hampton Research) at a protein concentration of 10 mg/mL and in the absence of the hexa-histidine purification tag. Despite low diffraction quality, a crystal grown in PEG 3350 (25% w/v) and 0.1 M Bis-Tris propane buffer (pH 5.5) and in the absence of added metals was used to measure an X-ray fluorescence spectrum at the Diamond Light Source in Didcot (see below). This condition was subsequently scaled up and optimised using 24-well Linbro-type crystallisation plates and the hanging drop vapour diffusion method. The best results were eventually obtained from crystallisation conditions containing PEG 3350 (26% w/v), 0.1 M Bis-Tris Propane buffer (pH 6.0), 1 mM ZnOAc and 250 μM phloretin (dissolved as a 500 mM stock solution in DMSO). Crystals were cryoprotected using the original mother liquor. The nature and percentage of the cryoprotectants was optimised by comparing diffraction qualities of crystals after the addition of 0-20 % (v/v) ethylene glycol or 0-20 % glycerol (v/v). All crystals were flash-cooled in liquid nitrogen prior to in-house testing.

5.2.10 Crystal testing, data collection and processing

Crystals grown in 96 well or 24 well plates were fished, cryoprotected, flash-cooled as described above and tested using an in-house X-ray source and a

MARRESEARCH MAR345 imaging plate detector. If a protein diffraction pattern was observed, crystals were stored in liquid nitrogen and sent to the Diamond Light Source in Didcot for further testing. For a native, wild type Phy crystal, a dataset was collected at a wavelength of 0.97631 Å on beamline I03, using a Pilatus detector. Data were evaluated and processed using the *xia2* automated processing software at Diamond and the *CCP4* suite of programs (152,160). Alternative models were built and used in molecular replacement and refinement by the *BALBES* molecular replacement pipeline (182). For crystals of the Phy^{SeMet} heavy atom derivative, three datasets were collected at wavelengths of 0.97949, 0.97961 and 0.97630 Å, corresponding to the selenium peak, inflection point and remote X-ray energy respectively.
All diffraction data for the selenomethione derivative were evaluated using the *xia2* automated processing software (152) and the *SHELXC/D/E* software package (183).

5.2.11 Measurement of an X-ray fluorescence spectrum

A weakly diffracting Phy crystal was obtained from the 96 well INDEX screen. This crystal had grown in the absence of any metals in the crystallisation solution (0.1 M Bis-tris propane (pH 5.5) and 25 % PEG 3350). It could thus be used to measure an X-ray fluorescence spectrum and detect metals tightly associated with the protein of interest. For this spectrum, the crystal was exposed to radiation at a wavelength of 0.9795 Å at beamline I04 of the Diamond Light Source synchrotron in Didcot. Emitted fluorescence was detected and measured on a Vortex™ fluorescence detector and evaluated using the *PYMCA* software package (184).

5.2.12 Spectrophotometric analysis of phloretin hydrolysis

To analyse hydrolysis of phloretin into phloroglucinol and phloretic acid, a spectrophotometric assay was developed. Making use of a drop in absorbance of phloretin at 288 nm, this assay allowed for a real-time observation of substrate turnover by the Phy^{wt}, Phy^{-10N} and Phy^{mut} proteins. Stock solutions of 1, 10 and 100 mM phloretin were prepared in dimethyl sulfoxide (DMSO) and added to 50 mM Tris buffer (pH 7.0) to give a final concentration range of 1 to 150 μM. After a flat baseline was observed, 1 μg (0.03 nmoles) of purified Phy enzyme was added

to the solution and the decrease in absorbance was measured for 60 s or 100 s for the wild type and mutant enzymes respectively. The molar extinction coefficient, ε, for phloretin was calculated by measuring the absorbance values of nine samples with a known substrate concentration of 5 – 80 μM and determined to be 11039 M^{-1} cm^{-1} using Beer's law. Subsequently, using the same law, the reduction in the each reaction sample's phloretin concentration could be determined from the measured drop in absorbance. Maximal substrate concentrations assayed varied between wild type and mutant proteins. Further measurements were generally abandoned once a plateau could be observed in the initial reaction velocities or absorbance values approached signal saturation (Phy^{wt}). All measurements were performed in triplicate and standard deviations were calculated for each datapoint. From initial reaction velocities determined in this experiment it was possible to calculate kinetic constants K_m, V_{max}, k_{cat} and k_{cat}/K_M for phloretin hydrolysis using the Grafit software package (185).

5.2.13 Determination of Phy substrate specificity

The substrate range of Phy was analysed at the University of Graz, using gas chromatography assays in which the both substrate and products of hydrolysis could be detected. Their identity was confirmed using standards and methods previously prepared and kindly provided by Dr. Elina Siirola. A total of six potential substrates were analysed: diacetylphloroglucinol (DAPG), monoacetylphloroglucinol (MAPG), 2-hydroxyacetophenone (2-HAP), 4-hydroxyacetophenone (4-HAP), 2,4-dihydroxyacetophenone (2,4-DHAP) as well as 2,6-dihydroxyacetophenone (2,6-DHAP). For the enzymatic reactions, an *E. coli* BL21 (DE3) expression strain was transformed by the YSBLIC3C plasmids encoding Phy^{wt} protein as well as the Ala126Ser and Phe218Tyr mutants designed to introduce additional hydrogen bond donors to the active site. 1 L of recombinant cells were cultured, disrupted and cell fractions separated as described for the Phy mutant library. The soluble supernatant was then flash cooled in liquid nitrogen and dried under vacuum for approximately 18 h before being stored at 4°C until further use. For determination of substrate specificity, the lyophilised lysates were resuspended to a final concentration of 20 mg/mL in 50 mM phosphate buffer, pH 7.5, and incubated for 30 min to equilibrate the enzyme solution. All putative substrates were dissolved as 5 M stock

solutions in DMSO and added to each reaction (with the exception of DAPG which was added directly) to give a final concentration of 50 mM substrate in 1 mL of buffer. Final solvent concentrations did thus not exceed 1% (v/v). Each reaction was prepared in duplicate and then incubated at 30°C while shaking at 120 rpm for approximately 18 h. After subsequent acidification using 2 drops of concentrated hydrochloric acid, all samples were extracted three times in 500 µL of ethyl acetate and the combined organic phases were dried over anhydrous Na_2SO_4. They were then analysed using an Agilent Technologies 7890A gas chromatograph and a HP-5 gas chromatography column. The injector temperature was set to 300°C, assay temperatures and retention times are listed below.

For all transformations, a PhlG reaction was run in parallel. The corresponding protein was prepared as described for Phy, using a recombinant plasmid kindly provided by Dr. Elina Siirola. For hydrolysis of DAPG and MAPG, this served as a positive and negative control respectively.

GC conditions and retention times:

MAPG and DAPG: 200°C, hold for 2 min, 10°C/min to 250°C, hold for 3 min. Retention times: 4.5 min (DAPG), 3.5 min (MAPG) and 2.2 min (phloroglucinol, PG). 2,4-DHAP and 2,6-DHAP: 190°C, hold for 5 min. Retention times: 2.4 min (2,4-DHAP and 2,6-DHAP), 1.7 min (1,3-dihydroxybenzene/resorcinol).

2-HAP: 140°C, hold for 5 min. Retention times: 2.5 min (2-HAP) and 1.9 min (phenol). 4-HAP: 140°C, hold for 7 min. Retention times: 5.6 min (2-HAP) and 1.9 min (phenol).

5.2.14 Primers used for LIC cloning of Phy full length and truncated proteins as well as for the preparation of an active site mutant library

Variant		Forward Primer	Reverse Primer
1	WT	ccagggaccagcaatggaagaagatttc aatatgagcacaccgggtg	gaggagaaggcgcgttattaaaaggtatctttttc ttctgcatacaggg
2	Phy^{-10N}	ccagggaccagcaggtgttaaagttggtg ttagcgaagaagaaaaaaaactg	gaggagaaggcgcgttattaaaaggtatctttttc ttctgcatacaggg
3	His123Ala	gatccggaagatgcgtattttgcacg	cgtgcaaaatacgcatcttccggatc
4	Glu154Gln	gtggatcattatattatgcaggatgttggtc cgggtcc	ggacccggaccaacatcctgcataatataatgat ccac
5	His251Ala	cgcagggtctgtttgcagcgaatattaaag aatttacc	ggtaaattctttaatattcgctgcaaacagaccct gcg
6	Glu255Gln	gcacataatattaaacagtttaccaatctgg cag	ctgccagattggtaaactgtttaatattatgtgc
7	Tyr115Phe	ggtagcgatctgcgttttaaaatttgggatc c	ggatcccaaattttaaaacgcagatcgctacc
8	His203Ala	ctgtgcagcagcaatgaccgcgaaatgg catccgtataaag	ctttatacggatgccatttcgcggtcattgctgctg cacag
9	Ala126Asn	cggaagatcattattttaatcgtgcatatcc ggcaagc	gcttgccggatatgcacgattaaaataatgatctt ccg
10	Ala126Ser	cggaagatcattattttagccgtgcatatcc ggcaagc	gcttgccggatatgcacggctaaaataatgatctt ccg
11	Phe218Tyr	ctgtttgaaagccgctattggattggttatc g	cgataaccaatccaatagcggctttcaaacag

Table 5.4: Primers used for the amplification of the wild type full length (1) and truncated *phy* (2) genes were optimised for LIC cloning. Rows 3-11 show primers used for the preparation of an active site single point mutant library. Nucleotides encoding the desired mutation are highlighted in red.

5.3 Results

5.3.1 Ligation independent cloning of the E. ramulus phy synthetic gene and the truncated construct Phy^{-10N}

To allow for efficient purification by metal affinity chromatography and subsequent cleavage of the affinity tag, the synthetic *phy* gene was sub-cloned into the YSBLIC3C vector. Within the vector, the gene is flanked by the NdeI and NcoI endonuclease restriction sites. A double digest of the YSBLIC3C:*phy* expression construct could thus confirm the presence of the gene of interest (see Fig. 5.4),

subsequent sequencing confirmed its completeness and identity with the deposited gene sequence. To remove a region of predicted protein disorder, the truncated construct Phy^{-10N} was prepared in addition, lacking the first ten N-terminal amino acids. The corresponding gene was amplified and subcloned analogous to *phy*^{wt}.

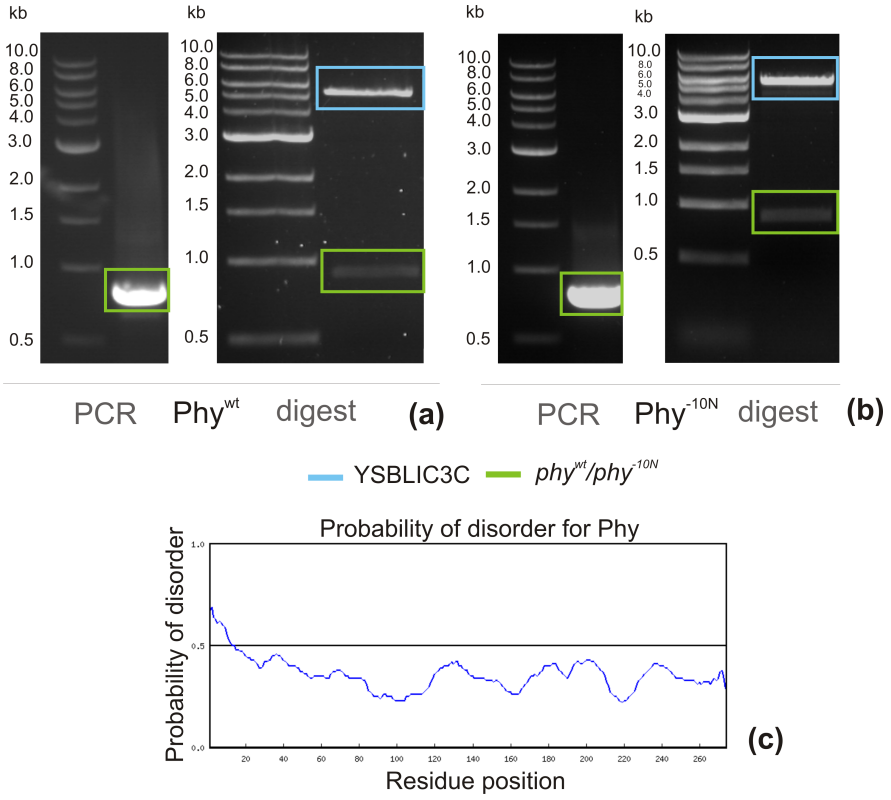


Figure 5.4: After PCR amplification and LIC cloning, an NcoI/NdeI restriction endonuclease double digest confirmed the presence of the *phy* gene (appr. 830 kb, green boxes) inserted into the YSBLIC3C vector (5.4 kb, blue box) (a). Protein disorder in the first ten N-terminal residues was predicted using the RONN server (178) (c) and prompted the amplification and subcloning of the truncated *phy*$^{-10N}$ construct (b).

5.3.2 Purification of Phy^{wt} *type, Phy*^{SeMet} *and Phy*$^{-10N}$ *proteins*

The native wild type Phy protein as well as its selenomethionine and truncated derivatives were produced as soluble enzymes by the *E. coli* recombinant hosts and could be purified to high homogeneity in two to three subsequent steps. Initial purification yields for the native protein were low using LB growth medium and IPTG based induction of gene expression (results not shown). As a consequence, recombinant cells were grown in autoinduction medium, leading to a 200-300%

increase in Phy production levels. Protein purity, although negatively influenced by the overall increase in cellular protein levels, improved significantly with every purification step. The enzymatic removal of the hexahistidine tag and repeated purification by nickel affinity chromatography in particular gave protein samples with few detectable impurities on SDS PAGE gels (see Fig. 5.5).

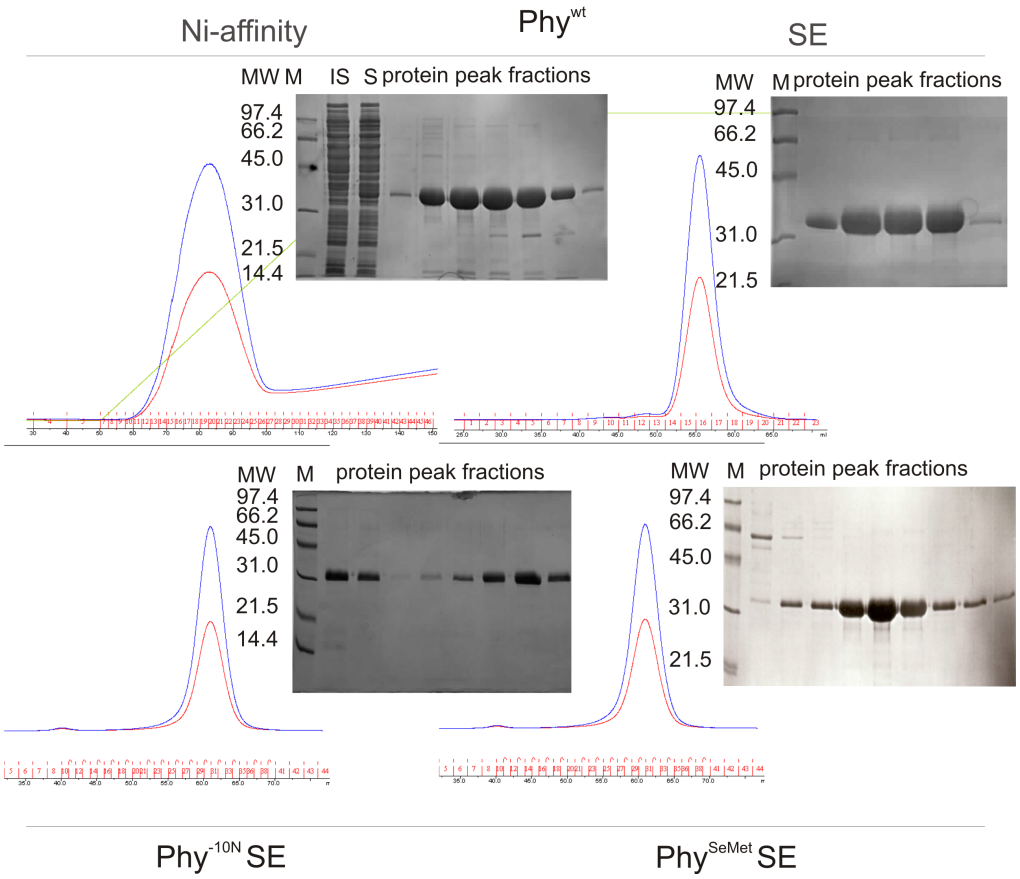


Figure 5.5: Soluble Phy^{wt}, Phy^{SeMet} and Phy^{-10N} protein was produced by the *E.coli* recombinant host and could be purified to high homogeneity using Ni^{2+}-affinity and size exclusion chromatography (SEC). FPLC chromatograms show elution of the proteins in single, narrow peaks corresponding to the protein fractions visible on 12 % SDS PAGE gels. The Phy^{SeMet} derivative, produced in minimal autoinduction medium, showed the highest of all protein production levels with over 60 mg of protein per litre of culture medium after metal affinity purification.

5.3.3 Computational modeling of the Phy 3-dimensional structure

In the absence of crystallographic data for Phy, experiments analysing the catalytic function of putative active site residues were based on a computational modelling

approach. Using the PHYRE^2 webserver (180), a three-dimensional model (Phy^{mod}) could thus be prepared. With respect to Phy, the only protein structure with significant sequence homology determined to date is that of diacetylphloroglucinol hydrolase (PhlG) (95). As a consequence, the phloretin hydrolase model produced by the PHYRE server strongly resembles the PhlG structure, particularly around the core of the protein. This core shows an arrangement of two orthogonal, anti-parallel β-sheets surrounding an internal cavity, similar to a β-sandwich. The sheets are surrounded on all four sides by α-helices and also close off both ends of the internal tubular cavity. The putative active site contains all residues previously proposed to be essential for catalysis in PhlG. All equivalent side chains in the Phy model are situated in a similar position when the two structures are superposed (see Fig. 5.6).

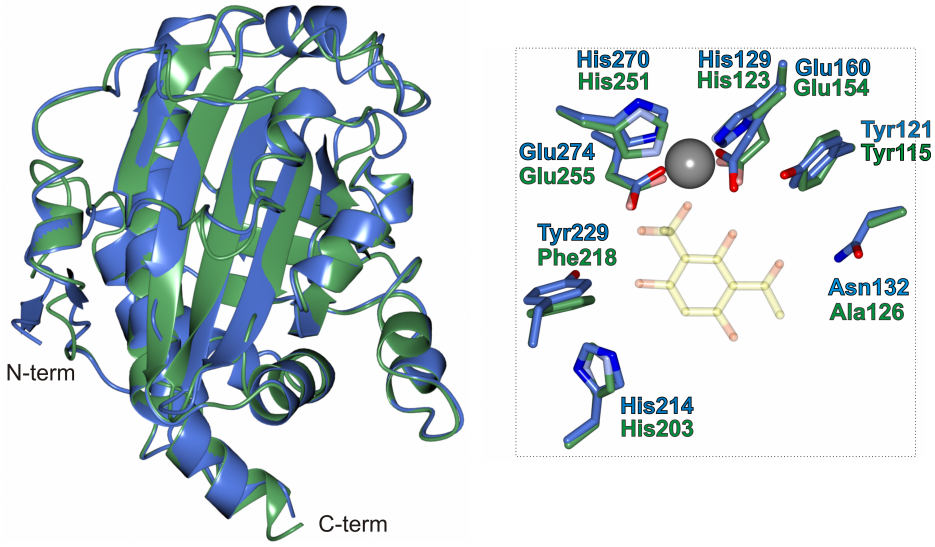


Figure 5.6: The superposition of Phy^{mod} with PhlG suggests high structural similarity between the two proteins (a). A superposition of the putative PhlG active site (including a modelled DAPG transition state) and the corresponding Phy^{mod} side chains highlights structural conservation and changes in this area (b).

The root mean square deviation (RMSD) between both structures is 0.64 Å over 258 residues. With the Phy model being based on structural data from PhlG, this value is however strongly biased.

5.3.4 Preparation, protein production and purification of a Phy mutant library

Based on previous work on the PhlG homologue (95) and the modelling approach described above, a Phy mutant library was prepared by site-directed mutagenesis.

Using this method, nine single amino acid alterations were successfully introduced into eight positions of the Phy primary sequence (see table 5.4). The alteration of an alanine in position 126 to an asparagine led to the production of insoluble protein. When the residue was mutated to an alternative hydrogen bond donor, a serine, the resulting protein was produced in its soluble form.
Of the remaining seven variant *phy* genes, five were expressed and translated into soluble enzymes. Protein production varied slightly between the mutants but was significantly increased for the His203Ala variant. The latter showed protein levels increased by up to a factor of ten compared to other mutants. Representative SDS PAGE gels of two mutant library proteins ($Phy^{His203Ala}$ and $Phy^{Tyr115Phe}$), purified by size-exclusion chromatography, are shown in Fig. 5.7.

Gene expression and protein production was assayed for mutants not producing a detectable peak during metal-affinity chromatography. When induced cultures of three *E. coli* expression systems, incubated at 37, 30 and 18 °C, were analysed by SDS PAGE, no significant Phy protein bands could be detected in the soluble cell fractions. Protein bands were however generally weak on the prepared gels. Significant amounts of insoluble protein at the Phy molecular weight could only be detected for the His251Ala mutant. The negative control shows that results remain ambiguous however as uninduced expression cells show a band of the same molecular weight as the Phy protein (see Fig. 5.7).

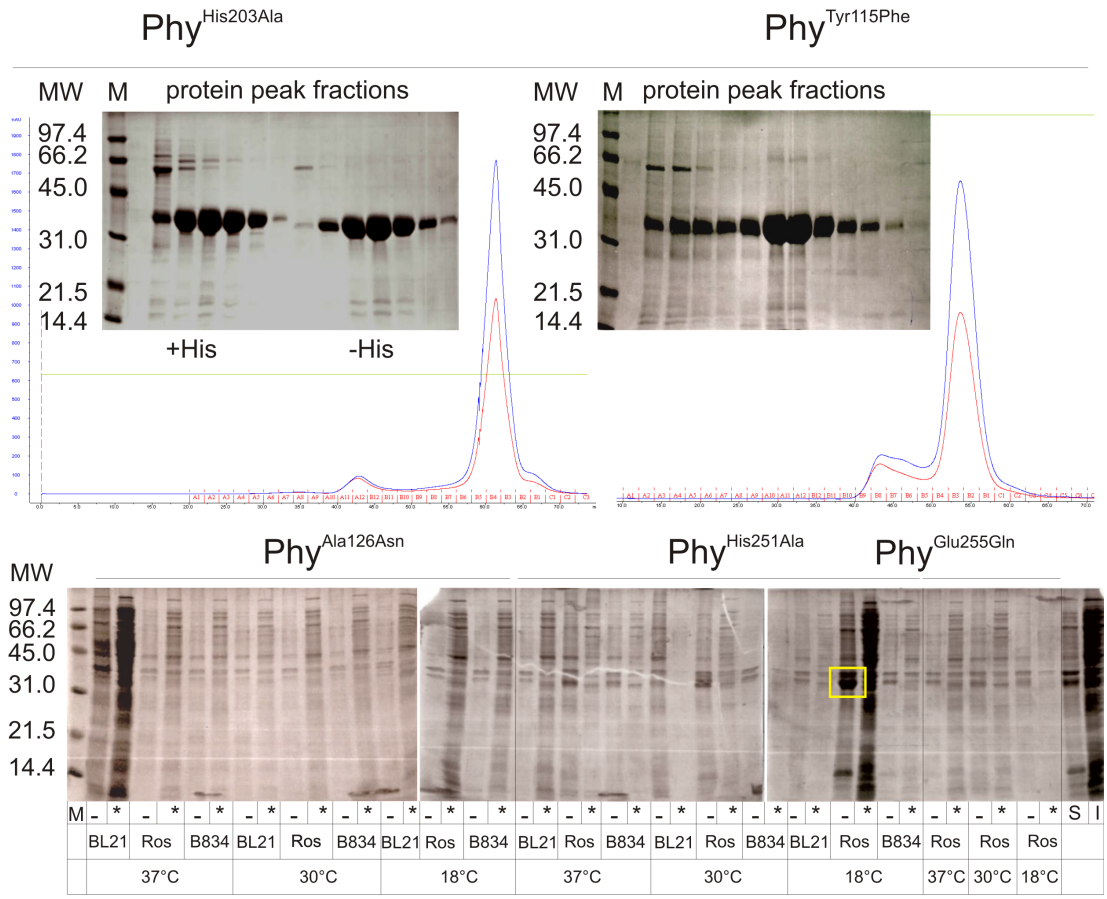


Figure 5.7: Size-exclusion chromatography results are shown for the $\text{Phy}^{\text{His203Ala}}$ and $\text{Phy}^{\text{Tyr115Phe}}$ variants, representing non-conservative and conservative amino acid alterations respectively (top). Mutants which could not be purified by Ni^{2+}-affinity chromatography were analysed for protein production using the *E.coli* expression strains BL21, Rosetta (Ros) and B834 (DE3) and incubation temperatures ranging from 18 - 37°C. Soluble (*) and insoluble (-) cell fractions are shown. A possible overproduction of insoluble $\text{Phy}^{\text{Glu255Gln}}$ protein is highlighted (bottom).

5.3.5 Crystallisation of Phy^{wt}, Phy^{SeMet}, Phy^{-10N} *and mutant proteins*

The purified Phy wild type protein as well as all of its selenomethionine, truncated and mutated derivatives were used in crystallisation experiments in the hope of obtaining a three dimensional structure of the enzyme. In several of the crystallisation screens carried out with the wild type protein, the formation of oval or spherulite shaped structures could be observed. These were present in various screens but particularly dominant with the INDEX conditions. From the latter, 24-well scale-up trays were prepared and presented larger structures of similar

morphology but, in parts, with detectable "crystalline" edges (see Fig. 5.8). When fished, flash cooled and exposed to synchrotron radiation at the Diamond Light Source in Didcot, these crystals showed the typical protein diffraction pattern to a resolution of approximately 4 Å. A second, almond shaped crystal form was detected in the INDEX screen, several months after its preparation, and diffracted to approximately 5 Å at the synchrotron. As this crystal had formed in the absence of metals in the crystallisation conditions, it was suitable for the measurement of a fluorescence spectrum (described below).

The apparent zinc binding properties of Phy, revealed by the fluorescence scan, were exploited in subsequent optimisation experiments with conditions producing the almond shaped crystal. The addition of zinc chloride or zinc acetate as well as the variation of polyethylene glycol (PEG) 3350 concentration and pH yielded single and reproducible, cube shaped Phy crystals (see Fig. 5.8). When exposed to X-ray radiation from an in-house source, the crystals diffracted to 5-8 Å. At Diamond Light Source, this resolution was improved to 3.5 Å for a crystal grown in the presence of 26 % PEG 3350, 0.1 M bis-tris propane buffer (pH 6.0) and 1 mM ZnOAc. The diffraction quality was generally higher for crystals grown in the presence of ZnOAc than from conditions containing ZnCl.

With the conditions described above, 24-well crystallisation screens were prepared for the Phy^{SeMet} derivative. After 24 h, large crystals of various morphologies could be observed. These were either single or appeared in larger structures of apparently overlapping crystals (see Fig. 5.8). When a single crystal was fished, diffraction to 5-6 Å was observed at the Diamond Light Source synchrotron.

In addition to the full length native Phy and its selenomethionine derivative, single crystals and crystalline precipitate were observed for the Phy^{-10N} truncated protein and several proteins of the mutant library. For the latter, the only hit which could be successfully reproduced and scaled up however was the formation of (originally) plate shaped crystals of the $Phy^{His203Ala}$ mutant in the PACT screen. Using a mother liquor containing 0.2 M MgCl, 25% PEG3350 and 0.1 M Bis tris buffer (pH 6.0, 6.5 or 7.0), rod shaped, twisted crystals were obtained after a 24 h in 24-well trays. None of the crystals diffracted when tested using a home X-ray source.

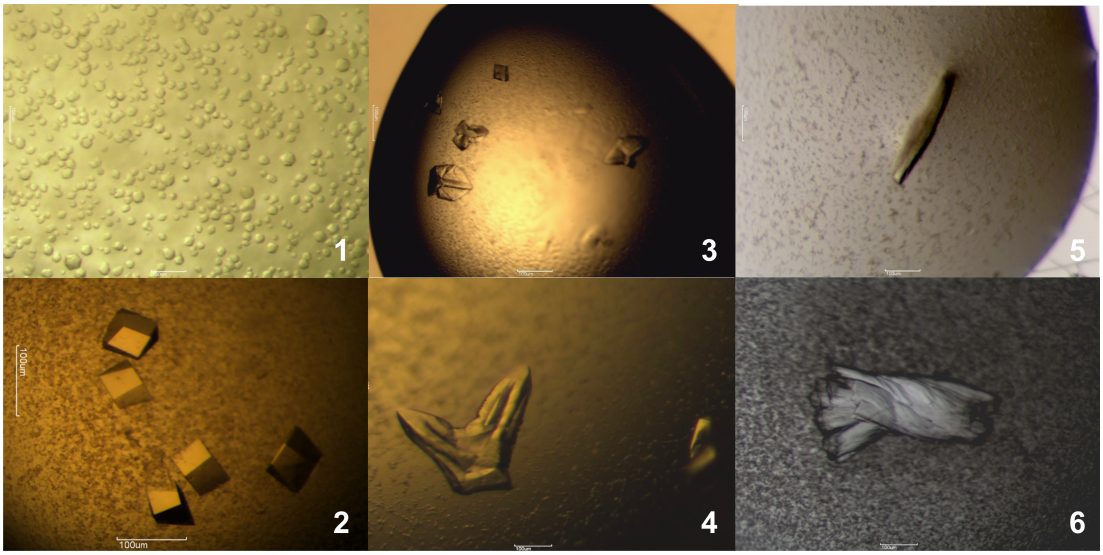


Figure 5.8: Crystals of varying morphologies were obtained from the Phy^{wt} protein in the absence and presence of zinc (1 and 2 respectively), Phy^{SeMet} (3 and 4), Phy^{-10N} (5) and the $Phy^{His203Ala}$ mutant enzymes.

5.3.6 Data collection and processing

Complete native and multiwavelength anomalous dispersion (MAD) datasets were collected for the Phy^{wt} and Phy^{SeMet} crystals respectively. Whereas the wild type protein crystal diffracted to 3.5 Å, the selenomethionine derivative did not exceed a resolution of 4.5 Å.

After reduction of the collected data, the *xia2* autoprocessing software (152) determined $P3_112$ to be the most likely space group for both crystals. The cumulative intensity distribution of the Phy^{wt} dataset indicated the presence of twinning in the crystal. Attempts to determine the native Phy structure by molecular replacement using the wild type PhlG coordinates (3HWP) failed. Alternative model building, molecular replacement and refinement was, in addition, attempted using the BALBES automated molecular replacement pipeline. This allowed for the calculation of electron density maps showing partial secondary structure features such as α-helical regions. A complete structure of the Phy protein could however not be determined.

As an alternative to molecular replacement, structure solution was attempted using the anomalous data collected from a Phy^{SeMet} crystal. Data analysis using the

SHELXC/D/E software package (183) showed a strong anomalous signal, indicating good replacement of methionine in the protein by its selenium derivative. However, the signal did not extend into resolution shells high enough to allow for the determination of a substructure from experimental phases. The overall diffraction of the Phy^{SeMet} crystals was also poor and, eventually, structure determination was not possible from this dataset.

5.3.7 Measurement of an X-ray fluorescence spectrum

In addition to the diffraction data described above, a fluorescence spectrum was collected from a Phy^{wt} crystal at the Diamond Light Source synchrotron. When evaluated using the PYMCA software (184), the spectrum revealed a single peak. This peak corresponds in emission energy to the fluorescence of zinc (approximately 8.6 kEV) and could be fitted to the metal's known signal (see Fig. 5.9) (93,184).

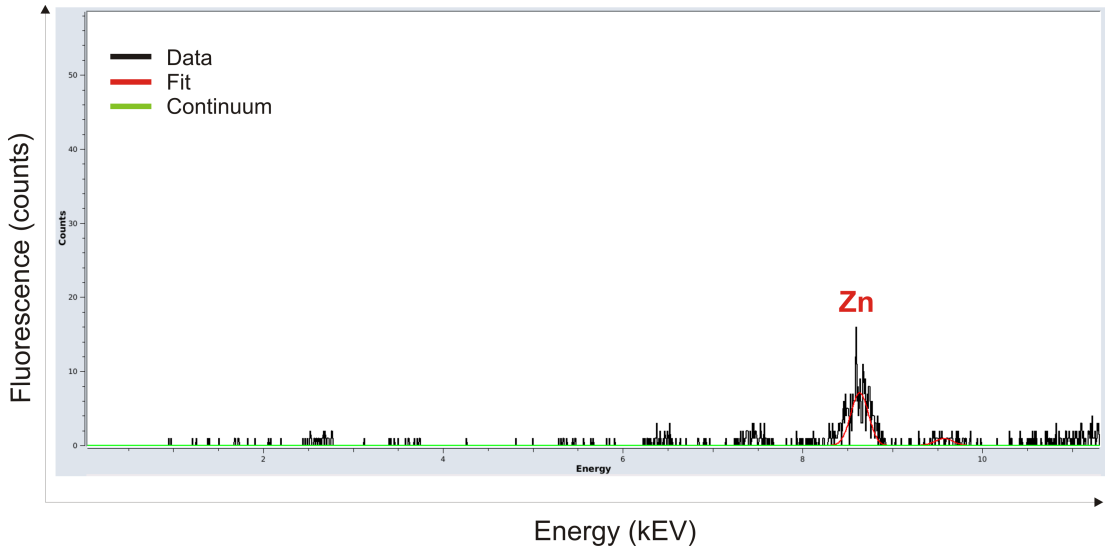


Figure 5.9: A fluorescence spectrum was measured on a Phy crystal grown in the absence of metals in either buffer or crystallisation conditions. The resulting peak corresponds in energy to the known fluorescence peak of zinc and was fitted to this accordingly (red line) (184).

5.3.8 Spectrophotometric analysis of phloretin hydrolysis

Using a spectrophotometric assay that allowed real-time observation of substrate turnover, individual reaction kinetics were analysed for the Phy^{wt} protein, Phy^{-10N}

and all soluble, purified proteins of the mutant library. Both wild type, full length and truncated protein showed clear substrate turnover within a period of 60 s. Kinetic constants could be determined for both enzymes although the quality of the data collected for the wild type was inferior with respect to the observed error margin. In addition, phloretin concentrations above 80 µM appeared to induce significant substrate inhibition for Phy^{wt}. Measurements in this area however also showed increased standard deviations and approached signal saturation. The turnover and catalytic efficiency of Phy^{-10N} was reduced by a factor of 2-3 compared to the wild type.

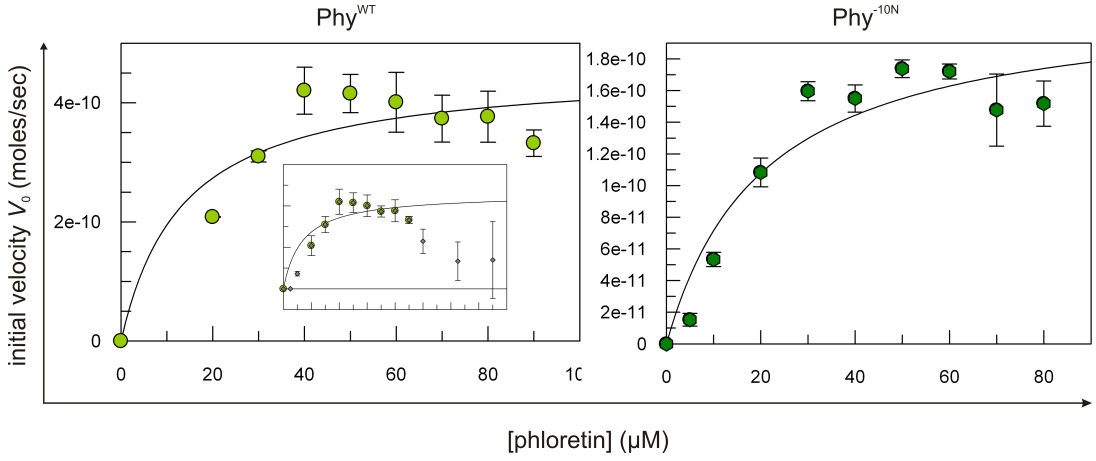


Figure 5.10: Despite data fluctuation and comparatively large error margins, clear substrate turnover was observed for the Phy^{wt} as well as the Phy^{-10N} hydrolase and was fitted to Michaelis Menten reaction kinetics. At higher phloretin concentrations, apparent substrate inhibition was detected for the wild type enzyme (inlet).

Of the six soluble Phy mutants, four showed detectable substrate turnover in spectrophotometric assays whereas two variants presented minimal or no phloretin hydrolysis (see fig. 5.11 and table 5.5). Alteration of the histidine in position 123 to alanine ($Phy^{His123Ala}$), designed to interrupt binding with the divalent metal, led to a significant loss in activity. The observed substrate turnover was minimal and measurements showed large error margins. The data did thus not allow for confident determination of kinetic constants. The His203Ala mutant did not show any activity when exposed to 5-100 µM phloretin solutions. The second soluble metal binding mutant, $Phy^{Glu154Gln}$, showed clear hydrolysis of phloretin within the course of the assay, albeit at reduced levels compared to the wild type. The Michaelis constant K_M

and catalytic efficiency k_{cat}/K_M were reduced by a factor of 2 and 3 respectively. The introduction of a novel hydrogen bond in position 126 ($Phy^{Ala126Ser}$) led to a similar reduction in the K_M but a four-fold decrease in catalytic efficiency. Two conservative mutations were introduced in position 115 and 218, changing tyrosine to phenylalanine and *vice versa* respectively ($Phy^{Tyr115Phy}$ and $Phy^{Tyr218Phe}$). Both alterations had the least effect on the reaction's Michaelis constant, reducing it by a factor of 0.5 ($Phy^{Tyr115Phe}$) and 0.6 ($Phy^{Phe218Tyr}$). Catalytic efficiencies, however, were determined to be a magnitude lower than for the wild type.

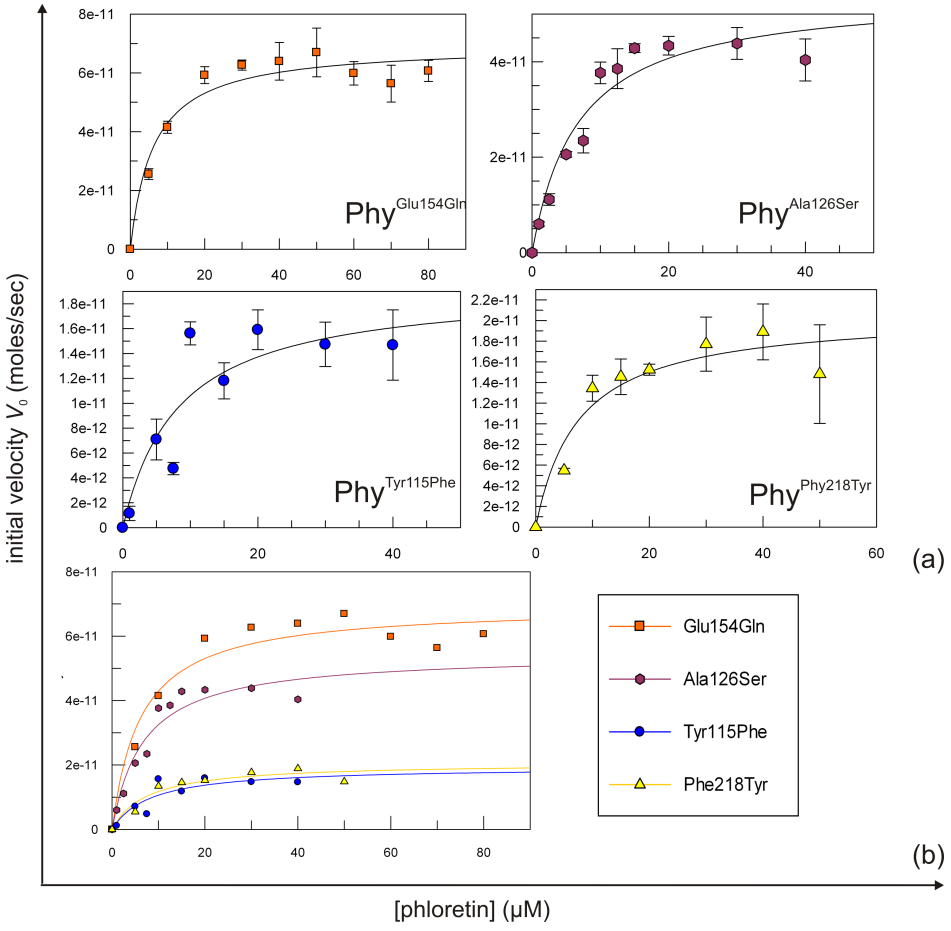


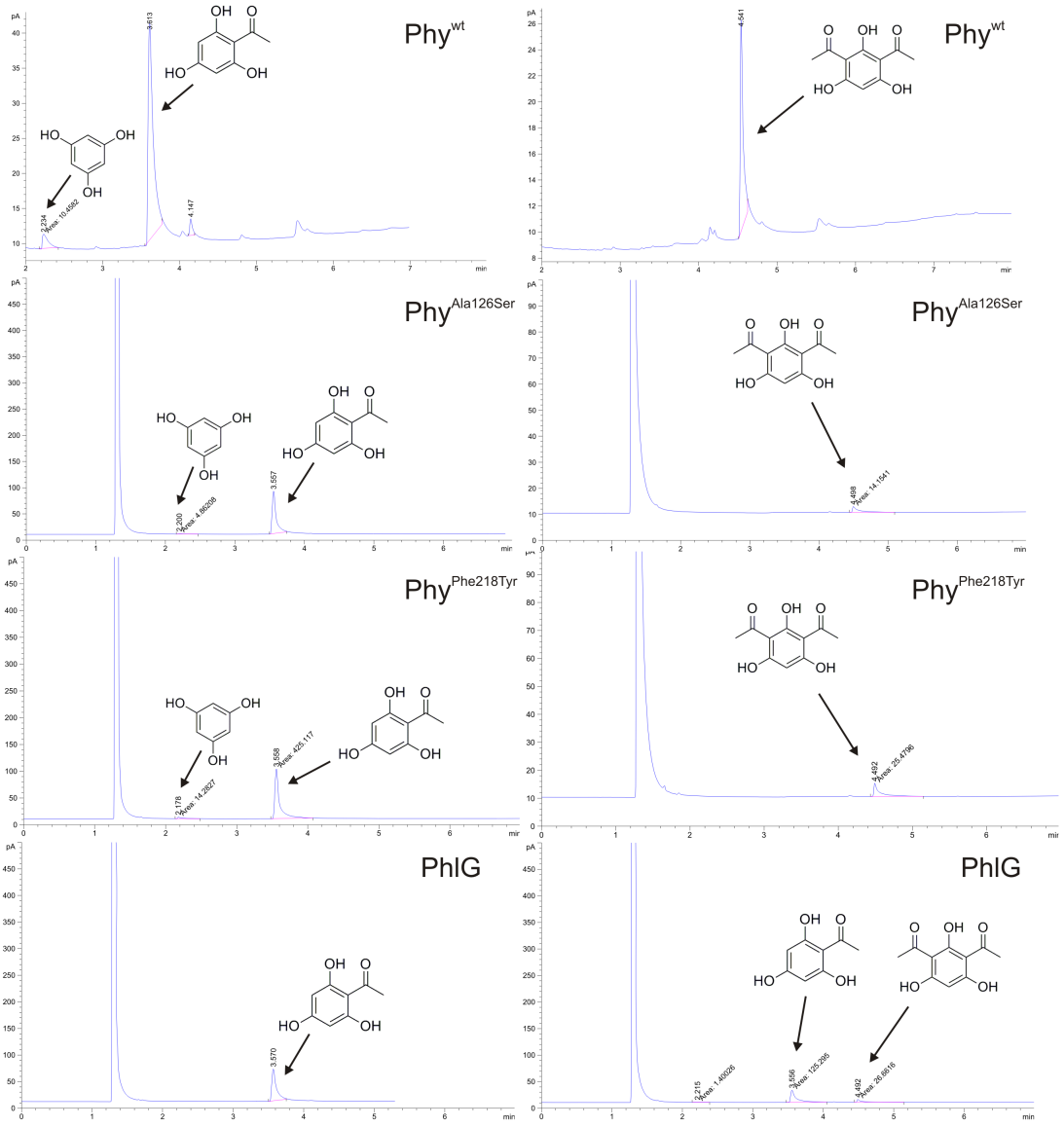
Figure 5.11: Using a spectrophotometric assay, phloretin hydrolysis was observed for four enzymes of the mutant library. Individual substrate dependent reaction velocities, to which standard Michaelis Menten kinetics were fitted, are shown. The bottom graph shows a superposition of all mutant reactions.

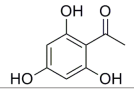
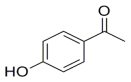
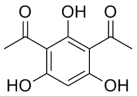
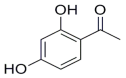
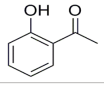
	K_M (μM)	V_{max} (moles/sec)	k_{cat} (sec^{-1})	k_{cat}/K_M ($sec^{-1}M^{-1}$)
Phy^{WT}	13.5	4.6×10^{-10}	8.7	6.4×10^{5}
Phy^{-10N}	20.6	2.2×10^{-10}	4.2	2.0×10^{5}
Phy mutation	Mutant library			
His123Ala	n.d.			
Glu154Gln	6.2	7.0×10^{-11}	1.3	2.1×10^{5}
His251Ala	Insoluble			
Glu255Gln	Insoluble			
Tyr115Phe	8.3	1.9×10^{-11}	0.4	4.3×10^{4}
His203Ala	n.d.			
Ala126Asn	Insoluble			
Ala126Ser	6.8	5.4×10^{-11}	1.0	1.5×10^{5}
Phe218Tyr	7.5	2.1×10^{-11}	0.4	5.3×10^{4}

Table 5.5: A summary of kinetic constants determined for the wild type Phy protein as well as its truncated and mutant library variants is shown. For two of the mutant proteins, no kinetic constants were determined (n.d.) as they showed minimal or non-detectable levels of hydrolysis.

5.3.9 Determination of Phy substrate specificity

To characterise the specificity of phloretin hydrolase, a variety of phloroglucinol and benzaldehyde derivatives were tested as potential substrates. In a GC based assay, both substrate turnover and formation of the hydrolytic product were determined.
Of the six molecules tested, only one was unambiguously accepted as a substrate by Phy^{wt} and both tested mutants, $Phy^{Ala126Ser}$ and $Phy^{Phe218Tyr}$. For monoacetylphloroglucinol (MAPG), a product peak was observed that corresponded in its retention time (2.2 min) to the expected hydrolytic product phloroglucinol (PG). Neither diacetylphloroglucinol (DAPG) nor any substrate with less than three hydroxyl substitutions on the aromatic ring were converted (see Fig. 5.12).



		Phy*	PhlG			Phy*	PhlG
	monoacetyl-phloroglucinol	✓	X		4-hydroxy-acetophenone	X	X
	diacetyl-phloroglucinol	X	✓		2,4-dihydroxy-acetophenone	X	X
	2-hydroxy-acetophenone	X	X		2,6-dihydroxy-acetophenone	X	X

* wild type and mutants

Figure 5.12: Six phloroglucinol and benzaldehyde derivatives were tested as potential substrates for the Phy wild type and mutant proteins as well as for PhlG. GC chromatograms for the reactions with MAPG and DAPG are shown. The table summarises formation (green tick) or absence (red cross) of hydrolytic products within the course of the assay.

5.4 Discussion

In recent years, a number of hydrolases have been characterised that are able to cleave carbon-carbon bonds, many taking part in the metabolism of aromatic compounds. A select number of these hydrolases catalyse the unusual enzymatic *retro* Friedel Crafts acylation and cleave the bond between an aryl and an acyl moiety (176). Whereas hydrolases acting on carbon-carbon bonds usually require an active side residue to act as a nucleophile, two of these recently characterised Friedel-Crafts hydrolases, diacetylphloroglucinol and phloretin hydrolase (PhlG and Phy respectively), do not appear to rely on this feature (95,96). Both enzymes show novel properties with respect to their reactions, sequence and fold but whereas the structure of PhlG has been determined to high resolution (95), little is known about the structure, reaction mechanism and substrate specificity of Phy.

When the synthetic *phy* gene was sub-cloned into an *E.coli* expression vector, soluble protein could be produced. The enzyme was purified at room temperature without significant losses in yield and remained in solution over many days (up to weeks) at 4 °C. Protein stability was thus assumed to be suitable for crystallisation. When initial crystallisation screening however remained unsuccessful, computational analysis of the secondary structure was aimed at identifying regions of disorder within the enzyme. The truncated construct Phy^{-10N}, lacking the first ten N-terminal residues, was designed, sub-cloned, purified and tested for crystallisation. Again, the formation of crystals or crystalline precipitate was minimal. A single crystal obtained with the truncated protein did not diffract beyond 4 Å. N-terminal protein disorder is thus unlikely to be the reason for a lack of crystal formation and poor diffraction quality. When, after several weeks under sitting-drop vapour diffusion conditions, a small full-length Phy crystal was obtained, this also diffracted poorly. As a fluorescence spectrum was measured on this crystal however, it presented a clear peak which could be fitted to the known signal of a divalent metal – zinc. When exposed to X-ray radiation, electrons in an atom get excited to a higher energy shell. The energy released on their return to the original, lower energy shell is detected as fluorescence and its level is unique to the atom type. The crystal used for measuring the spectrum is thus highly likely to contain zinc atoms. As it was grown in the absence of metals in either crystallisation condition or purification buffers, zinc was

most probably scavenged from the *E.coli* host by the Phy protein. To remain with the enzyme throughout purification and crystallisation, it is likely to be tightly attached to and highly coordinated by the amino acid side chains. Zinc is known to be of both functional and structural importance in certain proteins (186). In phloretin hydrolase, the metal may fulfil both. Addition of the chelating agent *o*-phenanthroline to a Phy solution was previously shown to significantly reduce enzymatic activity (96). When the original crystal was optimised in the presence of ZnOAc or ZnCl, crystals formed within 24 h as opposed to several days or weeks. Zinc may thus be essential or highly beneficial to protein folding and stability as well as activity.

Despite improvements in crystallisation, no structure could be determined for phloretin hydrolase within the course of this project. Even after exposure to small, high-intensity synchrotron beams, crystals did not diffract beyond 3.5 Å. The information contained in the corresponding datasets was insufficient to determine the enzyme's structure by molecular replacement. Despite the strict conservation of various active site residues, the overall sequence conservation between Phy and the only available, related structure (PhlG) is, at 25 %, comparatively low. At the given resolution, the two structures may thus differ too much to allow for the estimation of phases for Phy from the PhlG model alone.

Structure determination by experimental phasing was equally unsuccessful. A strong anomalous signal was present in the diffraction data collected from a Phy^{SeMet} crystal, indicating the good incorporation of selenomethionine into the protein. The signal did not, however, extend into the high enough resolution shells and overall diffraction did not extend beyond 4.5 Å. The quality of electron density maps is dependent on the available phase information. As a consequence, the initial maps prepared using experimental phases only extend to the resolution to which an anomalous signal is present. With the available data, phase information was not sufficient to build reliable electron density maps and determine the Phy structure.

Improvements in quality of either diffraction data (native and/or anomalous) or the availability of better search models may allow the determination of a Phy structure using experimental phasing or molecular replacement in the future.

In the absence of crystallographic data, a computational model of Phy (Phy^{mod}) was prepared. This was used as the basis for the preparation of an active site mutant library from which kinetic data were collected. The latter were hoped to give an

insight into the catalytic importance of individual residues, similarities between PhlG and Phy and a potential reaction mechanism.

The first four residues mutated were two glutamates and two histidines. In the PhlG structure all four were shown to surround and coordinate the bound metal. They are highly conserved in proteins showing sequence similarities to Phy and PhlG (95), indicating that these homologues may also be metalloproteins. The mutation of His251 and Glu255 to alanine and glutamine respectively resulted in a lack of protein production or insoluble polypeptides. Changing the histidine in position 123 to an alanine reduced activity to levels too low for the determination of kinetic constants. Only the conservative replacement of the glutamate in position 154 with a glutamine gave soluble, active protein albeit with rates of catalysis significantly reduced compared to the wild type. The reduced activity of zinc binding mutants is likely the consequence of an essential role the metal is thought to play in catalysis. In a potential mechanism, previously proposed for the PhlG homologue and observed for a variety of zinc binding enzymes, the metal lowers the pK_a of a closely positioned catalytic water, allowing its deprotonation and formation of a nucleophilic hydroxyl. The negative charge on the latter is stabilised by the positively charged divalent metal (95,186,187). The subsequent nucleophilic attack on the substrate leads to the formation of a tetrahedral intermediate and eventually cleavage of the scissile carbon-carbon bond (see Fig. 5.13) (95). Impaired coordination of the metal could thus reduce or even inhibit activation of the catalytic water. In addition, it was proposed that two of the metal coordinating residues, Glu160 and Glu274, may stabilise the tetrahedral intermediate through additional hydrogen bonds. This would further explain the reduced activity of the $Phy^{Glu154Gln}$ mutant (the mutated residue of which corresponds to Glu160 in PhlG, see Fig. 5.15).

The variant's Michaelis constant is reduced, which may be in indicator of increased substrate binding. This is surprising as the mutant's binding efficiency for the substrate should not be significantly altered compared to the wild type. A phloretin molecule would remain attached to the active site for longer, however, as coordination of the metal catalyst is impaired and turnover reduced. Furthermore, the Michaelis constant is not a universal indicator of binding affinity as it takes the turnover rate (k_{cat} or k_3) into consideration (188). Especially for reactions in which k_3 is significant with respect to dissociation of the complex into free enzyme and

substrate (k_2) (see section 2.5.1, equations 2.2 and 2.5), the K_M thus has to be evaluated in context of the reaction's k_{cat}.

In summary, these results support the idea that a metal ion is essential for both protein folding and function. The variations in solubility and activity between the mutants may be the consequence of varying contributions of the respective side chains to metal binding. Alterations of the two more N-terminal residues gives rise to soluble protein (Phy^{Glu154Gln} and Phy^{His123Ala}) whereas the two more C-terminal variants are either not fully translated or insoluble (Phy^{His251Ala} and Phy^{Glu255Gln}). The latter are thus likely to be most important to capturing and coordinating zinc and are thus essential for structural integrity.

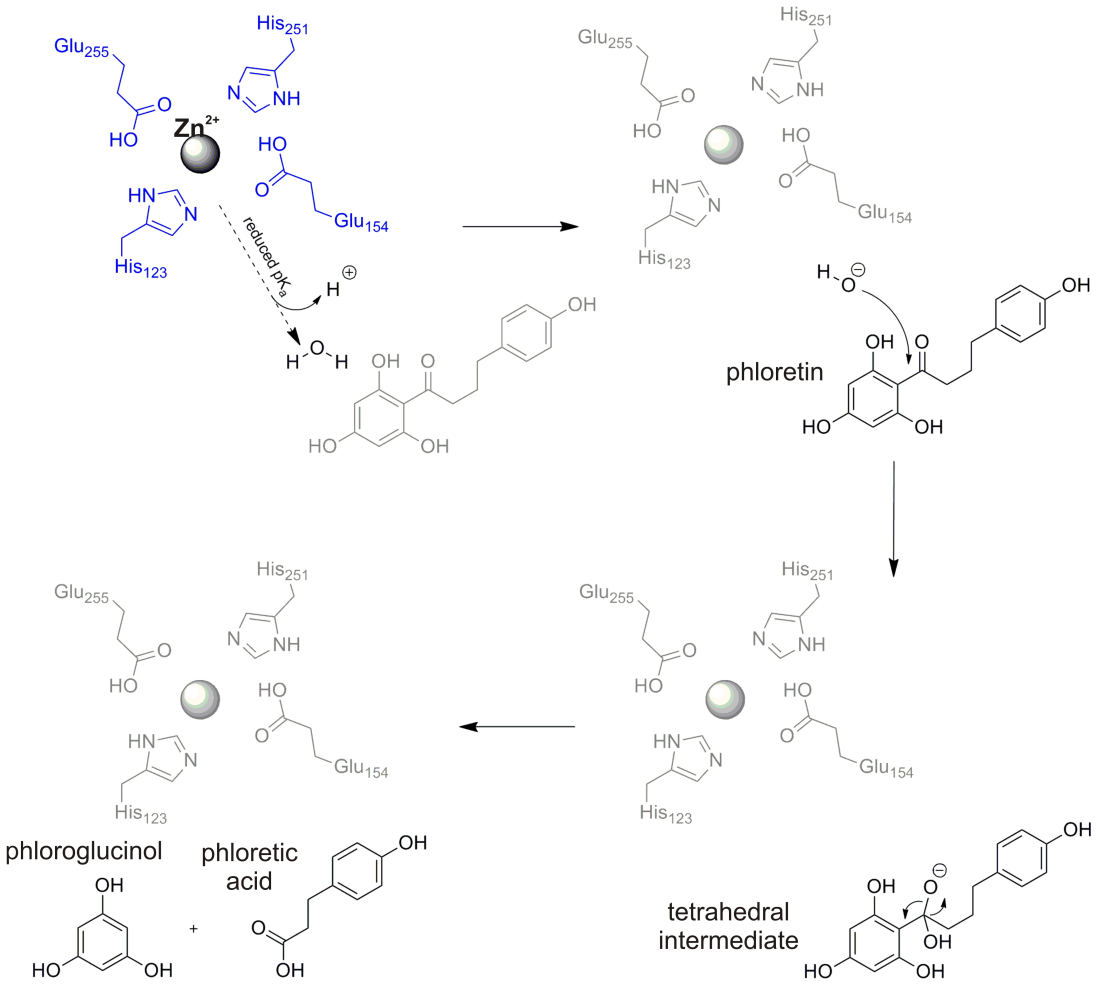


Figure 5.13: Analogous to the mechanism proposed for DAPG hydrolysis by PhlG (95), phloretin may be hydrolysed into phloretic acid and phloroglucinol by way of a tetrahedral intermediate. The divalent metal is essential to the mechanism as it lowers the pK_a of a catalytic water, allowing its activation and subsequent nucleophilic attack on the substrate.

In the PhlG homologue, three residues have been in implicated in hydrogen bond formation with the phloroglucinol substrate's hydroxyl groups: Tyr121, His214 and Tyr229 (see Fig. 5.2). A sequence comparison between PhlG and Phy reveals two of these residues to be fully conserved (Tyr115 and His203) whereas the most C-terminal tyrosine is replaced by a phenylalanine (Phe218). When a Phy mutant library was prepared, variants were designed in which this hydrogen bond formation was either interrupted ($Phy^{Tyr115Phe}$ and $Phy^{His203Ala}$) or introduced to allow for comparison of substrate binding between PhlG and Phy ($Phy^{Phe218Tyr}$). All mutant genes were translated into soluble proteins which could be purified. Spectrophotometric assays showed introduction of the two conservative mutations (phenylalanine to tyrosine and *vice versa*) to give rise to active proteins, albeit with reduced rates of hydrolysis. The phloroglucinol moiety in PhlG, and even more so, the full phloretin molecule in Phy, is highly substituted with polar groups. As a substrate, it is thus likely to be coordinated by an extensive hydrogen bonding network. Replacement or addition of a single bond donor may interfere with overall substrate coordination but would be unlikely to interrupt binding completely. In this context, it is surprising to see that a histidine to alanine mutation in position 203 abolished all enzymatic activity. Although the mutation is non-conservative and the structural change had the potential to influence the overall fold, the protein was highly soluble and purification yields exceeded those of all other mutants. It is thus possible that His203 has an as yet unknown function in either substrate binding or, even more likely, catalysis. Considering the differences in molecular structure between the diacetylphloroglucinol and phloretin, numerous additional interactions are likely to take place between the latter and Phy. Their identification however will require a three dimensional structure of the protein, ideally in complex with its substrate.

An analysis of the phloretin hydrolase substrate specificity has shown the enzyme to hydrolyse the PhlG product monoacetylphloroglucinol (MAPG) but not the diacetyl substrate DAPG (see also below). In PhlG, an asparagine side chain is thought to form a hydrogen bond to the second acetyl group; the corresponding residue in Phy however is an alanine. To assess the importance of bonding in this position and to test if Phy could be altered to accept DAPG as a substrate, the respective alanine in position 126 was mutated to an asparagine. The alteration gave rise to non-translated

or insoluble protein only. This may be the consequence of a change in protein fold caused by the larger side chain, changes to the active site environment and geometry and/or sterical clashes with other neighbouring residues. This is supported by the observation that replacing alanine with a smaller hydrogen bond donor, serine, gives rise to soluble protein. As for the $Phy^{Phe218Tyr}$ mutant, introduction of a new hydrogen bond donor lowers the overall enzymatic activity; the catalytic efficiency is reduced by a factor of 4. This is not surprising as the phloretin substrate does not possess a second acetyl moiety. Introduction of the serine in position 126 was designed to alter specificity, not to improve turnover of the native substrate. Impaired rather than enhanced phloretin hydrolysis was thus expected.

With respect to their catalytic activity, all mutants designed to alter substrate binding properties of Phy show a reduction in their Michaelis constant K_M. As the mutations should weaken protein-ligand interactions, this observation is, at first glance, surprising. In the context of the significant overall reduction however, a lowered K_M is not unusual for either of these mutants as described above for $Phy^{Glu154Gln}$. In addition, the Michaelis constants are derived from computational fitting. The measurements, at this substrate range and using the assay method described, allow for qualitative comparison of wild type and mutant proteins. They show an error margin which needs to be taken into consideration when evaluating total figures however. This is less significant for overall catalytic efficiencies (which differ by up to an order of magnitude) but important for the Michaelis constants, the differences of which are subtle.

To determine the substrate requirements for Phy and test if additional hydrogen bond donors would allow for hydrolysis of alternative substrates, the conversion of MAPG, DAPG and four acetophenone derivatives was tested. Using Phy^{wt}, $Phy^{Ala126Ser}$ and $Phy^{Phe218Tyr}$ as well as the PhlG homologue, neither of the acetophenones were converted. DAPG was exclusively turned over by PhlG and the MAPG hydrolysis product phloroglucinol was only observed for Phy wild type and mutant proteins. These results indicate that, in both Phy and PhlG, all three hydroxyl substituents are essential determinants of substrate specificity. Whereas PhlG requires the presence of two acetyl groups on the aromatic ring, the second group "inhibits" hydrolysis by Phy. The introduction of additional hydrogen bond donors in Phy, corresponding to similar functionalities in PhlG, does not alter the enzyme's

specificity. In case of the serine introduced in position 126, this may be the consequence of insufficient proximity between the serine side chain's OH group and the DAPG's hydroxyl. Even if the corresponding residues occupy comparable positions within the active site, as implied by the computational model, the serine side chain is shorter than the corresponding asparagine in PhlG (see Fig. 5.14). Its position may thus not allow for substrate binding and stabilisation by hydrogen bond formation. In addition, the Phy^{mod} structure shows an isoleucine in position 152 which, when superposed with the PhlG complex model, comes into very close proximity to the substrate's un-cleaved acetyl group. This potential steric clash could be an indicator for differences between the Phy model and the actual structure. Alternatively, if the model corresponds to the actual conformation of the active site, the phloretin substrate may be bound in an orientation different to the DAPG model.

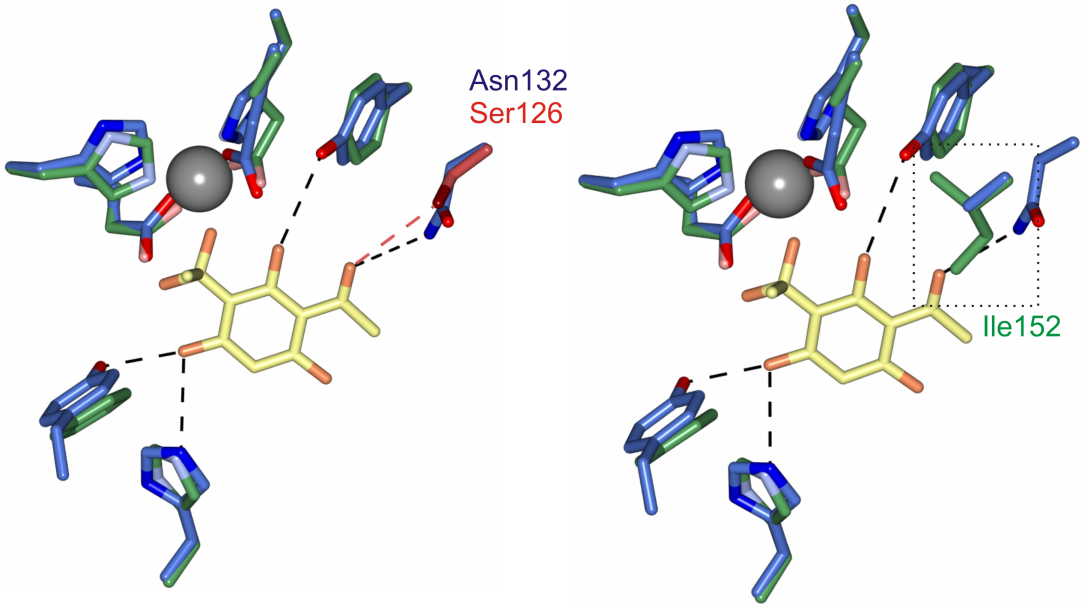


Figure 5.14: The addition of a serine hydrogen bond donor in position 126 (red, left image) does not change the substrate specificity of Phy. The side chain may either be too short compared to the corresponding asparagine in PhlG or in an unfavourable position for bond formation. Superposition of the modelled PhlG complex with the predicted Phy structure also shows an isoleucine residue in a possible steric clash with the DAPG transition state (right image), indicating that either modelling of the active site is incorrect in this position or the phloretin substrate is bound in a different conformation.

The alteration of a phenylalanine to a tyrosine in position 218 was designed to analyse the difference a hydrogen bond donor in this position makes to hydrolysis.

This functionality is present in PhlG and absent in Phy although both are thought to bind a phloroglucinol moiety's hydroxyl in this position. The alteration does not allow Phy to hydrolyse the PhlG substrate nor does it bind and stabilise less substituted benzaldehyde derivatives for catalysis. As with the histidine in position 203, assignment of a function to this residue thus awaits determination of the Phy three-dimensional (complex) structure.

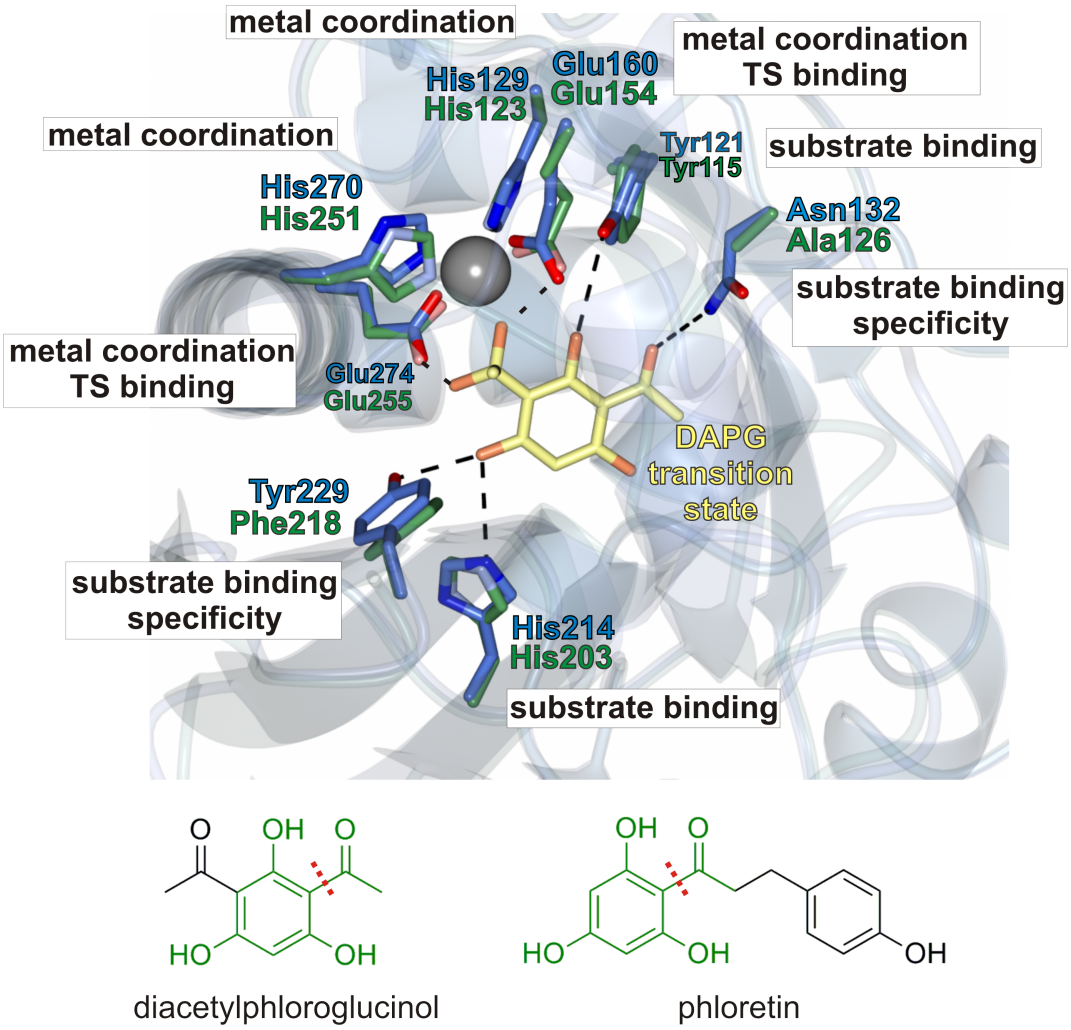


Figure 5.15: A superposition of the Phy computational model with the experimentally determined PhlG structure highlights the high similarities between model and template. All putative catalytic residues are shown. Their proposed functions in both proteins, which have been supported by kinetic data, are highlighted. The comparability of both enzymes is based on sequence conservation but also on the structural similarity of their substrates which share a common phloroglucinol core (highlighted in green) and are cleaved in the same position.

In summary, an insight has been gained into different aspects of carbon-carbon bond cleavage by phloretin hydrolase. The enzyme was found to be a metalloprotein in which a zinc atom may be of importance for structural integrity. Moreover, it appears essential for catalysis, possibly through activation of a catalytic water as proposed for the homologous PhlG protein. The similarities and differences in substrate specificity between both hydrolases are reflected in the potential to form hydrogen bonds with their respective ligands. Future protein engineering experiments may exploit these features to widen the Phy substrate scope. This will however require the three dimensional structure of the protein, ideally in complex with either substrate, reaction intermediate or products. More structural and mechanistic information may then even allow for a reversal of the phloretin hydrolase reaction and catalysis of an enzymatic Friedel-Crafts acylation.

Chapter 6: Purification and crystallisation of the oxidised polyvinyl alcohol hydrolase (OPH) from *Pseudomonas* sp. VM15C

6.1 Introduction

6.1.1 Biopolymer hydrolysis

In addition to the hydrolysis of large aromatic compounds, a number of β-diketone hydrolases were shown to cleave carbon-carbon bonds in high molecular weight polymer structures. This property makes the enzymes attractive targets for both synthetic and degradative biocatalytic applications.

A variety of polymers have been shown to be degraded by microorganisms or their respective isolated enzymes. These include polyesters, such as the bacterial polyhydroxyalkanoates, as well as polystyrenes, polyethylenes and polyvinyl alcohol. A common feature of enzymes involved in the breakdown of these polymers is the presence of a lipase box and a lipoprotein signal sequence in their primary amino acid sequence. Polymer breakdown is thus thought to occur on the cell surface and using a hydrolytic mechanism centred around the activation of a serine nucleophile for attack on the scissile bond (189,190).

6.1.2 Polyvinyl alcohol

Of the listed polymers, polyvinyl alcohol (PVA) has been of particular interest for research into bioremediation. The molecule is water soluble and thermostable and it is widely used as a fibre coating in the textile industry, as an adhesive or emulsifier as well as having a variety of other applications. As such it is released into wastewater streams in large quantities and requires efficient methods of breakdown (189). The linear backbone is a chain exclusively composed of carbon-carbon single bonds, substituted with a hydroxyl group in every second position (see Fig. 6.1). Its degradation is catalysed in two subsequent reactions. Stepwise, linear oxidation of the hydroxyl moieties leads to formation of β-diketone structures – oxidised polyvinyl alcohol (OPVA). This first reaction may be mediated by a dehydrogenase or an oxidase but, when quantified, was shown to take place on less than 4 % of the total vinyl alcohol units (89,189). As the β-diketone substrate is formed it may be

hydrolytically cleaved into a methylketone and a carboxylic acid by the action of a hydrolase. In *Alcaligenes faecalis,* a less common degradation involving formation of a β-hydroxyl rather than a diketone, followed by the aldol type cleavage into a monoketone and aldehyde, was also observed. OPVA hydrolysis takes places spontaneously and in the absence of any enzymes. It is however significantly enhanced by the catalysts.

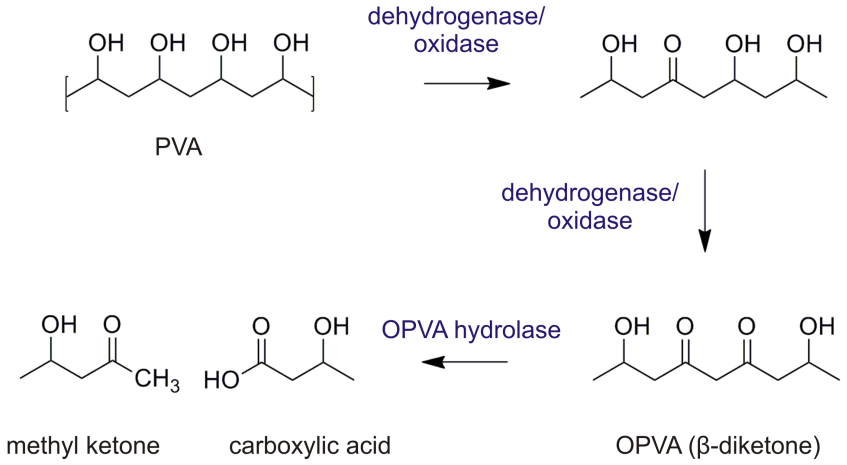


Figure 6.1: Enzymatic degradation of polyvinyl alcohol occurs *via* formation of a β-diketone which may be hydrolysed into a methyl ketone and a carboxylic acid. The latter can be used as a carbon source by the whole cell biocatalyst [adapted from (189)].

6.1.3 Oxidised polyvinyl alcohol from Pseudomonas VM15C

The microbial degradation of PVA is slow and generally requires the presence of mixed bacterial (predominantly *Pseudomonas*) cultures. Whereas one strain provides the oxidative and hydrolytic enzymes, a second is required for the supply of a pyrroloquinoline quinone (PQQ) cofactor essential for dehydrogenase activity (89,191,192).

The isolation of enzymatic catalysts is thus of particular interest for industrial applications. In 2000, Shimao and co-workers were able to identify, clone and characterise the oxidised polyvinyl alcohol hydrolase from *Pseudomonas* sp. VM15C (OPH) (89). The corresponding gene, *pvaB*, is encoded upstream of the polyvinyl alcohol dehydrogenase gene *pvaA* and both are thought to constitute an operon. When the gene was expressed in a recombinant *E. coli* host, the sequence translated into 379 amino acids assembling into a 40 kDa polypeptide. The translated protein

was however predominantly found in the insoluble membrane fraction. This feature has since been observed for other OPH enzymes, even after removal of the lipoprotein signal sequence (89,193). As a consequence, OPH purification has only succeeded from a native hosts to date (194,195). Using whole cell assays, the activity of *Pseudomonas* VM15C OPH could nonetheless be assayed and the enzyme was found to specifically enhance hydrolysis of oxidised polyvinyl alcohol. In addition to this native substrate, OPH was found to cleave the β-diketone moiety in 4,6-nonanedione, giving 2-pentanone and *n*-butyric acid at an equimolar ratio (see Fig. 6.2). This non-native substrate was however turned over at a considerably lower rate than OPVA (89). 2,4-pentanedione is not hydrolysed by the enzyme, the closely related OPH from *Pseudomonas vesicularis* however does accept this lower molecular weight diketone as a substrate(194). Non-oxidised PVA or monoketones were also not accepted as substrates by the *Pseudomonas* VM15C hydrolase which is in agreement with all other characterised OPH homologues (194,195).

6.1.4 Mechanism and structure of OPH

The sequence analysis and inhibition studies of OPH have given a first insight into potential reaction mechanisms. A lipase box indicates that the protein contains a catalytic serine nucleophile as described above. In accordance with this, all activity is inhibited in the presence of phenylmethanesulfonylfluoride (PMSF), a known serine hydrolase inihibitor. Metal chelating agents on the other hand did not appear to interfere with hydrolysis (89). The lipoprotein signal sequence and cell surface assays on related OPH enzymes suggest that catalysis takes place on the cell surface where the protein can come into direct contact with its high molecular weight substrate (189).

No three-dimensional structure has been determined for an OPH enzyme to date and sequence similarities to structurally characterised proteins are low. The closest related homologue of known structure is the fungal polyhydroxybutyrate depolymerase (PHBD) from *Penicillium funiculosum*. OPH and PHBP only share a sequence identity of 22 %. Both enzymes are however thought to be serine hydrolases and catalyse the fragmentation of high molecular weight polymers. PHBP shows a circularly permuted α/β-hydrolase fold and a highly conserved catalytic serine-histidine-aspartate triad. Analogous to the lipase family for which this fold

was first characterised, both PHBP and OPH have been proposed to catalyse hydrolysis by way of a tetrahedral intermediate and, possibly, a transient covalent acyl enzyme bond (84,89,196) (see Fig. 6.2). Apart from the similarities described above however, OPH is thought to be a highly unique, separately evolved enzyme (89).

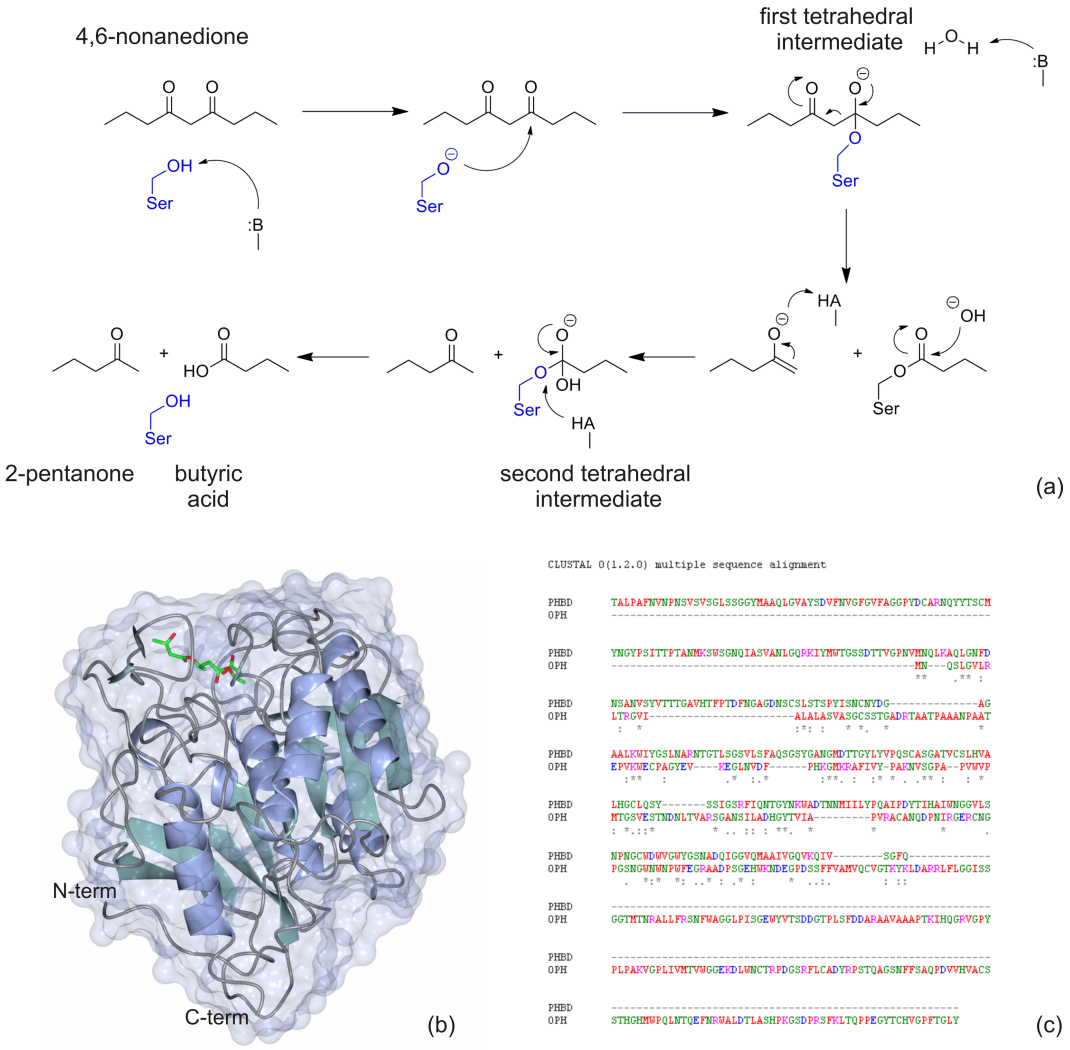


Figure 6.2: Hydrolysis of the OPH β-diketone substrate 4,6-nonanedione is thought to occur after nucleophilic attack by an active site serine residue and the consecutive formation of two tetrahedral intermediates [adapted from (84)] (a). The closest OPH homologue for which a structure has been determined is the fungal PHBD protein, showing a circularly permuted α/β hydrolase fold (b). The two enzymes share a sequence identity of 22 %, with all identical residues being centred around the core of significantly longer PHBD sequence.

6.1.5 Applications of OPH in biocatalysis

OPH-type enzymes have a considerable potential for industrial applications. They are the focus of research into bioremediation and their isolated use would significantly enhance the breakdown of soluble polymers in waste water streams (189). In addition, if the hydrolytic reaction could be reverted, the enzymes would be useful tools in the preparation of β-diketones. The latter are important building blocks for synthetic chemistry (84) and their preparation using biocatalysts would likely improve reaction specificities, e.g. with respect to acylation, over abiotic reactions.

6.1.6 Project aim

Due to the frequently reported problems with the production of correctly folded OPH protein, the major goal of this work was to obtain significant amounts of soluble enzyme for purification, characterisation and structure determination. Knowledge about the kinetic features of C-C bond hydrolysis, as well as the protein's fold, potential active site residues and enzyme-substrate interactions would then allow for the proposal of a detailed mechanism of β-diketone cleavage. This could, in the future, support protein engineering and experiments aimed at reversing the reaction into carbon-carbon bond formation.

6.2 Materials and Methods

6.2.1 Ligation independent cloning of the OPH synthetic gene

For production of the *Pseudomonas* VM15C OPH protein, a synthetic gene, optimised for recombinant expression in *E. coli* hosts systems, was kindly provided by Professor Wolfgang Kroutil from the University of Graz. To attach a hexahistidine-tag and allow for metal affinity purification of the translated protein, the gene was then inserted into the YSBLIC3C vector using a ligation independent cloning (LIC) protocol. LIC optimised primers were designed and used for amplification of the *oph* synthetic gene (103). Their sequences and amplification conditions are shown in table 6.1. After T4 polymerase treatment, both gene and

digested vector (kindly provided by Dr. Laila Roper) were annealed using the LIC optimised cloning procedure.

Variant		Forward Primer (5'-3')	Reverse Primer (5'-3')
1	OPH	ccagggaccagcaatgaatcagagcctgggt gttctgcgtc	gaggagaaggcgcgttattaatacagaccggta aacggaccaacatgac
2	OPH^{-LSS}	ccagggaccagcaatgggtgcagatcgtacc gcagcaacaccggcag	gaggagaaggcgcgttattaatacagaccggta aacggaccaacatgac

Reaction component	Concentration	Volume [ul]	final concentration
forward primer	20 pmol/µL	1	0.4 µM
reverse primer	20 pmol/µL	1	0.4 µM
template DNA	50 ng/uL	1	1 ng/µL
deoxyribonucleotides (dNTPS)	10 mM	2	0.4 mM
KOD® hot start DNA polymerase buffer	5x	5	1x
$MgSO_4$	25 mM	3	1.5 mM
DMSO	100 %	0.5	1 % (v/v)
KOD® hot start DNA polymerase	1.0U/ µL	1	0.02 U/ µL
deionised water	-	35.5	-

Cycles	Temperature [°C]	Time [min:s]	Reaction
1	94	2:00	initial denaturation
25	94	0:30	denaturation
	45	0:30	annealing
	72	0:30	extension
1	72	3:00	Final extension
-	4	∞	hold/storage

Table 6.1: The *oph* gene was amplified from its synthetic template in a PCR reaction using LIC specific primers (top). Reaction components (middle) and amplification conditions (bottom) are shown and correspond to LIC cloning reactions for both full length protein and the truncated OPH^{-LSS} described below.

The LIC3C:*oph* plasmid could eventually be isolated from an *E. coli* expression host transformed by the annealing product. A restriction endonuclease digest confirmed the presence and orientation of the gene of interest. Its sequence and completeness was verified by sending a small amount of DNA for sequencing with the University of York Technology Facility.
Detailed LIC procedures including the preparation of both vector and insert can also be found in chapter 5, section 5.2.

6.2.2 Analysis of OPH gene expression and protein production

After sequencing verified successful insertion of the *oph* gene into the YSBLIC3C vector, three *E.coli* (DE3) expression strains were transformed by the corresponding plasmid DNA: *E.coli* BL21, BL21 gold and Rosetta. Resulting colonies were picked and used for the preparation of 5 mL overnight cultures in LB medium. 10 mL LB expression cultures inoculated with this starter culture were grown to an optical density (OD) of 0.6, induced with 1 mM IPTG (final concentration). Kanamycin was used as the selective antibiotic, at a final concentration of 30 μg/mL for all solid and liquid *E. coli* cultures. The cells were incubated at 18°C for 18 h while shaking at 180 rpm. After harvesting by centrifugation at 13,000 rpm for 5 min, cells were resuspended in 1 mL of 50 mM Tris buffer containing 300 mM NaCl (hereafter named resuspension buffer) and disrupted by sonication at 14,000 amplitude microns for 3 x 3-4 s with 5 second cooling intervals. The crude cell extracts were subsequently centrifuged at 13,000 rpm for 30 min. 10 μL samples were taken from disrupted whole cells and the soluble supernatant after their centrifugation, representing soluble and total cell fractions respectively. All samples were eventually separated and visualised on a 12 % polyacrylamide gel using SDS PAGE chromatography.

6.2.3 OPH Ni-affinity purification

For purification of OPH from a recombinant host, an *E.coli* BL21 (DE3) expression strain was transformed by the YSBLIC:*oph* plasmid. Resulting colonies were picked and used for the preparation of starter cultures from which 2 L of LB medium were inoculated. The cultures were grown, induced with IPTG, incubated and harvested.

Cells were resuspended and disrupted by sonication after which separated, soluble fractions loaded could be loaded onto a 5 mL HiTrap™ Chelating HP Nickel affinity column. After washing of the column with resuspension buffer, an imidazole gradient was applied to allow for elution of the hexahistidine-tagged protein. All procedures correspond to those detailed for the *Bs*PAD protein and are also described in chapter 4, section 4.2.
Due to the absence of a chromatography peak or detectable protein bands after SDS PAGE chromatography, further purification steps were abandoned at this stage.

6.2.4 Solubility screening

A buffer based solubility screen was carried out after the OPH protein was found to be produced exclusively in its insoluble form in all expression strains. For this screen, induced cells were grown at 18°C for 18 h and harvested as described above for expression testing (section 6.2.2), in 30 aliquots of 10 mL each. Cell pellets were then resuspended in 1 mL of buffers, containing the detergents and additives listed in Fig. 6.5 (116). Protein solubility was assayed by taking samples of soluble and insoluble protein fractions and separating them by SDS PAGE chromatography.

6.2.5 Prediction and removal of the OPH lipoprotein signal sequence

The OPH amino acid sequence was entered into the Oxford Protein Production Facility's Protein Analysis Linker (OPAL) (197) and a lipoprotein signal sequence was detected, corresponding to previous reports (89). Consequently, primers for a truncated construct lacking the protein's first 29 N-terminal amino acids were designed and are listed in table 6.1. Cloning, isolation and sequencing of the shortened protein sequence were performed using the LIC methods described in chapter 5, section 5.2. The resulting plasmid and protein were designated YSBLIC:*oph*$^{-LSS}$ and OPH^{-LSS} respectively. Initial gene expression and protein production tests in *E. coli* BL21 (DE3) cells were performed as described for the full length construct,with additional samples being taken from a 37 °C incubation test, 1, 2, 3 and 18 h after induction. The samples were prepared in duplicate and resuspended in Tris or phosphate based resuspension buffers, either at a pH of 7.5.

6.2.6 Optimisation of OPH^{LSS} protein solubility

As, despite the lack of its lipoprotein signal sequence, the OPH^{-LSS} protein remained insoluble, 1 L of cells transformed by the YSBLIC:*oph*$^{-LSS}$ plasmids were grown, induced and incubated at 16 °C for 18 h. The culture was split into eight aliquots of 120-140 mL and cells were harvested as described above. The resulting pellets were resuspended in either Tris or phosphate based buffer, either of the two buffers containing 10 % glycerol, 10 % glycerol and 1 % triton X-100 or 10 % glycerol and 1% tween 20. After sonication and centrifugation, samples were taken from both soluble and resuspended insoluble cell fractions and visualised by SDS PAGE chromatography.

6.2.7 Variation of E. coli expression strains to enhance protein solubility

Two *E. coli* strains optimised to enhance solubility of the translated protein of interest were transformed by the YSBLIC:*oph*$^{-LSS}$ plasmids. Both strains, *E. coli* Lemo21 (DE3) (109,110) and *E. coli* Arctic Express™ (DE3) (111,112), were transformed by the YSBLIC:*OPH*^{LSS} plasmid and subsequently grown and induced according to the suppliers protocols. Whereas cells were incubated at reduced temperatures to induce chaperonin production in the Arctic Express strain, titration of rhamnose levels was aimed at regulating expression levels for the gene of interest in Lemo cells. As control reactions, both strains were transformed by empty YSBLIC3C plasmids, cultured, harvested and analysed under the same conditions as their YSBLIC:*OPH*^{LSS} expressing counterparts.

6.2.8 Preparation of OPH fusion constructs

Using the In-fusion cloning technique, the *oph*$^{-LSS}$ gene was fused to four different N-terminal protein solubility tags by the University of York's Technology facility. These tags include the glutathione S-transferase (GST), the maltose binding protein (MBP), immunoprotein 9 (IM9) and a green fluorescence protein (GFP). Each solubility tag contains an additional N-terminal, non-cleavable hexahistidine sequence and is linked to the OPH^{-LSS} protein sequence by a HRV3C protease cleavage site. The corresponding vectors are based on the YSBLIC3C system and

were thus designated YSBLIC3C:*oph:fus,* with *fus* corresponding to either the *gst, mbp, im9* or *gfp* genes.

6.2.9 Analysis of YSBLIC3C:oph:fus gene expression and protein production

To assess the effect the four protein tags have on OPH^{-LSS} solubility, *E. coli* BL21 (DE3) expression cells were transformed by the corresponding YSBLIC:*oph:fus* plasmids. Resulting colonies were picked and used for the preparation of 5 mL overnight cultures from which, in turn, 10 mL LB expression cultures could be inoculated, grown and induced as described in section 6.2.2. After shaking incubation at 18 °C for 18 h, cells were harvested by centrifugation, resuspended, disrupted and their soluble and insoluble fractions analysed by SDS PAGE chromatography as detailed above.

6.2.10 Purification of the OPH:GST fusion construct

As the OPH:GST fusion construct showed one of the most pronounced bands in the soluble cell fractions visualised by SDS PAGE, it was chosen as the first target for large scale protein purification. After transformation of an *E.coli* BL21 expression strain by the corresponding plasmid, resulting colonies were picked, starter cultures prepared and used for the inoculation of 2 L of LB medium as described above. After induction and incubation, cells were harvested, resuspended, disrupted and the cell free extract loaded onto a 5 mL HiTrap™ Chelating HP Nickel affinity chromatography column as detailed in section 6.2.3. In contrast to previous purification experiments, a chromatography peak could be detected during application of the 30-500 mM imidazole gradient. The corresponding peak fractions were pooled and mixed with purified HRV3C protease at a ratio of 1:10 (w/w) (protease : protein of interest). After incubation at 4 °C for 18 h, the protein/protease mix was concentrated to a final volume of 2 mL and loaded onto a 5 mL HiTrap™ nickel affinity chromatography column as described above. For elution of the cleaved protein, 50 ml of 50 mM Tris buffer (pH 7.5) containing 300 mM NaCl were applied to the column, followed by a 0-500 mM imidazole gradient over 100 ml for elution of the hexahistidine-tagged GST fusion protein. Protein peak fractions were subsequently pooled and concentrated to a total volume of 2 mL before being applied

to a Superdex™ S75 gel filtration column. From the latter, protein was eluted in resuspension buffer. During all stages of purification, 10 µL samples were taken to evaluate protein purity on a 12% SDS PAGE gel. After the final purification step, all protein was stored at 4 °C until further use in crystallisation experiments.

6.2.11 Crystallisation of purified OPH:GST protein

After purification, the OPH:GST construct did not appear to have been cleaved by the HRV3C protease. Nonetheless, structural experiments were carried out using the fused proteins.

For crystallisation, the purified OPH:GST was concentrated to 3 mg/mL. In three 96 well trays, the protein was assayed against the PACT, INDEX and CSS 1+2 screens using the sitting drop vapour diffusion method. For the latter, two separate pH variants were tested using MES Tris buffer at a pH of 5.6 and 8.0 respectively. A crystal hit found in the PACT screen's B4 condition (sodium malonate, imidazole, boric acid system, pH 7.0, 25 % PEG 1.5K) was subsequently scaled up using the hanging drop vapour diffusion method. Detergent concentrations of this optimisation tray were varied between 23 and 28 % PEG, pH values ranged from 6.5 to 7.5.

6.2.12 Crystal testing

Crystals obtained using the INDEX crystallisation screen were fished, flash-cooled in liquid nitrogen and tested for diffraction using an in-house X-ray source and a MARRESEARCH MAR345 imaging plate detector. A single crystal from the INDEX E5 condition was retained and sent to the Diamond Light Source synchrotron for further testing. No datasets were collected.

6.2.13 Prediction of protein disorder

To predict stretches of disordered residues in the protein structure, the OPH^{-LSS} amino acid sequence was entered into the disorder prediction server Regional Order Neural Network (RONN) from the Oxford Protein Production Facility (178).

6.3 Results

6.3.1 Ligation independent cloning of the oph and oph^{-LSS} synthetic genes

Using the ligation independent cloning method, the synthetic *oph* gene could be introduced into the YSBLIC3C plasmid. When analysed using the Oxford protein production facility Protein Analysis Linker (OPAL) (197), the OPH primary sequence was shown to include a lipoprotein signal peptide as suggested previously (89). Consequently, a second, truncated construct (OPH^{-LSS}) was designed, lacking the protein's first 29 N-terminal amino acids. It was equally amplified and inserted into the YSBLIC3C vector (see above, Fig. 6.3).

An endonuclease digest on two restriction sites flanking the genes of interest verified their insertion and correct orientation within the vector. When the gene sequences were determined, translated into the corresponding amino acids and aligned with the deposited OPH sequence, no mutations were observed. The signal sequence was absent in case of the YSBLIC3C:*oph*$^{-LSS}$ construct.

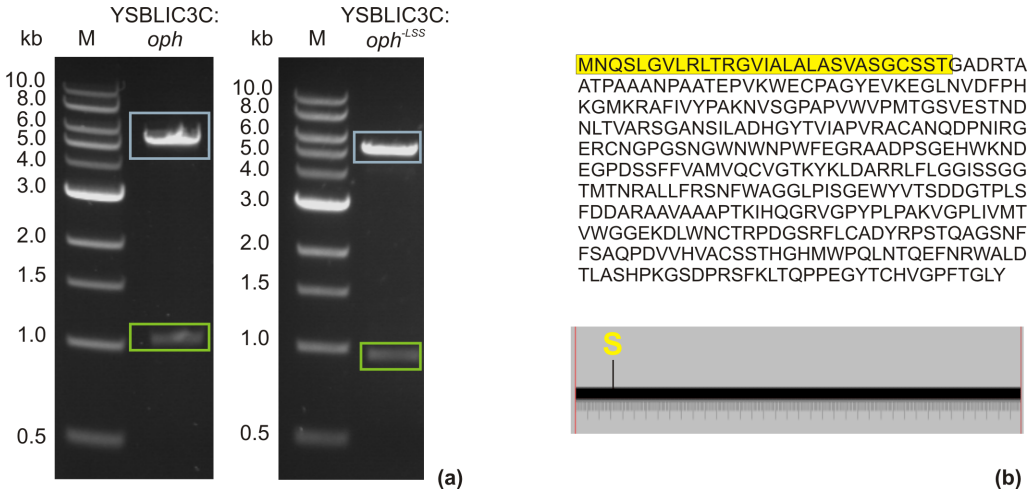


Figure 6.3: An endonuclease double digest verified the presence and orientation of the *oph* and *oph*$^{-LSS}$ genes in the YSBLIC3C vector. The plasmid DNA (blue boxes) can be seen at a height of 5.4 kilobases. The inserts (green boxes) are close to the 1.0 kb marker band as expected from their molecular weight of 1.1 and 1.0 kb (*oph* and *oph*$^{-LSS}$ respectively) (a). The lipoprotein signal peptide, as determined by the OPAL server, comprises the first 29 N-terminal OPH residues and is highlighted in yellow (b).

6.3.2 Analysis of OPH (full length) protein production and OPH Ni-affinity purification

After three *E. coli* strains were transformed by the YSBLIC3C:*oph* plasmid, gene expression and protein production were analysed under varying conditions. The overproduction of a protein with the correct molecular weight of approximately 40 kDa could be visualised using SDS PAGE. Irrespective of host strain or incubation temperature however, this overproduction was exclusively observed for the insoluble fractions of the disrupted cells (see Fig. 6.4). As a protein band appears at an equivalent position in uninduced samples or cells transformed by an empty YSBLIC3C plasmid, the result remained ambiguous. To determine if low amounts of OPH protein may be present in the soluble fraction, the culture volume was scaled up and soluble cell fractions loaded onto a Ni-affinity chromatography column. Application of a 30-500 mM imidazole gradient to the column did not result in the elution of any OPH protein, as shown by the lack of an FPLC absorbance peak or a protein band on the subsequently prepared SDS PAGE gel (results not shown).

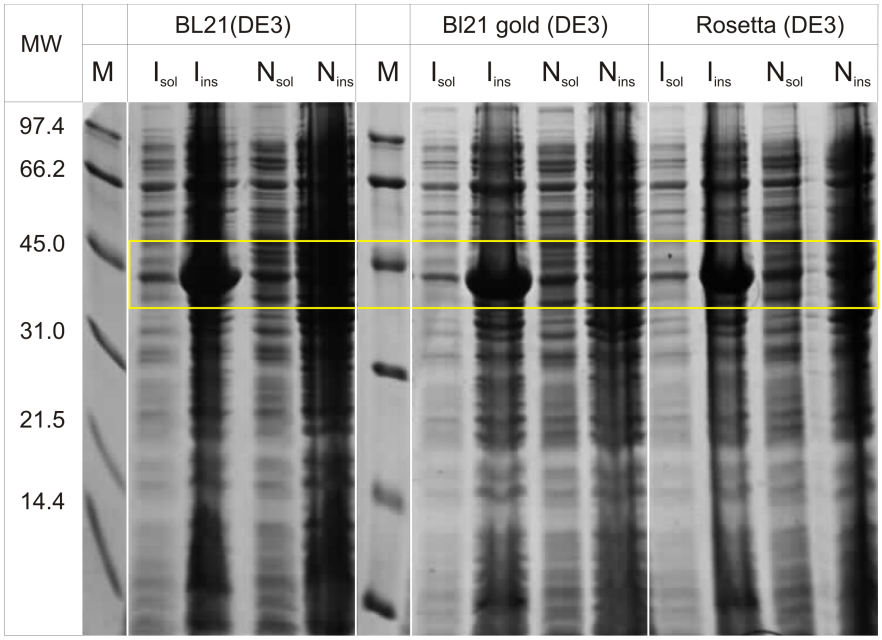


Figure 6.4: Initial expression tests showed the overproduction of a putative OPH full length protein with a molecular weight of approximately 40 kDa (yellow box). OPH produced exclusively in its insoluble form (ins), soluble fractions (sol) do not differ between induced (I) and non-induced (N) samples. An attempt at purification of the protein from a scaled up culture was unsuccessful and no protein signal could be observed (results not shown).

6.3.3 Solubility screening

In an attempt to enhance protein solubility and release OPH protein from the observed inclusion bodies, harvested cultures were resuspended and cells disrupted in buffers supplying a variety of pH values and additives (116).

1	100 mM Tris, 10% glycerol, pH 7.6
2	100 mM Tris, 50 mM LiCl, pH 7.6
3	100 mM HEPES, 50 mM $(NH_4)_2SO_4$, 10% glycerol, pH 7.0
4	100 mM HEPES, 100 mM KCl, pH 7.0
5	100 mM Tris, 50 mM NaCl, 10% isopropanol, pH 8.2
6	100 mM K_2HPO_4/KH_2PO_4, 50 mM $(NH_4)_2SO_4$, 1% Triton X-100, pH 6.0
7	100 mM triethanolamine, 100 mM KCl, 10 mM DTT, pH 8.5
8	100 mM Tris, 100 mM sodium glutamate, 10 mM DTT, pH 8.2
9	250 mM KH_2PO_4/K_2HPO_4, 0.1% CHAPS, pH 6.0
10	100 mM triethanolamine, 50 mM LiCl, 5 mM EDTA, pH 8.5
11	100 mM sodium acetate, 100 mM glutamine, 10 mM DTT, pH 5.5
12	100 mM sodium acetate, 100 mM KCl, 0.1% *n*-octyl-β-d-glucoside, pH 5.5
13	100 mM HEPES, 1 M $MgSO_4$, pH 7.0
14	100 mM HEPES, 50 mM LiCl, 0.1% CHAPS, pH 7.0
15	100 mM KH_2PO_4/K_2HPO_4, 2.5 mM $ZnCl_2$, pH 4.3
16	100 mM Tris, 50 mM NaCl, 5 mM calcium acetate, pH 7.6
17	100 mM triethanolamine, 50 mM $(NH_4)_2SO_4$, 10 mM $MgSO_4$, pH 8.5
18	100 mM Tris, 100 mM KCl, 2 mM EDTA, 1% Triton X-100, pH 8.2
19	100 mM sodium acetate, 1M $MgSO_4$, pH 5.5
20	100 mM Tris, 2M NaCl, 0.1% *n*-octyl-β-d-glucoside, pH 7.6
21	100 mM Tris, 1 M $(NH_4)_2SO_4$, 10 mM DTT, pH 8.2
22	100 mM sodium acetate, 50 mM LiCl, 5 mM calcium acetate, pH 5.5
23	100 mM HEPES, 100 mM sodium glutamate, 5 mM DTT, pH 7.0
24	100 mM triethanolamine, 100 mM sodium glutamate, 0.02% *n*-octyl-β-d-glucoside, 10% glycerol, pH 8.5
25	100 mM Tris, 50 mM NaCl, 100 mM urea, pH 8.2
26	100 mM triethanolamine, 100 mM KCl, 0.05% dextran sulfate, pH 8.5
27	100 mM KH_2PO_4/K_2HPO_4, 50 mM $(NH_4)_2SO_4$, 0.05% dextran sulfate, pH 6.0
28	100 mM HEPES, 50 mM LiCl, 0.1% deoxycholate, pH 7.0
29	100 mM Tris, 100 mM KCl, 0.1% deoxycholate, 25% glycerol, pH 7.6
30	100 mM potassium acetate, 50 mM NaCl, 0.05% dextran sulfate, 0.1% CHAPS, pH 5.5

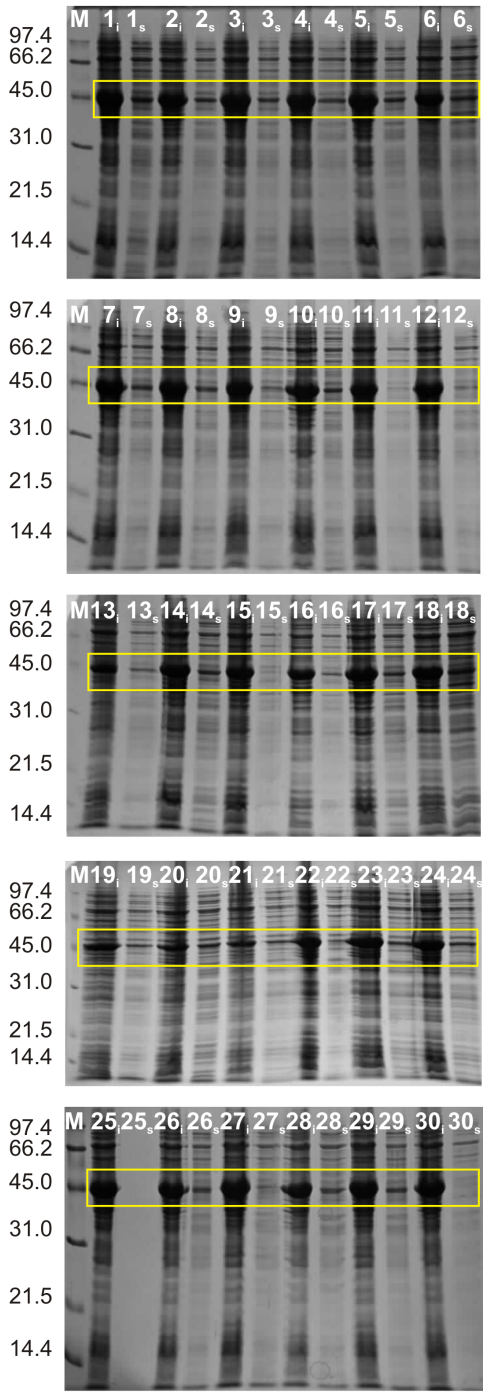


Figure 6.5: SDS PAGE gels show soluble and insoluble fractions of OPH protein cultures resuspended and lysed in one the 30 listed buffers. Significant OPH production (yellow box) can only be observed in the insoluble fractions.

Despite the clear overexpression of the protein of interest, no improvement in solubility could be observed (see Fig. 6.5).

6.3.4 Analysis of OPH^{LSS} protein production

As solubility problems prevailed with the OPH wild type, protein production and solubility were assayed for the construct lacking an N-terminal lipoprotein signal sequence. Initial tests using two buffer systems showed successful gene expression and translation of the recombinant OPH^{-LSS} protein, albeit exclusively in the form of insoluble inclusion bodies (see Fig. 6.6). Results thus resembled those of the full length hydrolase.

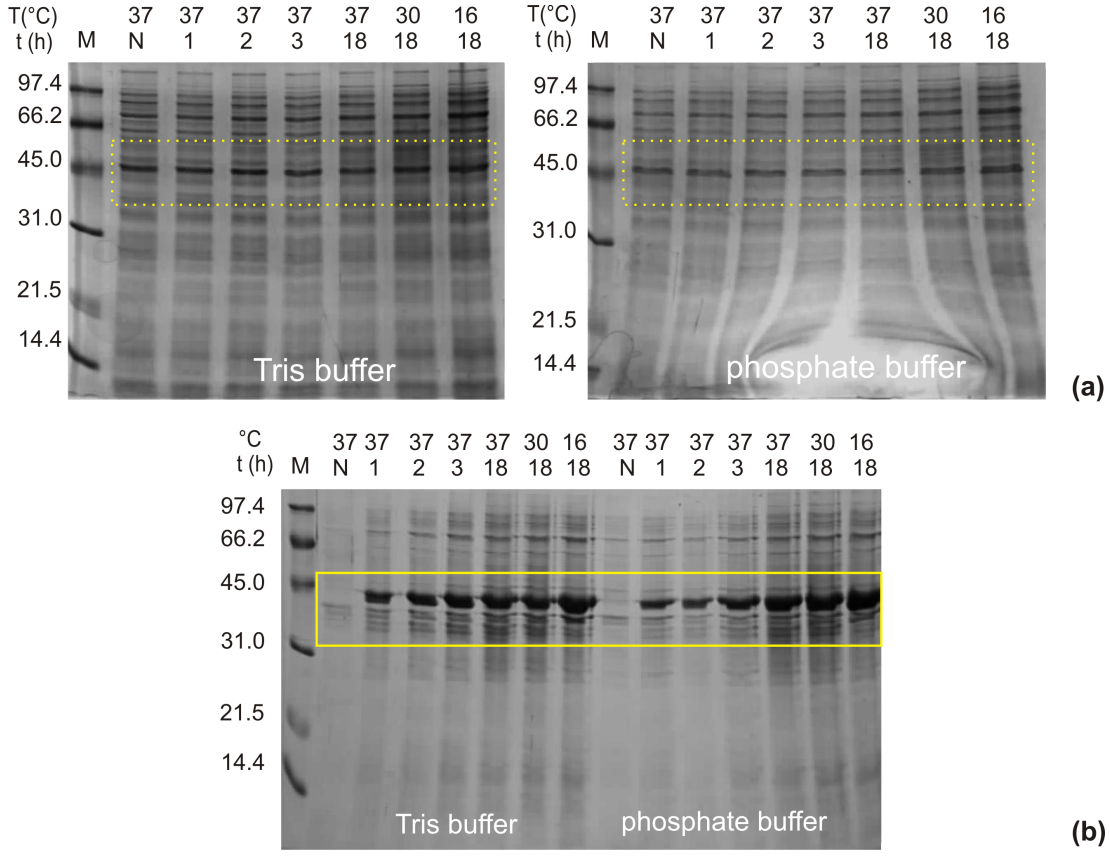


Figure 6.6: OPH^{-LSS} protein production was analysed under varying conditions with respect to incubation time and temperature as well as buffers used for resuspension of harvested pellets. SDS PAGE gels of both soluble (a) and insoluble (b) cell fractions are shown, including the non-induced control samples (N). The yellow box marks the expected position of soluble and insoluble OPH^{-LSS} (dashed and bold lines respectively).

6.3.5 Optimisation of OPH^{LSS} protein solubility

As the protein remained insoluble after removal of the lipoprotein signal sequence, attempts were made to solubilise OPH^{-LSS} in buffers containing the detergents glycerol, triton X-100 or tween 20. SDS PAGE analysis of soluble cell fractions did any noticeable solubilisation of the hydrolase (results not shown).

6.3.6 Variation of E. coli expression strains to enhance OPH^{LSS} protein solubility

Two alternative cell strains were transformed by the YSBLIC3C:*oph* plasmid. Both are based on the *E. coli* BL21(DE3) expression system but were modified to enhance solubility of recombinantly produced proteins. Whereas the Lemo strain allowed for titration of induction levels, a pair of temperature activated chaperonins were co-expressed in the Arctic express cells (109-112).

For the latter, production of OPH^{-LSS} protein in addition to at least one of the chaperones (Cpn60) was observed. The second chaperone (Cpn10) has a molecular weight of 10 kDa and was potentially visible at the bottom of the SDS gel. The protein of interest was however still insoluble. Experiments with the Lemo cells did not show any noticable production of recombinant OPH^{-LSS}, irrespective of the titrated induction levels (see Fig. 6.7).

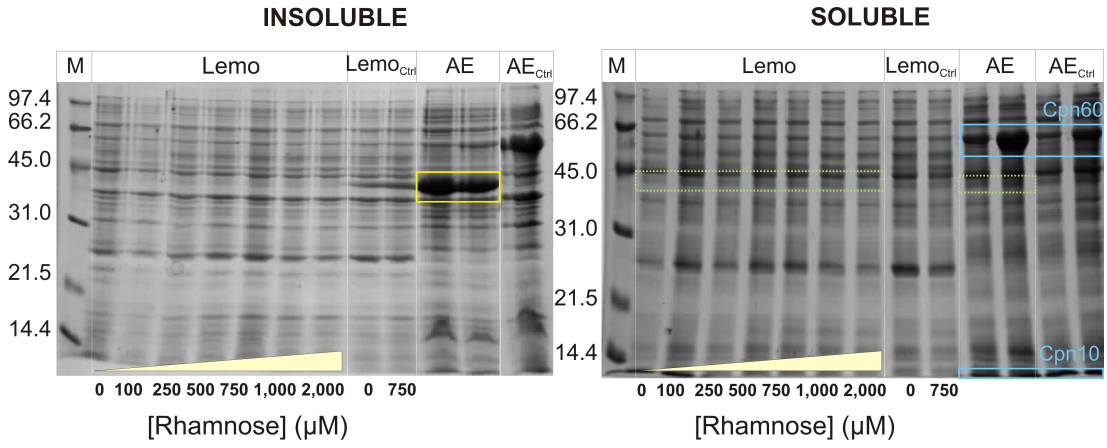


Figure 6.7: Alternative expression strains were tested for their ability to produce soluble OPH^{-LSS} protein. In Lemo cells, induction levels were titrated using L-rhamnose (white bar). In Arctic Express (AE) cells, cold-induced chaperones (blue boxes) were co-expressed. Cells transformed by an empty YSBLIC3C plasmid served as control reactions.

6.3.7 Preparation of OPH fusion constructs and analysis of protein production

Fusion constructs between OPH^{-LSS} and N-terminal GST, GFP, MBP and IM9 protein tags were prepared in an attempt to support correct folding of the hydrolase and make it accessible for purification. The truncated hydrolase gene was cloned into the appropriate fusion vectors by the University of York's Technology facility. Its successful insertion was verified by restriction endonuclease digest (results not shown).

When production of the fusion constructs was assayed in an *E. coli* BL21 (DE3) expression strain, significant overproduction of soluble protein could be observed for two out of four constructs, OPH^{-LSS}:GST and OPH^{-LSS}:MBP. Both presented strong bands on an SDS PAGE gel, corresponding to the added molecular weight of both hydrolase and fusion tag (see Fig. 6.8).

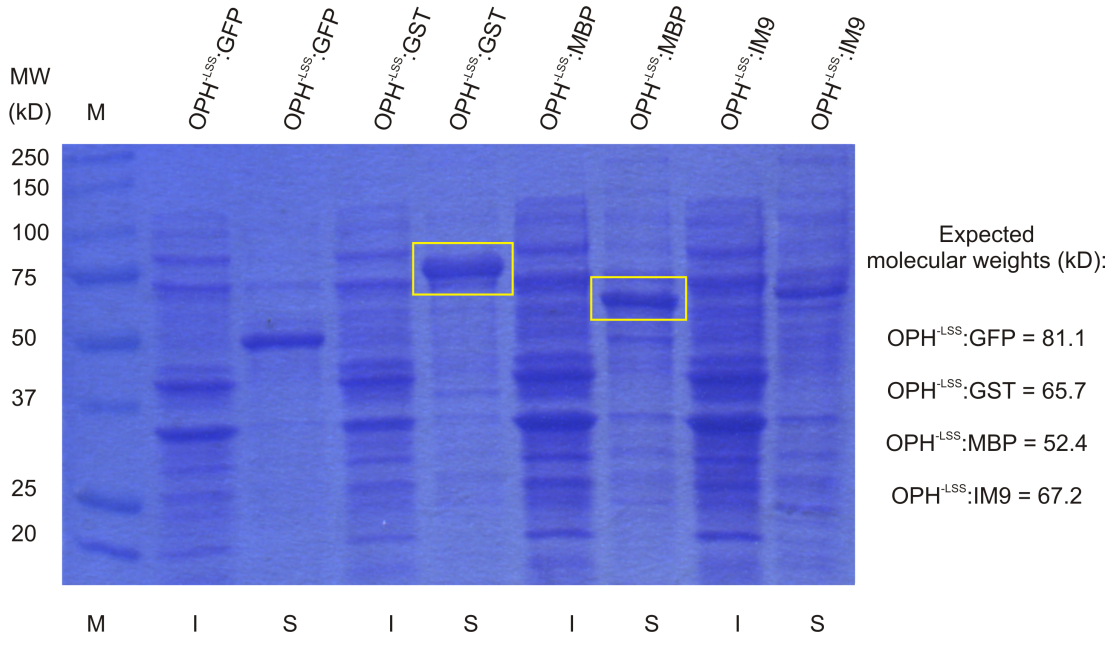


Figure 6.8: Protein production and solubility was assayed for *E. coli* cells transformed by the four fusion construct plasmids. Insoluble (I) and soluble (S) fractions of the lysed cells are shown. Yellow boxes mark soluble protein which corresponds in molecular weight to the expected fusion construct.

6.3.8Purification of the OPH^{LSS}:GST fusion construct

As the OPH^{-LSS} fusion construct presented the most significant overproduction of soluble protein, it was chosen as a first target for purification. Using Ni-affinity chromatography, an area with several peaks in absorbance could be separated from the soluble cell fraction. Analysis of the peak area using SDS PAGE showed dominant bands at the expected molecular weight of the fusion construct (65 kDa, see Fig. 6.9).

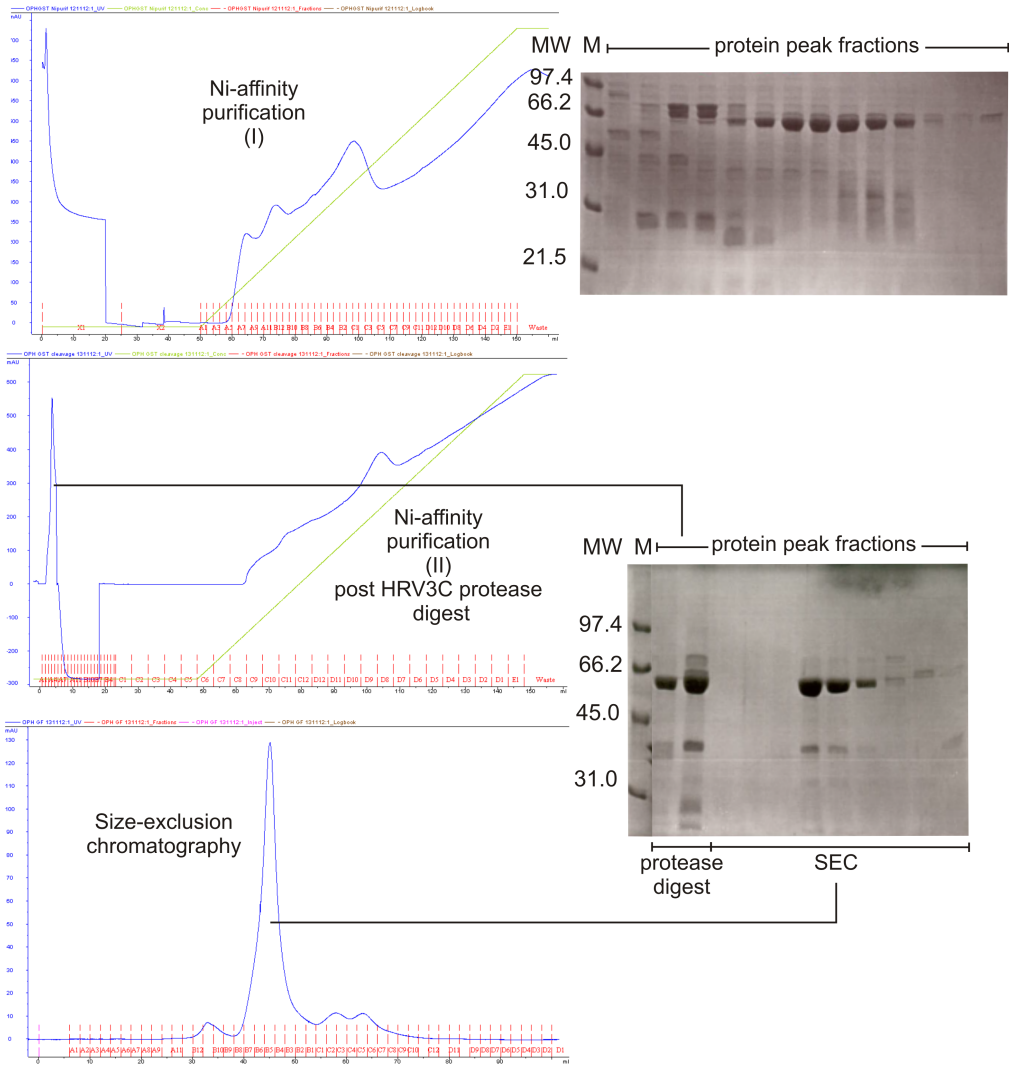


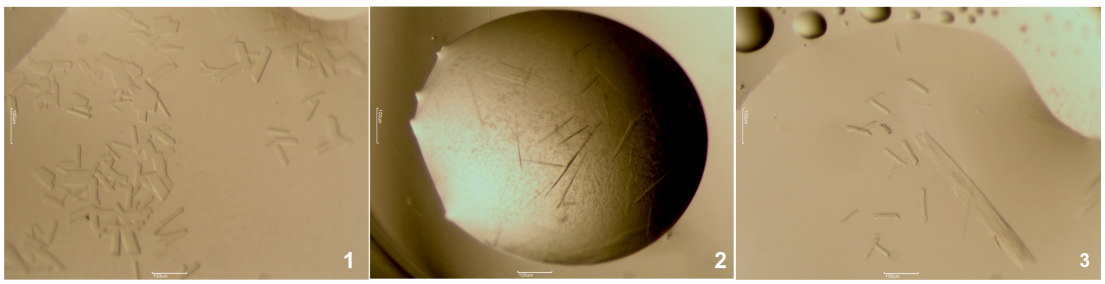
Figure 6.9: The OPH:GST fusion construct was purified in three steps. After Ni-affinity chromatography (top), a protease digest was carried out to separate the hydrolase and from its His-tagged fusion protein. SDS PAGE analysis of the second Ni-affinity chromatography (centre) however showed protein fractions corresponding predominantly to the fused construct. Subsequent size-exclusion chromatography then yielded an OPH:GST protein sample with only minor impurities (bottom).

Incubation of the pooled protein fractions with HRV3C protease initially appeared to allow for separation of the hydrolase from its fusion tag. A sharp absorbance peak was observed in the imidazole free buffer fractions of the second Ni-affinity purification step, followed by further peaks in the imidazole gradient. SDS PAGE analysis however showed the putative hydrolase fractions to correspond to a 65 kDa protein – the expected molecular weight of the OPH:GST fusion.

When the imidazole free peak fractions were pooled, a significant amount of protein remained in solution and could be concentrated for a final purification step using size exclusion chromatography. The latter resulted in a major peak in absorbance flanked by 3 smaller peaks. SDS PAGE analysis showed this peak area to correspond to the OPH:GST fusion in addition to minor, lower molecular weight impurities (see Fig. 6.9).

6.3.9 OPH:GST crystallisation and diffraction testing

To determine the OPH^{-LSS} three-dimensional structure, crystallisation experiments were carried using the purified fusion construct. With the hanging drop vapour diffusion method and three different screens, various crystal forms were observed after an incubation time of several days. The crystals varied in shape, forming needles or thin plates, and formed in isolation as well as in larger clusters (see Fig. 6.10). Hits were predominantly found in the PACT and INDEX screens.



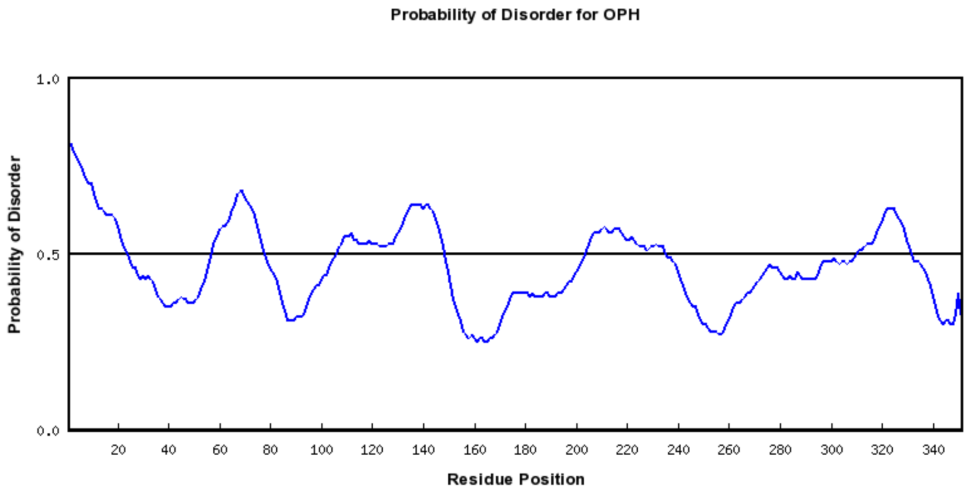
1. INDEX E2: 0.2 M Ammonium Acetate, 0.1 M Bis-Tris pH 5.5, 45% v/v MPD
2. INDEX E1: 0.2 M Ca Chloride, 0.1 M Bis-Tris pH 6.5, 45% v/v MPD
3. INDEX E5: 0.2 M Ammonium Acetate, 0.1 M Tris pH 8.5, 45% v/v MPD

Figure 6.10: When the purified OPH^{-LSS} protein was screened against 3 x 96 conditions, crystals of various morphologies were found. Hits were particularly abundant for the INDEX and PACT screens.

When three of the crystals (from INDEX screen E2 and E5, see above) were tested using an in-house X-ray source, no diffraction was observed. A single crystal from condition E5 was subjected to synchrotron radiation. Again, no significant diffraction could be detected (results not shown).

6.3.10 *Prediction of protein disorder*

Due to the absence of any perceivable diffraction, the OPH^{-LSS} amino acid sequence was analysed for structural disorder. Several amino acid stretches, distributed throughout the protein, were predicted to be disordered, with the most prominent area being the 23 most N-terminal residues.




```
1   MGADRTAATP AAANPAATEP VKWECPAGYE VKEGLNVDFP HKGMKRAFIV50
    ********** ********** ***_______ __________ __________
51   YPAKNVSGPA PVWVPMTGSV ESTNDNLTVA RSGANSILAD HGYTVIAPVR100
    ______**** ********** ******* ___ __________ __________
101  ACANQDPNIR GERCNGPGSN GWNWNPWFEG RAADPSGEHW KNDEGPDSSF150
    ______**** ********** ********** ********** ********__
151  FVAMVQCVGT KYKLDARRLF LGGISSGGTM TNRALLFRSN FWAGGLPISG200
    _______ _______ _______ _______ _______ _______ _______
201  EWYVTSDDGT PLSFDDARAA VAAAPTKIHQ GRVGPYPLPA KVGPLIVMTV250
     ******* ********** ********** ****_____ ______ _______
251  WGGEKDLWNC TRPDGSRFLC ADYRPSTQAG SNFFSAQPDV VHVACSSTHG300
    _______ _______ _______ _______ _______ _______ _______
301  HMWPQLNTQE FNRWALDTLA SHPKGSDPRS FKLTQPPEGY TCHVGPFTGL350
    __________ ********** ********** *_________ __________
351  Y
```

Figure 6.11 : Disorder in the OPH^{-LSS} three dimensional structure was predicted using the RONN server (178). Residues with a probability of disorder above 50 % are highlighted with stars in the amino acid sequence.

6.4 Discussion

The *pvaB* gene, encoding oxidised polyvinyl alcohol hydrolase from *Pseudomonas* sp. VM15C, has been optimised for expression in *E. coli* and, once subcloned into the pET28a vector, was supplied with an additional hexahistidine sequence for purification. However, any attempt to isolate the protein from cell free extracts has failed as OPH does not seem to be produced in its soluble form. This is in accordance with previous findings, for OPH from *Pseudomonas* as well as a homologue from *Sphingopyxis* (193). The only soluble protein isolated to this day was purified directly from a *Pseudomonas* strain (194).

The presence of a lipoprotein signal sequence appeared to be the most likely explanation for insolubility as it would target the protein to the membrane fraction and exclude it from cell free extracts used for protein purification. However, removal of the signal sequence does not circumvent this problem. Lyophilised whole cells or crude extracts of cells producing OPH without its signal sequence appear to have lost their hydrolytic activity. This is in contrast to results obtained using the full length protein (Ms Elina Siirola, personal communication). After translation, the signal sequence may thus, in addition to its targeting function, contribute to the protein's overall fold. The latter would also explain the lack of success in solubilising the protein by the addition of detergents to the resuspension buffer. Also the alteration of expression strains may not have an effect on this lack of structural support. The *E. coli* Lemo21 (DE3) strain allows for fine-tuning of T7 RNA polymerase expression and thus indirectly controls the Lac promoter in pET28-based YSBLIC3C plasmids (109). It is thus particularly well-suited for the overproduction of toxic proteins but may not overcome initial folding problems. Growth of the *E. coli* Arctic Express cells at low temperatures induces chaperonins which are subsequently hoped to aid in folding of the protein of interest (111,112). A prerequisite for this type of support however is the accessibility of the target peptide to the chaperones. Early aggregation of OPH^{-LSS} may prevent this interaction.

The attachment of protein tags has been shown to allow for solubilisation and subsequent purification of previously insoluble targets (118). When a variety of protein tags were fused to the OPH^{-LSS} hydrolase, two out four constructs appeared

as soluble proteins in the cytosolic fractions of lysed cells. The mechanisms by which proteinaceous tags allow for the solubilisation of their targets is not clear and might vary between the tags and targets. It has been suggested to be the consequence of stabilisation and enhancement of the native folding process. Alternatively, the tag may "transport" the target protein to a cellular location which makes it accessible to molecular chaperones. Assuming the hydrolase component of the GST:OPH fusion is contributing to full length construct's solubility and is thus itself correctly folded, the latter hypothesis would be a likely explanation for the observed production of soluble protein. The wild type, full length OPH protein appears in the form of inclusion bodies when cells are lysed and their soluble and insoluble fractions are visualised by SDS PAGE. Crude extracts of cells producing the protein however are active. It is possible that targeting of the protein through the cell plays an important role in its stabilisation. It may allow for close contact with chaperones which stabilise and/or catalyse the final folding process. It may then catalyse OPVA hydrolysis close to or incorporated within the membrane. It could even occur on the cell's surface as suggested previously (190). Removal of the signal sequence may thus lead to fast aggregation of the partially folded peptide, during or immediately after translation, precluding access to cytosolic chaperones. The fusion of OPH^{-LSS} to a protein solubility tag, which is itself highly soluble and stable, could enhance relocation and movement of the partially folded hydrolase in the cytosol. Here it might fold into a stable tertiary and quarternary structure, stabilised or assisted by chaperones.

If the solubility tag functions predominantly as a transporter, as described above, the OPH protein would thus be likely to remain in solution after its cleavage. If GST however has an essential stabilising function for the hydrolase, its separation may result in immediate precipitation of the hydrolase. If solubility can be retained by the hydrolase is unclear to date as attempts to separate both proteins by way of a targeted protease digest were unsuccessful. This may be the consequence of a protein fold or arrangement in which the 3C cleavage sequence is buried within one of the proteins or even"sandwiched" between them. Either case would restrict access of the protease to its recognition sequence.

Another stability indicator of the purified fusion construct is its apparent ability to crystallise. In the absence of diffraction and, consequently, structural data, it is

however unclear if the crystals have formed from the full length fusion protein. The crystallisation of targets with large protein tags is rare but has been observed previously (198). Alternatively, spontaneous proteolysis or even remaining 3C protease activity in the protein sample may have led to cleavage of the fusion and crystallisation of individual GST or OPH^{-LSS} protein. The predicted protein disorder of the latter would likely impair diffraction properties of the fusion and individual hydrolase crystals. The extremely low resolution obtained may be a consequence of this feature. In either case, the lack of isolated high resolution spots on the diffraction image indicates that the crystals have formed from protein rather than from salt contained in the screen solutions.

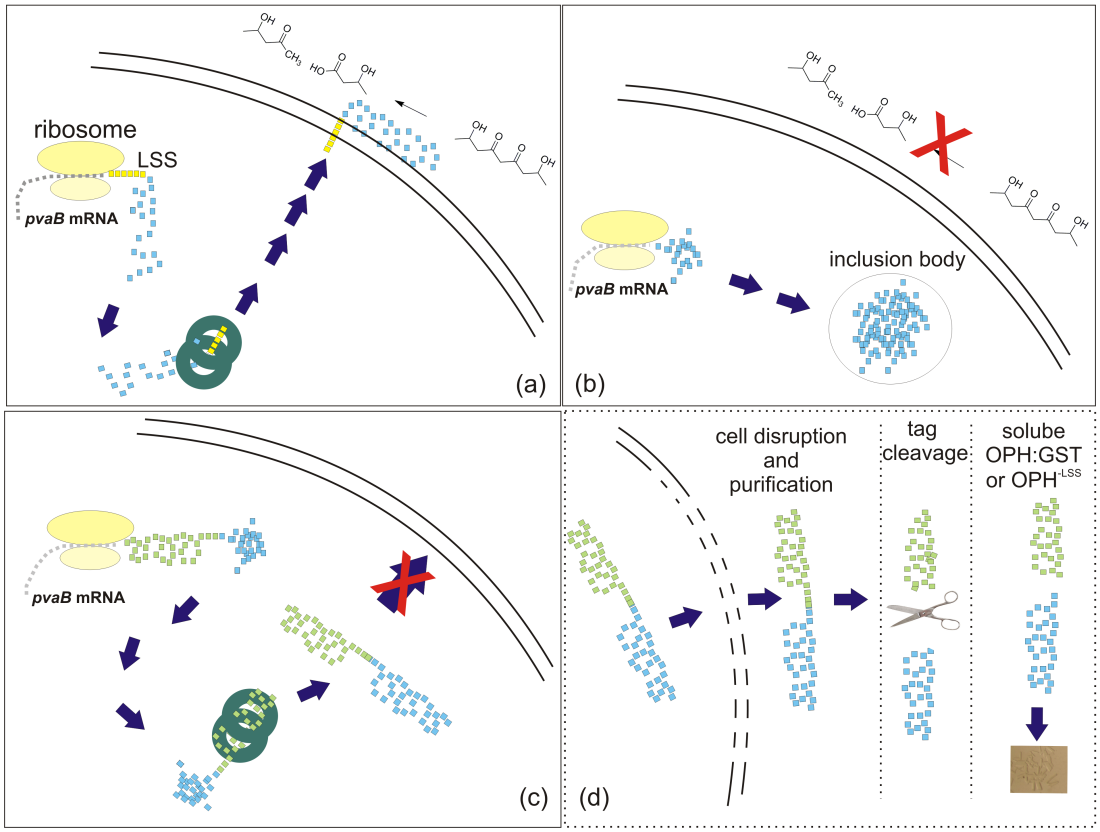


Figure 6.12: Protein tags have been suggested to enhance solubility of target proteins by allowing for interactions with molecular chaperones. For full length OPH, this interaction might occur spontaneously as the partially folded protein is channelled through the cytosol via the LSS (a). OPH^{-LSS} is likely to form insoluble aggregates after or during its translation, leading to formation of inclusion bodies and restricting access to chaperones (b). When fused to a highly soluble protein tag, OPH^{-LSS} is pulled into the cytosol where it may come into contact with chaperones (c). Once folded correctly it is stable, can be purified and concentrated. It remains to be determined if solubility is retained after removal of the protein tag (d).

In conclusion, an oxidised polyvinyl alcohol hydrolase fusion protein has been purified for the first time and could be crystallised. This may create the basis for structural and mechanistic investigations into this unusual family of β-diketone hydrolases. Future work should include the testing of alternative fusion constructs and, if required, their optimisation to allow for cleavage of the hydrolase from its protein tag. An extended linker sequence may overcome the apparent problem of accessibility to the protease recognition sequence. After purification of the isolated OPH^{-LSS} protein, the next priority will have to be the determination of an OPH three-dimensional structure, substrate range and a potential mechanism of hydrolysis. The identification and alteration of catalytic active site residues could alter the reaction with respect to efficiency and specificity. Eventually, experiments with an aim to revert the reaction could lead to the highly valuable enzymatic synthesis of β-diketones for synthetic applications.

Chapter 7: General conclusions and outlook

The work presented in the previous chapters gives some insight into the diversity of enzymatic reactions leading to the cleavage of carbon-carbon bonds. Members of the two major protein families involved in this process, the C-C hydrolases and lyases, have been analysed with respect to their structure, mechanism and/or substrate requirements.

The complex structure of a *Bs*PAD protein with its *p*-coumaric acid substrate, together with kinetic data obtained from an active site mutant library, gave an insight into the cofactor independent reactions of carbon-carbon bond lyases. Due to their effect on brewing and food fermentation processes, PAD enzymes have already (indirectly) been at the centre of biotechnological applications for centuries. More detailed understanding of the reactions involved however allows their use in more complex applications.

An example of this is the preparation of natural vanillin from plant derived ferulic acid. The process occurs naturally and was shown to be catalysed by a hydroxycinnamoyl hydratase lyase. It relies, however, on the substrate to initially form a thioester with coenzyme A (feruloylCoA). The latter is then hydrated prior to cleavage of the C-C bond (199). The use of a phenolic acid decarboxylase could now circumvent this coenzyme dependent reaction step if used in a "one pot" cascade reaction with an oxygenase (see Fig. 7.1). Initial experiments with a PAD enzyme from *Enterobacter* and a fungal alkene cleaving enzyme from *Trametes hirsuta* (200) have already shown promising results (Dr. Aashrita Rajagopalan, personal communication). If optimised, such processes could create an environmentally friendly method of turning plant derived waste products into high value flavour compounds.

In addition to the well characterised decarboxylation, PAD enzymes have been shown to catalyse carboxylation reactions and thus have potential applications in CO_2 fixation and the conversion a waste products into synthetic building blocks (13). An enzyme such as *Bs*PAD, which is stable, highly soluble, very reactive and amenable to mutations of active site residues is an ideal target for future work in the areas described above. Process and/or enzyme engineering could optimise the

catalyst even further for large scale applications and maybe even widen it substrate scope.

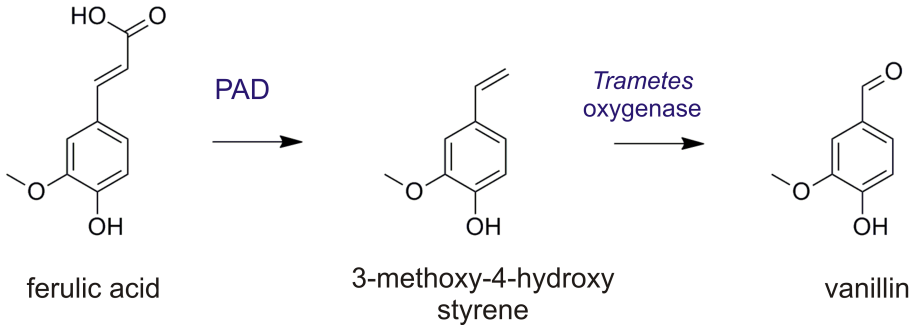


Figure 7.1: *Bs*PAD may be the future target for a cascade reaction with the alkene cleaving enzyme from *Trametes hirsuta*, producing high value vanillin from the renewable resource ferulic acid.

In contrast to decarboxylases, the enzyme-catalysed hydrolytic cleavage of carbon-carbon bonds is less well understood. With the work described in the previous chapters, a first insight has been gained into two, structurally and mechanistically, very different C-C hydrolases.

Phloretin hydrolase from *Eubacterium ramulus* (Phy) was shown to resemble a previously characterised diacetylphloroglucinol hydrolase (PhlG) in aspects of metal dependency and a potential reaction mechanism. Only small differences were found regarding the two enzymes' substrate specificities. With respect to structural features and catalytic motifs however, the two hydrolases differ significantly from the well characterised, serine dependent, α/β fold carbon-carbon hydrolases. Despite successful crystallisation of Phy, the protein's structure could not be determined to date. Efforts to produce crystals of higher diffraction quality and/or an improved anomalous signal should thus be central to future work on this hydrolase. Once a three-dimensional structure is obtained, Phy may be a fascinating target for further protein engineering. The enzyme accepts lower and higher molecular weight substrates (MAPG *vs*. phloretin). This feature that could be exploited for modelling it according to ligand requirements as described previously for other biocatalysts (41). Ultimately, reverting the Phy reaction into an enzymatic Friedel Crafts acylation

would be an unprecedented biocatalytic reaction and allow for the synthesis of high value chemical building blocks.

The oxidised polyvinyl alcohol hydrolase analysed as part of this project is a similarly exciting target for biocatalytic applications. If its reaction could be reverted, OPH would catalyse the formation of β-diketones for downstream synthetic applications. A truncated version of the enzyme, lacking its lipoprotein signal sequence, has been purified (and possibly crystallised) in the form of a soluble fusion construct for the first time. This now creates the basis for fundamental research into the OPH structure and function. If the enzyme is active in the presence of its fusion tag, an activity screen using a variety β-diketones could give important insights into possible substrate requirements. Determination of the enzyme's structure, ideally in complex with one of its substrates or an analogue thereof, would allow for an insight into a potential mechanism. Although OPH contains the lipase box motif and is likely to make use of a serine nucleophile, it remains to be determined if the catalytic mechanism involves formation of a covalent acyl enzyme intermediate. With respect to large scale applications, engineering of additional fusion constructs may be required. These should be designed to increase protein production levels and, potentially, allow cleavage of the hydrolase from its fusion tag. As for Phy (and *Bs*PAD), a thorough understanding of the wild type enzyme's structure and function may inspire future protein engineering to improve activity, change selectivity or even enable a reversal of the hydrolytic reaction into carbon-carbon bond formation.

In summary, all catalysts described are important targets for biocatalytic applications. They could, in time, become a "green" alternative to their abiotic counterparts and may be used for both degradative and synthetic reactions. Whereas the decarboxylase is already a fascinating target for novel applications, more fundamental research is required for the C-C hydrolases described. With the appropriate enzyme and process engineering, all proteins may one day be important catalysts in large scale industrial applications.

LIST OF ABBREVIATIONS	
Amino acids	
Ala	Alanine
Arg	Arginine
Asn	Asparagine
Gln	Glutamine
Glu	Glutamate
His	Histidine
Lys	Lysine
Met	Methionine
Phe	Phenylalanine
SeMet	Selenomethionine
Ser	Serine
Thr	Threonine
Trp	Tryptophan
Tyr	Tyrosine
PAD	Phenolic acid decarboxylase
Proteins	
AADC	Aromatic amino acid decarboxylase
ABDH	*Anabaena β*–diketone hydrolase
AMDase	α–aryl malonate decarboxylase
BphD	2-hydroxy-6-keto-6-phenylhexa-2,4-dienoic acid hydrolase
*Bp*PAD	Bacillus pumilus phenolic acid decarboxylase
*Bs*PAD	Bacillus subtilis phenolic acid decarboxylase
DDC	L-3,4-dihydroxyphenylalanine decarboxylase
DHPONH	2,6-dihydroxy-*pseudo*-oxynicontine hydrolase
*EP*AD	*Enterobacter* sp. Px6-4 phenolic acid decarboxylase
FAH	fumarylacetoacetate hydrolase
Fus	Fusion protein (“tag”)
GFP	Green fluorescence protein
GST	Glutathione S-transferase
HCHL	Hydroxycinnamoyl-CoA hydratase lyase

ICL	Isocitrate lyase
Im9	Immunity protein 9
*Ko*PAD	*Klebsiella oxytoca* phenolic acid decarboxylase
*Lp*PAD	*Lactobacillus plantarum* phenolic acid decarboxylase
LSS	Lipoprotein signal sequence
MBP	maltose binding protein
MhpC	2-hydroxy-6-keto-nona-1,9-dienoic acid 5,6 hydrolase
MSAD	malonate semialdehyde decarboxylase
NeuA	*N*-acetylneuraminic acid lyase
NusA	N-utlilization substance A
OCH	oxocamphor hydrolase
OMPDC	orothidine 5'-monophosphate decarboxylase
OMP	outer membrane protease
OPH	Oxidised polyvinyl alcohol hydrolase
PDC	pyruvate decarboxylase
PhlG	Diacetylphloroglucinol hydrolase
Phy	Phloretin hydrolase
TrxA	Thioredoxin A
VAO	Vanillyl-alcohol oxidase
PHBD	Polyhydroxybutyrate depolymerase
Cofactors	
CoA	Coenzyme A
PLP	Pyridoxal phosphate
PQQ	Pyrroloquinoline quinone
ThDP	Thiamine diphosphate
TPP	Thiamine pyrophosphate
Nucleic acids	
dATP	deoxyadenosine triphosphate
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide
dTTP	deoxythymidine triphosphate
mRNA	Messenger ribonucleic acid
RNA	Ribonucleic acid

tRNA	Transfer ribonucleic acid
Methods	
GC	Gas chromatography
HPLC	High-performance liquid chromatography
IMAC	Immobilised metal affinity chromatography
LIC	Ligation independent cloning
NMR	Nuclear magnetic resonance
PCR	Polymerase chain reaction
SDM	Site-directed mutagenesis
SDS PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
UV	Ultraviolet
UV-VIS	Ultraviolet-visible
Reagents	
DMSO	Dimethylsulfoxide
EDTA	Ethylenediaminetetraacetic acid
IPTG	β-D-1-thiogalactopyranoside
LB	Lysogeny broth
MES	2-(N-morpholino)ethanesulfonic acid 2-(N-morpholino)ethanesulfonic acid
Ni-NTA	Ni-nitriloacetic acid
PEG	Polyethylene glycol
PMSF	Phenylmethanesulfonylfluoride
Units	
Å	Ångstrom
bp	Basepair
c	Concentration
h	Hour
kDa	Kilodalton
L	Litre
M	Molar
mg	Milligram

min	Minute
µg	Microgram
µL	Microlitre
µm	Micrometre
µM	Micromolar
mL	Millilitre
mm	Millimetre
mM	Millimolar
ng	Nanogram
nm	Nanometre
nmoles	Nanomoles
OD	Optical density
rpm	Revolutions per minute
s	Second
v/v	Volume/volume
w/v	Weight/volume
w/w	Weight/weight
Reaction substrates and products	
2-HAP	2-hydroxyacetophenone
2,4-DHAP	2,4-dihydroxyacetophenone
2,6-DHAP	2,6-dihydroxyacetophenone
4-HAP	4-hydroxyacetophenone
carbidopa	(2*S*)-3-(3,4-dihydroxyphenyl)-2-hydrazinyl-2-methyl-propanoic acid
CEHPOBA	4-[(2-carboxyethyl)hydroxyphosphynil]-3-oxobutyric acid
DAPG	Diacetylphloroglucinol
DHP	2,6-dihydroxypyridine
DOPA	L-3,4-dihydroxyphenylalanine
dopamine	L-3,4-dihydroxyphenylamine
HODA	2-hydroxy-6-oxo- dienoic acid
HPD	2-hydroxypenta-2,4-dienoic acid
HOPDA	2-hydroxy-6-keto-6-phenylhexa-2,4-dienoic acid

MAB	γ-N-methylaminobutyrate
MAPG	Monoacetylphloroglucinol
MCP	*Meta*-cleavage product
OHCU	2-Oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline
OPVA	Oxidised polyvinyl alcohol
PG	Phloroglucinol
PVA	polyvinylalcohol
R-PAC	(*R*)-phenyl acetylcarbinol
TAPG	Triacetylphloroglucinol
TCA	Tricarboxylic acid
Crystallography	
CSS	Clear strategy screen
I	Intensity
MAD	Multiwavelength anomalous dispersion
NCS	Non-crystallographic symmetry
RMSD	Root-mean-square deviation
R_{merge}	Merging R factor
R_{pim}	precision-indicating merging R factor
Other	
EC	Enzyme commission
Fig.	Figure
HRV	human rhinovirus
n.d.	Not determined
NP	Non-polar
P	Polar
PDB	Protein Data Bank
pI	Isoelectric point
TM	Trademark
wt	Wild type

References

1. Anastas, P., and Warner, J. (2000) *Green Chemistry: Theory and Practice*, Oxford University Press, USA
2. Grogan, G. (2009) *Practical Biotransformations: a beginner's guide*, 1 ed., Wiley
3. Fessner, W. D. (1998) Enzyme mediated C-C bond formation. *Current Opinion in Chemical Biology* **2**, 85-97
4. Sukumaran, J., and Hanefeld, U. (2005) Enantioselective C-C bond synthesis catalysed by enzymes. *Chemical Society Reviews* **34**, 530-542
5. Breuer, M., and Hauer, B. (2003) Carbon–carbon coupling in biotransformation. *Current Opinion in Biotechnology* **14**, 570-576
6. Breuer, M., and Hauer, B. (2003) Carbon-carbon coupling in biotransformation. *Current Opinion in Biotechnology* **14**, 570-576
7. Sprecher, M., Berger, R., and Sprinson, D. B. (1964) Stereochemical Course of the Isocitrate Lyase Reaction. *Journal of Biological Chemistry* **239**, 4268-4271
8. Dunn, M. F., Ramírez-Trujillo, J. A., and Hernández-Lucas, I. (2009) Major roles of isocitrate lyase and malate synthase in bacterial and fungal pathogenesis. *Microbiology* **155**, 3166-3175
9. Serrano, J. A., Camacho, M., and Bonete, M. J. (1998) Operation of glyoxylate cycle in halophilic archaea: presence of malate synthase and isocitrate lyase in Haloferax volcanii. *Federation of European Biochemical Societies Letters* **434**, 13-16
10. Yoshida, T., Fujita, K., and Nagasawa, T. (2002) Novel reversible indole-3-carboxylate decarboxylase catalyzing nonoxidative decarboxylation. *Bioscience, Biotechnology & Biochemistry* **66**, 2388-2394
11. He, Z., and Wiegel, J. (1995) Purification and characterization of an oxygen-sensitive reversible 4-hydroxybenzoate decarboxylase from Clostridium hydroxybenzoicum. *European Journal of Biochemistry* **229**, 77-82
12. He, Z., and Wiegel, J. (1996) Purification and characterization of an oxygen-sensitive, reversible 3,4-dihydroxybenzoate decarboxylase from Clostridium hydroxybenzoicum. *Journal of Bacteriology* **178**, 3539-3543
13. Wuensch, C., Glueck, S. M., Gross, J., Koszelewski, D., Schober, M., and Faber, K. (2012) Regioselective Enzymatic Carboxylation of Phenols and Hydroxystyrene Derivatives. *Organic Letters* **14**, 1974-1977
14. Vanpoelje, P. D., and Snell, E. E. (1990) PYRUVOYL-DEPENDENT ENZYMES. *Annu. Rev. Biochem.* **59**, 29-59
15. Wise, E. L., Yew, W. S., Gerlt, J. A., and Rayment, I. (2004) Evolution of Enzymatic Activities in the Orotidine 5‘-Monophosphate Decarboxylase Suprafamily: Crystallographic Evidence for a Proton Relay System in the Active Site of 3-Keto-l-gulonate 6-Phosphate Decarboxylase†,‡. *Biochemistry* **43**, 6438-6446
16. Rodriguez, H., Landete, J. M., Curiel, J. A., de Las Rivas, B., Mancheno, J. M., and Munoz, R. (2008) Characterization of the p-coumaric acid decarboxylase from Lactobacillus plantarum CECT 748(T). *Journal of Agricultural & Food Chemistry* **56**, 3068-3072

17. Huang, Z., Dostal, L., and Rosazza, J. P. (1994) Purification and characterization of a ferulic acid decarboxylase from Pseudomonas fluorescens. *Journal of Bacteriology* **176**, 5912-5918
18. Cavin, J. F., Dartois, V., and Divies, C. (1998) Gene cloning, transcriptional analysis, purification, and characterization of phenolic acid decarboxylase from Bacillus subtilis. *Applied & Environmental Microbiology* **64**, 1466-1471
19. Ward, O. P., and Singh, A. (2000) Enzymatic asymmetric synthesis by decarboxylases. *Current Opinion in Biotechnology* **11**, 520-526
20. Percudani, R., and Peracchi, A. (2003) A genomic overview of pyridoxal-phosphate-dependent enzymes. *European Molecular Biology Organization Reports* **4**, 850-854
21. Eliot, A. C., and Kirsch, J. F. (2004) PYRIDOXAL PHOSPHATE ENZYMES: Mechanistic, Structural, and Evolutionary Considerations. *Annu. Rev. Biochem.* **73**, 383-415
22. Alexander, F. W., Sandmeier, E., Mehta, P. K., and Christen, P. (1994) Evolutionary relationships among pyridoxal-5′ -phosphate-dependent enzymes. *European Journal of Biochemistry* **219**, 953-960
23. Christen, P., and Mehta, P. K. (2001) From cofactor to enzymes. The molecular evolution of pyridoxal-5′ -phosphate-dependent enzymes. *The Chemical Record* **1**, 436-447
24. Hayashi, H. (1995) Pyridoxal Enzymes: Mechanistic Diversity and Uniformity. *Journal of Biochemistry* **118**, 463-473
25. Mozzarelli, A., and Bettati, S. (2006) Exploring the pyridoxal 5′ -phosphate-dependent enzymes. *The Chemical Record* **6**, 275-287
26. Zhu, M. Y., and Juorio, A. V. (1995) Aromatic l-amino acid decarboxylase: Biological characterization and functional role. *General Pharmacology: The Vascular System* **26**, 681-696
27. Jansonius, J. N. (1998) Structure, evolution and action of vitamin B6-dependent enzymes. *Current Opinion in Structural Biology* **8**, 759-769
28. Ford, G. C., Eichele, G., and Jansonius, J. N. (1980) Three-dimensional structure of a pyridoxal-phosphate-dependent enzyme, mitochondrial aspartate aminotransferase. *Proceedings of the National Academy of Sciences* **77**, 2559-2563
29. Burkhard, P., Dominici, P., Borri-Voltattorni, C., Jansonius, J. N., and Malashkevich, V. N. (2001) Structural insight into Parkinson's disease treatment from drug-inhibited DOPA decarboxylase. *Nature Structural Molecular Biology* **8**, 963-967
30. Giardina, G., Montioli, R., Gianni, S., Cellini, B., Paiardini, A., Voltattorni, C. B., and Cutruzzolà, F. (2011) Open conformation of human DOPA decarboxylase reveals the mechanism of PLP addition to Group II decarboxylases. *Proceedings of the National Academy of Sciences*
31. Ishii, S., Mizuguchi, H., Nishino, J., Hayashi, H., and Kagamiyama, H. (1996) Functionally Important Residues of Aromatic L-Amino Acid Decarboxylase Probed by Sequence Alignment and Site-Directed Mutagenesis. *Journal of Biochemistry* **120**, 369-376
32. Aminoff, M. J. (1994) Treatment of Parkinsons-disease. *West. J. Med.* **161**, 303-308
33. Kaye, A. M. (1984) Ornithine decarboxylase. Purification and properties of ornithine decarboxylase. *Cell Biochemistry and Function* **2**, 2-6

34. Duggleby, R. G. (2006) Domain Relationships in Thiamine Diphosphate-Dependent Enzymes. *Accounts of Chemical Research* **39**, 550-557
35. Bugg, T. (2009) Enzymatic Carbon–Carbon Bond Formation. in *Introduction to Enzyme and Coenzyme Chemistry*, Blackwell Publishing Ltd. pp 156-192
36. Jordan, F. (2003) Current mechanistic understanding of thiamin diphosphate-dependent enzymatic reactions. *Natural Product Reports* **20**, 184-201
37. Andrews, F. H., and McLeish, M. J. (2012) Substrate specificity in thiamin diphosphate-dependent decarboxylases. *Bioorganic Chemistry* **43**, 26-36
38. Pei, X.-y., Erixon, K. M., Luisi, B. F., and Leeper, F. J. (2010) Structural Insights into the Prereaction State of Pyruvate Decarboxylase from Zymomonas mobilis. *Biochemistry* **49**, 1727-1736
39. Versées, W., Spaepen, S., Wood, M. D. H., Leeper, F. J., Vanderleyden, J., and Steyaert, J. (2007) Molecular Mechanism of Allosteric Substrate Activation in a Thiamine Diphosphate-dependent Decarboxylase. *Journal of Biological Chemistry* **282**, 35269-35278
40. Meyer, D., Neumann, P., Parthier, C., Friedemann, R., Nemeria, N., Jordan, F., and Tittmann, K. (2010) Double Duty for a Conserved Glutamate in Pyruvate Decarboxylase: Evidence of the Participation in Stereoelectronically Controlled Decarboxylation and in Protonation of the Nascent Carbanion/Enamine Intermediate. *Biochemistry* **49**, 8197-8212
41. Gocke, D., Walter, L., Gauchenova, E., Kolter, G., Knoll, M., Berthold, C. L., Schneider, G., Pleiss, J., Müller, M., and Pohl, M. (2008) Rational Protein Design of ThDP-Dependent Enzymes—Engineering Stereoselectivity. *ChemBioChem* **9**, 406-412
42. Rogers. (1997) Biotransformation for L-ephedrine production. *Advances in Biochemical Engineering/Biotechnology* **56**, 33-59.
43. Pohl, M., Sprenger, G. A., and Müller, M. (2004) A new perspective on thiamine catalysis. *Current Opinion in Biotechnology* **15**, 335-342
44. Dobritzsch, D., König, S., Schneider, G., and Lu, G. (1998) High Resolution Crystal Structure of Pyruvate Decarboxylase from Zymomonas mobilis : Implications for substrate activation in pyruvate decarboxylases. *Journal of Biological Chemistry* **273**, 20196-20204
45. Li, T., Huo, L., Pulley, C., and Liu, A. (2012) Decarboxylation mechanisms in biological system. *Bioorganic Chemistry* **43**, 2-14
46. Begley, T. P., and Ealick, S. E. (2004) Enzymatic reactions involving novel mechanisms of carbanion stabilization. *Current Opinion in Chemical Biology* **8**, 508-515
47. Anand, R., Dorrestein, P. C., Kinsland, C., Begley, T. P., and Ealick, S. E. (2002) Structure of Oxalate Decarboxylase from Bacillus subtilis at 1.75 Å Resolution†,‡. *Biochemistry* **41**, 7659-7669
48. Reinhardt, L. A., Svedruzic, D., Chang, C. H., Cleland, W. W., and Richards, N. G. J. (2003) Heavy Atom Isotope Effects on the Reaction Catalyzed by the Oxalate Decarboxylase from Bacillus subtilis. *Journal of the American Chemical Society* **125**, 1244-1252
49. Tanner, A., Bowater, L., Fairhurst, S. A., and Bornemann, S. (2001) Oxalate Decarboxylase Requires Manganese and Dioxygen for Activity: Overexpression and characterization of bacillus subtilis YvrK AND YoaN. *Journal of Biological Chemistry* **276**, 43627-43634

50. Mäkelä, M., Hildén, K., and Lundell, T. (2010) Oxalate decarboxylase: biotechnological update and prevalence of the enzyme in filamentous fungi. *Applied Microbiology and Biotechnology* **87**, 801-814
51. Goto, M., Hayashi, H., Miyahara, I., Hirotsu, K., Yoshida, M., and Oikawa, T. (2006) Crystal Structures of Nonoxidative Zinc-dependent 2,6-Dihydroxybenzoate (γ-Resorcylate) Decarboxylase from Rhizobium sp. Strain MTP-10005. *Journal of Biological Chemistry* **281**, 34365-34373
52. Silverman, R. (2002) Organic Chemistry of Enzyme-Catalyzed Reactions, Revised Edition. 2 Ed., Elsevier
53. Martynowski, D., Eyobo, Y., Li, T., Yang, K., Liu, A., and Zhang, H. (2006) Crystal Structure of α-Amino-β-carboxymuconate-ε-semialdehyde Decarboxylase: Insight into the Active Site and Catalytic Mechanism of a Novel Decarboxylation Reaction†,‡. *Biochemistry* **45**, 10412-10421
54. Liu, A., and Zhang, H. (2006) Transition Metal-Catalyzed Nonoxidative Decarboxylation Reactions†. *Biochemistry* **45**, 10407-10411
55. Yew, W. S., Wise, E. L., Rayment, I., and Gerlt, J. A. (2004) Evolution of Enzymatic Activities in the Orotidine 5‘-Monophosphate Decarboxylase Suprafamily: Mechanistic Evidence for a Proton Relay System in the Active Site of 3-Keto-l-gulonate 6-Phosphate Decarboxylase†. *Biochemistry* **43**, 6427-6437
56. Okrasa, K., Levy, C., Hauer, B., Baudendistel, N., Leys, D., and Micklefield, J. (2008) Structure and Mechanism of an Unusual Malonate Decarboxylase and Related Racemases. *Chemistry – A European Journal* **14**, 6609-6613
57. Poelarends, G. J., Johnson, W. H., Murzin, A. G., and Whitman, C. P. (2003) Mechanistic Characterization of a Bacterial Malonate Semialdehyde Decarboxylase: Identification of a new activity in the tautomerase superfamily. *Journal of Biological Chemistry* **278**, 48674-48683
58. Chan, K. K., Wood, B. M., Fedorov, A. A., Fedorov, E. V., Imker, H. J., Amyes, T. L., Richard, J. P., Almo, S. C., and Gerlt, J. A. (2009) Mechanism of the Orotidine 5′ -Monophosphate Decarboxylase-Catalyzed Reaction: Evidence for Substrate Destabilization. *Biochemistry* **48**, 5518-5531
59. Benning, M. M., Haller, T., Gerlt, J. A., and Holden, H. M. (2000) New Reactions in the Crotonase Superfamily: Structure of Methylmalonyl CoA Decarboxylase from Escherichia coli†,‡. *Biochemistry* **39**, 4630-4639
60. Rodríguez, H., Angulo, I., de las Rivas, B., Campillo, N., Páez, J. A., Muñoz, R., and Mancheño, J. M. (2010) p-Coumaric acid decarboxylase from Lactobacillus plantarum: Structural insights into the active site and decarboxylation catalytic mechanism. *Proteins: Structure, Function, and Bioinformatics* **78**, 1662-1676
61. Miyamoto, K., and Ohta, H. (1992) Purification and properties of a novel arylmalonate decarboxylase from Alcaligenes bronchisepticus KU 1201. *European Journal of Biochemistry* **210**, 475-481
62. Okrasa, K., Levy, C., Wilding, M., Goodall, M., Baudendistel, N., Hauer, B., Leys, D., and Micklefield, J. (2009) Structure-Guided Directed Evolution of Alkenyl and Arylmalonate Decarboxylases. *Angewandte Chemie International Edition* **48**, 7691-7694
63. Obata, R., and Nakasako, M. (2010) Structural Basis for Inverting the Enantioselectivity of Arylmalonate Decarboxylase Revealed by the Structural Analysis of the Gly74Cys/Cys188Ser Mutant in the Liganded Form. *Biochemistry* **49**, 1963-1969

64. Matoishi, K., Ueda, M., Miyamoto, K., and Ohta, H. (2004) Mechanism of asymmetric decarboxylation of α-aryl-α-methylmalonate catalyzed by arylmalonate decarboxylase originated from Alcaligenes bronchisepticus. *Journal of Molecular Catalysis B: Enzymatic* **27**, 161-168
65. Frank, A., Eborall, W., Hyde, R., Hart, S., Turkenburg, J. P., and Grogan, G. (2012) Mutational analysis of phenolic acid decarboxylase from Bacillus subtilis (BsPAD), which converts bio-derived phenolic acids to styrene derivatives. *Catalysis Science & Technology* **2**, 1568-1574
66. Gideon, G. (2002) β-Diketone hydrolases. *Journal of Molecular Catalysis B: Enzymatic* **19-20**, 73-82
67. Bugg, T. (2009) Enzymatic Hydrolysis and Group Transfer Reactions. in *Introduction to Enzyme and Coenzyme Chemistry*, Blackwell Publishing Ltd. pp 81-120
68. Horsman, G. P., Bhowmik, S., Seah, S. Y. K., Kumar, P., Bolin, J. T., and Eltis, L. D. (2007) The Tautomeric Half-reaction of BphD, a C-C Bond Hydrolase. *Journal of Biological Chemistry* **282**, 19894-19904
69. Henderson, I. M. J., and Bugg, T. D. H. (1997) Pre-Steady-State Kinetic Analysis of 2-Hydroxy-6-keto-nona-2,4-diene-1,9-dioic Acid 5,6-Hydrolase: Kinetic Evidence for Enol/Keto Tautomerization†. *Biochemistry* **36**, 12252-12258
70. Hamed, R., Batchelar, E., Clifton, I., and Schofield, C. (2008) Mechanisms and structures of crotonase superfamily enzymes – How nature controls enolate and oxyanion reactivity. *Cellular and Molecular Life Sciences* **65**, 2507-2527
71. Li, C., Hassler, M., and Bugg, T. D. H. (2008) Catalytic promiscuity in the alpha/beta-hydrolase superfamily: hydroxamic acid formation, C--C bond formation, ester and thioester hydrolysis in the C--C hydrolase family. *Chembiochem* **9**, 71-76
72. Siirola, E., Grischek, B., Clay, D., Frank, A., Grogan, G., and Kroutil, W. (2011) Tolerance of β-diketone hydrolases as representatives of the crotonase superfamily towards organic solvents. *Biotechnology and Bioengineering* **108**, 2815-2822
73. Ollis, D. L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S. M., Harel, M., Remington, S. J., Silman, I., Schrag, J., Sussman, J. L., Verschueren, K. H. G., and Goldman, A. (1992) The α/β hydrolase fold. *Protein Engineering* **5**, 197-211
74. Heikinheimo, P., Goldman, A., Jeffries, C., and Ollis, D. L. (1999) Of barn owls and bankers: a lush variety of alpha/beta hydrolases. *Structure* **7**, R141-146
75. Blow, D. M., Birktoft, J. J., and Hartley, B. S. (1969) Role of a Buried Acid Group in the Mechanism of Action of Chymotrypsin. *Nature* **221**, 337-340
76. Bugg, T. D. H. (2012) Hydrolytic and Group Transfer Enzymes. in *Introduction to Enzyme and Coenzyme Chemistry*, John Wiley & Sons, Ltd. pp 77-114
77. Harwood, C. S., and Parales, R. E. (1996) The β-ketoadipate pathway and the biology of self-identity. *Annual Review of Microbiology* **50**, 553-590
78. van der Meer, J. R., de Vos, W. M., Harayama, S., and Zehnder, A. J. (1992) Molecular mechanisms of genetic adaptation to xenobiotic compounds. *Microbiological Reviews* **56**, 677-694

79. D. H. Bugg, T., and J. Winfield, C. (1998) Enzymatic cleavage of aromatic rings: mechanistic aspects of the catechol dioxygenases and later enzymes of bacterial oxidative cleavage pathways. *Natural Product Reports* **15**, 513-530
80. Lam, W. W. Y., and Bugg, T. D. H. (1997) Purification, Characterization, and Stereochemical Analysis of a C−C Hydrolase: 2-Hydroxy-6-keto-nona-2,4-diene-1,9-dioic Acid 5,6-Hydrolase. *Biochemistry* **36**, 12242-12251
81. Fleming, S. M., Robertson, T. A., Langley, G. J., and Bugg, T. D. H. (2000) Catalytic Mechanism of a C−C Hydrolase Enzyme: Evidence for a Gem-Diol Intermediate, Not an Acyl Enzyme†. *Biochemistry* **39**, 1522-1531
82. Horsman, G. P., Ke, J., Dai, S., Seah, S. Y. K., Bolin, J. T., and Eltis, L. D. (2006) Kinetic and Structural Insight into the Mechanism of BphD, a C−C Bond Hydrolase from the Biphenyl Degradation Pathway†. *Biochemistry* **45**, 11071-11086
83. Ruzzini, A. C., Ghosh, S., Horsman, G. P., Foster, L. J., Bolin, J. T., and Eltis, L. D. (2012) Identification of an Acyl-Enzyme Intermediate in a meta-Cleavage Product Hydrolase Reveals the Versatility of the Catalytic Triad. *Journal of the American Chemical Society* **134**, 4615-4624
84. Grogan, G. (2005) Emergent mechanistic diversity of enzyme-catalysed beta-diketone cleavage. *Biochemical Journal* **388**, 721-730
85. Pokorny, D., Steiner, W., and Ribbons, D. W. (1997) β-Ketolases-forgotten hydrolytic enzymes? *Trends in Biotechnology* **15**, 291-296
86. Fujii, I., Yasuoka, Y., Tsai, H.-F., Chang, Y. C., Kwon-Chung, K. J., and Ebizuka, Y. (2004) Hydrolytic Polyketide Shortening by Ayg1p, a Novel Enzyme Involved in Fungal Melanin Biosynthesis. *Journal of Biological Chemistry* **279**, 44613-44620
87. Schleberger, C., Sachelaru, P., Brandsch, R., and Schulz, G. E. (2007) Structure and Action of a CC Bond Cleaving α/β-Hydrolase Involved in Nicotine Degration. *Journal of Molecular Biology* **367**, 409-418
88. Sachelaru, P., Schiltz, E., Igloi, G. L., and Brandsch, R. (2005) An α/β-Fold C—C Bond Hydrolase Is Involved in a Central Step of Nicotine Catabolism by Arthrobacter nicotinovorans. *Journal of Bacteriology* **187**, 8516-8519
89. Shimao, M., Tamogami, T., Kishida, S., and Harayama, S. (2000) The gene pvaB encodes oxidized polyvinyl alcohol hydrolase of Pseudomonas sp. strain VM15C and forms an operon with the polyvinyl alcohol dehydrogenase gene pvaA. *Microbiology* **146**, 649-657
90. Timm, D. E., Mueller, H. A., Bhanumoorthy, P., Harp, J. M., and Bunick, G. J. (1999) Crystal structure and mechanism of a carbon–carbon bond hydrolase. *Structure* **7**, 1023-1033
91. Bateman, R. L., Ashworth, J., Witte, J. F., Baker, L. J., Bhanumoorthy, P., Timm, D. E., Hurley, T. D., Grompe, M., and McClard, R. W. (2007) Slow-onset inhibition of fumarylacetoacetate hydrolase by phosphinate mimics of the tetrahedral intermediate: kinetics, crystal structure and pharmacokinetics. *Biochemical Journal* **402**, 251-260
92. Bateman, R. L., Bhanumoorthy, P., Witte, J. F., McClard, R. W., Grompe, M., and Timm, D. E. (2001) Mechanistic Inferences from the Crystal Structure of Fumarylacetoacetate Hydrolase with a Bound Phosphorus-based Inhibitor. *Journal of Biological Chemistry* **276**, 15284-15291
93. (2012) Fluorescence Spectra Analysis in Diamond MX Village PyMca User Guide. (Source, D. L. ed., Didcot

94. Bottiglieri, M., and Keel, C. (2006) Characterization of PhlG, a hydrolase that specifically degrades the antifungal compound 2,4-diacetylphloroglucinol in the biocontrol agent Pseudomonas fluorescens CHA0. *Applied and environmental microbiology* **72**, 418-427
95. He, Y.-X., Huang, L., Xue, Y., Fei, X., Teng, Y.-B., Rubin-Pitel, S. B., Zhao, H., and Zhou, C.-Z. (2010) Crystal Structure and Computational Analyses Provide Insights into the Catalytic Mechanism of 2,4-Diacetylphloroglucinol Hydrolase PhlG from Pseudomonas fluorescens. *Journal of Biological Chemistry* **285**, 4603-4611
96. Schoefer, L., Braune, A., and Blaut, M. (2004) Cloning and expression of a phloretin hydrolase gene from Eubacterium ramulus and characterization of the recombinant enzyme. *Applied & Environmental Microbiology* **70**, 6131-6137
97. Grogan, G., Graf, J., Jones, A., Parsons, S., Turner, N. J., and Flitsch, S. L. (2001) An Asymmetric Enzyme-Catalyzed Retro-Claisen Reaction for the Desymmetrization of Cyclic β-Diketones. *Angewandte Chemie International Edition* **40**, 1111-1114
98. Grogan, G., Roberts, G. A., Bougioukou, D., Turner, N. J., and Flitsch, S. L. (2001) The Desymmetrization of Bicyclic β-Diketones by an Enzymatic Retro-Claisen Reaction: A new reaction of the crotonase superfamily. *Journal of Biological Chemistry* **276**, 12565-12572
99. Whittingham, J. L., Turkenburg, J. P., Verma, C. S., Walsh, M. A., and Grogan, G. (2003) The 2-Å Crystal Structure of 6-Oxo Camphor Hydrolase: New structural diversity in the crotonase superfamily. *Journal of Biological Chemistry* **278**, 1744-1750
100. Bennett, J. P., Whittingham, J. L., Brzozowski, A. M., Leonard, P. M., and Grogan, G. (2006) Structural Characterization of a β-Diketone Hydrolase from the Cyanobacterium Anabaena sp. PCC 7120 in Native and Product-Bound Forms, a Coenzyme A-Independent Member of the Crotonase Suprafamily†,‡. *Biochemistry* **46**, 137-144
101. Leonard, P. M., and Grogan, G. (2004) Structure of 6-Oxo Camphor Hydrolase H122A Mutant Bound to Its Natural Product, (2S,4S)-α-Campholinic Acid: Mutant structure suggests an atypical mode of transition state binding for a crotonase homolog. *Journal of Biological Chemistry* **279**, 31312-31317
102. Li, C., Hassler, M., and Bugg, T. D. H. (2008) Catalytic Promiscuity in the α/β-Hydrolase Superfamily: Hydroxamic Acid Formation, C C Bond Formation, Ester and Thioester Hydrolysis in the C C Hydrolase Family. *ChemBioChem* **9**, 71-76
103. Aslanidis, C., and de Jong, P. J. (1990) Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Research* **18**, 6069-6074
104. Fogg M, W. A. (2008) - Higher-throughput approaches to crystallization and crystal structure determination. *Biochemical Society Transactions* **36**, 771-775
105. Clontech. (2012) In-Fusion® HD Cloning Kit User Manual. (Clontech ed.
106. Doherty, A. J., Ashford, S. R., Brannigan, J. A., and Wigley, D. B. (1995) A superior host strain for the over-expression of cloned genes using the T7 promoter based vectors. *Nucleic Acids Research* **23**, 2074-2075
107. Novagen. (2003) pET System Manual. (Novagen ed.

108. Brinkmann, U., Mattes, R. E., and Buckel, P. (1989) High-level expression of recombinant genes in Escherichia coli is dependent on the availability of the dnaY gene product. *Gene* **85**, 109-114
109. Wagner, S., Klepsch, M. M., Schlegel, S., Appel, A., Draheim, R., Tarry, M., Högbom, M., van Wijk, K. J., Slotboom, D. J., Persson, J. O., and de Gier, J.-W. (2008) Tuning Escherichia coli for membrane protein overexpression. *Proceedings of the National Academy of Sciences* **105**, 14371-14376
110. David Vikström, M. K., Samuel Wagner, James C. Samuelson and Jan-Willem de Gier. (2011) E. coli Lemo21(DE3) A T7 RNA Polymerase-based protein overexpression platform for routine and difficult targets.
111. Agilent. (2012) ArcticExpress Competent Cells and ArcticExpress (DE3) Competent Cells.
112. Ferrer, M., Chernikova, T. N., Yakimov, M. M., Golyshin, P. N., and Timmis, K. N. (2003) Chaperonins govern growth of Escherichia coli at low temperatures. *Nature Biotechnology* **21**, 1266-1267
113. Porath, J., Carlsson, J. A. N., Olsson, I., and Belfrage, G. (1975) Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature* **258**, 598-599
114. Bornhorst, J. A., and Falke, J. J. (2000) Purification of proteins using polyhistidine affinity tags. in *Methods in Enzymology* (Jeremy Thorner, S. D. E. J. N. A. ed.), Academic Press. pp 245-254
115. GE. (2010) Gel Filtration Principles and Methods. (GE-Healthcare ed.
116. Lindwall, G., Chau, M.-F., Gardner, S. R., and Kohlstaedt, L. A. (2000) A sparse matrix approach to the solubilization of overexpressed proteins. *Protein Engineering* **13**, 67-71
117. Cacace, M. G., Landau, E. M., and Ramsden, J. J. (1997) The Hofmeister series: salt and solvent effects on interfacial phenomena. *Quarterly Reviews of Biophysics* **30**, 241-277
118. Walls, D., and Loughran, S. (2011) Tagging Recombinant Proteins to Enhance Solubility and Aid Purification. in *Protein Chromatography* (Walls, D., and Loughran, S. T. eds.), Humana Press. pp 151-175
119. Kapust, R. B., and Waugh, D. S. (1999) Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Science* **8**, 1668-1674
120. Raran-Kurussi, S., and Waugh, D. S. (2012) The Ability to Enhance the Solubility of Its Fusion Partners Is an Intrinsic Property of Maltose-Binding Protein but Their Folding Is Either Spontaneous or Chaperone-Mediated. *PLoS ONE* **7**, e49589
121. Hutchison, C. A., Phillips, S., Edgell, M. H., Gillam, S., Jahnke, P., and Smith, M. (1978) Mutagenesis at a specific position in a DNA sequence. *Journal of Biological Chemistry* **253**, 6551-6560
122. Weiner, M. P., Costa, G. L., Schoettlin, W., Cline, J., Mathur, E., and Bauer, J. C. (1994) Site-directed mutagenesis of double-stranded DNA by the polymerase chain reaction. *Gene* **151**, 119-123
123. Stratagene. (2006) Quikchange Site-Directed mutagenesis Kit - Instruction Manual.
124. Johnson, K. A., and Goody, R. S. (2011) The Original Michaelis Constant: Translation of the 1913 Michaelis–Menten Paper. *Biochemistry* **50**, 8264-8269

125. Rogers, A., and Gibon, Y. (2009) Enzyme Kinetics: Theory and Practice. in *Plant Metabolic Networks* (Schwender, J. ed.), Springer New York. pp 71-103
126. Stryer, L. (1995) Enzymes: Basic Concepts and Kinetics. in *Biochemistry* (Stryer, L. ed.), W.H. Freeman and Co. pp 181-206
127. Owen, T. (2000) Fundamentals of modern UV-visible spectroscopy.
128. Waters. (2013) HPLC - High Performance Liquid Chromatography.
129. Hinshaw, J. (2006) The flame ionization detector in *Chromatography Online*
130. Agilent. (2010) Agilent 7890A Gas Chromatograph Operating Guide.
131. Rodriguez, H., Curiel, J. A., Landete, J. M., de las Rivas, B., Lopez de Felipe, F., Gomez-Cordoves, C., Mancheno, J. M., and Munoz, R. (2009) Food phenolics and lactic acid bacteria. *International Journal of Food Microbiology* **132**, 79-90
132. Dai, J., and Mumper, R. J. (2010) Plant Phenolics: Extraction, Analysis and Their Antioxidant and Anticancer Properties. *Molecules* **15**, 7313-7352
133. Barthelmebs, L., Divies, C., and Cavin, J. F. (2001) Expression in Escherichia coli of native and chimeric phenolic acid decarboxylases with modified enzymatic activities and method for screening recombinant E. coli strains expressing these enzymes. *Applied & Environmental Microbiology* **67**, 1063-1069
134. Gu, W., Yang, J., Lou, Z., Liang, L., Sun, Y., Huang, J., Li, X., Cao, Y., Meng, Z., and Zhang, K.-Q. (2011) Structural Basis of Enzymatic Activity for the Ferulic Acid Decarboxylase (FADase) from <italic>Enterobacter</italic> sp. Px6-4. *PLoS ONE* **6**, e16262
135. Hashidoko, Y., and Tahara, S. (1998) Stereochemically specific proton transfer in decarboxylation of 4-hydroxycinnamic acids by 4-hydroxycinnamate decarboxylase from Klebsiella oxytoca. *Archives of Biochemistry & Biophysics* **359**, 225-230
136. Bennett, J. P., Bertin, L., Moulton, B., Fairlamb, I. J. S., Brzozowski, A. M., Walton, N. J., and Grogan, G. (2008) A ternary complex of hydroxycinnamoyl-CoA hydratase-lyase (HCHL) with acetyl-CoA and vanillin gives insights into substrate specificity and mechanism. *Biochemical Journal* **414**, 281-289
137. van den Heuvel, R. H. H., van den Berg, W. A. M., Rovida, S., and van Berkel, W. J. H. (2004) Laboratory-evolved Vanillyl-alcohol Oxidase Produces Natural Vanillin. *Journal of Biological Chemistry* **279**, 33492-33500
138. Narbad, A., and Gasson, M. J. (1998) Metabolism of ferulic acid via vanillin using a novel CoA-dependent pathway in a newly-isolated strain of Pseudomonas fluorescens. *Microbiology* **144**, 1397-1405
139. van den Heuvel, R. H., Fraaije, M. W., Mattevi, A., and van Berkel, W. J. (2000) Asp-170 is crucial for the redox properties of vanillyl-alcohol oxidase. *Journal of Biological Chemistry* **275**, 14799-14808
140. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005) Protein identification and analysis tools on the ExPASy server. in *The Proteomics Protocols Handbook* (Walker, J. M. ed.), Humana Press. pp 571-607
141. Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., Grosdidier, A., Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti, R., Moretti, S.,

Mostaguir, K., Redaschi, N., Rossier, G., Xenarios, I., and Stockinger, H. (2012) ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research* **40**, W597-W603
142. Eborall, W. (2009) Structural and Mechanistic Studies of Hydroxycinnamic Acid Decarboxylase from Bacillus Subtilis. University of York, York
143. Bugg, T. (2009) Enzymes are Wonderful Catalysts. in *Introduction to Enzyme and Coenzyme Chemistry*, Blackwell Publishing Ltd. pp 29-50
144. Matte, A., Grosse, S., Bergeron, H., Abokitse, K., and Lau, P. C. K. (2010) Structural analysis of Bacillus pumilus phenolic acid decarboxylase, a lipocalin-fold enzyme. *Acta Crystallographica Section F* **66**, 1407-1414
145. Galdiero, S., Galdiero, M., and Pedone, C. (2007) beta-Barrel membrane bacterial proteins: structure, function, assembly and interaction with lipids. *Current Protein and Peptide Science* **8**, 63-82
146. Flower, D. R., North, A. C. T., and Attwood, T. K. (1993) Structure and sequence relationships in the lipocalins and related proteins. *Protein Science* **2**, 753-761
147. Storch, J., and McDermott, L. (2009) Structural and functional analysis of fatty acid-binding proteins. *Journal of Lipid Research* **50**, S126-S131
148. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948
149. Lesley, S., and Wilson, I. (2005) Protein Production and Crystallization at the Joint Center for Structural Genomics. *Journal of Structural and Functional Genomics* **6**, 71-79
150. Brzozowski, A. M., and Walton, J. (2001) Clear strategy screens for macromolecular crystallization. *Journal of Applied Crystallography* **34**, 97-101
151. Newman, J., Egan, D., Walter, T. S., Meged, R., Berry, I., Ben Jelloul, M., Sussman, J. L., Stuart, D. I., and Perrakis, A. (2005) Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the PACT/JCSG+ strategy. *Acta Crystallographica Section D* **61**, 1426-1431
152. Winter, G. (2010) xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography* **43**, 186-190
153. Kabsch, W. (2010) XDS. *Acta Crystallographica Section D* **66**, 125-132
154. Vagin, A., and Teplyakov, A. (1997) MOLREP: an Automated Program for Molecular Replacement. *Journal of Applied Crystallography* **30**, 1022-1025
155. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallographica Section D* **53**, 240-255
156. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D* **60**, 2126-2132
157. Schuttelkopf, A. W., and van Aalten, D. M. F. (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallographica Section D* **60**, 1355-1363
158. Matthews, B. W. (1968) Solvent content of protein crystals. *Journal of Molecular Biology* **33**, 491-497
159. Kantardjieff, K. A., and Rupp, B. (2003) Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein–nucleic acid complex crystals. *Protein Science* **12**, 1865-1871

160. Potterton, E., McNicholas, S., Krissinel, E., Cowtan, K., and Noble, M. (2002) The CCP4 molecular-graphics project. *Acta Crystallographica Section D* **58**, 1955-1957
161. Dauter, Z. (1999) Data-collection strategies. *Acta Crystallographica Section D* **55**, 1703-1717
162. Diederichs, K., and Karplus, P. A. (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nature Structural Molecular Biology* **4**, 269-275
163. Weiss, M. (2001) Global indicators of X-ray data quality. *Journal of Applied Crystallography* **34**, 130-135
164. Middleton, E., Kandaswami, C., and Theoharides, T. C. (2000) The Effects of Plant Flavonoids on Mammalian Cells:Implications for Inflammation, Heart Disease, and Cancer. *Pharmacological Reviews* **52**, 673-751
165. Nowakowska, Z. (2007) A review of anti-infective and anti-inflammatory chalcones. *European Journal of Medicinal Chemistry* **42**, 125-137
166. Rahman, M. (2011) Chalcone: A Valuable Insight into the Recent Advances and Potential Pharmacological Activities. *Chemical Sciences Journal* **2011**
167. Williams, R. J., Spencer, J. P. E., and Rice-Evans, C. (2004) Flavonoids: antioxidants or signalling molecules? *Free Radical Biology and Medicine* **36**, 838-849
168. Hijova, E. (2006) Bioavailability of chalcones. *Bratisl Lek Listy* **107**, 80-84
169. Singh, N. (2012) Biological Potentials of Chalcones: A Review. *International Journal of Pharmaceutical & Biological Archives 2012 1298-1303* **3**, 1298-1303
170. Rezk, B. M., Haenen, G. R. M. M., van der Vijgh, W. J. F., and Bast, A. (2002) The antioxidant activity of phloretin: the disclosure of a new antioxidant pharmacophore in flavonoids. *Biochemical and Biophysical Research Communications* **295**, 9-13
171. Schneider, H., and Blaut, M. (2000) Anaerobic degradation of flavonoids by Eubacterium ramulus. *Archives of Microbiology* **173**, 71-75
172. Schneider, H., Schwiertz, A., Collins, M. D., and Blaut, M. (1999) Anaerobic transformation of quercetin-3-glucoside by bacteria from the human intestinal tract. *Archives of Microbiology* **171**, 81-91
173. Moore, W. E., and Holdeman, L. V. (1974) Human fecal flora: the normal flora of 20 Japanese-Hawaiians. *Applied Microbiology* **27**, 961-979
174. Chatterjee, A. K., and Gibbins, L. N. (1969) Metabolism of phloridzin by Erwinia herbicola: nature of the degradation products, and the purification and properties of phloretin hydrolase. *Journal of Bacteriology* **100**, 594-600
175. Minamikawa, T., Jayasankar, N. P., Bohm, B. A., Taylor, I. E., and Towers, G. H. (1970) An inducible hydrolase from Aspergillus niger, acting on carbon-carbon bonds, for phlorrhizin and other C-acylated phenols. *Biochemical Journal* **116**, 889-897
176. Siirola, E., Frank, A., Grogan, G., and Kroutil, W. (2013) C C Hydrolases for Biocatalysis. *Advanced Synthesis & Catalysis* **355**, 1677-1691
177. Studier, F. W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expression and Purification* **41**, 207-234
178. Yang, Z. R., Thomson, R., McNeil, P., and Esnouf, R. M. (2005) RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics* **21**, 3369-3376

179. Ward, J. J., McGuffin, L. J., Bryson, K., Buxton, B. F., and Jones, D. T. (2004) The DISOPRED server for the prediction of protein disorder. *Bioinformatics* **20**, 2138-2139
180. Kelley, L. A., and Sternberg, M. J. E. (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols* **4**, 363-371
181. Needleman, S. B., and Wunsch, C. D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology* **48**, 443-453
182. Long, F., Vagin, A. A., Young, P., and Murshudov, G. N. (2008) BALBES: a molecular-replacement pipeline. *Acta Crystallographica Section D* **64**, 125-132
183. Sheldrick, G. (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallographica Section D* **66**, 479-485
184. Solé, V. A., Papillon, E., Cotte, M., Walter, P., and Susini, J. (2007) A multiplatform code for the analysis of energy-dispersive X-ray fluorescence spectra. *Spectrochimica Acta Part B: Atomic Spectroscopy* **62**, 63-68
185. Leatherbarrow, R. J. (2009) GraFit Version 7. Erithacus Software Ltd., Horley, U.K.
186. McCall, K. A., Huang, C.-c., and Fierke, C. A. (2000) Function and Mechanism of Zinc Metalloenzymes. *The Journal of Nutrition* **130**, 1437S-1446S
187. Christianson, D. W., and Cox, J. D. (1999) Catalysis by metal-activated hydroxide in zinc and manganese metalloenzymes. *Annu. Rev. Biochem.* **68**, 33-57
188. Bugg, T. D. H. (2012) Methods for Studying Enzymatic Reactions. in *Introduction to Enzyme and Coenzyme Chemistry*, John Wiley & Sons, Ltd. pp 50-76
189. Shimao, M. (2001) Biodegradation of plastics. *Current Opinion in Biotechnology* **12**, 242-247
190. Yamatsu, A., Matsumi, R., Atomi, H., and Imanaka, T. (2006) Isolation and characterization of a novel poly(vinyl alcohol)-degrading bacterium, <i>Sphingopyxis</i> sp. PVA3. *Applied Microbiology and Biotechnology* **72**, 804-811
191. Sakazawa, C., Shimao, M., Taniguchi, Y., and Kato, N. (1981) Symbiotic Utilization of Polyvinyl Alcohol by Mixed Cultures. *Applied and environmental microbiology* **41**, 261-267
192. Shimao, M., Yamamoto, H., Ninomiya, K., Kato, N., Adachi, O., Ameyama, M., and Sakazawa, C. (1984) Pyrroloquinoline Quinone as an Essential Growth Factor for a Poly(vinyl alcohol)-degrading Symbiont, Pseudomonas sp. VM15C. *Agricultural and Biological Chemistry* **48**, 2873-2876
193. Yang, Y., Zhang, D., Liu, S., Jia, D., Du, G., and Chen, J. (2012) Expression and fermentation optimization of oxidized polyvinyl alcohol hydrolase in <i>E. coli</i>. *Journal of Industrial Microbiology & Biotechnology* **39**, 99-104
194. Kawagoshi, Y., and Fujita, M. (1997) Purification and properties of the polyvinyl alcohol-degrading enzyme 2,4-pentanedione hydrolase obtained from Pseudomonas vesicularis var. povalolyticus PH. *World Journal of Microbiology and Biotechnology* **14**, 95-100

195. Klomklang, W., Tani, A., Kimbara, K., Mamoto, R., Ueda, T., Shimao, M., and Kawai, F. (2005) Biochemical and molecular characterization of a periplasmic hydrolase for oxidized polyvinyl alcohol from Sphingomonas sp. strain 113P3. *Microbiology* **151**, 1255-1262
196. Hisano, T., Kasuya, K.-i., Tezuka, Y., Ishii, N., Kobayashi, T., Shiraki, M., Oroudjev, E., Hansma, H., Iwata, T., Doi, Y., Saito, T., and Miki, K. (2006) The Crystal Structure of Polyhydroxybutyrate Depolymerase from Penicillium funiculosum Provides Insights into the Recognition and Degradation of Biopolyesters. *Journal of Molecular Biology* **356**, 993-1004
197. Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8**, 785-786
198. Smyth, D. R., Mrozkiewicz, M. K., McGrath, W. J., Listwan, P., and Kobe, B. (2003) Crystal structures of fusion proteins with large-affinity tags. *Protein Science* **12**, 1313-1322
199. Gasson, M. J., Kitamura, Y., McLauchlan, W. R., Narbad, A., Parr, A. J., Parsons, E. L., Payne, J., Rhodes, M. J., and Walton, N. J. (1998) Metabolism of ferulic acid to vanillin. A bacterial gene of the enoyl-SCoA hydratase/isomerase superfamily encodes an enzyme for the hydration and cleavage of a hydroxycinnamic acid SCoA thioester. *Journal of Biological Chemistry* **273**, 4163-4170
200. Mang, H., Gross, J., Lara, M., Goessler, C., Schoemaker, H. E., Guebitz, G. M., and Kroutil, W. (2006) Biocatalytic Single-Step Alkene Cleavage from Aryl Alkenes: An Enzymatic Equivalent to Reductive Ozonization. *Angewandte Chemie International Edition* **45**, 5201-5203

Bibliography

The following publications have resulted from the work presented in this thesis:

Annika Frank, William Eborall, Ralph Hyde, Sam Hart, Johan P. Turkenburg and Gideon Grogan (2012) Mutational analysis of phenolic acid decarboxylase from *Bacillus subtilis* (BsPAD), which converts bio-derived phenolic acids to styrene derivatives, *Catalysis, Science and Technology*, **2**(8): 1568-1574.

Annika Frank, Elina Siirola, Wolfgang Kroutil, Gideon Grogan (2013) Mutational analysis of the C-C bond Cleaving Enzyme Phloretin Hydrolase from *Eubacterium ramulus*, Topics in Catalysis (*manuscript accepted*)

BIOTRAINS collaborations

In addition to the work described in this thesis, a number of collaborative projects have been carried out with other members of the Marie Curie BIOTRAINS network. This work involved the structural and mechanistic characterisation of various biocatalytically relevant enzymes such as a monoamine oxidase (MAO), a novel imine reductase (IR), various alcohol dehydrogenases as well as enzymes involved in carbon-carbon bond catalysis (some of which are described in this thesis).

In collaboration with Professor Nicholas Turner's group at the University of Manchester, the structure of a monoamine oxidase mutant from *Aspergillus niger* (MAO-N) was determined in 2011. This structure was part of large-scale protein engineering project in which a MAO-N library, prepared by directed evolution, was shown to oxidise a wide variety of primary, secondary and tertiary amines to give high value, chiral products. The crystallographic data on MAO-N D11C, in complex with its FAD cofactor, provided a structural insight into the observed substrate- and enantioselectivity of the mutant.

The work was published in the Journal of the American Chemical Society in 2013:

Diego Ghislieri, Anthony Green, Marta Pontini, Simon Willies, Ian Rowles, Annika Frank, Gideon Grogan, Nicholas Turner (2013) Engineering an enantioselective amine oxidase for the synthesis of pharmaceutical building blocks and alkaloid natural products, Journal of the American Chemical Society

Another collaboration with the University of Manchester as well as with the University of Oviedo involved the characterisation of a novel imine reductase from *Streptomyces kanamyceticus*. The enzyme catalyses the reduction of the asymmetric imine 2-methyl-1-pyrroline to *(R)*-2-methylpyrrolidine and is thus a promising catalyst for the production of valuable chiral amines. During an exchange visit of Dr. Maria Rodriguez Mata, the imine reductase gene was cloned into the YSBLIC3C vector, the recombinant protein purified to high homogeneity and screened against a variety of crystallisation conditions.
After structure determination and mechanistic analysis of two active site point mutants by Elisabeth Wells, the results were published in the ChemBioChem Journal in 2013:

Maria Rodriguez Mata, Annika Frank, Elizabeth Wells, Friedemann Leipold, Nicholas J.Turner, Sam Hart, Johan P. Turkenburg and Gideon Grogan (2013) Structure and activity of NADPH- dependent (R)-selective imine reductase Q1EQE0 from *Streptomyces kanamyceticus*,ChemBioChem.

A long-term collaboration, throughout the course of this PhD project, involved work with Professor Wolfgang Kroutil's group from the University of Graz. In particular, the work on carbon-carbon hydrolases, described in chapters 5 and 6 of this thesis, was carried out in close cooperation with Dr. Elina Siirola. This also involved two exchange visits between York and Graz. This collaborative work is included in a review published in the journal Advanced Synthesis and Catalysis in 2013:

Elina Siirola, Annika Frank, Gideon Grogan, Wolfgang Kroutil (2013) C-C hydrolases for Biocatalysis, Review, Advanced Synthesis & Catalysis, 355(9): 1677-1691

In addition to the work described above, the exchange visits of two Marie Curie Early Stage Reasearchers, Dr. Justyna Kulig and Kinga Kedziora, involved projects on the structural and mechanistic characterization of alcohol dehydrogenases. Their work has been submitted for publication.