Ivan Hajnal

Characterisation, Identification of Novel Functions and Engineering of Substrate Specificity of Cupin Superfamily Enzymes

DISSERTATION

Zur Erlangung des Akademischen Grades eines

Doktors der Naturwissenschaften, Dr.rer.nat.

An der Technischen Universität Graz

Durchgeführt am Institut für Molekulare Biotechnologie

in den Laboren der ACIB GmbH

unter der Betreuung von

Univ.-Prof. Dipl.-Ing. Dr.techn. Helmut Schwab

Graz, 2013

Deutsche Fassung:

EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, am ……………………………   …………………………………………………..
                                                                              (Unterschrift)

Englische Fassung:

STATUTORY DECLARATION

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

……………………………   …………………………………………..
          date   (signature)

ABSTRACT

This thesis takes a broad perspective at proteins belonging to the cupin superfamily towards their potential for the discovery and engineering of novel, useful biocatalytic properties. Starting from available literature, database searches, and previous discoveries made in our group, a set of interesting cupins has been characterised in detail.

The thesis contributes the first detailed biochemical characterisation of a cupin-type hydroxynitrile lyase. The enzyme was shown to be metal-dependent and to prefer manganese. This constitutes the first description of a manganese-dependent hydroxynitrile lyase. In addition to the establishment of the metal dependence, the enzyme has been characterised through site-directed mutagenesis and residues important for activity but not for metal-binding and folding or stability have been identified. The thesis thus contributes work towards solving the reaction mechanism of this novel and unusual class of hydroxynitrile lyases which are dissimilar to known enzymes which catalyse the same reaction.

In addition to this characterisation of an enzymatic activity known previously in cupins, a part of this thesis constitutes the discovery and identification of an interesting oxidative activity in a cupin from *Thermotoga maritima* which was not known in cupins earlier. Since no biochemical function was known for this cupin to date, this thesis provides new knowledge on the catalytic diversity found within cupin type proteins.

 A further part of the thesis concerns the biochemical and mutational analysis of two cupin ring –cleavage dioxygenases active on highly similar aminohydorxybenzoic acid substrates, but neither of which has detectable activity towards the substrate of the other. The analysis has yielded valuable information towards a rational understanding of the very high substrate specificity in these cupin dioxygenases. This work thus contributes the engineering of a highly-substrate specific cupin dioxygenase towards a broadened substrate scope.

KURZZUSAMMENFASSUNG

Diese Dissertation nimmt einen weiten Blick auf Proteine aus der Cupin Überfamilie in Hinblick auf deren Potential für die Entdeckung und Engineering von neuartigen, nützlichen biokatalytischen Eigenschaften. Ausgehend von vorhandener Literatur, Suchen in Datenbanken, und früheren Entdeckungen aus unserer Gruppe, wurde ein Set aus interessanten Cupinen im Detail Charakterisiert.

Die Dissertation enthält die erste detaillierte biochemische Charakterisierung einer Cupin-Hydroxynitrillyase. Es konnte gezeigt werden, dass das Enzym Metall-abhängig ist mit einer Bevorzugung für Mangan. Dies ist die erste Beschreibung einer Mangan-abhängigen Hydroxynitrillyase. Zusätzlich zur Darstellung der Metallabhängigkeit, wurde das Enzym durch zielgerichtete Mutagenese charakterisiert, und Aminosäurereste welche für die Aktivität, aber nicht für die Metallbindung und Faltung oder Stabilität wichtig sind, wurden identifiziert. Die Dissertation trägt daher bei zur Lösung des Reaktionsmechanismus dieser interessanten neuartigen Klasse von Hydroxynitrillyasen welche keinerlei Ähnlichkeit zu bekannten Enzymen welche die gleiche Reaktion katalysieren aufweisen.

Zusätzlich zu dieser Charakterisierung einer Enzymaktivität welche bereits in Cupinen bekannt war, behandelt ein Teil dieser Dissertation die Entdeckung und Identifizierung einer interessanten oxydativen Aktivität in einem Cupin aus *Thermotoga maritima*, welche bisher in Cupinen nicht bekannt war. Da für dieses Cupin keine biochemische Funktion bekannt war, bereitet diese Dissertation neues Wissen über die Vielfältigkeit von katalytischen Eigenschaften der Cupine.

Ein weiterer Teil dieser Dissertation beschäftigt sich mit der biochemischen und mutationellen Analyse von zwei Cupinen mit Ring-spaltender Dioxygenaseaktivität welche an zwei sehr ähnlichen Aminohydroxybenzoesäure Substraten aktiv sind, wobei keines von beiden detektierbare Aktivität mit dem natürlichen Substrat des jeweils anderen aufweist. Die Analyse brachte wertvolle Informationen in Richtung eines rationalen Verständnisses der sehr hohen Substratspezifität in den besagten Cupin Dioxygenasen. Diese Arbeit beinhaltet somit das Engineering einer sehr substrat-spezifischen Cupin Dioxygenase in Richtung eines verbreiterten Substratspektrums.

CONTENTS

VI

VIII

1. INTRODUCTION

1.1 CUPINS

Cupins comprise a large superfamily of very diverse small β-barrel proteins (Agarwal et al., 2009; Dunwell, 1998; Khuri et al., 2001; Uberto and Moomaw, 2013). They include enzymatic and non-enzymatic members. The non-enzymatic members include plant storage proteins such as vicilins, and small-molecule binding proteins such as auxin-binding and sucrose-binding proteins, also found in plants (Dunwell et al., 2004). Enzymatic members are diverse and include isomerases, oxidases, decarboxylases, dioxygenases and hydroxynitrile lyases (Dunwell et al., 2004; Hussain et al., 2012). Many cupins are metalloproteins, which coordinate a divalent metal cation inside the typical cupin-type β-barrel. The metal may or may not be directly involved in enzymatic activity of the cupins in which one is found, and not all cupins are metalloproteins (Agarwal et al., 2009; Uberto and Moomaw, 2013). In recent years, cupins have gained increased attention from the scientific community, and there is possibly a quite considerable potential for their biotechnological application. This potential is due to their small size and easy handling of the encoding genes, very high stability in some cases, and a range of interesting activities found (Dunwell et al., 2004). As an indication of this diversity, discoveries of cupins with activities not known in this fold before have been reported with some regularity (Hussain et al., 2012; Jaroszewski et al., 2004; Straganz et al., 2003), and there are recent reports of promising promiscuous activities in cupins (Leitgeb and Nidetzky, 2010). In fact, it has been proposed that cupins might be one of the most functionally diverse protein superfamilies known so far (Dunwell et al., 2004). These findings imply a possibly even richer repertoire for the future biotechnological utilisation of proteins from this fascinating, very versatile protein superfamily.

1.1.1 History and Discovery

Originally, cupins have been annotated as a protein superfamily starting from the sequence of an unusual non-flavin oxalate oxidase from plants, termed germin (Dunwell, 1998). The discovery of a number of so called germin-like proteins had preceded the discovery of cupins and the significance of this discovery was not fully understood at the time. The discovery of this group of germin-like proteins led to the designation of a "germin box" sequence (PH(T/I)HPRATE) (Lane et al., 1991; Dunwell and Gane, 1998). The germin-box in turn was found by comparing these early sequences of germin-like proteins to spherulins, a group of storage proteins from plants without known enzymatic activity (Lane et al., 1991). Following an increase in the number of available sequences at the time, Dunwell and Gane were able to show that the germin box is actually part of a larger motif - $G(X)_5HXH(X)_{11}G$, which in these proteins was followed after usually 15AA by a second motif - $G(X)_5P(X)_4H(X)_3N$ (Dunwell and Gane, 1998). Already in this seminal study they were able to annotate members with many different enzymatic and non-enzymatic functions to the nascent protein superfamily. They have also discovered two-domain proteins where these two motifs were both found together in each of the two domains, indicating a gene-duplication event. Proteins of this two-domain type include an oxalate decarboxylase from *Collybia velutipes*, which was proposed to be a duplication of an ancestor highly similar its most similar sister protein, the single-domain cupin oxalate oxidase. Additional proteins also annotated as cupins during this initial study included auxin-binding proteins, epimerases involved in cell wall synthesis, a phosphomannose isomerase and a class of iron-containing eukaryotic dioxygenases (Dunwell and Gane, 1998). These first members contained additional motifs in addition to what would later be recognised as the typical cupin-type β-barrel. Afterwards, members of the new fold were discovered which were smaller than the initially discovered representatives and did not contain any α-helix elements. Since these newly discovered cupins were not much more than a singular small β-barrel, Dunwell proposed to name the new proteins cupins, after the Latin word cupa which means a small barrel (Dunwell, 1998). Interestingly, already at that time it was apparent that many additional proteins containing both cupin motifs may be found among progressively smaller hypothetical or cryptic genes possibly encoding proteins with unknown function. Thus the data indicated from the beginning that a potentially very large novel superfamily has been discovered (Dunwell, 1998).

1.1.2 Structure

As the story of their discovery indicates, the very name of the cupin superfamily is owed to their barrel-like structure. Thus, cupins are small, predominantly β-barrel proteins. In the structural classification of proteins database (SCOP, Andreeva et al., 2007; Murzin et al., 1995) they are classified as the RmlC-like cupins superfamily [51182] within the double-stranded β-helix fold. Currently 24 families are recognised within this superfamily in the SCOP database. The namesake protein of this superfamily, RmlC, is a dTDP-6-deoxy-d-xylo-4-hexulose 3, 5 epimerase (EC 5.1.3.13), and its structure has been solved at a time when the cupin superfamily was a very recent discovery (Giraud et al., 2000). RmlC displays a jelly-roll fold, as is illustrated by the topology diagram in Figure 1a, with a barrel structure composed of two β-sheets arranged as a β sandwich, as is illustrated in the cartoon depiction of the cupin barrel of RmlC shown in Figure 1b. The barrel contains the active site, and both β-sheets contribute active-site residues (Giraud et al., 2000). In this regard RmlC is typical and highly representative of cupins.

A   **B**
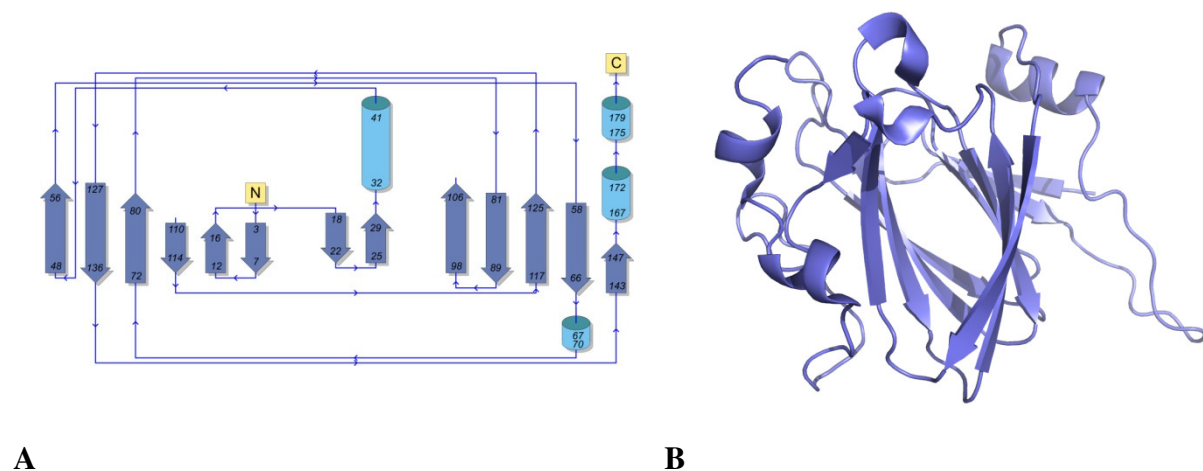
Figure 1: A: Topology diagram of RmlC showing the jelly roll-like topology typical of cupins. The topology diagram was generated using PDBsum (Laskowski et al., 1997). **B:** A monomer of RmlC showing the typical cupin β-barrel structure. Figure generated using PyMOL from PDB entry 1DZR.

Somewhat less typically, there seem to be no indications of a metal in the RmlC barrel (Giraud et al., 2000), whereas many cupins contain a metal at the centre of the barrel (Agarwal et al., 2009; Uberto and Moomaw, 2013). In other cupins, the metal bound in the barrel is often involved in the catalytic activity, or at least contributes to ligand binding in non-enzymatic members (Agarwal et al., 2009; Uberto and Moomaw, 2013). In many cupins, including germin, which was among first to be discovered, two histidines and a glutamate from the cupin motif 1 and a third histidine from cupin motif 2 bind an active site metal inside the cupin barrel (Dunwell et al., 2004). It should be noted that these metal-binding residues are not strictly conserved, and may partly or entirely be missing. Concerning their tertiary structure, cupin barrels have been observed to form dimers such as RmlC, trimers such as in glycinine, or hexamers as in germin (Dunwell et al., 2004). In some cases they are present as single or multiple fused cupin domains as part of various multi-domain proteins (Dunwell et al., 2004).

1.2   FUNCTIONAL DIVERSITY IN THE CUPIN SUPERFAMILY

As a highest-order division of function one can differentiate cupin superfamily proteins which are enzymes, from such members which do not have an enzymatic function or whose enzymatic function is secondary to another function such as cellular signalling.

1.2.1   Cupins with a non-enzymatic main function

Many non-enzymatic members of the cupin superfamily are plant storage proteins such as vicilins, spherulins and proteins related to legumin (Bernier and Berna, 2001). Some cupins from this class are relevant not only as constituents of the human diet and animal feed, but also as a major class of food-borne allergens (Breiteneder and Mills, 2005; Radauer et al., 2008). Non-enzymatic cupins are of course not limited to plant storage proteins even if these are highly relevant to humans. Well-characterised examples of non-enzymatic cupins which are not storage proteins are found among auxin-binding proteins (Timpte, 2001). Auxin-binding proteins have important roles in plant development and seem to be widely distributed in plants, even though the exact mechanism of their biological function has not been determined yet (Timpte, 2001). A crystal structure of auxin-binding protein 1 from *Zea mays*

has been determined in complex with auxin, and it shows auxin bound in the cupin barrel, with its carboxylic acid moiety in direct contact with a zinc ion. The zinc ion in turn is coordinated by three histidines and one glutamate from the typical cupin-type motifs (Woo et al., 2002). While the auxin-binding proteins are putatively involved in cellular signalling without a clear mechanism described so far, in some cases a cupin domain is part of a classical DNA-binding receptor protein. In fact, the cupin fold has been shown to be one of only a handful of very ancient protein folds from which the great majority of small molecule binding domains in the very important helix-turn-helix family of transcription factors have arisen (Aravind et al., 2005). In other cases, the cupin domain is part of a protein which couples an enzymatic function to a role in cellular signalling. The oxygen-responsive peptide-modifying dioxygenases are such an example (Iyer et al., 2010). An example of these are the JmjC jumonji-like family of transcription regulators, and for many of them cupin-type architecture has been deduced from sequence similarity (Clissold and Ponting, 2001). An example of similar cupins involved in signalling in humans is the nuclear protein pirin, and for it cupin architecture has been ascertained based on a crystal structure (Pang et al., 2003). Another interesting example is the hypoxia-inducible factor (HIF), which partakes in the signalling pathway induced by lack of oxygen (Hewitson et al., 2002). This enzyme has been shown conclusively to belong to the cupin family, and its enzymatic function involves the hydroxylation of specific asparaginyl and prolyl residues in a 2-oxoglutarate dependent fashion, which in turn leads to the transmission of the cellular signal (Hewitson et al., 2002). This is only a short view of the scope of diversity found in cupins whose main function is non-enzymatic. While these proteins are diverse and fascinating in their own right, perhaps most cupins are simply metabolic or defensive enzymes (Dunwell et al., 2004), and these are the ones which are most directly relevant for the scope of this thesis.

1.3 ENZYMES OF THE CUPIN SUPERFAMILY

The following chapter introduces the diversity of enzymatic activities so far described in cupins and offers a view of the potential within cupins for the engineering and utilisation of interesting, unusual and possibly even completely novel enzymes.

1.3.1   Oxalate degrading cupins – oxidases and decarboxylases

Among the first assigned cupin proteins, which led to the discovery of cupins as a superfamily, was a fairly unusual enzyme from plants termed germin (Dunwell, 1998). Germin is a fascinating protein which had been studied for a long time as a marker of early plant development. It oxidises oxalic acid at the expense of molecular oxygen, and produces hydrogen peroxide in the process (Lane et al., 1993). The protein has been shown to be involved in plant defence, especially against fungal parasites. This harsh evolutionary task of fighting fungi is what likely led to the extreme stability of germin towards hydrogen peroxide and acidic pH. The stability of germin has been likened to that of prions (Lane, 1994). Even though early research indicated the possibility of germin being a flavoenzyme, it was later definitely shown that this is not the case, and manganese has been identified as the sole cofactor necessary for activity (Requena and Bornemann, 1999). As the crystal structure of germin illustrates (Figure 2), the high stability is at least partly due to a very tight packing as a homohexamer, which in turn is composed of three homodimers of two typical cupin barrels each. One manganese atom is bound at the centre of each cupin barrel in the crystal structure (Woo et al., 2000). Interestingly, in addition to oxalate oxidase activity, germin was shown to possess a superoxide dismutase side-activity which represents one of a handful of cases of promiscuous activity found in cupins (Woo et al., 2000).
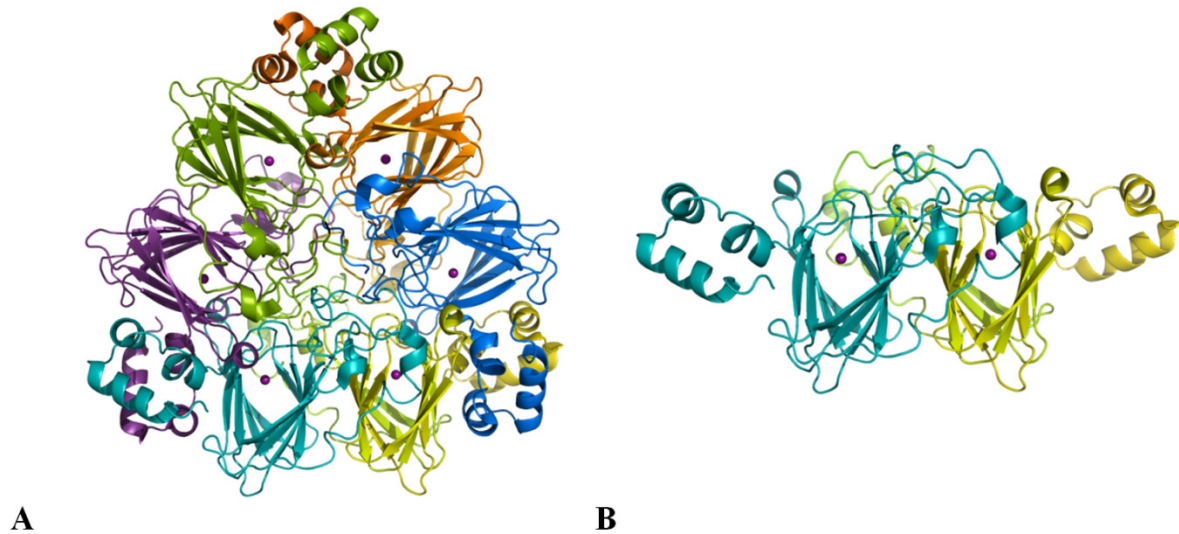
Figure 2: A: An illustration of the tight germin homohexamer comprising six cupin barrels (coloured cartoon depiction), with one manganese atom at the centre of each barrel (purple spheres). The full homohexamer is a trimer of homodimers. **B:** A single germin homodimer extracted from the hexameric structure. Figure generated using PyMOL from PDB entry 1FI2.

From a broader biotechnology perspective, oxalate oxidase has been used to generate transgenic plants with an increased resistance to oxalate-producing fungal parasites (Dias et al., 2006; Dong et al., 2008). On the other hand, the isolated enzyme is finding applications in the paper pulp industry as a means to reduce oxalate content of wastewater. This is done in order to help reduce scaling and other problems associated with oxalate accumulation in pulping plants (Larsson et al., 2003). In a somewhat more specialised, niche application, oxalate oxidase has been used for a long time in kits for the highly specific determination of oxalate content in the urine of human patients (Li and Madappally, 1989). Both wheat and barley germins have been successfully overexpressed in *Pichia pastoris*, which might make their biotechnological applications even more broadly accessible in the future (Pan et al., 2007; Whittaker and Whittaker, 2002).

One group of cupin enzymes structurally closely related to germins are the oxalate decarboxylases. These enzymes decarboxylate oxalate, forming formate and CO_2. They are produced under acidic conditions, probably as a means to eliminate excess oxalate, for example by *Bacillus subtilis* (Tanner and Bornemann, 2000). For this class of enzymes it has

also been demonstrated that they rely on manganese for activity. Interestingly, even though the reaction performed on oxalate by the decarboxylases is not an oxidation per se, the oxalate decarboxylase from *Bacillus subtilis* nevertheless requires dioxygen for activity (Tanner et al., 2001). The crystal structure shows that the protein forms a hexamer very similar to that of oxalate oxidase. However, in this case two cupin domains are fused in each monomer which classifies this protein as a bicupin. This architecture is most likely a consequence of a gene duplication event (Anand et al., 2002). Both active sites within one bicupin monomer show very high similarity to each other in their metal-binding residues. Additionally, the oxalate decarboxylase itself shows very high similarity in and around the metal-binding site to oxalate oxidase. Due to the very similar overall active-site geometry, it was an interesting question just what determines the oxalate decarboxylase versus oxalate oxidase activity in these proteins. After a crystal structure was solved of a closed conformation of oxalate decarboxylase, the identity of the active site within the bicupin domain was determined. An active site lid and a loop residue important for a deprotonation implicated in the mechanism have been identified (Just, 2004). In fact, the importance of this active site lid and the similarity of oxalate decarboxylase to oxalate oxidase could be established experimentally by converting the oxalate decarboxylase to an oxidase by a specificity factor of up to 282 000, by mutating residues in the lid alone (Burrell et al., 2007). In biotechnology, genes encoding oxalate decarboxylase have been used similarly to ones encoding oxalate oxidase, to confer resistance against the fungus *Collybia velutipes* in transgenic tobacco and tomato (Kesarwani et al., 2000).

1.3.2   Cupin dioxygenases

The oxalate degrading cupins described above employ either an oxidative or a decarboxylase mechanism within a very similar metal-binding site. Of the two activities, the oxidative reaction is perhaps more typical of cupins, as many different oxidative enzymes are found within the cupin fold. One group with a large number of described members are the cupin dioxygenases, which catalyse the oxidation of diverse substrates by the incorporation of both atoms of a molecule of dioxygen (Fetzner, 2012). In some cases they are involved in the oxidation of substrates which are easily oxidisable, making a reaction which proceeds spontaneously even more efficient. On the other hand, some dioxygenases are involved in the biological degradation of aromatic compounds which are difficult to oxidise. The dioxygenase

step is the one in which the aromatic ring is cleaved, producing a non-aromatic substrate. This is significant since aromatic substances are often very stable due to the delocalised nature of their ring electrons (Vaillancourt et al., 2006). Due to the importance of aromatics degradation pathways in nature, cupins represent only one of a number of protein folds in which dioxygenase activities have emerged through convergent evolution. They often utilize a metal ion such as iron, copper, nickel or manganese in the active site and are present in all three domains of life (Fetzner, 2012; Vaillancourt et al., 2006). Even though cupins constitute only a part of the known ring-cleavage dioxygenases, a number of cupins with ring-cleaving activity have been described, and most of them appear to be active on substituted benzoic acids containing at least one hydroxy group (Fetzner, 2012). Representative cupin dioxygenases active on various aromatic and non-aromatic substrates are introduced in this chapter.

1.3.3 Acireductone dioxygenases

Acireductone dioxygenases (ARD) act on the easily oxidised substrate 1,2-dihydroxy-3-keto-5-(methylthio)pentene (Figure 3), also called acireductone. The dioxygenase reaction is part of the methionine salvage pathway from methyl thioadenosine, as has been described in *Klebsiella* (Dai et al., 2001). ARD enzymes display a very unusual behaviour in which the same protein backbone yields what may be viewed as two enzymes, possessing differing enzymatic activities, depending on the type of the metal bound in the active site. Both enzyme forms act on acireductone, whereby the iron-containing enzyme catalyses a 1,2 – dioxygenation, leading to a β-keto acid precursor of methionine (Figure 3), and the same protein backbone with nickel or cobalt in the active site, catalyses a 1,3 dioxygenation which yields an off-pathway product with unknown function, together with carbon monoxide and formic acid (Figure 3; Dai, 1999). Interestingly, it was possible to separate the two protein forms using ion exchange chromatography or hydrophobic-interaction chromatography (Dai, 1999). The first ARD structure to be solved was that of the Ni-containing enzyme. The structure was solved using protein NMR spectroscopy, and it was this data that demonstrated that the enzyme indeed belongs to the cupin superfamily (Pochapsky et al., 2002). In another study, the structures of the iron form and the nickel form of the enzyme have been solved by a combination of x-ray crystallographic analysis and NMR analysis, and it could be shown that there is a change in the protein structure depending on the bound metal. It has been postulated

that it is indeed this change in structure which leads to the change in the catalytic activity, by inducing a change in substrate orientation in the active site (Ju et al., 2006). The enzymes were first discovered in *Klebsiella*, which was used as a model organism in part because the complete methionine salvage pathway was known previously. On the other hand, ARDs are also found in mammals, and a crystal structure has been solved for the mouse ARD, which also displays the cupin fold (Sakai et al., 2006).
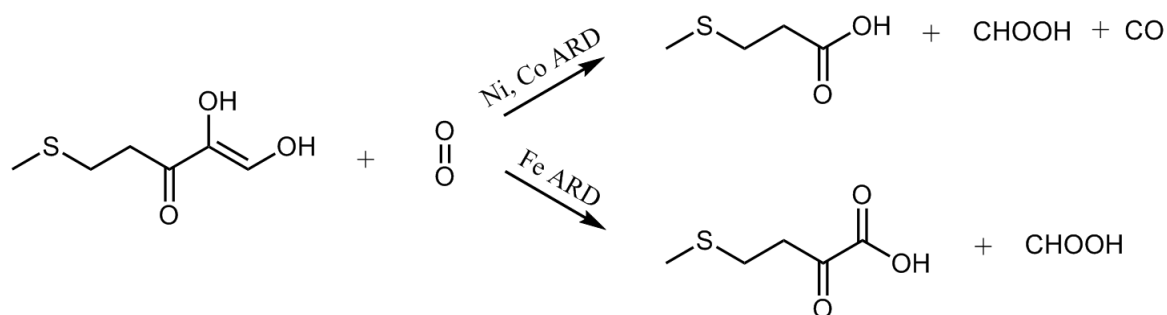
Figure 3: The acireductone dioxygenase reaction as is catalysed by the iron-containing enzyme yields a β-keto acid precursor of methionine and formic acid at the expense of molecular oxygen. The same enzyme backbone with nickel or cobalt in the active site produces a cleavage product with unknown function.

1.3.4 Thiol dioxygenases

Cupins with thiol dioxygenase activity are involved in the metabolism of organic sulphur compounds such as cysteine and cysteamine and are essential for the biosynthesis of taurine (Stipanuk et al., 2010). The mammalian cysteine dioxygenase catalyses the irreversible oxidation of the sulfhydryl group of cysteine to yield cysteine sulfinate (Figure 4).
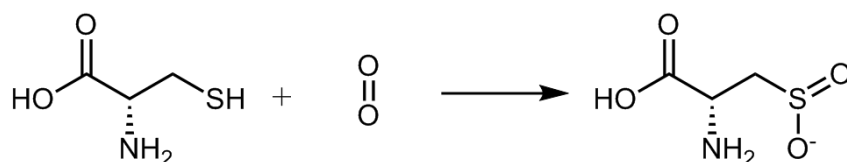
Figure 4: Reaction diagram of the oxidation of cysteine to cysteine sulfinate catalysed by cysteine dioxygenase.

This reaction forms a crucial step in the degradation pathway of cysteine to pyruvate and sulphate, which is important in providing inorganic sulphur to the metabolism (Simmons et al., 2006). The crystal structure of the mouse enzyme, which shows 91% sequence identity to the human homolog, has been solved and the enzyme was shown to contain iron, nickel and zinc, with increased iron concentration correlating with increased activity (McCoy et al., 2006).

Only one additional mammalian thiol dioxygenase has been discovered so far. This enzyme is active on cysteamine (2-aminoethanethiol). Cysteamine is formed by decarboxylation of cysteine during the biosynthesis of Coenzyme A. The oxidation of cystaemine by this dioxygenase is a step which leads to taurine. Taurine in turn is significant as one of the most abundant non-proteinogenic amino acids in mammalian tissues (Dominy et al., 2007).

1.3.5 Diketone dioxygenase

A cupin dioxygenase capable of cleaving the c-c bond of a diketone substrate, termed diketone dioxygenase (Dke1), has been described in some detail. The dioxygenase reaction catalysed by the native Dke1 cleaves a c-c bond in acetylacetone using atmospheric oxygen and enables the organism *Acinetobacter johnsonii* to utilize this toxic substance as sole carbon source (Straganz et al., 2003). The reaction occurs at an interesting three-histidine non-heme Fe^{II} – centre. The active centre has very specific geometry, since the replacement of any of the histidines for glutamate or asparagine, which is known in other cupin metal-binding sites, leads to a complete disruption of activity (Leitgeb et al., 2009). This enzyme also displays what is perhaps an even more surprising switch in the catalytic activity than in the case of acireductone dioxygenase. The β-diketone dioxygenase has been shown to be active as an esterase towards activated para-nitrophenyl esters if reconstituted with zinc (Leitgeb and Nidetzky, 2010). It is also reported in the same paper that there is a promiscuous esterase activity in the enzyme's native Fe^{II}-form, however the esterase activity of the zinc enzyme is about ten times higher. This esterase side-activity can perhaps be rationalised in the context of the metal-binding site of Dke1, which is comprised of three histidines residues (PDB ID 3BAL (Stranzl et al., to be published). The active site of Dke1 is thus quite similar, in principle, to the active site of zinc carbonic anhydrase (Eriksson et al., 1988), which itself is

an enzyme known for a very long time to catalyse the hydrolysis of para-nitrophenyl esters (Pocker and Stone, 1967).

1.3.6   Quercetinases – cupins with flavone dioxygenase activity

In contrast to the enzymes illustrated above, in which a switch in the active-site metal leads to a striking change or a loss of activity, one group of cupin dioxygenases are quite versatile regarding their metal preferences, while performing the same reaction on the same flavone substrate. This group, which displays versatility regarding the active site metal, are the cupin flavone dioxygenases, which are more commonly known as quercetinases. While quercetin is a flavone which contains multiple aromatic rings, the dioxygenase step in this case involves the cleavage of the oxygen containing heterocyclic flavonol ring (Figure 5), yielding 2-protocatechuoylphloroglucinolcarboxylic acid and releasing carbon monoxide in the process (Steiner et al., 2002).
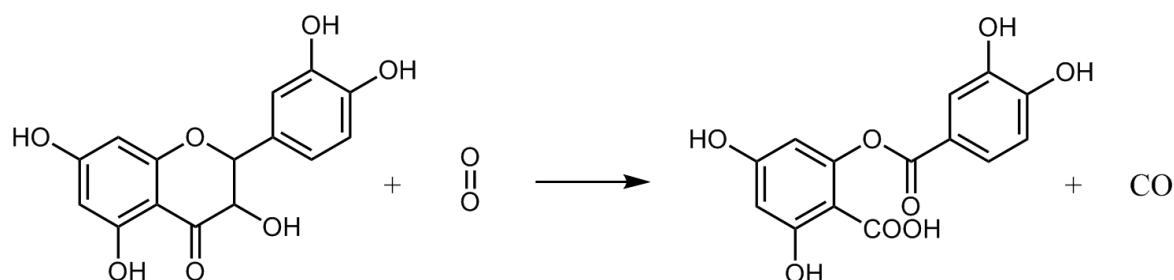
Figure 5: Reaction diagram of the reaction catalysed by quercetinases. The oxygen-containing ring of quercetin is cleaved in an oxidative manner with atmospheric oxygen while releasing carbon monoxide.

Quercetin is found in large amounts in the biosphere mostly in plant material, and also including human food (Formica and Regelson, 1995). Due to this abundance of quercetin in the biosphere, it is perhaps no surprise that many quercetinases are fungal enzymes (Oka and Simpson, 1971; Tranchimand et al., 2008). The fungal quercetinases are unusual in that they are the only firmly established copper dioxygenases known to date, even though their copper dependence has been known for a long time (Oka and Simpson, 1971). The crystal structure

of the quercetinase from Aspergillus japonicas has been solved, and the protein displays bicupin geometry with two cupin domains separated by a linker of 60 amino acids. These individual cupin domains are highly similar to each other to the extent that they can be superimposed (Fusetti et al., 2002). Crystal structures of substrate bound to the *Aspergillus* enzyme have been solved, and quercetin is found complexed to the active-site copper, firmly establishing the metal in the cupin barrel as the site of enzymatic activity (Fusetti et al., 2002; Steiner et al., 2002).

While the unusual, fungal quercetinases which were discovered early are copper-dependent, bacterial quercetinases with promiscuity towards the active site metal have been described recently. When the gene YxaG from *Bacillus subtilis* was cloned and overexpressed in *E. coli*, it was found to have quercetinase activity and to contain iron, which was in contrast to the copper-dependent fungal quercetinases (Barney et al., 2004; Bowater et al., 2004). A crystal structure of this enzyme has also been solved and it displays the same bicupin geometry as the *Aspergillus* enzyme, with two fused cupin domains and one active site in each domain (Gopal et al., 2005). Surprisingly, in later studies this enzyme was found to be even more active with metals other than iron, further broadening the range of known metals in quercetinases. Schaab et al. have produced metal-substituted variants of YxaG by expressing the protein in *E. coli* in minimal medium with the addition of various metals. They have found that the protein efficiently takes up manganese, and that the manganese protein was in fact almost 40-fold more active than the iron protein. In addition to manganese, the enzyme was active if the cultures were supplemented with copper and cobalt (Schaab et al., 2006).

In addition to these bicupin-type quercetinases a monocupin with this activity has also been discovered. The gene was originally cloned fro*m Streptomyces sp.,* and it was shown to have quite a broad flavone dioxygenase activity not limited to quercetin (Merkens et al., 2007). Interestingly, this enzyme seems to show a preference for nickel and cobalt, with no increase in activity with manganese or copper. In this case the metals have also been added to the medium. The metal preference of this enzyme is in contrast to the *Bacillus* enzyme expressed in the same host (Merkens et al., 2008).

In a somewhat surprising discovery, a low quercetinase activity has been reported in an *E. coli* protein with high structural similarity to human pirin (Adams and Jia, 2005). Pirin is a human nuclear protein as has been mentioned in the chapter on cupins with a non-enzymatic main function. It is unclear why *E. coli* should possess a protein with high similarity to human

pirin, and why such a protein should have quercetinase activity. This is again an illustration of the wide distribution and still largely unexplored nature of cupins.

1.3.7 Ring-cleavage dioxygenases of the cupin superfamily

Cupins constitute a major group of enzymes capable of degrading aromatics other than the flavonoids described above. These aromatics-degrading enzymes are ring-cleavage dioxygenases and they dearomatise the benzene ring of a variety of aromatic substrates by the incorporation of both atoms of molecular oxygen into the substrate. While this type of activity has developed through convergent evolution in a number of different protein folds (Vaillancourt et al., 2006), this chapter concentrates only on ring-cleavage dioxygenases of the cupin superfamily. A large number of dioxygenases are non-heme iron-dependent enzymes (Bugg and Ramaswamy, 2008), and the majority of cupin ring-cleavage dioxygenases also depend on a single iron cofactor bound in the canonical metal-binding site within the cupin barrel (Fetzner, 2012). This active-site metal is also the site of substrate binding and conversion.

1.3.8 Dioxygenases acting on hydroxylated benzoic acid derivatives

Many dioxygenases act on catechcholic and non-catecholic hydroxybenzoic acids (Bugg and Ramaswamy, 2008; Fetzner, 2012). Dioxygenases acting on gentisic acid (2,5 dihydroxybenzoic acid) have been purified and characterised relatively early, and the gentisate dioxygenase reaction itself is known since at least the 1970s (Crawford et al., 1975). The ring cleavage of gentisate proceeds between the carboxylic group in position 1 and the hydroxy group in position 2, yielding maleylpyruvate which is further degraded by the cells. Gentisate dioxygenases (GDOs) from *Klebsiella* and *Pseudomonas* have been purified and it has been established early on that the enzyme binds iron, as well as that the substrate is coordinated to the iron in the active site (Harpel and Lipscomb, 1990; Suárez, et al., 1996). The structures of GDOs from *Silicibacter pomeroy* and the human pathogen *Escherichia coli* O157:H7 have been solved recently, and both enzymes display a bicupin two-domain architecture with a typical cupin-barrel in each domain (Adams et al., 2006; Chen et al., 2008). In general, similar GDOs seem to be widespread in bacteria and these cupins play an

important role in the degradation pathways of many aromatics and xenobiotics (Hirano et al., 2007). Enzymes with high similarity to gentisate dioxygenases have also been described recently. A salicylate 1, 2 dioxygenase from *Pseudaminobacter salicylatoxydans* has been cloned and its crystal structure solved (Matera et al., 2008). It displays a bicupin-architecture highly similar to gentisate dioxygenases. The catalytic centre has been found to contain iron which is also typical for many cupin dioxygenases. The enzyme is unusual in that it seems to have a broader substrate spectrum than other similar hydroxybenzoic acid dioxygenases, cleaving different salicylates, gentisates and even the bulky substrate 1-hydroxy-2-naphthoate with similarly high activity (Matera et al., 2008). Another dioxygenase active on the bulky, singly hydroxylated substrate 1-hydroxynaphtoate-2-naphthoate, isolated from the phenanthrene-degrading *Nocardioides sp.* is similar to GDOs in that it also cleaves an aromatic ring containing a carboxylic and a hydroxyl group (Iwabuchi and Harayama, 1998). However, the amino-acid sequence deduced for this enzyme apparently differs significantly from those of other ring-cleavage dioxygenases which act on hydroxylated aromatic rings (Iwabuchi and Harayama, 1998). This enzyme was discovered before the cupins were firmly established as a protein superfamily, but it was clearly demonstrated to belong to the cupins later (Dunwell et al., 2001). It has recently been reported that it is possible to generate a 2-hydroxy-1-naphthoate dioxygenase by rationally mutating a salicylate dioxygenase, further underscoring the similarity of these enzymes (Ferraroni et al., 2012).

1.3.9 Hydroxyaminobenzoic acid dioxygenases

Cupin dioxygenases acting on benzoic acids with amino substituents in addition to hydroxyl groups are also known and some have been characterised in considerable detail. A number of highly similar dioxygenases acting on 2-amino-3-hydroxybenzoic acid, also called 3-hydroxyanthranilic acid (3-HAA, Figure 6) have been discovered in organisms ranging from bacteria and yeasts to mammals, including humans (Đilović et al., 2009; Kucharczyk et al., 1998; Malherbe et al., 1994; Zhang et al., 2005). In mammals, 3-HAA dioxygenases are part of the kynurenine pathway of tryptophane degradation, and the ring cleavage product of this enzyme can spontaneously cyclise to yield an endogenous neurotoxin – quinolinic acid (Calderone et al., 2002; Malherbe et al., 1994).
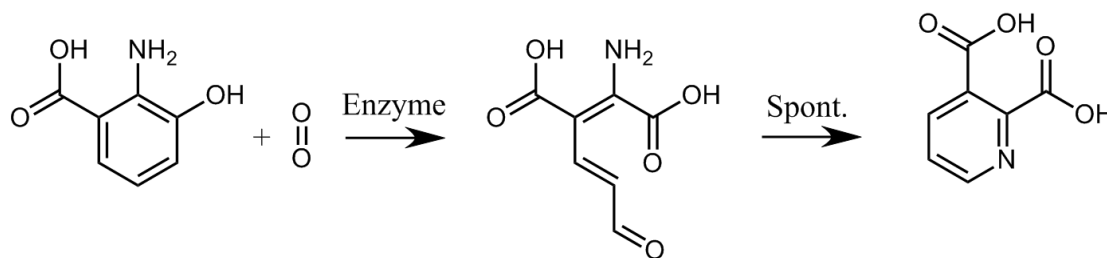
Figure 6: Reaction diagram of the 3-hydroxyanthranilate dioxygenase reaction. The enzyme cleaves the ring at the position adjacent to the hydroxy group, producing a semialdehyde product, which can spontaneously recyclise, yielding quinolinic acid.

Interestingly, it is this same reaction yielding quinolinic acid which is employed by some microorganisms in the production of nicotinamide in an alternative pathway via 3-HAA, which is again derived from the kynurenine pathway (Kurnasov et al., 2003; Zhang et al., 2005). In addition, a dioxygenase of this type has also been observed in a degradation pathway of the xenobiotic 2-nitrobenzoate in a strain of *Pseudomonas fluorescens* (Muraki et al., 2003). This is an interesting example of the evolutionary repurposing of the same reaction and same enzyme in different pathways multiple times.

The overall structure of the 3-HAA dioxygenase enzymes differs between various taxa. The bacterial and yeast proteins are monocupin homodimers, whereby an additional rubredoxin-like domain, with an additional iron atom ligated by four cysteines, is fused to the iron-containing cupin domain in each monomer (Li et al., 2006; Zhang et al., 2005). The mammalian enzymes are bicupin monomers in solution, with an additional cupin-like domain fused to the active-site harbouring barrel (Đilović et al., 2009). In contrast to the differences in overall structure, the catalytic center is fully conserved between all the enzymes described so far, with two histidines and a glutamate as the metal binding residues. Within the metal-binding site, Fe^{II} has been shown to be necessary for activity in all 3-HAA dioxygenases described thus far (Calderone et al., 2002; Đilović et al., 2009; Li et al., 2006; Zhang et al., 2005). A different enzyme acting on a rather similar substrate, 4-amino-3-hydroxybenzoic acid (4A3HBA) has been described in *Bordetella sp.* 10d, a strain capable of growth on 4A3HBA (Takenaka et al., 2002). The enzyme cleaves 4A3HBA adjacent to the hydroxy group, in a manner highly similar to 3-HAA dioxygenase (Figure 7).
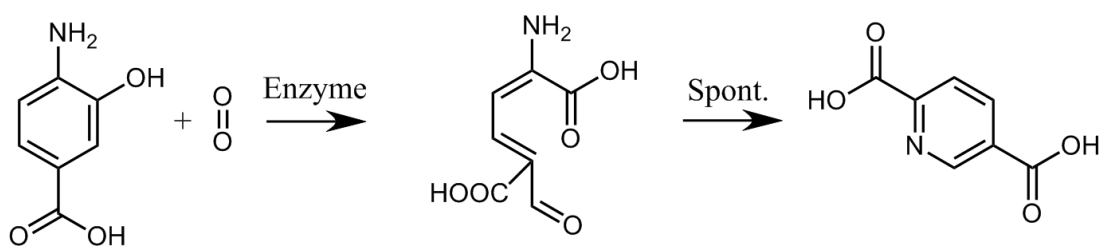
Figure 7: Reaction diagram of the 4-amino-3-hydroxybenzoic acid dioxygenase reaction. The ring-cleavage reaction proceeds at the position adjacent to the hydroxy group, in nalogy to the 3-HAA dioxygenase reaction.

The gene encoding this enzyme has been cloned and the protein purified, and it could be shown that this 4A3HBA dioxygenase is a homodimer of monocupin monomers, with highest sequence similarity to hydroxybenzoic acid dioxygenases (Murakami et al., 2004; Takenaka et al., 2002). The enzyme is apparently part of a complete pathway of 4A3HBA degradation in the source organism (Orii et al., 2006), and it appears to be highly specialised as it is highly specific for this particular substrate (Takenaka et al., 2002). Mutagenesis studies and biochemical characterisation of two cupin hydroxyaminobenzoate dioxygenases, acting on 3-HAA and 4A3HBA respectively, including an inquiry into the determinants of substrate specificity forms a part of this thesis.

1.3.10 Cupin enzymes acting on sugars and sugar-nucleotides

Very early after the discovery of cupins, a novel and apparently convergently evolved phosphoglucose isomerase has been discovered in the hyperthermophilic archaeon *Pyrococcus furiosus*. The enzyme was purified, its N-terminus sequenced, and upon the identification of the coding gene it became apparent that the enzyme must belong to the cupin superfamily (Verhees et al., 2001). It is fascinating that an activity as central to metabolism as the isomerisation of glucose and fructose phosphates could have arisen a second time through what must have been convergent evolution. This is perhaps also an illustration of the versatility of the cupin fold, with a member possibly having replaced an ancient and highly efficient enzyme by convergent evolution. The structure of the *P. furiosus* enzyme has been

solved in the presence of an inhibitor, and an unusual metal-dependent direct hydride-transfer mechanism has been proposed for this manganese-containing cupin (Berrisford et al., 2004). However, the same group has later produced evidence that an enediol mechanism is most likely, since they were able to show an exchange of hydrogen between the substrate and solvent via NMR. Another indication of the second mechanism is the fact that a stable analogue of an enediol intermediate is a very potent inhibitor of this enzyme (Berrisford et al., 2006). A cupin-type phosphoglucose isomerase has also been found in another hyperhtermophilic archaeon – *Thermococcus litoralis*. In this case, the enzymatic activity could be increased by the addition of ferrous iron, was inhibited by zinc and EDTA, and was lost completely if metal-binding residues are mutated, which strongly indicates the involvement of the metal in catalysis (Jeong et al., 2003). Cupin-type phosphoglucose isomerases have subsequently been found not only in archaea but also in bacteria such as *Salmonella enterica* serovar Typhimurium and *Ensifer meliloti*, and they have been demonstrated to represent a novel metal-dependent family for this type of enzymes (Hansen et al., 2005).

Additional cupins active on substrates have been discovered as part of biosynthesis pathways of somewhat more specialised carbohydrates. One such enzyme is RmlC which has been described already as the namesake of the RmlC-like cupins. RmlC is a dTDP-6-deoxy-d-xylo-4-hexulose 3,5 epimerase (EC 5.1.3.13), and thus it acts on a sugar-nucleotide substrate to produce rhamnose nucleotides for cell-wall biosynthesis (Giraud et al., 2000). The rhamnose pathway is not present in humans, and as such it has aroused great interest as a potential target for antibiotics development. The structure of a similar RmlC protein from *Streptococcus suis*, which has low sequence identity but high structure similarity to other bacterial enzymes of this class, has been solved in the presence of inhibitors. The active site has been confirmed to be located within the cupin barrel (Dong et al., 2003). A further interesting example of a cupin acting on unusual sugars has been found in the pathogenic *E. coli* strain O157:H7. The enzymatic function was originally unknown, and its gene annotated in the genome as "hypothetical". The overexpressed enzyme was shown to act on the rare pentose d-lyxose, converting it to mannose in a reversible reaction (van Staalduinen et al., 2010). As has been demonstrated in the the case of the lyxose isomerase, there is also potential for the discovery of novel sugar-converting enzymes within cupins with as of yet unknown function.

1.3.11 Cupin lyases

Even though there seems to be a scarcity of reports of lyases displaying the cupin fold, interesting examples have been described recently. Lyases acting on the algal competitive solute dimethylsulfoniopropionate, producing dimethylsulfide and acrylate, have recently been discovered in a wide range of bacteria such as *Alcaligenes*, *Arcobacter* and *Shewanella* (Curson et al., 2011). Also, a novel dimethylsulfoniopropionate lyase from marine *Roseobacter* species has been proposed to belong to the cupin superfamily based on the sequence-based prediction of a cupin-type metal-binding pocket (Todd et al., 2011).

Recently, the discovery of hydroxynitrile lyases (HNLs) of the cupin superfamily has been reported by our group (Hussain et al., 2012). HNLs catalyse the reversible cleavage of cyanohydrins to yield aldehydes or ketones and defensive HCN, and are usually found in plants (Dadashipour and Asano, 2011; Winkler et al., 2012). The cupin HNLs are unusual in that they have been isolated from bacteria, and do not show apparent structural or sequence similarity to any of the previously known hydroxynitrile lyases (Winkler et al. 2012). The absence of similarity to known enzymes strongly indicates a unique reaction mechanism, and a cupin hydroxynitrile lyase has been characterised in detail as part of this thesis.

1.4 Cupins with unknown function

As has been illustrated to some extent by this introductory chapter, cupins are indeed distributed widely within the tree of life, from bacteria to plants and mammals, and they fulfil a large number of very diverse functions. According to the SCOP database there are almost 100000 recognised cupins (Andreeva et al., 2007; Murzin et al., 1995). Many have not been assigned a function, especially not one supported by hard biochemical data.

Possibly due to their small size and stable folding, cupins have been crystallised numerous times in high-throughput projects such as the structural genomics project of *Thermotoga maritima* (Lesley et al., 2002), and a number of structures are also available for cupins with unassigned function (Jaroszewski et al., 2004; McMullan et al., 2004). An attempt to assign cupins to novel functions based on structural similarity has been undertaken (Agarwal et al., 2009), and computational studies which use similarity networks to cluster similar cupins have been published recently (Uberto and Moomaw, 2013).

While structure-based activity prediction methods offer a great promise for future development, pure in-silico approaches have some limitations. At the very least, biochemical data are necessary to clarify any structure-based activity predictions.

This thesis includes the discovery of a novel activity in a cupin from *Thermotoga maritima*, for which no biochemical activity was known prior to this work.

2. Aim of the thesis

As has been discussed in the introductory chapter, the cupin superfamily of proteins is characterized by a large functional diversity (Agarwal et al., 2009; Dunwell et al., 2004; Uberto and Moomaw, 2013). Enzymatic members of the cupin superfamily, include enzymes performing reactions found in other, non-related and dissimilar proteins (Hansen et al., 2005; Hussain et al., 2012). These are examples of the repurposing of the cupin scaffold through convergent evolution, for activities already found in other groups of proteins. Cupins have also been shown to display catalytic promiscuity when acting on the same substrate, in dependence on the incorporated metal (Dai, 1999). Some cupins also display catalytic promiscuity catalysing divergent reactions on dissimilar substrates (Leitgeb and Nidetzky, 2010). This diversity and apparent plasticity, in addition to their small size, easy handling and in some cases spectacular stability (Woo et al., 2000), prompts the question whether hitherto unknown members of this superfamily can be made accessible for diverse biotechnological and biocatalytic applications. This thesis constitutes an attempt towards a characterisation of a part of the diversity of enzymatic functionalities found in the cupin superfamily of proteins. The approach involves characterisation and engineering of known functions as well as the investigation and search for novel enzymes. One part of this thesis contributes biochemical work towards the clarification of the potential of cupins for the discovery and utilisation of novel enzymatic activities. Cupins with unknown function were taken as a basis, and broad based activity assays have been applied. The thesis includes the discovery of a biochemical function in a cupin where no function was known previously, underscoring the potential of cupins for the discovery of novel enzymatic activities. Such a broad-based biochemical characterisation forms the basis to make cupins generally accessible to engineering and possibly rational design in the future.

3. MANUSCRIPTS AND PUBLICATIONS

The manuscripts listed in this chapter comprise the most significant part of the thesis. Presented here are the most relevant scientific results, in the form of manuscripts which have either already been published in peer-reviewed journals or are intended for such publication. The first manuscript has been published in the FEBS Journal and presents the biochemical and structural of a novel and unusual manganese-dependent hydroxynitrile lyase which belongs to the cupin superfamily of proteins. The second manuscript presents the discovery of a peroxide-driven alkene cleavage activity in a cupin from the marine bacterium *Thermotoga maritima* for which no biological role or biochemical activity was previously known. The final manuscript presents the investigation and engineering of substrate-specificity in two cupin dioxygenases acting on highly similar hydroxyaminobenzoic acid substrates with high substrate specificity.

3.1 Biochemical and structural characterisation of a novel bacterial manganese-dependent hydroxynitrile lyase

Ivan Hajnal[1,2], Andrzej Łyskowski[1], Ulf Hanefeld[2], Karl Gruber[1,3], Helmut Schwab[1,4], and Kerstin Steiner[1,*]

[1]ACIB, Austrian Centre of Industrial Biotechnology GmbH, Petersgasse 14, 8010 Graz, Austria

[2]Gebouw voor Scheikunde, Technische Universiteit Delft, Julianalaan 136, 2628BL Delft, The Netherlands

[3]Institute of Molecular Biosciences, University of Graz, Humboldtstraße 50, 8010 Graz, Austria

[4]Institute of Molecular Biotechnology, TU Graz, Petersgasse 14, 8010 Graz, Austria

Author contribution:

I have contributed the experimental data regarding the biochemical characterisation of *Gt*HNL including mutagenesis, protein purification for parts not concerning the crystallisation, metal-exchange experiments, kinetic measurements and cyanohydrin synthesis reactions. I have also written the parts of the manuscript which describe this part of the work. The crystal structure was solved by Dr Łyskowski and Dr Steiner, using protein purified by Dr Steiner, and these two authors have also written the parts of the manuscript regarding the crystal structure. The seinor authors have contributed scientific direction and critical reading of the manuscripts.

*Corresponding author: Dr Kerstin Steiner, Austrian Centre of Industrial Biotechnology GmbH, Petersgasse 14/4, 8010 Graz, Austria, phone: +43-316-8739346, fax: +43-316-8739302, email: kerstin.steiner@acib.at

3.1.1 ABSTRACT

Hydroxynitrile lyases (HNLs), which catalyse the decomposition of cyanohydrins, are found mainly in plants. *In vitro*, they are able to catalyse the synthesis of enantiopure cyanohydrins, which are versatile building blocks in the chemical industry. Recently, HNLs have also been discovered in bacteria. Here, we report on the detailed biochemical and structural characterization of a hydroxynitrile lyase from *Granulicella tundricola* (*Gt*HNL), which was successfully heterologously expressed in *Escherichia coli*. The crystal structure was solved at a crystallographic resolution of 2.5 Å and exhibits a cupin fold. As *Gt*HNL does not show any sequence or structural similarity to any other HNL and does not contain conserved motifs typical of HNLs, cupins represent a new class of HNLs. *Gt*HNL is metal-dependent, as confirmed by inductively coupled plasma/optical emission spectroscopy, and in the crystal structure, manganese is bound to three histidine and one glutamine residue. *Gt*HNL displayed a specific activity of $1.74\ \text{U}\cdot\text{mg}^{-1}$ at pH 6 with (*R*)-mandelonitrile, and synthesized (*R*)-mandelonitrile with 90% enantiomeric excess at 80% conversion using 0.5 m benzaldehyde in a biphasic reaction system with methyl tertiary butyl ether.

3.1.2 INTRODUCTION

Hydroxynitrile lyases (HNLs) are enzymes that catalyse the reversible cleavage of cyanohydrins, yielding HCN and aldehydes or ketones. In the last two decades, the enzymes have been investigated extensively, especially regarding their potential use as catalysts in organic chemistry (Fesko and Gruber-Khadjawi, 2013; Gruber-Khadjawi et al., 2012; Winkler et al., 2012). In chemical synthesis, reversal of the natural cleavage reaction is employed. Cyanohydrins constitute important building blocks for fine chemicals, as their synthesis creates a new stereocentre and adds a versatile functional group (the nitrile group) at the same time (Hilterhaus and Liese, 2012; Holt and Hanefeld, 2009; Purkarthofer et al., 2007). HNLs comprise a diverse group of enzymes that vary in terms of their substrate specificity, enantioselectivity and the need for a co-factor. This may account mainly for the fact that they belong to different protein families with no significant sequence similarity. The structures solved so far show similarity to α/β-hydrolases, oxidoreductases, carboxypeptidases or Zn^{2+}-dependent alcohol dehydrogenases (Gruber and Kratky, 2004; Winkler et al., 2012).

Interestingly, members of the α/β-hydrolase fold show both (*R*)- and (*S*)-selectivity (Andexer et al., 2012; Guterl et al., 2009). Most HNLs have been discovered in plants, and several have been cloned successfully and expressed in *Escherichia coli* or yeasts such as *Pichia pastoris* (Dadashipour and Asano, 2011; Sharma et al., 2005). Very recently, the discovery of hydroxynitrile lyases in endophytic bacteria has been described by our group (Hussain et al., 2012). Interestingly, they are not related to any of the HNLs characterized so far, but show sequence similarity to proteins of the cupin superfamily, which comprises a large number of small β-barrel-fold proteins with diverse functionalities (Agarwal et al., 2009; Dunwell et al., 2001, 2004). As HNL activity has not previously been described for proteins of this superfamily, cupins represent a novel class of these industrially useful enzymes. Although cupins are structurally conserved and usually contain two conserved motifs, G-(X)5-H-X-H-(X)3,4-E-(X)6-G (motif 1) and G-(X)5-P-X-G-(X)2-H-(X)3-N (motif 2), the overall sequence identity is low among members of this superfamily (Agarwal et al., 2009). The two motifs also include the residues for metal binding. Most cupins are metal-binding proteins that bind divalent cations such as iron, zinc, manganese, copper, nickel or cadmium (Chavez et al., 2011). The metal is usually involved in the enzymatic reaction either directly in the reaction mechanism or at least via an interaction with the substrate. In this paper, we describe the biochemical and structural characterization of the manganese-dependent hydroxynitrile lyase from the acidobacterium *Granulicella tundricola* (*Gt*HNL), which shows high sequence similarity to two previously described cupins with HNL activity (Hussain et al., 2012) and was heterologously expressed in *E. coli*.

3.1.3   RESULTS

Based on sequence similarities to recently discovered bacterial hydroxynitrile lyases with a cupin fold from *Pseudomonas mephitica* (*Psm*HNL) and *Burkholderia phytofirmans* (*Bp*HNL) (Hussain et al., 2012), several highly similar proteins were discovered using a BLASTp search on the UniProt (Universal Protein Resource) server. Most of them originate from bacterial genome or metagenome sequencing projects and are not annotated, thus their natural function is unknown. Some of them were obtained as synthetic genes, cloned into expression vectors, successfully expressed as soluble proteins in *E. coli*, and tested for HNL activity. One of them, a protein originating from *Granulicella tundricola* with 79.1%

sequence similarity to *Bp*HNL, showed higher cyanolysis activity in preliminary screenings and was chosen for detailed biochemical and structural analyses.

3.1.3.1 Mandelonitrile cyanolysis activity of GtHNL is dependent on manganese

*Gt*HNL was expressed at a very high yield as soluble protein in *E. coli* (data not shown), with yields of > 50% of total soluble protein. In initial experiments, the cyanolysis activity of *Gt*HNL was assessed using mandelonitrile as substrate. The enzyme cleaved (*R*)-mandelonitrile to benzaldehyde and HCN, as indicated by a sensitive colony-based filter assay for HCN (data not shown), as well as by the increase in absorption at 280 nm attributed to released benzaldehyde measured in a spectrophotometric assay using cleared lysate (Fig. 1). The enzymatic activity is clearly distinguishable from the background reaction in the same buffer.
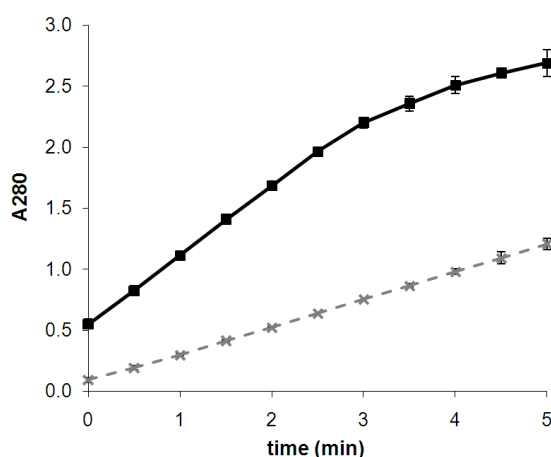
Figure 1. Release of benzaldehyde from 18 mm (*R*)-mandelonitrile at pH 5.5 (80 mm MES, 80 mm sodium oxalate buffer), catalysed by 50 μg manganese-reconstituted *Gt*HNL apoprotein (black squares, solid line) at 25–27 °C, compared to the buffer background (grey crosses, dashed line).

3.1.3.2 Metal dependence

Most cupins are metal-binding proteins that bind Mn^{2+}, Fe^{2+}, Ni^{2+}, Cu^{2+} or Zn^{2+} in the active site (Chavez et al., 2011). The characteristic cupin domain comprises two conserved motifs, G-(X)5-H-X-H-(X)3,4-E-(X)6-G (motif 1) and G-(X)5-P-X-G-(X)2-H-(X)3-N (motif 2), which are both involved in metal ion binding (Dunwell et al., 2004). The two His residues and the Glu/Gln residue in motif 1 as well as the His residue in motif 2 act as ligands for binding of the metal ion in the active site. These conserved motifs, including the putative metal-binding amino acids, are conserved in the sequence of *Gt*HNL, indicating that *Gt*HNL is also metal-dependent. Preliminary experiments using a colony-based filter assay on colonies grown on LB agar supplemented with various divalent cations (Fe^{2+}, Zn^{2+}, Mn^{2+}, Ni^{2+}, Co^{2+} or Cu^{2+}) showed that HNL activity increased in the presence of Mn^{2+} in the medium, compared to colonies without supplementation (data not shown). The results of a detailed metal analysis are described below.

3.1.3.3 pH dependence

The stability and activity of HNLs at acidic pH is especially important as the non-specific background reaction is reduced at low pH. Purified protein, which was expressed in the presence of 100 μm $MnCl_2$, catalysed the cyanolysis of (*R*)-mandelonitrile in aqueous sodium oxalate buffer (which was identified as the most appropriate buffer) at pH 4.5, with a K_m of 10.3 ± 2 mm and a k_{cat} of 0.03 s^{-1}, corresponding to an enzymatic activity of 0.12 ± 0.009 U·mg^{-1} at V_{max} (Fig. S1). Unsurprisingly, the enzyme is more active at less acidic pH, and the activity of the protein increased to 1.74 U·mg^{-1} at pH 6. Protein reconstituted with manganese *in vitro* showed an even higher activity (> 4 U·mg^{-1} at pH 6.0), which may be explained by the higher manganese loading achieved *in vitro*. Overall, the specific activity increased ~ 40-fold from pH 4.5 to pH 6.0 (Fig. 2). The activity was not measured at higher pH, as the enzyme catalysis becomes indistinguishable from the base-catalysed chemical background reaction at pH values above 7. The pH stability of *Gt*HNL at pH 4.0 was confirmed by circular dichroism spectroscopy (Fig. S2A). No unfolding was observed within 1 h, indicating that the overall fold is stable at low pH.
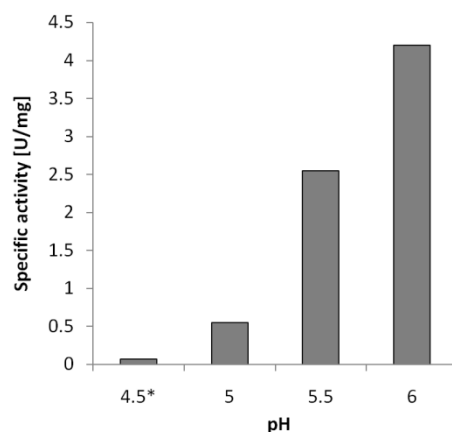
Figure 2. Specific activity ($U \cdot mg^{-1}$) of *Gt*HNL (50 μg Mn^{2+}-reconstituted apoprotein) at various pH values (80 mm MES, 80 mm sodium oxalate buffer, 18 mm (*R*)-mandelonitrile, 25–27 °C); the chemical background reaction at the various pH values was subtracted. *Sodium oxalate at pH 4.5 without MES.

3.1.3.4 Metal analysis

As described above, addition of manganese to the expression medium led to an increase in HNL activity. To prevent interference of the His-tag with the metal content and analysis, the protein was purified without the tag. Surprisingly, the first purification step using anion exchange chromatography resulted in almost pure protein (Fig. S3), and the purity was further improved by size-exclusion chromatography. The yield of purified protein was up to 50–60 mg per litre of culture, corresponding to ~ 20 mg of purified protein per g wet cell paste. The metal contents of purified protein samples, which were either grown in LB medium supplemented with various metals or deprived of metal using the chelating agent 2,6-pyridinedicarboxylic acid and reconstituted with various metals, were analysed by inductively coupled plasma/optical emission spectroscopy (ICP-OES). As *Gt*HNL is a monocupin with one proposed metal-binding site, one metal ion is expected per subunit. This was also confirmed by X-ray crystallography (see below).

When the protein was expressed in LB medium without addition of metals, only traces of zinc and iron were detected during analysis (3.8 mol% Zn^{2+} and 5.3 mol% Fe^{2+} per metal-binding site). These two metals are generally present in LB medium, which was confirmed by ICP-OES analysis of the medium. As Zn^{2+} and Fe^{2+} were incorporated into the protein in only very low amounts, and the expression in minimal medium without any metal was significantly lower than in LB medium, further experiments were performed in complex medium. Interestingly, when 100 μm $MnCl_2$ (final concentration) was added to the medium during induction with isopropyl thio-β-d-galactoside, the resulting enzyme showed a significantly increased manganese loading, corresponding to 58.9 mol% manganese per monomer. Addition of 100 μm $ZnSO_4$ or $FeSO_4$ to the expression medium did not lead to an increase in incorporation of these metals into the protein *in vivo*. In another set of experiments, apoprotein was produced by dialysing purified protein against a buffer containing the strong chelating agent 2,6-pyridinedicarboxylic acid. Circular dichroism spectroscopy confirmed that the apoprotein was still folded correctly, even after incubation for 1 h at pH 4.0 (Fig. S2B). According to ICP-OES measurements, the apoprotein contained residual metal (mainly iron) amounting to < 10% of all metal-binding sites. Subsequently, the apoprotein was incubated with $MnCl_2$, $ZnSO_4$ and $FeSO_4$. In addition, purified protein expressed without addition of metal ions to the medium was also loaded with the same metals *in vitro*, as it contains almost no metal (data not shown). In both cases, the protein efficiently took up manganese, zinc or iron (Fig. 3A). Interestingly, only the protein reconstituted with manganese regained clearly measurable activity with respect to mandelonitrile cyanolysis (Fig. 3B). The activity of apoprotein reconstituted with Zn^{2+} or Fe^{2+} was slightly higher than that of the metal-free apoprotein, but was not clearly distinguishable from the high background reaction.
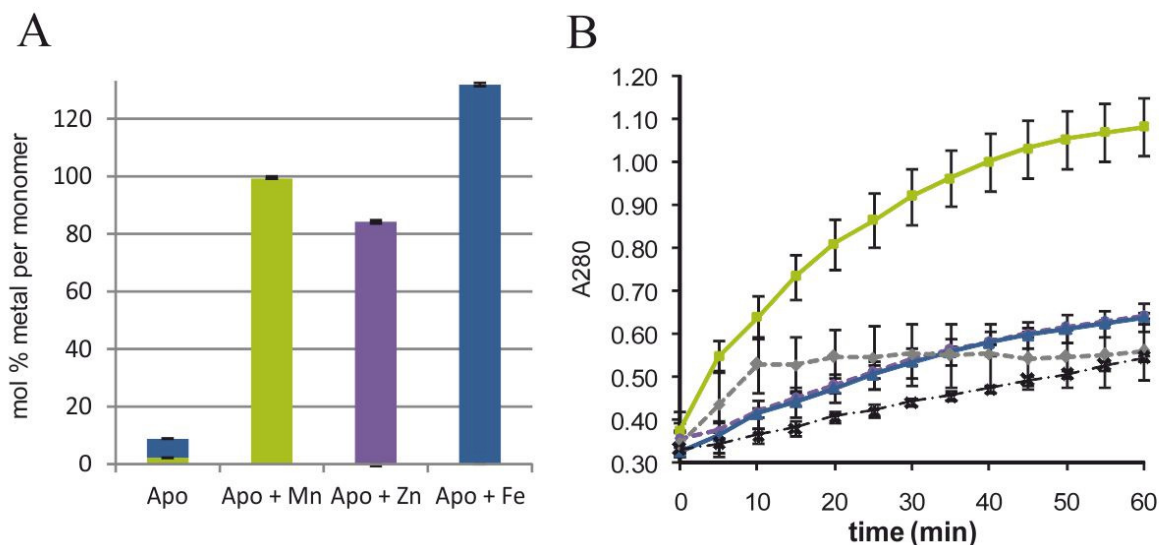
Figure 3. (A) Metal contents (determined by ICP-OES) of apoprotein after treatment with chelating agent, and subsequent incubation with MnCl$_2$, ZnSO$_4$ or FeSO$_4$: green bars, manganese; purple bars, zinc; blue bars, iron. **(B)** Cyanolysis assay (67 mm sodium oxalate buffer pH 4.5, 18 mm (*R*)-mandelonitrile, 25 °C) with 50 μg of purified apoprotein (grey diamonds, dashed line) or apoprotein reconstituted with either manganese (green squares, solid line), zinc (purple circles, dashed line) or iron (blue triangles, solid line). Black crosses represent the buffer control.

3.1.3.5 Enantioselective synthesis of mandelonitrile

The ability of the enzyme to catalyse the synthesis of mandelonitrile from benzaldehyde and HCN was examined using a two-phase system consisting of benzaldehyde and HCN dissolved in methyl tertiary butyl ether (MTBE) as the organic phase, and an aqueous phase comprising concentrated cleared lysate, acidified to a final pH of 4 using sodium acetate buffer. The reaction was tested at 15 °C and 5 °C. As the reactions at 5 °C yielded a significantly higher enantiomeric excess (data not shown), all further reactions were performed at 5 °C. The enzyme expressed in the presence of manganese is highly enantioselective, yielding (*R*)-mandelonitrile with 90% enantiomeric excess (*ee*) at 80% conversion of the initial 0.5 m benzaldehyde after 6 h. As expected, when cleared lysate of cells grown without addition of manganese to the expression medium was used, the yield and enantiomeric excess were markedly lower: < 50% conversion and ~ 80% *ee* (Fig. 4).
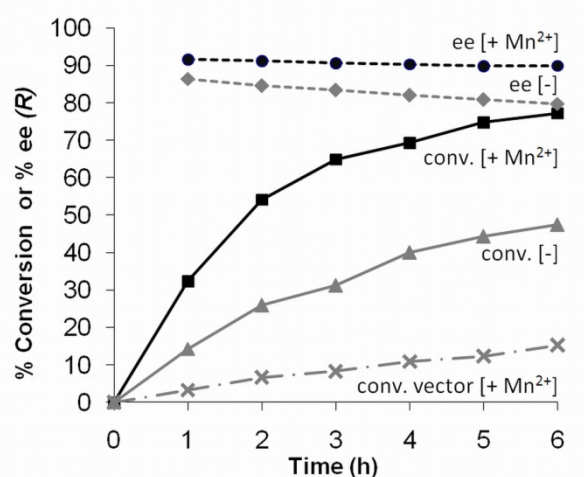
Figure 4. Mandelonitrile synthesis in a two-phase system consisting of 500 μL of cleared cell lysates expressed with and without Mn^{2+} addition (50 mg·mL^{-1} total protein in 100 mm sodium acetate, pH 4.0) as the aqueous phase and 1 mL MTBE containing 0.5 m benzaldehyde and 2 m HCN as the organic phase, shaken at 1000 rpm and 5 °C. The curves show the percentage enantiomeric excess (*ee*) of (*R*) for the enzyme expressed with Mn^{2+} addition (black circles, dashed line) and without Mn^{2+} addition (grey diamonds, dashed line), and conversions for the enzyme expressed with Mn^{2+} addition (black squares, solid line) and without Mn^{2+} addition (grey triangles, solid line) and the empty vector control expressed with Mn^{2+} addition (grey crosses, dotted/dashed line).

As it is known that a chemical background reaction proceeds readily in aqueous medium (Holt and Hanefeld, 2009), a buffer control, as well as cleared lysate of *E. coli* without *Gt*HNL (empty vector control), were used as controls. The lysate of the empty vector control showed 15% conversion after 6 h, compared with 80% for the lysate expressing *Gt*HNL, and the control reaction showed no enantioselectivity. Purified *Gt*HNL had lower apparent activity (data not shown) and consequently yielded mandelonitrile with lower enantiomeric excess even when added at approximately the same amount as present in the cleared lysate. A possible reason for this may be the stabilizing effect of *E. coli* proteins in the cleared lysate, especially considering the harsh reaction conditions of pH 4 in the aqueous phase, the contact to MTBE, and the high concentrations of benzaldehyde and HCN at the interphase. In another cyanohydrin synthesis study, it was shown that addition of albumin stabilized the hydroxynitrile lyase from *Hevea brasiliensis* in the presence of high concentrations of HCN in

buffer-saturated solvent (Costes et al., 2001). On the other hand, it has been reported that an HNL from *Arabidopsis thaliana*, which was expressed in *E. coli* and purified, retained considerable activity even though it precipitated in a buffer-saturated organic system (Okrob et al., 2011).

Additionally, apoprotein and reconstituted proteins were used in mandelonitrile synthesis reactions (Table 1). The apoprotein was clearly less active and less enantioselective compared to the metal-containing proteins (22% conversion, 40% *ee*). The zinc-reconstituted protein showed only slightly higher conversion than the apoprotein (32%), but the product had a significantly higher *ee* of 74%. The iron-reconstituted protein (81% conversion, 87% *ee*) showed comparable activity and enantioselectivity to the manganese-reconstituted aliquot (76% conversion, 89% *ee*), although the difference in their cyanolysis activities was more pronounced (Fig. 3B). *Gt*HNL did not accept aromatic ketones, such as acetophenone and propiophenone, as substrates.

Table 1. Conversions and enantiomeric excess of mandelonitrile synthesis reactions with purified *Gt*HNL either as apoprotein or reconstituted with Mn^{2+}, Zn^{2+} and Fe^{2+} after 6 h and 24 h, respectively. The two-phase system consisted of 500 µL of aqueous phase (20 mg mL^{-1} protein in 100 mM Na-acetate, pH 4.0) and 1 mL MTBE containing 0.5 M benzaldehyde and 2 M HCN as organic phase, 1000 rpm, 5°C.

Sample	% Conversion		% *ee* (R)	
	6 h	24 h	6 h	24 h
*Gt*HNL Apo	11 ± 1.5	22 ± 1.8	40 ± 1.5	40 ± 0.1
Apo + Mn	35 ± 1.7	76 ± 1.4	90 ± 0.1	89 ± 0.1
Apo + Zn	11 ± 1.0	32 ± 1.0	71 ± 0.0	74 ± 0.1
Apo + Fe	39 ± 8.9	81 ± 8.4	88 ± 0.3	87 ± 0.4

3.1.3.6 Overall three-dimensional structure of GtHNL

The His-tagged protein was purified efficiently using affinity chromatography and subsequent size-exclusion chromatography as described previously (Łyskowski et al., 2012). The protein crystals diffracted to ~ 2.5 Å, and contained two homotetramers in the asymmetric unit (Fig. 5A). The electron density of the polypeptide chain backbones was well-ordered in all subunits containing all 131 amino acid residues (see Table 2).
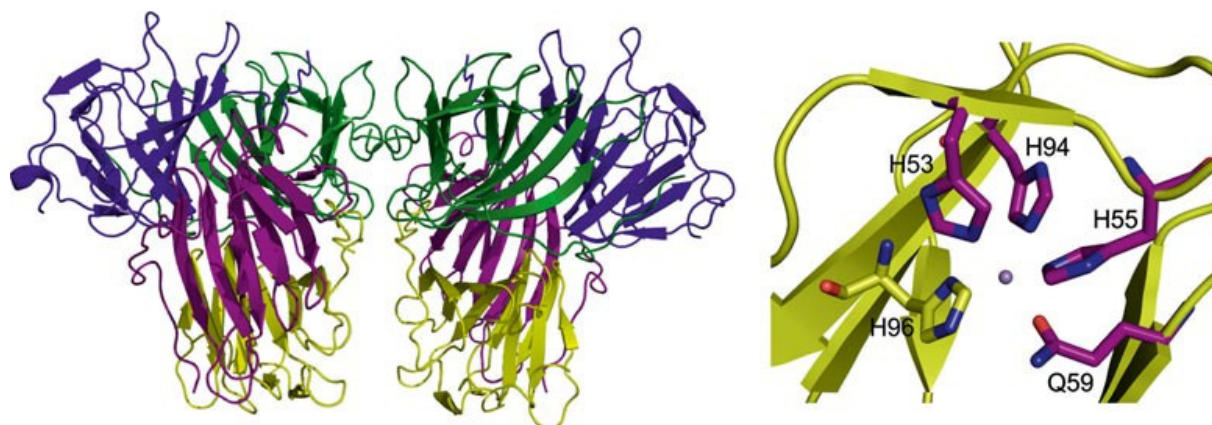
Figure 5. (A) Cartoon depiction of the overall structure of the two homotetramers of *Gt*HNL. The subunits are coloured in blue, magenta, light and dark green. **(B)** Stick representation of the amino acids in the metal-binding site. Magenta, metal-binding residues; purple, metal ion; light green, amino acids close to the metal-binding site. The figures were prepared using the PyMOL Molecular Graphics System (version 1.5.0.4, Schrödinger LLC), Portland, OR, USA.

Table 2. Data collection and refinement statistics.

Data processing statistics	
Beam line	ESRF/ID23-1
Wavelength [Å]	1.0329
Resolution range (outer shell) [Å]	37.72- 2.46 (2.52- 2.46)
Space group	C 2 2 21
Unit cell	a=126.31 b=254.78 c=82.47 $\alpha=\beta=\gamma=90.0$
R_{merge}	0.121 (0.58)

(continued) Data processing statistics	
R$_{meas}$	0.131 (0.227)
Mean((I)/sd(I))	11.3 (3.6)
No. of reflections	352885 (26703)
No. of unique reflections	48423 (3563)
Completeness [%]	99.3 (99.8)
Multiplicity	7.3 (7.5)
Refinement statistics	
Resolution range [Å]	2.46 – 37.72
Protein residues	1048
No. atoms:	8768
- non-solvent	8061
- solvent	707
- ligands	8
R-factor	0.1767
R-free	0.2329
Deviation from ideal values:	
- bonds	0.015
- angles	1.216
- chirality	0.080
- planarity	0.006
- dihedral	16.183
Min. nonbonded distance	1.856
Molprobity statistics	
- all-atom clashscore	6.55
Ramachandran plot:	
- favoured (%)	97.58
- allowed (%)	1.94
- outliers (%)	0.48
Rotamers outliers (%)	6.24
C-beta deviations	1.00
Wilson B-factors	36.67

The α-carbons of the subunits superimpose very well, with an RMSD (using chain A as the reference) of 0.23 Å. Each monomer comprises eleven β-strands and does not contain any helices. It forms a jellyroll β-sandwich with a topology that is characteristic of the cupin-barrel fold. The name cupin is derived from the Latin term 'cupa', which means small barrel and refers to the shape of the structure. The tetramer is formed by two dimers (mean interface

area for A to B, C to D, E to F, G to H = 1554 Å^2), whereby the N-terminus of chain A interacts with a β-sheet (G83–F88) from chain B and *vice versa*. The same holds for chain C and D. The two dimers are then twisted by 90° relative to each other (with chain A mainly interacting with chain C and chain B mainly interacting with chain D; mean total interface area of dimer AB to CD = 619.2 Å^2). Each monomer has a metal-binding site at the base of the β-barrel (Fig. 5B). The metal ion interacts with three His residues and one Gln, a metal-binding site that is rare but has previously been reported in other cupins (Chavez et al., 2011; Rajavel et al., 2009). The metal-binding residues are His53, His55, Gln59 and His94, with distances to the metal of 2.2, 2.2, 2.0 and 2.4 Å, respectively (calculated after alignment of all chains to the reference chain A as the distance between the metal ion and the mean position of metal-coordinating atoms). They align well with metal-binding residues in other cupins in terms of both sequence and structure. Interestingly, another histidine residue, His96, is located within binding distance to the metal (mean 2.4 Å). However, a metal-binding site comprising five binding residues has not been described in cupins previously, and His96 is located on the opposite side of the residues that are usually involved. The remaining coordination sites of the metal ion may interact with the substrate. The refinement with manganese as the metal ion supports the biochemical data and produces the best refinement statistics. Unfortunately, no substrate could be co-crystallized or soaked into the active site so far. Docking experiments with mandelonitrile resulted in several possible binding modes, most likely due to the large cavity.

A structural similarity search for *Gt*HNL using the DALI server (Holm and Rosenstrom, 2010) showed that the closest structure is 2F4P, an uncharacterized monocupin protein from *Thermotoga maritima*, which was also used as a template for molecular replacement (Z-score 21.9, RMSD 1.7 over 125 amino acids). However, it has only 40% sequence identity to *Gt*HNL. The other structures showed even lower similarity [1VJ2: Z-score 14.5, RMSD 1.9, 20% sequence identity; 1O4T: Z-score 14.5, RMSD 2.0, sequence identity 21% (both uncharacterized monocupins from *Thermotoga maritima*); 3HT2: Z-score 14.5, RMSD 2.4, 14% sequence identity (a zinc-containing polyketide cyclase RemF from *Streptomyces resistomycificus* (Silvennoinen et al., 2009). The gene corresponding to 2F4P was obtained as a synthetic gene and re-cloned, and the protein was expressed as described for *Gt*HNL, but showed only very weak hydroxynitrile lyase activity, which was not investigated any further (data not shown).

3.1.3.7 Structure-guided mutagenesis

In addition to the amino acids of the metal-binding site, amino acids in and around the main cavity, which harbours the active site in all cupins characterized so far, were identified by structural analysis and mutated by site-directed mutagenesis (Fig. 6).
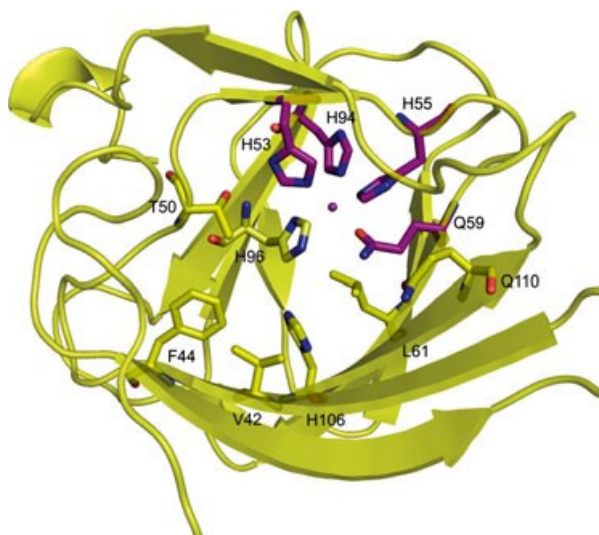
Figure 6. Stick representation of several amino acids in or close to the active site. Purple, metal-binding residues and metal ion; light green, amino acids in or close to the active site. The figure was prepared using the PyMOL Molecular Graphics System (version 1.5.0.4, Schrödinger LLC).

The effect of mutations on the cyanolysis reaction was qualitatively analysed using a colony-based assay (see Fig. 7A for metal-binding sites). Several of the mutations had no effect on the activity (F44A, L61A, Q110A) and were not investigated any further. The most interesting variants were also examined in the synthesis direction, analogous to the wild-type protein (Table 3). As it was shown that manganese is important for catalysis, the presumed metal-binding residues His53, His55, Gln59 and His94 were all mutated to alanine. In addition, Gln59 was mutated to glutamate, as Glu is more abundant in the metal-binding sites of cupins than Gln. However, no difference in expression or activity was observed (data not shown). All four alanine variants were expressed at wild-type level (Fig. 7A, upper row). Moreover, CD spectroscopy confirmed that the overall fold of the variants did not change (Fig. S4 and Table S3). Nevertheless, three of these alanine variants lost cyanolysis and

cyanohydrin synthesis activity completely (Fig. 7A and Table 3). Interestingly, although mutant H94A was still active with respect to cyanolysis (approximately the same activity, 0.13 U·mg^{-1}, as the wild-type enzyme in the photometric assay at pH 4.5), it showed a dramatic loss in synthesis activity and enantioselectivity with respect to mandelonitrile (Table 3). Metal analysis by ICP-OES showed that this variant was able to bind manganese efficiently, in contrast to mutants of the other three residues (Fig. 7B). To rule out stability problems for variant H94A due to contact with organic solvents in the synthesis reaction in contrast to the aqueous cyanolysis conditions, acidified cleared lysate (pH 4.0) containing the enzyme (mutant H94A and wild-type as control) was shaken with MTBE for 1 h. The subsequent filter assay showed no apparent loss of cyanolysis activity, indicating that the two-phase reaction system was not responsible for the lack of activity.

Table 3. Conversions and *ee* of mandelonitrile in synthesis reactions with cleared lysates of *Gt*HNL variants after 6 h, pH 4.0 in aqueous phase, 5°C.

Variant	% Conversion	% ee (R)
WT	64	86
T50A	57	91
H96A	6	n.d.
H96D	5	n.d.
H106A	5	n.d.
H106D	26	54
H55A	2	n.d.
H94A	11	21

n.d. not determined due to the low conversion

Among the residues in or close to the active site, Thr50 and two histidines, His96 and His106, were considered as candidate residues involved in catalysis. In the HNLs from *Hevea brasiliensis* (*Hb*HNL) and *Prunus amygdalus* (*Pa*HNL), a histidine is proposed to be involved in catalysis (Dreveny et al., 2009; Gruber et al., 2004). In *Pa*HNL, the histidine may act directly as a general base, abstracting the proton from the mandelonitrile hydroxyl group, whereas in *Hb*HNL, the histidine is part of a catalytic triad formed by Asp, His and Ser, in

which the serine deprotonates the cyanohydrin (Gruber et al., 2004). Thr50, which is located in the second shell of the active site, may form a hydrogen bond with the vicinal His96 in the active site. His106 is located further away from Thr50 in the solved structure. However, the mutant T50A did not lose activity in any of the reactions. It showed a comparable conversion of ~ 58% and an even higher *ee* value (91.5%) than the wild-type protein (64% conversion, 86.6% *ee*), indicating that this residue is not essential for mandelonitrile synthesis (Table 3).
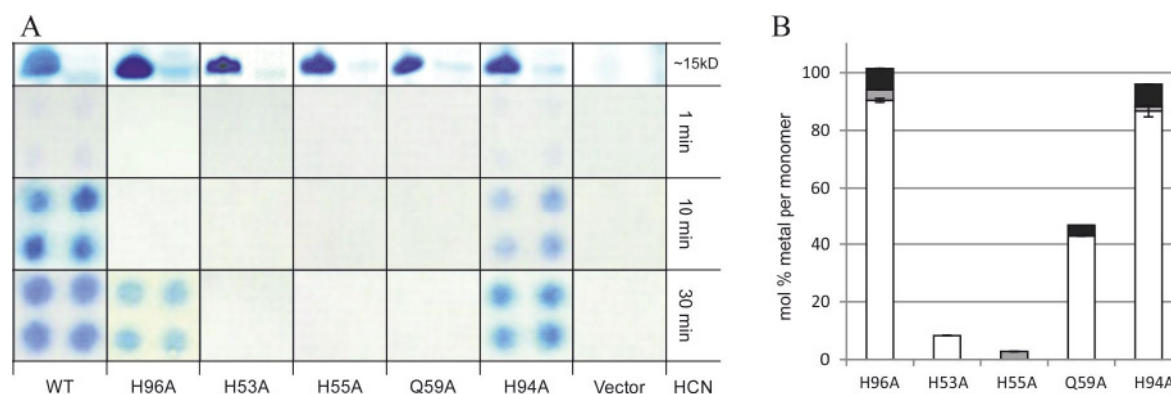
Figure 7. (A) Colony-based filter assay for the mutants H96A, H53A, H55A, Q59A and H94A, wild-type and vector control for mandelonitrile cyanolysis. Blue colour indicates the release of HCN. The incubation time in minutes is shown on the right. The top row shows the band at ~ 15 kDa obtained upon SDS/PAGE of the soluble and insoluble protein fractions of wild-type and each variant, respectively. **(B)** Metal contents (analysed by ICP-OES) for purified variant proteins expressed in LB medium with addition of 100 μm $MnCl_2$; white bars, manganese; grey bars, zinc; black bars, iron.

His96, which is located within binding distance of the metal ion in the structure of *Gt*HNL, was mutated to alanine. This variant was efficient at binding manganese (Fig. 7B) and folded correctly as determined by CD spectroscopy (Fig. S4), but it nevertheless lost most of its cyanolysis activity, based on the colony assay (Fig. 7A). Moreover, purified H96A had no detectable cyanolysis activity even at pH 6.0 with oxalate buffer in the standard cuvette assay (data not shown). In addition, His96 was mutated to lysine or arginine to introduce a positive charge or aspartate to introduce a negative charge. The H96R mutation resulted in insoluble protein, most likely due to clashes with the metal-binding site, while a lysine or aspartate at this position led to a soluble but inactive enzyme. All variants at position His96 were inactive

in terms of mandelonitrile synthesis (Table 3). Variants of the other histidine residues in the active site, H106A and H106K, were expressed as soluble proteins in amounts comparable to the wild-type, but were inactive in the photometric mandelonitrile assay and the synthesis reaction (Table 3). Some weak activity was detectable in the more sensitive colony-based assay for cyanolysis. Interestingly, mutant H106D showed almost wild-type activity in the colony-based filter assay, and converted 26% of benzaldehyde to mandelonitrile (clearly above the background reaction), but with a low *ee* (Table 3).

3.1.4 DISCUSSION

We have cloned, crystallized and characterized a novel cupin hydroxynitrile lyase biochemically as well as by site-directed mutagenesis. To the best of our knowledge, the data shown here represent the first crystal structure, as well as the first detailed biochemical characterization of a cupin hydroxynitrile lyase. The enzyme has measurable, but relatively low, cyanolysis activity compared to previously described plant hydroxynitrile lyases (Dadashipour and Asano, 2011). Specific activity reached more than 4 U·mg^{-1} at pH 6.0 in the cyanolysis direction (Fig. 2). However, HNLs are distinguished by their ability to catalyse cyanohydrin synthesis from carbonyl and cyanide enantioselectively, rather than by their activity in the cyanolysis direction (Gruber-Khadjawi et al., 2012; Purkarthofer et al., 2007). The enzyme catalyses mandelonitrile formation from benzaldehyde and HCN with high stereoselectivity (Fig. 4). The enzyme thus represents a novel (*R*)-selective hydroxynitrile lyase. The activity was shown to be manganese-dependent in the cyanolysis direction. The manganese content and the activity increased significantly when the protein was expressed in medium supplemented with manganese. Zinc and iron did not significantly bind to the protein *in vivo* during expression in *E. coli*, although the two metal ions are abundant in the medium, but availability in the cytoplasm of *E. coli* is tightly regulated (Andrews et al., 2003; Outten and O'Halloran, 2001). Interestingly, zinc and iron bind to the protein *in vitro* after incubation of purified metal-free protein with either metal. Additionally, if manganese was removed *in vitro* from purified protein using chelating agents, the activity was completely lost. The same result was obtained by mutations in the metal-binding site. Activity was restored by loading the wild-type apoprotein with manganese. Under the significantly different conditions of the synthesis reaction, with high substrate loading and much longer reaction times, a metal

dependence without clear preference for manganese was detected. CD spectra showed that all proteins and variants were correctly folded, indicating that the metal ion is not necessary for overall folding of the protein. However, small changes in the metal-binding site cannot be assessed by this method, and detailed X-ray crystallography and NMR studies are ongoing. Previously, it was reported that the manganese-dependent oxalate decarboxylase from *Bacillus subtilis*, which is a cupin, requires addition of manganese salts to the expression medium when over-expressed in *E. coli* (Tanner et al., 2001). However, in the case of the *Bacillus* enzyme, the manganese salts were required for proper folding. While the data strongly suggest that manganese is the preferred metal for *Gt*HNL, this may only be taken as an indication, as the natural intracellular concentrations of the relevant metals are not known for *Granulicella tundricola*. In nature, cells tightly regulate the concentrations of various metals in their cytoplasm in order to survive, and utilize elaborate mechanisms *in vivo* to ensure that the various metalloproteins incorporate the correct metal (Ma et al., 2009; Waldron and Robinson, 2009). It has been shown previously that a marine cyanobacterium controls incorporation of manganese versus copper in two cupins with identical metal-binding amino acid residues by folding the protein in either the cytoplasm or the periplasm (Tottey et al., 2008). The organism thus achieves selectivity even though the metal-binding site of each enzyme is capable of accepting both metals *in vitro*. This manganese dependence of a hydroxynitrile lyase represents a new catalytic role for manganese, which is present in a wide range of enzymes, such as hydrolases, dismutases, ligases, oxidoreductases and isomerases (Smith et al., 2010). From a chemical perspective, manganese is known to act as a Lewis acid, and has been applied in organometallic catalysis for activation of benzaldehyde (Horike et al., 2008). However, zinc is also known to act as a Lewis acid in enzymatic reactions on carbonyls, e.g. in alcohol dehydrogenases (Parkin, 2004). Moreover, a Zn^{2+}-dependent HNL from *Linum usitatissimum* (*Lu*HNL) with high sequence similarity to alcohol dehydrogenases was discovered many years ago (Trummler et al., 1998). However, no exact reaction mechanism has yet been published. *Lu*HNL and the cupin HNLs do not share any conserved motifs. Even the metal-binding site is different, as the Zn^{2+} ion binds to two Cys and one His residue in *Lu*HNL. The exact reason for the metal preference in *Gt*HNL and the apparent difference in the synthesis and cyanolysis direction remains to be elucidated. In addition to manganese, two histidine residues appear to be necessary for efficient cyanolysis, both of which are found near the metal-binding site. It has been postulated that HNLs require a

positive charge and a general base for activity (Dreveny et al., 2009; Gruber et al., 2004) – functions that may both be fulfilled by histidines under different protonation states.

3.1.5 CONCLUSION

The cupin-fold HNLs appear to constitute a mechanistically distinct class of HNLs with no obvious similarity to previously described enzymes. Further research is necessary to elucidate the exact mechanism of this unusual, manganese-dependent, novel type of hydroxynitrile lyase, as well as to identify the full substrate range of this enzyme.

3.1.6 EXPERIMENTAL PROCEDURES

All chemicals were purchased from Sigma-Aldrich (St Louis, MO, USA) unless specified otherwise.

3.1.6.1 Cloning, mutagenesis and expression

The sequence encoding AciX9_0562 (gene ID 322434201) was obtained as a synthetic gene (GeneArt/Life Technologies, Carlsbad, CA, USA) that was codon-optimized for *E. coli*. The coding region was directly flanked by *Nde*I and *Hin*dIII restriction sites (ThermoScientificBio, Waltham, MA, USA), which were used to clone the gene into the expression vector pET26b(+) (Novagen/Merck, Darmstadt, Germany). Mutations were introduced by overlap-extension PCR, and the PCR products were ligated into vector pET28a(+) using the *Nco*I and *Hin*dIII restriction sites (see Table S1 for primer sequences). The constructs were confirmed by sequencing (LGC Genomics, Berlin, Germany). *E. coli* BL21 -Gold(DE3) (Stratagene, La Jolla, CA, USA) was used as the expression host. All cells were grown in LB (lysogeny broth, Lennox) medium (Carl Roth GmbH, Karlsruhe, Germany) supplemented with kanamycin sulfate (40 mg·L^{-1} final concentration). The protein was routinely expressed from pre-cultures diluted to an attenuance at 600 nm of ~ 0.1, and grown in baffled flasks at 37 °C and 120 rpm until an attenuance at 600 nm of ~ 0.6–0.8 was reached, after which the cultures were cooled to 25 °C, and expression was induced by addition of 0.1 mm isopropyl thio-β-d-galactoside (Carl Roth GmbH). When indicated,

100 μm of MnCl$_2$, FeSO$_4$ or ZnSO$_4$ was added concomitantly with the induction. The induced cultures were harvested after 20–21 h at 25 °C. Protein expression and localization in cell extract fractions were monitored by SDS/PAGE. The percentage of expressed protein in cleared lysates was determined by quantifying the respective band on a Coomassie-stained SDS/PA gel using a G-box HR16 device (Syngene, Synoptics, Cambridge, UK), the software gene snap version 7.05 and Gene Tools version 4.00, using rolling disc baseline correction set at 30. Cells were disrupted by sonification using a Branson sonifier S-250 (80% duty cycle, output control 7), twice for 3 min each, and cooled on ice. The crude lysates were cleared by centrifugation for 1 h at 48 250 *g* and 4 °C.

3.1.6.2 Colony-based filter assay

A colony-based filter assay suitable for high-throughput screening was performed as described previously (Krammer et al., 2007). Four identical spots of each clone were stamped adjacent to each other on a nylon membrane (Biodyne A, 0.2 μm, PALL Life Sciences, Port Washington, NY). The membranes were placed on LB agar (Lennox) plates supplemented with 40 mg·L^{-1} kanamycin, and the cells were grown overnight at 37 °C. The membranes with the colonies were then transferred onto LB agar plates containing 40 mg·L^{-1} kanamycin, 0.1 mm isopropyl thio-β-d-galactoside and 100 μm MnCl$_2$, and the protein was expressed overnight at 20 °C. The colonies were pre-incubated with 100 mm citrate buffer, pH 3.5, after which buffer with 12 mm (*R*)-mandelonitrile (97%) was added. A piece of plastic mosquito net was placed on top of the membranes to prevent wetting, and overlaid by a filter paper (Whatman No. 1, GE Healthcare, Uppsala, Sweden) soaked with copper-(II) ethyl acetoacetate and 4,4′-methylenebis(*N*,*N*-dimethyl-aniline). Development of a blue colour indicates release of HCN. In the case of cleared lysates, the assay was adapted by placing a detection filter paper on top of a 96-well microtitre plate containing reaction mixtures comprising 50 μL cleared lysate and 100 μL 18 mm (*R*)-mandelonitrile in sodium oxalate buffer, pH 4.5.

3.1.6.3 Protein purification

Crude lysates of cells expressing *Gt*HNL (calculated pI = 5.74, according to Protparam (Gasteiger et al., 2005) were produced by lysing cells by sonication (see above) in buffer A (50 mm Bis-Tris/HCl, pH 6.8, with 50 mm NaCl), and cleared by centrifugation (1 h, 48 250 *g*, 4 °C). Wild-type *Gt*HNL and variants were purified in a first step on a Q-Sepharose anion-exchange column (HiTrap™ Q FF, 5 mL, GE Healthcare, Uppsala, Sweden). The column was loaded with ~ 40–50 mg total protein per ml matrix volume. *Gt*HNL was eluted in a single step using 10% buffer B (50 mm Bis-Tris/HCl, pH 6.8, containing 1 m NaCl). The fractions were analysed by SDS/PAGE. Positive fractions were pooled, concentrated using Vivaspin 20 centrifugal filter units (10 000 Da molecular mass cut-off; Sartorius, Göttingen, Germany), and subsequently loaded onto a Superdex 200 16/60 size-exclusion column (GE Healthcare) pre-equilibrated with 20 mm Tris/HCl, pH 7.5, containing 200 mm NaCl.

3.1.6.4 Protein concentration determination

Protein concentrations of cleared lysates were routinely determined using the Bradford assay (Bio-Rad, Hercules, CA, USA). Concentrations of purified proteins were determined using a Nanodrop spectrophotometer (model 2000c, Peqlab, Erlangen, Germany), using an absorbance of 24.3 for a 1% solution (10 $g \cdot L^{-1}$) at 280 nm, as calculated based on the amino acid sequence using Protparam (Gasteiger et al., 2005).

3.1.6.5 Metal exchange

The apoprotein was obtained by dialysing ~ 70–100 mg *Gt*HNL (in ~ 20–50 mL of 20 mm Tris/HCl, pH 7.5, 200 mm NaCl) twice overnight against 500 mL chelating buffer (100 mm sodium acetate, 10 mm 2,6-pyridinedicarboxylic acid, pH 5.5, 200 mm NaCl) and once overnight against 4 L of the original buffer at 4 °C. The protein was reconstituted with metal by incubating 0.15–0.3 mm of purified protein in 20 mm Tris/HCl, pH 7.5, 200 mm NaCl with a tenfold molar excess (1.5–3 mm) of either $MnCl_2$, $FeSO_4$ or $ZnSO_4$ for at least 4 h at room temperature (22 °C). Afterwards, the protein was transferred into metal-free buffer

using PD-10 desalting columns (GE Healthcare). The samples were concentrated using Vivaspin 20 centrifugal filter units.

3.1.6.6 Metal analysis

The metal content was analysed by ICP-OES (Spectro Ciros Vision EOP, Spectro Analytical Instruments, Kleve, Germany) at the Institute of Analytical Chemistry and Food Chemistry, Technische Universität Graz, Austria. Purified water (18 MΩ·cm^{-1}, Barnstead Nanopure, Thermo Fisher Scientific) and high purity acids (HCl and HNO$_3$, Suprapur, Merck) were used throughout. ICP-OES calibration solutions (0.04–10 mg·L^{-1}) were prepared from a 100 mg·L^{-1} multi-element stock (28-element ICP standard solution, Carl Roth) in 3% v/v HNO$_3$. Enzyme samples weighing ~ 0.5 g (corresponding to 500 μL of 10–20 mg·mL^{-1} purified protein solutions in 20 mm Tris/HCl, pH 7.5, 200 mm NaCl) were digested with 1 mL HCl, 2 mL HNO$_3$ and 2 mL H$_2$O by means of a pressurized microwave digestion system (Multiwave 3000, Anton Paar, Graz, Austria). The microwave power was increased to 1400 W over a period of 10 min. During this period, the maximum working pressure of the PFA digestion vessels (HF rotor, 4 MPa) was reached. After a further 20 min of continued heating, the content of the vessels was cooled and diluted to a final volume of 20 mL. Scandium was added to the solutions as an internal standard. The concentrations of Cr, Cu, Fe, Mn, Ni and Zn were determined under robust plasma conditions (Mermet, 1991) by ICP-OES. The ICP was operated at 1350 W RF power, 12.5 L·min^{-1} cooling gas flow, 0.6 L·min^{-1} auxiliary gas flow and 0.83 L·min^{-1} nebulizer gas flow. A standard torch with a 2.5 mm ID injector and a cross-flow nebulizer with a Scott-type spray chamber were used. The following emission lines were used: Cr, 205.552 nm; Cu, 219.226 nm; Fe, 238.204 nm; Mn, 259.373 nm; Ni, 231.604 nm; Sc, 361.384 nm; Zn, 206.191 nm. Metal loading was calculated based on the molar protein concentrations and the measured molar concentrations of the metals. As controls, LB medium, buffer and the flow-through of the Vivaspin 20 centrifugal filter units were analysed.

3.1.6.7 Mandelonitrile cyanolysis assay

The mandelonitrile cyanolysis assay at pH 4.5 was performed in microtitre plates using various final concentrations (3–48 mm) of (*R*)-mandelonitrile as substrate, 50–100 µg of purified protein per well in 50 µL of 20 mm Tris/HCl, pH 7.5, 200 mm NaCl and 100 µL of 100 mm sodium oxalate, pH 4.5, giving a final volume of 150 µL per well. The kinetic data were recorded at 25 °C on a Synergy MX SMATBLD microplate reader (Biotek, Winooski, VT, USA), running gen 5 software version 1.11. The increase in the reaction product, benzaldehyde, was followed at 280 nm and used for the calculation of activities (ε_{280} nm = 1.376 L·mmol^{-1}·cm^{-1}) (Krammer et al., 2007; Okrob et al., 2011). K_m and V_{max} values were calculated using SigmaPlot 11.0 (systat Software, Chicago, IL, USA) with the module 'Enzyme Kinetics 1.3′ based on the Michaelis–Menten model. For determination of the cyanolysis activity at higher pH values, a two-component buffer with 100 mm oxalic acid and 100 mm MES was used, which was adjusted to pH 5.0, 5.5 or 6.0 with sodium hydroxide. The photometric assays at pH 5–6 were performed using 50 µg of purified protein in 500 µL (final volume) in 1 mL quartz cuvettes (1 cm pathlength) using a Beckman DU800 spectrophotometer (Beckman Coulter, Pasadena, CA, USA) without temperature control (the temperature in the cuvettes in the photometer was measured using a standard thermometer at several time points and revealed a constant temperature between 25 and 27 °C). The linear range of the photometer was probed by a standard curve with benzaldehyde, showing that the absorbance is perfectly linear up to a value of 3 ($R^2 = 0.998$).

The substrate stock (90 mm (*R*)-mandelonitrile) was prepared in 3 mm oxalic acid, and 100 µL thereof were added to the cuvettes to start the reactions. The cuvettes were sealed immediately to prevent release of HCN. For all buffers and pH values, the corresponding chemical cyanolysis background values were subtracted. As the assay is monitored at 280 nm, at which proteins also absorb, variants with low activity could not be measured, as the detection limit of the photometer was reached due to the high amount of protein required in the assay.

One mandelonitrile cyanolysis unit was defined as the amount of enzyme that catalyses the release of 1 µmol of benzaldehyde from mandelonitrile dissolved in aqueous buffer in 1 min at 25 °C, under specified buffer and pH conditions.

3.1.6.8 Mandelonitrile synthesis reaction

All experiments involving the possible release of HCN were performed in a well-ventilated hood, in which an HCN detector (Dräger PACIII, Lübeck, Germany) was placed for continuous monitoring. Mandelonitrile synthesis was performed in a two-phase system with a 0.5 mL aqueous phase containing the enzyme (cleared lysate with 40–50 mg·mL^{-1} of total protein or alternatively ~ 20 mg·mL^{-1} of purified enzyme at pH 4.0 in 100 mm sodium acetate buffer) and a 1 mL organic phase comprising 0.5 m freshly distilled benzaldehyde (molar ratio enzyme to benzaldehyde 1 : 1000), 2% v/v triethylbenzene or triisopropylbenzene as internal standard, and 2 m HCN in MTBE. The HCN/MTBE solution was prepared via *in situ* extraction of HCN released from sodium cyanide with HCl, as described previously (Okrob et al., 2011). The reaction mixtures were cooled to 5 °C in a thermomixer and shaken at 1000 rpm. Samples comprising 10 μL of the organic phase were withdrawn at appropriate time points through rubber septa to prevent the release of HCN, and diluted 1 : 100 in HPLC eluent (*n*-heptane/isopropanol/trifluoroacetic acid, 95 : 5 : 0.1). Chiral analysis by HPLC using a Chiracel OB-H column (4.5 μm × 250 mm, Chiral Technologies Europe SAS, Cedex, France) was performed as described previously (Paravidino et al., 2010). Alternatively, a Daicel OD-H column (Chiral Technologies Europe SAS) was used (OD-H 25 0 mm× 4.6 mm × 5 μm; *n*-hexane/isopropanol/trifluoroacetic acid, 96 : 4 : 0.1 ; 20 °C; 1 mL·min^{-1}; 4 MPa; retention times 3.4 min for triisopropylbenzene, 6.7 min for benzaldehyde, 24.7 min for (*S*)-mandelonitrile and 26.2 min for (*R*)-mandelonitrile). The conversion was calculated based on the decrease in the benzaldehyde peak area compared to the peak area of the internal standard. A sample was taken from the organic phase before addition of the aqueous phase, and the peak areas were taken as time point zero. No mandelonitrile was detectable in the starting samples. The enantiomeric excess was calculated by comparing the peak areas corresponding to (*R*)- and (*S*)-mandelonitrile. The analyte content in the aqueous phase compared to the organic phase was assumed to be negligible, and this was confirmed by running a blank reaction with only buffer as the aqueous phase. All experiments were performed in triplicate. Acetophenone (97%) and propiophenone (99%) (both from ACROS Organics/Thermo Fisher, Waltham, MA, USA) were used at 0.2 m concentration (tenfold molar excess of HCN), and the reactions were monitored for conversion in analogy to benzaldehyde.

3.1.6.9 Crystallization and structure determination

*Gt*HNL was purified and crystallized as described previously (Łyskowski et al., 2012). In brief, the initial crystallization condition H02 (10% w/v poly(ethylene glycol) 8000, 20% v/v ethylene glycol, 0.02 m each of the amino acids d,l-Ala, d,l-Lys, d,l-Ser, l-Glu and Gly, 0.1 m MES/imidazole buffer, pH 6.5, Morpheus crystallization screen, Molecular Dimensions, Newmarket, UK (Gorrec, 2009) and the purified protein sample ($\sim 10 \, \text{mg·mL}^{-1}$) were dispensed using an Oryx8 protein crystallization robot (Douglas Instruments Ltd, Hungerford, UK) onto a SWISSCi two-well crystallization plate (Molecular Dimensions). A serial dilution grid of both components was generated using the XStep program (Douglas Instruments Ltd) in order to improve the crystal quality. The best diffracting crystals were harvested under conditions comprising ~ 30% original protein solution and ~ 35% original precipitant solution. For data collection, crystals were frozen without cryoprotection. Three native data sets were collected at the European Synchrotron Radiation Facility (Grenoble, France) at beam lines ID23-2 and ID23-1 (see Table S2). The data were processed using program xia2 (Winter, 2009) in semi-automatic mode using 3d mode (XDS, XSCALE, LABELIT and CCP4 backends (Kabsch, 2010; Winn et al., 2011)), analysed and converted to mtz format using the program POINTLESS (Evans, 2006). The processing statistics as well as data quality analysis were performed using the program Xtriage of the PHENIX suite (Adams et al., 2010). The data processing results for the best dataset are presented in Table 2. Molecular replacement was performed using BALBES (Long et al., 2008), an automated molecular replacement pipeline (version 1.1.5_DB_1.1.5). In total, eight molecules were placed in the asymmetric unit based on template 2F4P, which is deposited in the Protein Data Bank as a cupin-like protein from *Thermotoga maritima* (tm1010). Initial automated model building was performed in PHENIX suite (versions 1.8.1-1168 and 1.8.2-1309) using the AutoBuild wizard. The final model was built manually using the program Coot (version 0.7) (Emsley and Cowtan, 2004) and refined using the programs REFMAC (Murshudov et al., 2011) (version 5.7.0032) and phenix.refine from the PHENIX suite. The atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB ID 4BIF).

3.1.7 Acknowledgements

3.1.8 Supporting Information

3.1.8.1 Material and Methods

Circular dicroism (CD) Spectroscopy: CD spectra were recorded on a Jasco J-715 spectropolarimeter (Jasco Europe, Cremella, Italy). The wavelength range of 190 to 260 nm was scanned with 10 nm min^{-1} and a sensitivity setting of 100 mdeg, at 1 nm bandwidth with a data pitch of 0.2 nm, in a 0.01 cm path length 121.000 QS cylindrical cuvette. The sample comprised 1 mg mL^{-1} purified protein, in 100 mM Tris/Acetate, pH 7.5. For each sample ten spectra were recorded sequentially and the data accumulated. The spectra were analysed using the DICHROWEB analysis webserver [1], using the CDSSTR protocol and reference set 4 (see Table S3) [2].

For assessment of the protein stability at pH 4.0, purified *Gt*HNL (1 mg mL^{-1}) either expressed with the addition of 100 µM MnCl$_2$, or as apoprotein obtained by *in vitro* removal of the manganese, was incubated in 90 mM Na-acetate, pH 4.0. The spectra were followed for 1 h, with a scan every 5 min, without data accumulation.

3.1.8.2 Supplemental Tables

Table S1. Sequences of primers used for cloning and mutagenesis in this study. Sequences of restriction sites are underlined, start and stop codon are indicated with boxes and mutated bases are highlighted in bold letters.

Primer Name (mutation)	Sequence
*Gt*HNL_for (*Nco*I)	AATCA<u>CC</u>ATGGAAATTAAACGTGTTGGTAGCCAGG
*Gt*HN_ rev (*Hind*III)	AATCA<u>AAGCTT</u>TTAGCGACGATACTGTTCATCG
*Gt*HNL-F44V_for	GCAAGCGTTACC**G**TTGAACCGGGTG
*Gt*HNL-F44V_rev	CACCCGGTTCAA**C**GGTAACGCTTGC
*Gt*HNL-T50A_for	CTTTGAACCGGGTGCACGT**GCA**GCATGGCATACCCATCCG
*Gt*HNL-T50A_rev	CGGATGGGTATGCCATGC**TGC**ACGTGCACCCGGTTCAAAG
*Gt*HNL-L61A_for	CATCCGCTGGGTCAGACC**TTT**ATTGTTACCGCAGGTTG
*Gt*HNL-L61A_rev	CAACCTGCGGTAACAAT**AAA**GGTCTGACCCAGCGGATG
*Gt*HNL-H53A_for	CACGTACCGCATGG**GCA**ACCCATCCGCTGGG
*Gt*HNL-H53A_rev	CCCAGCGGATGGGT**TGC**CCATGCGGTACGTG

Primer Name (mutation)	Sequence
*Gt*HNL-H55A_for	CGCATGGCATACC**GCA**CCGCTGGGTCAGAC
*Gt*HNL-H55A_rev	GTCTGACCCAGCGG**TGC**GGTATGCCATGCG
*Gt*HNL-Q59A_for	CCCATCCGCTGGGT**GCA**ACCCTGATTGTTACC
*Gt*HNL-Q59A_rev	GGTAACAATCAGGGT**TGC**ACCCAGCGGATGGG
*Gt*HNL-H94A_for	GTTTAGTCCGGGTGAAAAA**GCA**TGGCATGGTGCAGCACCG
*Gt*HNL-H94A_rev	CGGTGCTGCACCATGCCA**TGC**TTTTTCACCCGGACTAAAC
*Gt*HNL-H96A_for	GGGTGAAAAACATTGG**GCA**GGTGCAGCACCGACC
*Gt*HNL-H96A_rev	GGTCGGTGCTGCACC**TGC**CCAATGTTTTTCACCC
*Gt*HNL-H96D_for	GGTGAAAAACATTGG**GAT**GGTGCAGCACCGACC
*Gt*HNL-H96D_rev	GGTCGGTGCTGCACC**ATC**CCAATGTTTTTCACC
*Gt*HNL-H96K_for	GGGTGAAAAACATTGG**AAA**GGTGCAGCACCGACC
*Gt*HNL-H96K_rev	GGTCGGTGCTGCACC**TTT**CCAATGTTTTTCACCC
*Gt*HNL-H96R_for	GGTGAAAAACATTGG**CGT**GGTGCAGCACCGACC
*Gt*HNL-H96R_rev	GGTCGGTGCTGCACC**ACG**CCAATGTTTTTCACC
*Gt*HNL-H106A_for	GACCACCGCAATGACC**GCA**CTGGCAATTCAGGAAC
*Gt*HNL-H106A_rev	GTTCCTGAATTGCCAG**TGC**GGTCATTGCGGTGGTC
*Gt*HNL-H106D_for	GACCACCGCAATGACC**GAT**CTGGCAATTCAGGAAC
*Gt*HNL-H106D_rev	GTTCCTGAATTGCCAG**ATC**GGTCATTGCGGTGGTC
*Gt*HNL-H106K_for	GACCACCGCAATGACC**TTT**CTGGCAATTCAGGAAC
*Gt*HNL-H106K_rev	GTTCCTGAATTGCCAG**AAA**GGTCATTGCGGTGGTC
*Gt*HNL-Q110A_for	CAATGACCCATCTGGCAATT**GCA**GAACGTCTGGACGGTAAAG
*Gt*HNL-Q110A_rev	CTTTACCGTCCAGACGTTCTGCAAT**TGC**CAGATGGGTCATTG

Table S2. Data collection statistics. The structure was refined against the highest resolution dataset (Native-2).

Data processing statistics				
Data set	Native-1	Native-2	Native-3	Combined
Beam line	ESRF/ ID23-2	ESRF/ ID23-1	ESRF/ ID23-1	-
Wavelength [Å]	0.8726	1.0329	1.0334	-
Resolution range (outer shell) [Å]	70.35- 2.53 (2.60- 2.53)	37.72- 2.46 (2.52- 2.46)	42.65- 2.59 (2.66- 2.59)	70.48- 2.46 (2.52- 2.46)
Space group	C 2 2 21	C 2 2 21	C 2 2 21	C 2 2 21
Unit cell	a=125.32 b=255.05 c=83.11 $\alpha=\beta=\gamma=90.0$	a=126.31 b=254.78 c=82.47 $\alpha=\beta=\gamma=90.0$	a=127.31 b=255.91 c=82.71 $\alpha=\beta=\gamma=90.0$	a=126.31 b=254.78 c=82.47 $\alpha=\beta=\gamma=90.0$
R_{merge}	0.195 (0.847)	0.121 (0.58)	0.143 (0.87)	0.324 (0.595)
R_{meas}	0.210 (0.975)	0.131 (0.227)	0.057 (0.342)	0.333 (0.639)
Mean((I)/sd(I))	8.900 (2.1)	11.3 (3.6)	12.4 (2.7)	10.62 (3.01)
No. of reflecions	311792 (14767)	352885 (26703)	309487 (23288)	967206 (26687)
No. of unique	44697 (3130)	48423 (3563)	43208 (3109)	48657 (3516)
Completeness [%]	99.6 (95.9)	99.3 (99.8)	99.9 (100)	99.8 (99.8)
Multiplicity	7.0 (4.7)	7.3 (7.5)	7.2 (7.5)	19.9 (7.6)

Table S3. Predicted secondary structure composition, calculated from the CD spectra in Fig. S5 using the DICHROWEB web-based analysis tool.

Protein	% Alpha	% Beta	% Turns	% Unordered	% Coverage
*Gt*HNL WT LB + 100μM MnCl$_2$	6	36	24	34	100
*Gt*HNL WT LB no metal addition	4	32	26	36	98
*Gt*HNL Apoprotein	6	37	22	33	98
H96A	7	46	20	26	99
H53A	6	38	22	34	100
H55A	2	40	23	33	98
Q59A	4	39	24	32	99
H94A	4	39	23	33	99

3.1.8.3 Supplemental Figures
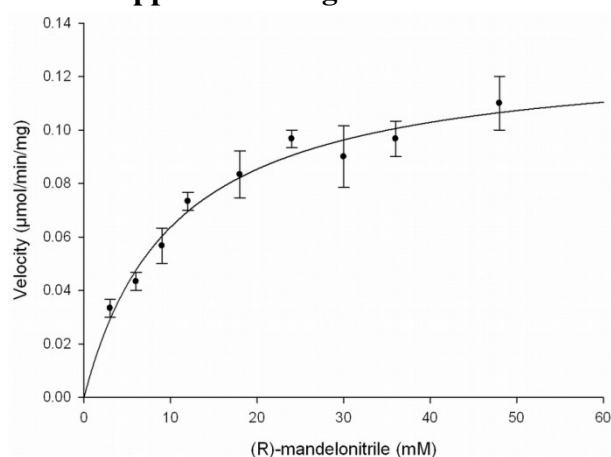
Fig. S1: A: Michaelis-Menten kinetics of purified *Gt*HNL expressed with addition of 100 µM MnCl$_2$. The enzymatic reaction was performed in triplicates in the range of 3 to 48 mM *(R)*-mandelonitrile, and a buffer blank with the same concentration subtracted.

A

B
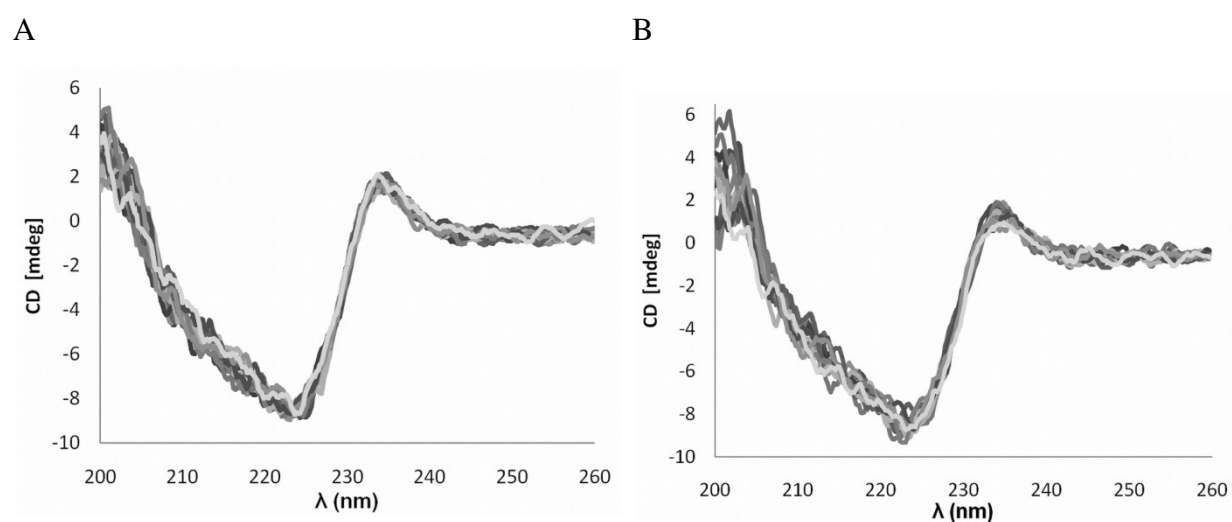
Fig. S2. CD spectra displaying the pH stability of purified *Gt*HNL protein at pH 4.0. Superposition of spectra taken every 5 minutes during incubation at pH 4.0 for 1 h at room temperature. A: *Gt*HNL expressed in the presence of 100 µM MnCl$_2$. B: Apoprotein obtained through *in vitro* removal of the incorporated manganese.
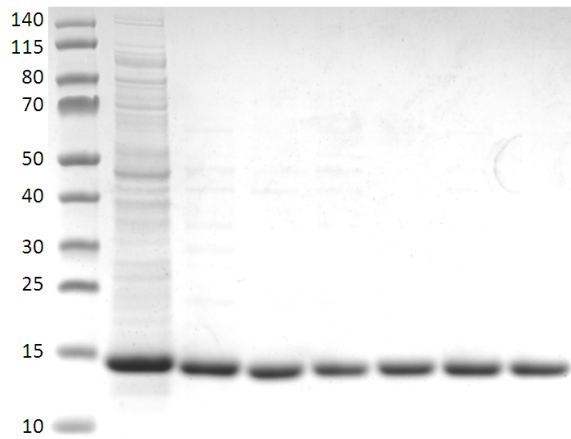
Fig. S3. Cleared lysate (~8 µg, lane 2) and purified protein (~4 µg) of wild-type *Gt*HNL (lane 3) and the mutants H96A, H53A, H55A, Q59A and H94A (lane 4 to 8, respectively). Standard (lane 1): Page Ruler Prestained Protein Ladder (Fermentas, ThermoScientificBio, Waltham, MA, USA), molecular weights in kDa.
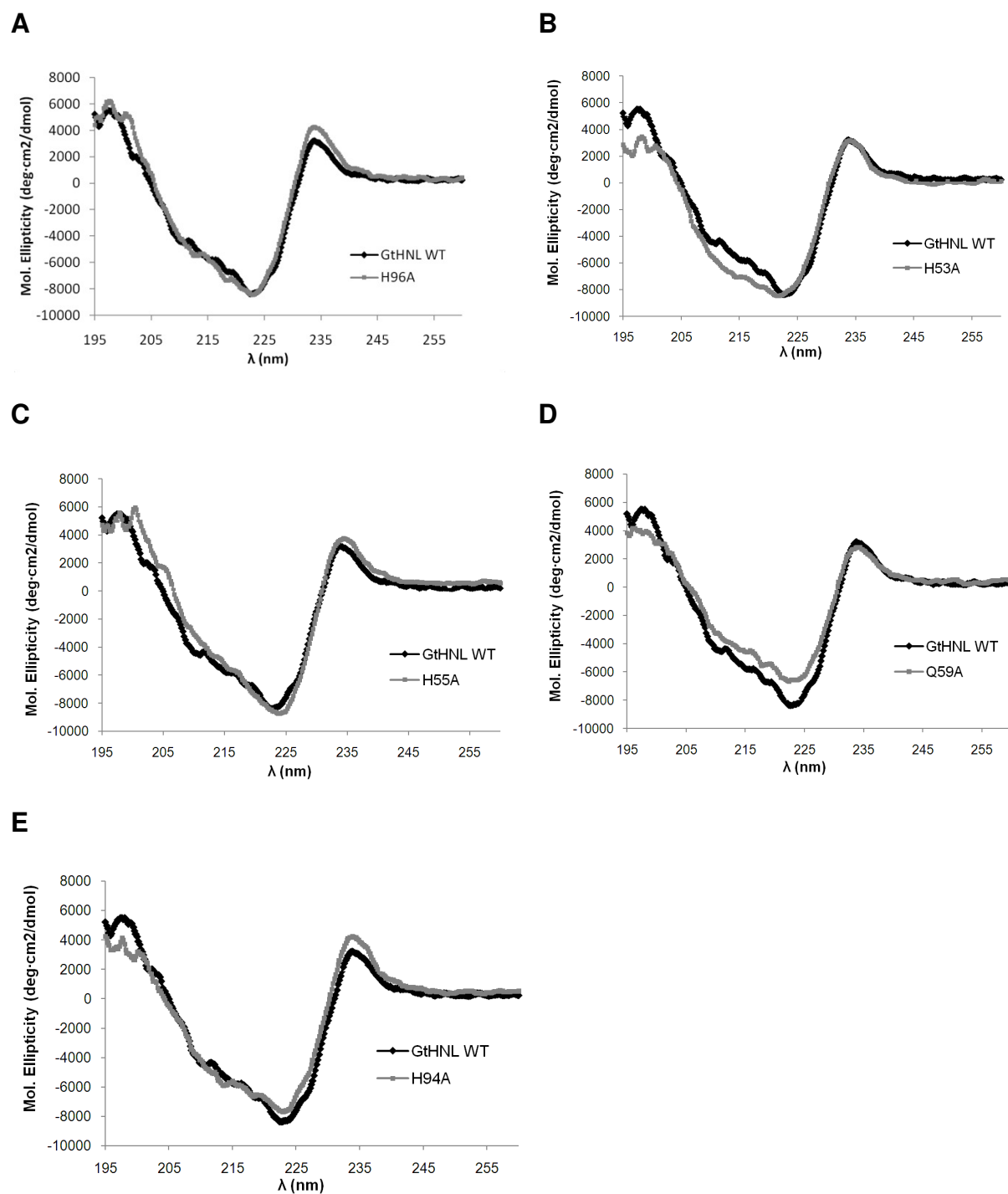
Fig. S4. CD spectra of *Gt*HNL variants superimposed on the CD spectrum of the wild-type protein. **A** H96A (adjacent to active-site metal) **B:** H53A **C:** H55A **D:** Q59A and **E:** H94A (B-E metal-binding).

Supplemental References

1 Whitmore L, & Wallace BA (2008) Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers* **89**, 392-400.

2 Sreerama N, & Woody RW (2000) Estimation of protein secondary structure from circular dichroism spectra: Comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal Biochem* **287**, 252-260.

3.2 Discovery and characterisation of alkene-cleavage activity in the cupin TM1459 from *Thermotoga maritima*

Ivan Hajnal[1], Kurt Faber[1,2], Helmut Schwab[1,3], Mélanie Hall[2,*] and Kerstin Steiner[1,*]

[1]ACIB, Austrian Centre of Industrial Biotechnology GmbH, Petersgasse 14, 8010 Graz, Austria

[2]Institute of Chemistry, University of Graz, Heinrichstraße 28, 8010 Graz, Austria

[3]Institute of Molecular Biotechnology, TU Graz, Petersgasse 14, 8010 Graz, Austria

Author contribution: I have contributed the entire experimental data presented in this paper. A part of the protein purifications was performed by a technician, Karin Reicher, metal contents was analysed by Dr Helmar Wiltsche at the institute of analytical chemistry TU Graz and NMR data were interpreted by Dr Mélanie Hall. The manuscript was written by me.

*Corresponding authors: Dr Mélanie Hall, Institute of Chemistry, University of Graz, Heinrichstraße 28/2, 8010 Graz, Austria phone: +43-316-3805479, e-mail: melanie.hall@uni-graz.at; Dr Kerstin Steiner, Austrian Centre of Industrial Biotechnology GmbH, Petersgasse 14/4, 8010 Graz, Austria, phone: +43-316-8739346, fax: +43-316-8739302, email: kerstin.steiner@acib.at

3.2.1 ABSTRACT

A novel peroxide-dependent alkene-cleavage activity has been discovered in a cupin from *Thermotoga maritima* (TM1459) for which no enzymatic acitivity was known to date. The protein is a metalloenzyme and the reaction rate was increased significantly when manganese was present in the protein or added to reactions, while the addition of iron did not increase activity. The enzyme cleaves a range of substituted styrene derivatives to aldehydes and ketones. The reaction conditions have been optimised regarding temperature, pH, cosolvents as well as oxidant and substrate concentrations. The enzyme produced up to 87% 4-Cl-acetophenone from 50 mM 4-Cl-α-methylstyrene in a two-phae system with ethyl acetate and aqueous buffer. The novel biocatalyst was stable, easy to produce and handle, and this work contributes a potentially very useful novel member for the biocatalytic C=C cleavage toolbox.

3.2.2 INTRODUCTION

Oxidative cleavage of alkenes is a very useful transformation in organic chemistry. It was employed successfully in total synthesis of natural products; however, its use requires extreme conditions involving strong cooling and specialised equipment (Paterson et al., 2005). As with many transformations in organic chemistry, it is possible to use enzymes to conduct reactions under much milder conditions than would otherwise be possible. Enzymatic alkene cleavage activity with molecular oxygen was described and studied in detail in a dioxygenase from *Trametes hirsuta* (Lara et al., 2008; Mang et al., 2006, 2007). In a different system, alkene cleavage was shown to occur as a side-reaction in enzymatic alkene epoxidations with fungal peroxidases using hydroperoxide as oxidant (Tuynman et al., 2000). To our knowledge, the oxidative cleavage of alkene double bonds has not been reported in cupins, while cupins are known to catalyse different oxidative reactions. For example, cupins which oxidatively cleave aromatic rings are well known (Fetzner, 2012). Also, a cupin, which oxidatively cleaves the C-C bond of acetylacetone, was characterised in detail (Straganz et al., 2003). Cupins comprise a large and very diverse superfamily of small, β-barrel proteins with diverse non-enzymatic and enzymatic functions (Agarwal et al., 2009; Uberto and Moomaw, 2013). There is also a very large uncharacterised diversity in cupins. According to the Uniprot database (The UniProt Consortium, 2012), cupin domains were found either alone or in combination with other domains in 99646 proteins as of 2013. Numerous cupins were crystallised in high-

throughput genomics-based crystallisation projects. One such structural genomics project was conducted on the genome of the hyperthermophilic bacterium *Thermotoga maritima* (Lesley et al., 2002). Among the structures solved in this project were also novel cupin proteins with unknown or putative function (Jaroszewski et al., 2004; McMullan et al., 2004). In our work, we looked at the potential of some of these unassigned cupins for biocatalysis. We present the discovery of peroxidase-like alkene-cleavage activity in a previously biochemically uncharacterised manganese-containing cupin from *Thermotoga maritima*, TM1459, whose structure was solved as part of the above-mentioned high-throughput project (Jaroszewski et al., 2004). In peroxidases, alkene cleavage represents an interesting ostensible enzymatic promiscuity and has been described as heme-based (Mutti et al., 2010). Synthetic peroxidase-like sulfoxidation activity has been introduced into streptavidin by the incorporation of manganese-salen complexes (Pordea et al., 2009), which is a powerful approach to introduce novel activities into protein scaffolds (Ringenberg and Ward, 2011). Similar oxidative activity could also be introduced into xylanase by a similar manganese-complex incorporation approach, with alkene epoxidation as main activity and C=C bond cleavage as a minor reaction (Allard et al., 2012). In a different approach, manganese-loaded carbonic anhydrase also displayed peroxidase-like activity, including especially the epoxidation of styrenes (Fernández-Gacio et al., 2006; Okrasa and Kazlauskas, 2006). However, in one of the studies on manganese-substituted carbonic anhydrase, the authors explicitly reported a lack of C=C bond cleavage products (Okrasa and Kazlauskas, 2006). The enzyme TM1459 showed a higher selectivity towards alkene cleavage versus epoxidation when compared both to known peroxidase systems and manganese-protein based systems (Allard et al., 2012; Fernández-Gacio et al., 2006; Okrasa and Kazlauskas, 2006; Tuynman et al., 2000), and hence constitutes a novel alkene cleavage enzyme. The novel cupin biocatalyst was stable and easy to produce in high yield in the simple-to-handle host *E. coli*, in comparison to many typical peroxidases (Conesa et al., 2002), which makes it a very useful addition to the biocatalytic C=C cleavage toolbox.

3.2.3 RESULTS AND DISCUSSION

3.2.3.1 Sequence similarity of TM1459 to other proteins

A protein blast search (Altschul et al., 2005) with the amino acid sequence of TM1459, yielded only few matches with high sequence identity to this protein (Figure S1). The proteins with the highest sequence identity originate from organisms closely related to *Thermotoga maritima*, and none of these has a known function. The closest match with any predicted function is a putative carbohydrate binding protein from *P. furiosus* DSM 3638 with 44% identity, and the closest match with a known, biochemically characterised function is the polyketide cyclase RemF from *Streptomyces resistomycificus* with 37% sequence identity. It can thus be said that this protein has no close relatives with a known biochemical function, and the data presented in this work may also contribute to the elucidation of functions in other cupins.

3.2.3.2 TM1459 exhibits oxidative alkene cleavage

Initial experiments showed that crude lysate of *E. coli* cells expressing TM1459 cleaved 4-Cl-α-methylstyrene in the presence of *tert*-butyl hydroperoxide yielding 4-Cl-acetophenone, with significantly higher conversion than the background reaction in the presence of cleared *E. coli* lysate without TM1459. Since this result was obtained without any optimisation, the reaction conditions and enzyme preparations were further optimised as described below. The background conversion was small but clearly measurable. In order to further suppress the background reaction and to use better defined conditions, the protein was purified (Figure S2) and all further experiments were performed with purified protein.

3.2.3.3 Effect of temperature and hydroperoxide concentration

Thermotoga maritima, which is the natural source of this enzyme, is an obligate hyperthermophile and initial experiments were conducted at 50°C. Since heterologously expressed and purified proteins often behave differently than the same protein in a living cell, and protein stability and activity do not necessarily correlate, the effect of temperature was tested at 30°C, 50°C and 70°C. The best conversion of 4-Cl-α-methylstyrene to 4-Cl-acetophenone was observed at 30°C (Figure 1). The *tert*-butyl hydroperoxide (*t*-BuOOH)

concentration was varied between a concentration equimolar to the substrate and a double and triple molar excess over the substrate. A significant improvement was observed between the equimolar and the double molar hydroperoxide concentration, and a small further increase with triple molar excess of hydroperoxide (Figure 1). The highest conversion was achieved at 30°C with a triple molar excess of hydroperoxide, with a product peak area of **1b** of 89% of total peak areas.
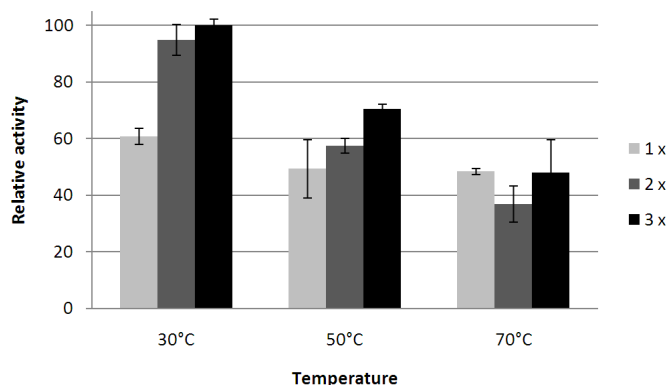
Figure 1: Effect of temperature and hydroperoxide concentration on the enzymatic cleavage of **1a** to **1b**. Relative activity (relative peak area of product **1b**, highest activity as 100%, product peak 89% of total peak areas) of TM1459 (1 mg mL^{-1}, ~0.08 mM) in 50 mM NaPi pH 7.0 at 30, 50 and 70°C, in dependence of hydroperoxide concentration. Conditions: 15 mM substrate (2 µL neat substance) **1a**; 15, 30 and 45 mM *t*-BuOOH (1 x, light grey bars; 2 x, grey bars; or 3x, black bars, molar excess over substrate), the reactions were shaken at 1000 rpm for 20 h.

The background reaction was also tested with 0.1 mM MnCl$_2$ in the same buffer as the enzymatic reactions. The conversion was significantly lower than with the enzyme. The background reaction was somewhat higher at higher temperatures whereas the enzymatic reaction was highest at lower temperatures (Figure S3), which is consistent with a purely chemical versus an enzymatic reaction. The reactions with 0.1 and 0.2 mM MnCl$_2$ did not show a significant difference (Figure S3), which further indicated that manganese salts alone did not act as efficient catalysts for the reaction under the described conditions.

3.2.3.4 Optimisation of buffer pH

The effect of the pH was tested in 50 mM sodium phosphate buffers with pH values of 5 to 9. The relative enzymatic activity increased strongly between pH 5 and 7, without significant further change between pH 7 and 9 (Figure 2). The enzyme thus preferred neutral or slightly basic pH, which is in contrast to the pH profile reported for the promiscuous activity for alkene cleavage described in horseradish peroxidases (Mutti et al., 2010), which was highest at low pH values, underscoring the difference between the two systems. The background reaction with 0.1 mM $MnCl_2$ was low and was not significantly affected by pH values within the tested range (data not shown). All further experiments were conducted in phosphate buffer at pH 7.
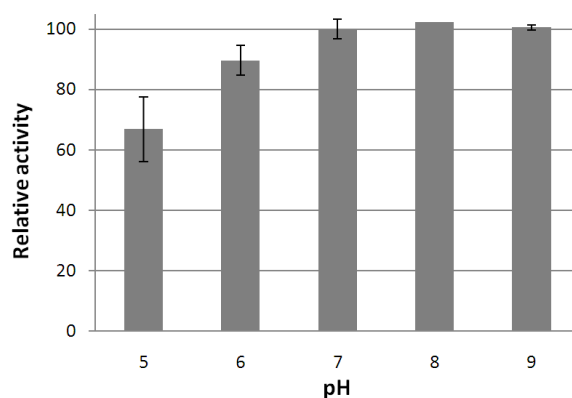
Figure 2: Effect of pH on the enzymatic cleavage of **1a** to **1b**. Relative activity (relative peak area of product **1b**, pH 7.0 as 100%, product peak 86% of total peak areas) of TM1459 (1 mg mL^{-1}, ~0.08 mM) in 50 mM NaPi at different pH values. Conditions were as follows: 10 mM substrate **1a** (as 10 µL of 1 M stock in ethanol) 30 mM t-BuOOH, shaken at 30°C and 1000 rpm for 20 h.

3.2.3.5 Effect of organic solvents

To test the effect of organic solvents on the enzymatic reaction, and also to increase the concentrations of the highly hydrophobic substrates, the reactions were conducted in different two-phase systems with the addition of cyclohexane, or ethyl acetate. The substrate concentration of 4-Cl-α-methylstyrene was varied at 10, 50 and 100 mM. Ethyl acetate and cyclohexane were varied at 25 and 50 vol%. The ethyl acetate two-phase system had

consistently better conversions and was further tested at 10 and 15 vol% organic phase. The best conversion was obtained with 15% ethyl acetate and 50 mM substrate (Figure 3), and this combination was used to test further substrates as described below.
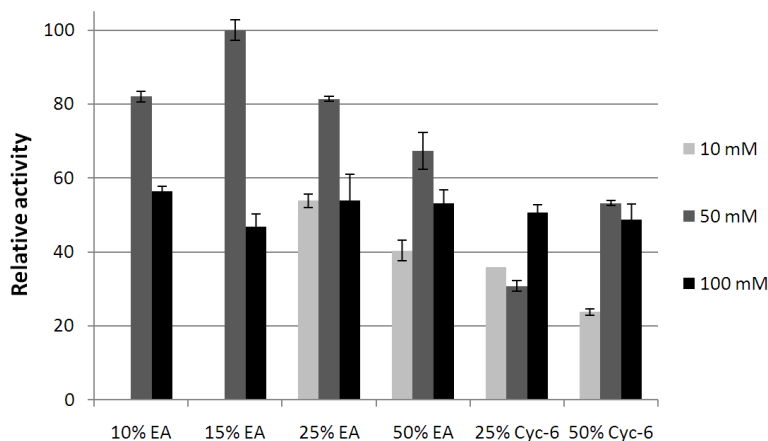
Figure 3. Relative activities (relative peak area of product **1b**, highest activity as 100%, product peak 78% of total peak areas) of 1 mg mL^{-1} TM1459 in two-phase systems in the presence of ethyl acetate (EA) or cyclohexane (cyc-6) at the indicated percentage of total volume (lower axis). Substrate was tested at a concentration of 10 mM (light grey), 50 mM (grey), or 100 mM (black bars). The *tert*-butyl hydroperoxide concentration was always three times than that of the substrate. The reactions were shaken at 1000 rpm and 30°C for 20 h.

3.2.3.6 Oxidant preference

The enzyme was not active with molecular oxygen at atmospheric concentration or hydrogen peroxide at the same concentration as *tert*-butyl hydroperoxide (data not shown). It is interesting that the enzyme was not active with H_2O_2, since in many cases reactions including alkene cleavage and epoxidation on highly similar substrates, proceeded more or less efficiently with this well-established oxidant (Fernández-Gacio et al., 2006; Okrasa and Kazlauskas, 2006; Tuynman et al., 2000). Activity was observed only with *tert*-butyl hydroperoxide. The enzyme was very stable in the presence of this oxidant, and was active in up to 150 mM *t*-BuOOH in a two-phase system. To rule out the involvement of atmospheric oxygen, the reaction was performed under argon, and the GC-MS data confirmed a similar conversion of **1a** to **1b** with high selectivity (data not shown). This strongly indicated that molecular oxygen was not needed for the reaction in stoichiometric amounts. Small traces of

oxygen cannot be excluded with our setup, but since the result was comparable to the standard reactions and since no reaction was observed without hydroperoxide, molecular oxygen is most likely not needed for the reaction.

3.2.3.7 Metal dependence

The authors of the TM1459 crystal structure assigned the protein as a manganese-containing cupin with one metal-binding site per monomer based on crystallographic features (Jaroszewski et al., 2004). To investigate manganese binding analytically, we expressed the enzyme with or without the addition of 100 µM $MnCl_2$ to the expression medium, and the purified proteins were analysed by ICP-OES. The enzyme expressed without manganese bound only 22% zinc and 13% iron, with no other metals detected (Figure 4). The addition of manganese to the expression medium was necessary for efficient manganese uptake, and the protein expressed in LB medium with 100 µM $MnCl_2$ bound 37 ± 4 % manganese atoms per enzyme monomer (Figure 4). The total metal loading after purification is less than 50%, which is significantly lower than full occupancy, but is in line with what has been reported previously for cupins highly overexpressed in *E. coli* (Barney et al., 2004; Hajnal et al., 2013; Merkens et al., 2008). We analysed the flow-through of 10 kDa cut-off concentrators, and no metals were detectable. This strongly indicated that the metal was bound to the protein. Interestingly, the enzyme expressed in the absence of manganese accumulated significantly more zinc, whereas the addition of manganese drastically reduced zinc binding. Taken together, these data indicate that the protein indeed prefers manganese to other intracellular metals if sufficient amounts are provided.
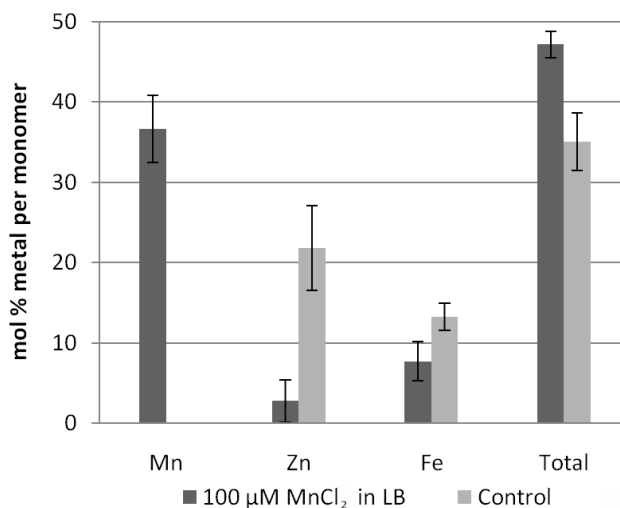
Figure 4. Metal loading (determined by ICP-OES) of purified protein, which was expressed in LB medium either with (dark grey) or without (light grey bars) addition of 100 µM MnCl$_2$.

The cleavage of 4-Cl-α-methylstyrene (**1a**) with *tert*-butyl hydroperoxide to 4-Cl-acetophenone (**1b**) proceeded significantly more efficiently with 1 mg mL^{-1} purified protein (~ 0.08 mM) expressed in the presence of manganese in the medium than with an equal amount of protein expressed without metal addition (Figure 5). The protein expressed without manganese was activated by the addition of 0.1 mM manganese (II) chloride or manganese (III) acetate to the reactions, but not by the addition of 0.1 mM iron (II) sulphate or iron (III) chloride (Figure 5). Additionally, none of the metal salts in buffer without protein, or in combination with 1 mg mL^{-1} BSA (bovine serum albumin) catalysed the reaction under the tested conditions (Figure S4). The experiment was performed in parallel with all components diluted from the same stock solutions for all reactions. The results indicated that both the cupin protein and manganese were needed in combination to catalyse the reaction. Under the mild conditions used in the experiment, both a simple metal catalysis and non-specific protein catalysis could be excluded. The results of these control experiments were especially important considering that manganese and other transition metals are well known catalysts in chemical catalysis of oxidative reactions, using for example hydrogen peroxide (Lane and Burgess, 2003). Since the total metal loading was comparable in both cases but the activity was significantly higher with manganese, it can be concluded that the enzyme clearly shows higher activity with manganese compared to iron and zinc.
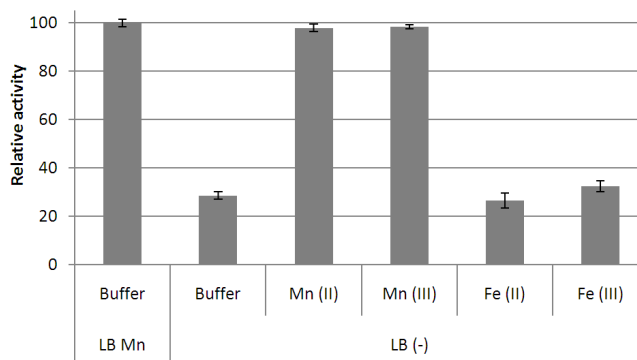
Figure 5. Relative activities (relative peak area of product **1b**, highest activity as 100%, product peak 91% of total peak areas) of 1 mg mL^{-1} purified ($\sim$ 0.08 mM) TM1459 expressed with and without addition of 100 μM MnCl$_2$ to the expression medium (LB Mn and LB (-), respectively), as well as with the protein expressed without metal addition, with the addition of 0.1 mM MnCl$_2$, Mn (III) acetate, FeSO$_4$ or FeCl$_3$ to the reaction mixtures. Conditions were as follows: 10 mM substrate, 30 mM t-BuOOH, 50 mM NaPi, pH 7.0; the reactions were shaken at 30°C and 1000 rpm for 20 h.

3.2.3.8 Substrate scope

The enzyme cleaved various styrene derivatives - alkenes with a double bond conjugated to an aromatic system (**1a to 10 a**), to their aldehyde or ketone derivatives (**1b to 10b**; Scheme 1) with varying activity (Table 1).
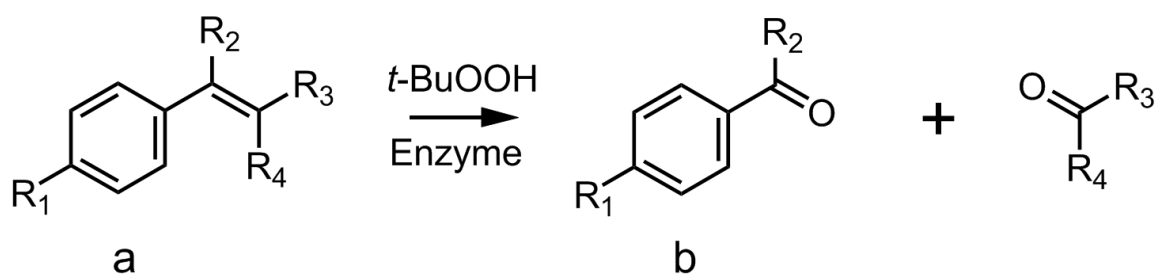
Scheme 1. TM1459 catalysed alkene cleavage of styrene substrates **a** to aldehydes and ketones **b**

The alkene cleavage proceeded without detectable side-products in the case of 4-Cl-α-methylstyrene **1a** and α-methyl styrene **2a**, yielding 4-Cl-acetophenone **1b** and acetophenone **2b**, respectively, with a high conversion reaching up to 87 %. In the case of styrenes without substituents at the vinyl moiety (**3a-7a**), the most significant products were the aldehydes **3b-7b**, with a significant over-oxidation to the acid. The over-oxidation to the acid was significantly reduced (up to 10-fold) in the two-phase system as compared to the aqueous microemulsion system (Table 1). While some of the more activated substrates showed detectable conversion in the control reactions without enzyme, this conversion was significantly lower (Table S3). Importantly, the enzyme was also able to catalyse alkene cleavage with substrates where no background conversion was detectable in our setup (Table S3).

Table 1: The product distribution of the oxidative conversions of various substituted styrenes **a** used in this study are listed as mol% of product based on mol initial substrate. Data are given for an aqueous microemulsion system (1-phase) with 10 mM substrate and for a two-phase system with 15% ethyl acetate with 50 mM substrate (2-phase).

	Substituents				% Carbonyl		% Acid	
Substrate (a)	R1	R2	R3	R4	1-phase	2-phase	1-phase	2-phase
1	Cl	Me	H	H	48	87	n.d.	n.d.
2	H	Me	H	H	25	76	n.d.	n.d.
3	H	H	H	H	1	10	7	2
4	Cl	H	H	H	2	16	19	1
5	MeO	H	H	H	2	9	n.d.	n.d.
6	NH2	H	H	H	n.d.	n.d.	5	1
7	Me	H	H	H	n.d.	14	n.d.	1
8	H	H	Me	H	2	4	10	n.d.
9	MeO	H	Me	H	1	6	22	2
10	H	H	Me	Me	1	14	14	n.d.
11	H	H	NO2	H	n.d.	n.d.	n.d.	n.d.
12	H	H	NO2	Me	n.d.	n.d.	n.d.	n.d.

*n.d. - none detected

Interestingly, a methyl substituent at the β-position of the styrene, as in 1-phenyl-1-propene **8a**, led to a shift in the chemoselectivity towards the production of epoxide of substrate **8**. The same was true in the case of two methyl substituents at the β-position, as in substrate **10a**. However, while the conversion to epoxide was clearly detectable on GC-MS, it was very low. As additional side-reaction, there is a non-negligible loss of substrate from the mass balance as measured in GC that can be explained as a likely polymerisation of the styrene substrate in the presence of hydoperoxide over the long reaction time. A loss of substrate was observed both in the enzymatic reactions and the controls. This side-reaction was less pronounced in the two-phase system (data not shown). With methyl cinnamate, a small amount of product with excellent GC-MS data correlation to its epoxide was detected in the enzymatic reactions and not in the control reactions. However, the product amount was very low and this conversion was not pursued further. A nitro group at position R3, as in *trans*-β-nitrostyrene and *trans*-β-methyl-β-nitrostyrene completely impaired conversion with the enzyme. Aliphatic alkenes were not suitable substrates for the enzyme, even in the case of the alicyclic substrate 4-vinylcyclohexene, the size and structure of which are very similar to styrene. However, aliphatic alkenes are often more difficult substrates than conjugated aromatic alkenes, even in purely chemical conversions (Lane and Burgess, 2003), so this result was not surprising. The enzyme showed no activity towards phenylacetylene, which was also not surprising due to the high stability of the C≡C triple bond. Illustrative examples of substrates not converted by the enzyme are shown in Figure 6. Further substrates, which also showed no conversion, are listed in the supporting material (Table S4).
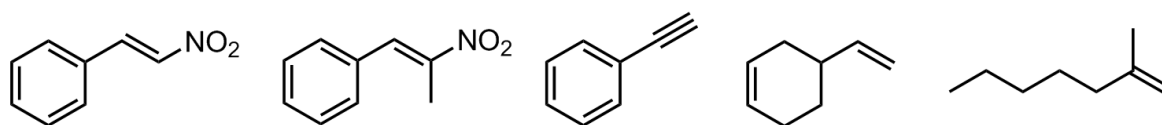
Figure 6. Examples of substrates not suitable for TM1459: The enzyme showed no activity on styrenes with bulky nitro substituents, on aliphatic and alicyclic alkenes or on the triple-bond of phenylacetylene.

In a system with a manganese-containing carbonic anhydrase, there are no reports on aldehydes and ketones as side-products (Okrasa and Kazlauskas, 2006, Fernández-Gacio et al., 2006). On the other hand, in the heme-containing *Coprinus cinereus* peroxidase, benzaldehyde was observed as a major product from styrene in conversions with hydrogen peroxide (Tuynman et al., 2000). Okrasa and Kazlauskas speculated that the observed lack of aldehyde side-products in manganese-containing carbonic anhydrase-catalysed alkene epoxidation stems from a difference in mechanism compared to typical heme peroxidases (Okrasa and Kazlauskas, 2006). In our system, which also contained a protein-bound manganese ion, we observed aldehydes or ketones as major products. Large differences can be reasonably expected between the catalytic mechanisms of a heme peroxidase and a manganese-metalloenzyme, regardless of whether the manganese enzyme is synthetic or natural. It is perhaps the difference in product distribution between manganese-containing TM1459 and the manganese-carbonic anhydrase that is more striking than the obvious difference to heme peroxidase. This is especially interesting considering that in both cases the presumed active site contains a manganese ion coordinated only by histidine residues however with three and four ligands to the metal, respectively (Eriksson et al., 1988; Jaroszewski et al., 2004). In the second study of manganese-containing carbonic anhydrases as epoxidation catalysts, the authors reported high epoxide yields, much higher than what we observed with some of the same substrates. The enzyme reported here thus showed a high propensity towards alkene cleavage compared to both known peroxidases and artificial manganese-containing epoxidising enzymes. This sets it apart as a novel peroxide-driven alkene cleavage enzyme.

3.2.4 CONCLUSION

We report the discovery of peroxidase-like activity towards the cleavage and epoxidation of alkenes in a so far uncharacterised cupin protein. The enzyme was active towards a range of substituted styrenes with *tert*-butyl hydroperoxide as oxidant. The alkene cleavage was highest with 4-Cl-α-methylstyrene with up to 87% conversion with 50 mM substrate and 150 mM hydroperoxide. The enzyme is small, stable, and easily purified in high yield. Thus, the discovery of an easy to handle alkene-cleaving cupin contributes a useful new biocatalyst and furthers our understanding of both biocatalytic alkene cleavage and of the diversity of reactions catalysed by cupins.

3.2.5 Experimental Section

3.2.5.1 Cloning, expression and purification

The sequence encoding TM1459 from *Thermotoga maritima* (Swissprot: Q9X1H0, European nucleotide archive: EHA59553.1) was ordered as synthetic gene, codon optimised for *E. coli* (GeneArt, Life Technologies, Carlsbad, CA, USA). The synthetic gene was cloned into the expression vector pET26b(+) (Novagen, Merck KGaA, Darmstadt, Germany) using restriction sites for the enzymes *Nde*I and *Hin*dIII (ThermoScientificBio, Waltham, MA, USA) and the correct cloning was confirmed by sequencing (LGC Genomics, Berlin, Germany). *E. coli* BL21-Gold (DE3) was used as expression host (Stratagene, La Jolla, CA, USA). The cells were grown in Lysogeny Broth Lennox medium (Carl Roth GmbH, Karlsruhe, Germany), supplemented with 40 mg L^{-1} kanamycin sulphate (Carl Roth). For protein expression, overnight cultures were diluted to an OD_{600} of ~0.1 in fresh LB medium with 40 mg L^{-1} kanamycin, and grown in baffled flasks at 37°C and 120 rpm to an OD_{600} of ~0.6-0.8. Expression was induced by the addition of 0.1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG, Carl Roth) after cooling the cultures to 25°C, and 100 μM of $MnCl_2$ was added to the LB medium where indicated. The induced cultures were harvested after 20-21 h at 25°C. Protein expression and solubility were monitored by SDS-PAGE. For initial screenings, lysates of cells expressing TM1459 and of cells carrying the control vector were produced by sonicating the cells after harvesting by centrifugation (20 min, 2800 x g) in 10 mM NaPi, pH 7.0, two times for 3 min on ice (Branson Sonifier S-250, set to 80% duty cycle, and output control 7). The crude lysates were cleared by centrifugation (1 h, 48000 x g, 4°C).

The purification procedure was adapted from a protocol published by our group for an unrelated cupin (Hajnal et al., 2013), whereby only the pH of the buffer was changed due to the different isoelectric point (pI) of TM1459 (calculated pI of 5.57 according to Protparam, http://web.expasy.org/protparam/ (Gasteiger et al., 2005)). Cells were lysed by sonication (see above) in Buffer A (50 mM Bis-Tris/HCl, pH 6.6, 50 mM NaCl), and cleared by centrifugation (see above). TM1459 was eluted with 10% buffer B (50 mM Bis-Tris/HCl, pH 6.6, 1 M NaCl). Size exclusion chromatography was performed on a Superdex 75 HiLoad 16/600 size-exclusion column (GE Healthcare, Uppsala, Sweden) pre-equilibrated with 50 mM NaPi, pH 7.0, 100 mM NaCl. The purest peak fractions of the S75 size-exclusion step were pooled and brought to a protein concentration of 20 mg mL^{-1} using the same Vivaspin

20 Centrifugal Filter Units as above, yielding the final enzyme preparation used in the experiments.

3.2.5.2 General Procedure for Bioconversions in an Aqueous Microemulsion System

Purified TM1459 protein (20 mg mL^{-1}, 50 µL) in the storage buffer (50 mM NaPi, pH 7.0, 100 mM NaCl) was diluted to a final concentration of 1 mg mL^{-1} with 50 mM NaPi, pH 7.0, with *tert*-butyl hydroperoxide pre-dissolved in the buffer to yield a final concentration of 30 mM. The reactions were started by adding a 10 µL drop of 1 M substrate stock in ethanol to a final reaction volume of 1 mL and a final concentration of 10 mM and starting the agitation. The reactions were conducted for 20 h at 30°C and 1000 rpm in a thermomixer (comfort, Eppendorf, Hamburg, Germany). Reactions were stopped by turning off the agitation and extracted two times with 500 µL ethyl acetate. The combined organic phases were dried over Na$_2$SO$_4$, and used directly as samples for the GC and GC-MS analysis.

3.2.5.3 General Procedure for Bioconversions in an Aqueous – Ethyl acetate two-phase System

Purified TM1459 protein (20 mg mL^{-1}, 50µL) in the storage buffer (50 mM NaPi, pH 7.0, 100 mM NaCl) in 2 mL microreaction tubes was diluted with 800 µL reaction buffer (50 mM NaPi, pH 7.0. The reactions were started by adding 150 µL of ethyl acetate containing dissolved substrate and *t*-butyl hydroperoxide to yield a final reaction volume of 1 mL and final concentration of 50 mM substrate and 150 mM hydroperoxide, and starting the agitation. The reactions were conducted in the same way as for the aqueous microemulsion system (shaken for 20 h at 30°C and 1000 rpm). Reactions were stopped by quenching unreacted hydroperoxide with 50-100 mg of sodium bisulfite, adding 350 µL of ethyl acetate to yield 500 µL of organic phase, and extracting a second time with another 500 µL ethyl acetate. GC-MS and GC analysis was conducted in the same way as described for the general aqueous microemulsion system.

3.2.5.4 Synthesis of Standard Reference Material for the Epoxides of 8 and 10.

The epoxide standard materials for epoxides of substrates **8** and **10** were prepared by an adaptation of a previously published procedure (Mang et al., 2006). To a heterogeneous stirred mixture of alkenes **8a** and **10a** (1 mmol) and Na_2HPO_4 (2 mmol) in CH_2Cl_2 (5 mL in 10 mL round-bottom flask) cooled on ice, powdered 3-Cl-perbenzoic acid (67%) was added in portions over 30 min. The reaction was stirred at room temperature for further 3 h. Afterwards, $NaHSO_3$ (1.5 mmol) was added to destroy residual non-reacted 3-chloroperbenzoic acid and the reaction mixture was washed with $NaHCO_3$ (saturated aqueous solution, 5 mL) and extracted with CH_2Cl_2 (3 x 10 mL). The combined organic phases were dried over Na_2SO_4 and finally evaporated, yielding the crude epoxide as slightly yellow oil or a waxy solid. The epoxide product was purified by flash chromatography (Silica Gel Merck 20, petroleum ether/EtOAc 100:1) and the purity and identity confirmed by GC-MS and NMR. **2,2-Dimethyl-3-phenyl-oxirane**: ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.38-7.25 (m, 5H, Ar-C*H*), 3.89 (s, 1H, $(CH_3)_2$CC*H*), 1.51 (s, 3H, C*H$_3$*), 1.10 (s, 3H, C*H$_3$*') ppm. **2-Methyl-3-phenyl-oxirane**: ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.42-7.19 (m, 5H, Ar-C*H*), 3.60 (d, 1H, CH_3C*H*CH, J=1.9 Hz), 3.10-3.02 (qd, 1H, CH_3CHC*H*, J=5.1 Hz, J= 2.1 Hz), 1.48 (d, 3H, C*H$_3$*, J=5.1 Hz) ppm.

3.2.5.5 Metal Analysis

The analysis of metal content was conducted by an external laboratory at the Institute of Analytical Chemistry and Food Chemistry, TU Graz. The metals were analysed using ICP-OES (inductively coupled plasma – optical emission spectroscopy), using exactly the same setup as has been described previously (Hajnal et al., 2013). Metal loading was calculated using the molar concentration of the purified proteins (Extinction of a 1% (10 g L^{-1}) solution at 280nm of 6.47 according to Protparam), and the metal concentrations measured with ICP-OES. As controls the flow-through fractions of the Vivaspin 20 Centrifugal Filter Units were analysed.

3.2.6 Acknowledgements

3.2.7 Supporting information

3.2.7.1 Chemicals

All chemicals except the epoxides of **8** and **10** were purchased from commercial sources and were used without further purification. Styrene (99%), 4-chlorostyrene (97%), α-methylstyrene (99%), *trans*-anethole (99%), 2-methyl-1-phenylpropene (99%), benzaldehyde (99%), 4-chlorobenzaldehyde (97%), 4-methoxybenzaldehyde (98%), acetophenone (99%), isoeugenol (98%), methyl-cinnamate (99%), *trans*-β-methyl-β-nitrostyrene (99%), 4-vinylcyclohexene (99%), 3-chloroperbenzoic acid (67%) and *tert*-butyl hydroperoxide (70% aqueous) were purchased from Aldrich (Sigma-Aldrich, St. Louis, MO, USA); 4-methylstyrene (99%), *trans*-1-phenyl-1-propene (98%), phenylacetylene (97%) and β-carotene (99%) were obtained from Fluka (Sigma-Aldrich); 4-chloroacetophenone (97%), 4-chloro-α-methylstyrene (95%), 4-methoxystyrene (98%), *para*-aminostyrene (97%) and 4-acetoxystyrene (95%) were purchased from Lancaster (Alfa Aesar, Ward Hill, MA, USA).

3.2.7.2 Optimisation of Temperature and Hydroperoxide Concentration

To a solution of purified protein (1 mg mL^{-1}) in 50 mM NaPi, pH 7.0, with 15, 30 or 45 mM *tert*-butyl hydroperoxide, a final concentration of 15 mM 4-Cl-α-methylstyrene was added as a drop of neat substrate without solvent to start the reactions. The reactions were conducted for 20 h at 30, 50, or 70°C and 1000 rpm in a thermomixer (comfort, Eppendorf). Background controls were performed with the same amount of buffer (50 mM NaPi, pH 7.0, 100 mM NaCl) with the addition of either 0.1 mM or 0.2 mM MnCl$_2$ in the reaction mixture. GC analysis was conducted as described below.

3.2.7.3 Screening of conditions for two-phase systems

Purified TM1459 protein (20 mg mL^{-1}, 50 µL) in the storage buffer (50 mM NaPi pH 7.0 100 mM NaCl) in 2 mL microreaction tubes was diluted with an appropriate volume of reaction buffer (50 mM, NaPi pH, 7.0). The reactions were started by adding 100-500 µL of substrate and *t*-butyl hydroperoxide either pre-dissolved in ethyl acetate, hexane, or cyclohexane, to yield a final reaction volume of 1 mL and final concentrations of 10, mM or 100 mM of substrate and 30, 150 or 300 mM of hydroperoxide, respectively, and starting the agitation. The reactions were conducted in the same way as for the aqueous microemulsion system (shaken 20 h at 30°C and 1000 rpm). Reactions were stopped by quenching unreacted hydroperoxide with 50-100 mg of sodium bisulfite, adding an appropriate amount of ethyl acetate to yield 500µL organic phase, and extracting a second time with 500 µL of ethyl acetate and the combined organic phases were dried and used for GC analysis.

3.2.7.4 Analytical Procedures

For qualitative identification of products, GC-MS analysis was carried out on a 7890A GC System (Agilent Technologies, Santa Clara, CA, USA), equipped with 5975C inert XL MSD mass spectrometer (Agilent Technologies); the following temperature methods were used for GC-MS: A: 0.5 min equilibration time; 100°C for 0.5 min, ramp of 10°C/min to 300°C, and final 300°C 0.5 min; B: 0.5 min equilibration time; 50°C for 1 min, ramp of 20°C/min to 270°C, hold 270°C for 1 min, and final 270°C for 0.5 min; C: 0.5 min equilibration time; 200°C for 0.5 min, ramp of 5°C/min to 300°C, hold 300°C for 2 min, and final for 300°C 0.5 min; D: 0.5 min equilibration time; 40°C for 2 min, ramp of 20°C/min to 180°C, hold 180°C for 1 min, and final 180°C for 0.5 min. GC-MS retention times and methods for each compound are given in Table S1. For quantification purposes substrate and product peaks were quantified on a GC or an HPLC instrument as described below. NMR spectra were recorded on a Bruker Avance III 300 MHz NMR spectrometer, chemical shifts are reported relative to TMS (δ= 0.00 ppm).

3.2.7.5 Determination of conversions

Conversions were determined by comparing substrate and product peak areas to external calibration curves prepared using authentic reference materials, for each substrate and product, respectively. All styrenes, aldehydes, and ketones were quantified using GC FID analysis on a 7890A GC System (Agilent Technologies), with helium gas at the front SS inlet and an FID detector running on H_2 gas at 30 mL min^{-1} and 300 °C, using an HP-5 column (5% phenyl methyl siloxane, 30 m x 320 µm x 0.25 µm, J&W Scientific, Agilent Technologies). Quantifications were done by comparing peak areas to calibration curves prepared using between 1 and 10 mM authentic standard material, with the analytical samples diluted to correspond to 10 mM starting substrate concentration where necessary. As volume and injection control, 10 µL of a 1 M 1-naphthol solution in ethanol was added to each sample and reference standard. The following temperature program was used: 0.5 min equilibration time; 100°C for 0.5 min, ramp of 10°C/min to 280°C, and final 280°C for 2 min. All benzoic acids were quantified using an HPLC system equipped with an SPD-M20A diode array detector (Shimadzu, Kyoto, Japan), using a Luna C18 (2) 100A reversed-phase column (250 x 4.60 mm x 5 µm, Phenomenex, Torrance, CA, USA) and the following conditions: H_2O with 0.1 % TFA as eluent A, acetonitrile with 0.1 % TFA as eluent B; 1 mL min^{-1}, 0% eluent B to 100% eluent B in 20min. The benzoic acids were quantified by comparing to calibration curves prepared using between 0.1 and 1 mM authentic standard material. GC retention times are given in Table S1 and HPLC analysis wavelengths and HPLC retention times are given in Table S2.

3.2.7.6 Supplemental Tables

Table S1: GC-MS methods, GC and GC-MS retention times.

Compound	GC-MS method	GC-MS rt (min)	GC rt (min)
4-Cl-α-methylstyrene	A	5.2	3.79
styrene	B	4.63	2.02
4-Cl-styrene	B	6.25	3.02
4-methoxystyrene	A	5	3.65
p-aminostyrene	A	5.92	4.42
p-methylstyrene	B	5.59	2.52
α-methylstyrene	B	5.45	2.42
trans-1-phenyl-1-propene	B	5.84	2.68
trans-anethole	A	6.55	4.97
2-methyl-1-phenylpropene	B	6.31	3.07
benzaldehyde	B	5.28	2.33
p-Cl-benzaldehyde	A	6.71	3.46
p-methoxybenzaldehyde	A	6.21	4.67
p-methylbenzaldehyde	B	6.32	2.99
acetophenone	B	6.19	2.96
4-Cl-acetophenone	A	5.99	4.46
trans-β-nitrostyrene	B	8.76	-
trans-β-methyl-β-nitrostyrene	B	9.1	-
4-acetoxystyrene	A	6.56	-
isoeugenol	A	8.54	-
methyl cinnamate	A	7.75	-
4-ethylphenol	A	5	-
bromobenzene	B	5.04	-
ergosterol	C	20.2	-
limonene	B	5.87	-
phenylacetone	A	4.77	-
2-cyclohexene-1-one	B	5.02	-
phenylacetylene	D	5.79	-
4-vinylcyclohexene	B	4.13	-
2-methylheptene	D	4.74	-

rt: retention time

Table S2: HPLC analytical wavelengths and retention times

Compound	Analytical wavelength (nm)	HPLC rt (min)
benzoic acid	227	12.30
4-Cl-benzoic acid	237	13.70
4-methoxybenzoic acid	254	12.50
4-aminobenzoic acid	220	8.70
4-methylbenzoic acid	237	13.10

rt: retention time

Table S3: Conversions in control reactions without enzyme, but with 0.1 mM $MnCl_2$ in the same buffer, conditions were as described in the experimental section.

		Substituents			Background Reaction % Conversion to **b**	
Substrate (**a**)	R1	R2	R3	R4	1-phase	2-phase
1	Cl	Me	H	H	3	26
2	H	Me	H	H	n.d.	17
3	H	H	H	H	n.d.	n.d.
4	Cl	H	H	H	n.d.	n.d.
5	MeO	H	H	H	n.d.	2
6	NH2	H	H	H	n.d.	n.d.
7	Me	H	H	H	n.d.	3
8	H	H	Me	H	n.d.	4
9	MeO	H	Me	H	n.d.	5
10	H	H	Me	Me	1	6
11	H	H	NO2	H	n.d.	n.d.
12	H	H	NO2	Me	n.d.	n.d.

Table S4: Substrates not suitable for TM1459

Compound
trans-β-nitrostyrene
trans-β-methyl-β-nitrostyrene
4-acetoxystyrene
4-ethylphenol
bromobenzene
ergosterol
limonene
phenylacetone
2-cyclohexene-1-one
phenylacetylene
4-vinylcyclohexene
2-methylheptene

3.2.7.7 Supplemental Figures
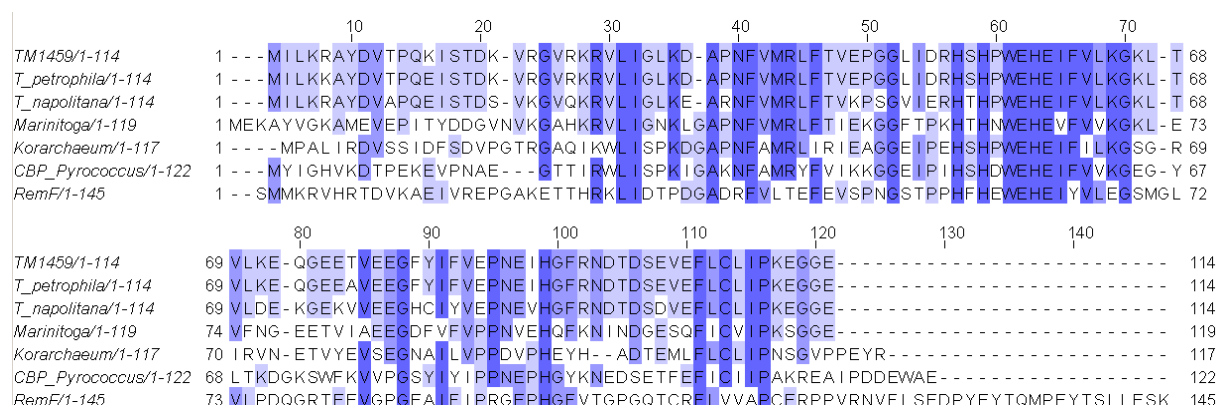
Figure S1: Sequence alignment (ClustalW2) of TM1459 with BLAST search results most similar to it and the closest matching sequences with a known function.
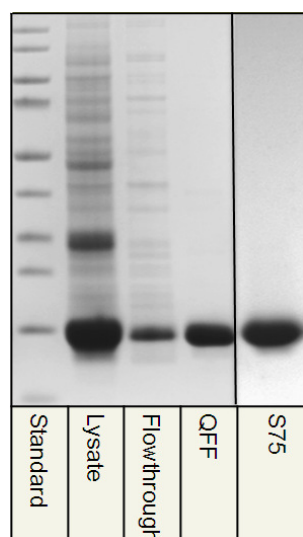
Figure S2: SDS-PA gel of the purification of TM1459. Standard: PageRuler prestained protein ladder (Fermentas), Lysate: Cleared lysate, Flowthrough of QFF column. QFF: Fractions pooled after QFF purification. S75: pooled fractions from gel filtration.
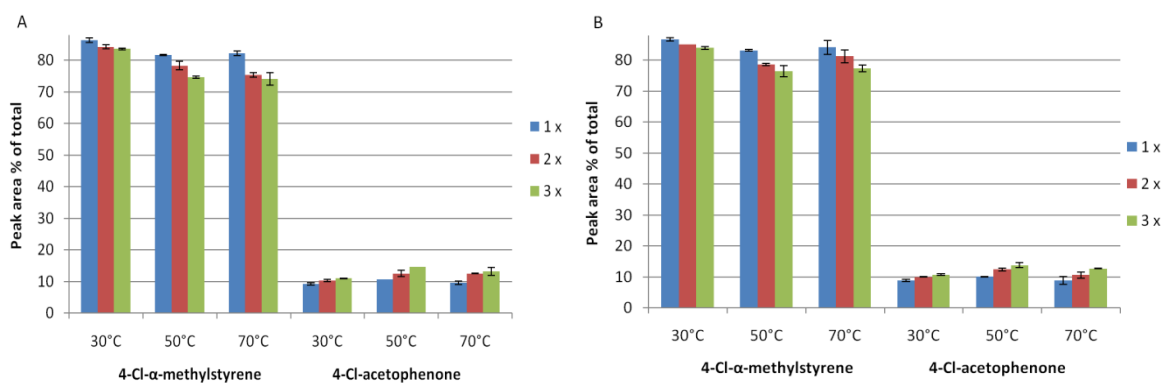
Figure S3: Background reaction under tested conditions for the conversion of 4-Cl-α-methylstyrene **1a** to 4-Cl-acetophenone **1b**. A: reactions with 0.1 mM $MnCl_2$ B: control reactions with 0.2 mM $MnCl_2$. The reactions were conducted with 1, 2 or 3-fold molar excess of *tert*-butyl hydroperoxide, shaken at 30, 50 or 70°C and 1000 rpm for 20 h.
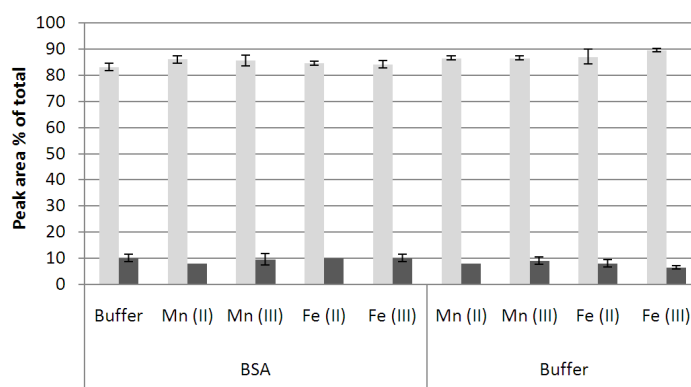
Figure S4: Background conversions of 4-Cl-α-methylstyrene **1a** (light grey bars) to 4-Cl-acetophenone **1b** (dark grey bars), in the presence of 1 mg mL^{-1} BSA (bovine serum albumin) in buffer, or in buffer with the addition of 0.1 mM of different metals. Additionally, the metal salts were tested without BSA. Reaction conditions: 10 mM substrate, 30 mM *t*-BuOOH, in 50 mM NaPi, pH 7.0, shaken at 30°C and 1000 rpm for 20 h. Substrate and product peak areas are given as percentage of total peak areas in GC for each sample.

3.3 Characterisation and engineering of two hydroxy amino benzoic acid dioxygenases towards a broadened substrate range

Ivan Hajnal[1], Helmut Schwab[1,3], and Kerstin Steiner[1,*]

[1]ACIB, Austrian Centre of Industrial Biotechnology GmbH, Petersgasse 14, 8010 Graz, Austria

[2]Institute of Molecular Biotechnology, TU Graz, Petersgasse 14, 8010 Graz, Austria

Author contribution: I have contributed the experimental data including biochemical charcterisation and kinetic measurements. A part of the expression constructs and mutated genes were made by a diloma student, Claudia Pany, and a technician, Julia Midl. All mutations and mutagenic primers were designed by Dr Steiner. I have written the manuscript.

*Corresponding author: Dr Kerstin Steiner, Austrian Centre of Industrial Biotechnology GmbH, Petersgasse 14/4, 8010 Graz, Austria, phone: +43-316-8739346, fax: +43-316-8739302, email: kerstin.steiner@acib.at

3.3.1 ABSTRACT

A 3-hydroxyanthranilic acid dioxygenase from *Burkholderia phytofirmans* PsJN (*Bp*3HAO) has been overexpressed in *E. coli* and characterised biochemically and through site-directed mutagenesis. Buffer, pH and temperature optima and kinetic data of the enzyme have been determined. Mutational analysis afforded information on residues important for activity and substrate-binding. Additionally a 4-amino-3-hydroxybenzoic acid dioxygenase from *Bordetella sp.* (*Bsp*4A3HBA23D), which acts on a substrate highly similar to the substrate of *Bp*3HAO, was characterised using site-directed mutagenesis and important residues were identified. Both enzymes are highly specific for their native substrate and do not accept the substrate of the other. The data gleaned from mutagenesis have been used to engineer *Bsp*4A3HBA23D to accept the native substrate of *Bp*3HAO, which contributes towards the understanding of substrate specificity in cupin dioxygenases.

3.3.2 INTRODUCTION

Cupins comprise a large and diverse superfamily of small β-barrel proteins and many have been described as dioxygenases (Dunwell et al., 2004). Many of the cupin dioxygenases are ring-cleavage dioxygenases which play an important role in the metabolism of a large number of aromatic compounds (Fetzner, 2012). Among these, a number of dioxygenases active on 2-amino-3-hydroxybenzoic acid, also termed 3-hydroxyanthranilic acid (3-HAA, Scheme 1a), have been identified in different organisms. They are particularly interesting because they act on aromatic substrates containing an amino group adjacent to a hydroxy group, whereas most substrates of cupin ring-cleavage dioxygenases contain one or two hydroxy groups at the site of ring-cleavage by oxygen and lack amino groups (Fetzner, 2012). 3-HAA dioxygenases (3-HAOs) are known from mammals, including humans, where they are part of the kynurenine pathway, and they have clinical significance due to their involvement in the production of the endogenous neurotoxin quinolinic acid (Calderone et al., 2002; Malherbe et al., 1994). In microorganisms, 3-HAOs primarily play a role in the alternative microbial metabolic pathway towards nicotinamide (Kurnasov et al., 2003; Zhang et al., 2005), and in some cases in the degradation pathways of xenobiotic compounds (Muraki et al., 2003). While the overall domain arrangement differs somewhat between members, the active sites are conserved between bacteria, yeast and mammals, and contain an active site divalent iron coordinated by two conserved histidines and a glutamate residue inside a cupin barrel (Đilović et al., 2009; Li

et al., 2006; Zhang et al., 2005). A dioxygenase acting on a highly similar substrate – 4-amino-3-hydroxybenzoic acid (4A3HBA, Scheme 1b) has been discovered in *Bordetella sp.* (Takenaka et al., 2002), and was named accordingly 4A3HBA-2,3-dioxygenase (*Bsp*4A3HBA23D). In spite of the fact that both substrates are benzoic acid derivatives with a hydroxy group adjacent to an amino moiety, both enzymes appear to be highly selective for their native substrates and show no detectable activity towards the substrate of the other. While cupin dioxygenases are often very substrate specific, there are reports in the literature of successful switching of substrate specificities through site-directed mutagenesis. For example, the conversion of an extradiol to an intradiol dioxygenase active on the same substrate requires only a single amino acid substitution (Groce and Lipscomb, 2003). In another report, a 2-hydroxy-1-naphthoate dioxygenase has been generated from a salicylate dioxygenase using only a small number of amino acid substitutions (Ferraroni et al., 2012). In this case, however, the relative position of functional groups of the substrates to each other remains the same. This work reports the biochemical and mutational characterisation of two hydroxy aminobenzoic acid dioxygenases and the introduction of activity towards the native substrate of *Bp*3HAO into the *Bsp*4A3HBA23D dioxygenase scaffold, where it is not found naturally.
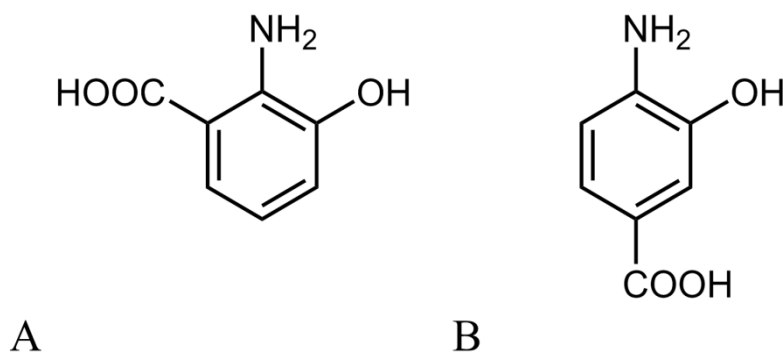
Scheme 1: The hydroxyaminobenzoic acid substrates of the dioxygenases used in this work: **A:** 3-hydroxyanthranilic acid (3-HAA); **B:** 4-amino-3-hydroxybenzoic acid (4A3HBA).

3.3.3 RESULTS AND DISCUSSION

3.3.3.1 Biochemical characterisation of wild-type *Bp*3HAO

While 3-HAOs from eukaryotes and especially mammals have been known for a long time (Koontz and Shiman, 1976), and were shown to have clinical significance in humans (Schwarcz et al., 1988), the discovery of prokaryotic 3-HAOs was reported more recently (Muraki et al., 2003). *Bp*3HAO is a protein from *Burkholderia phytofirmans* PsJN with high sequence similarity to known 3-HAOs from other bacteria and from yeast (Figure 1), such as the 3-HAO from *Cupriavidus metallidurans* (syn. *Ralstonia metallidurans*), the structure of which has been solved (Zhang et al. 2005). All residues known to be important for activity in other 3-HAOs are conserved, and have been analysed with site-directed mutagenesis along with other residues of interest.
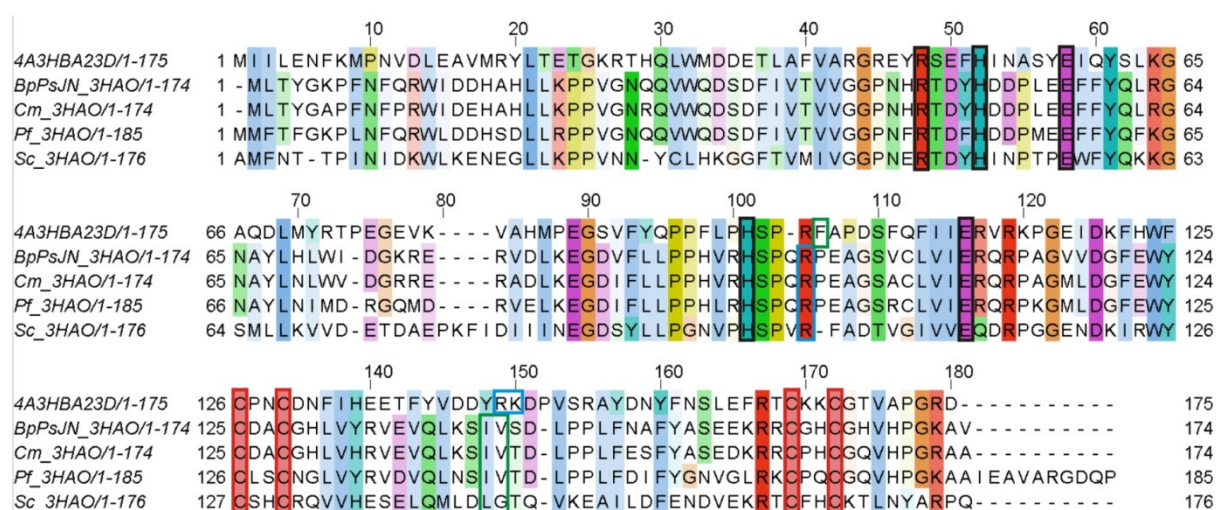
Figure 1: Sequence alignment of *Bp*3HAO with the 3-HAO from *B. phytofirmans* PsJN (BpPsJN_3HAO), and the 3-HAOs from *C. metallidurans* (Cm_3HAO), *Pseudomonas fluorescens* (Pf_3HAO) and from the yeast *Saccharomyces cerevisiae* (Sc_3HAO). The conserved active-site resides including the two histidines and one glutamate involved in metal binding, as well as the conserved argininie and glutamate in direct proximity of the metal (R47 and E110 in 1YFY) are shown in black boxes. The conserved cysteines of the rubredoxin-like secondary iron center are shown in red boxes. Hydrophobic and positively charged residues in corresponding positions of a *Bsp*4A3HBA23D model and 1YFY are shown in green and blue boxes, respectively.

The protein was expressed efficiently in *E. coli* BL21-Gold (DE3) (Figure S2). The enzymatic function as 3-HAO was confirmed and high activity was demonstrated using the assay published previously for the *C. metallidurans* enzyme (Zhang et al., 2005). The enzyme appeared to be unstable when exposed to air, which is not surprising since instability has been reported for HAOs earlier (Malherbe et al., 1994). Subsequently, different storage conditions were tested. As can be seen in Figure 2, the enzyme retained most activity when DTT was added to the buffer and atmospheric oxygen was displaced by N_2 in the headspace of the storage vial (13.6 ± 1 U·mg^{-1}). The activity retained after one week of storage with DTT and N_2 was more than ten-fold higher then when stored without DTT and with atmospheric oxygen concentration. In order to avoid inactivation, the wild-type enzyme and mutants were stored as filtered cleared lysates with the addition of 10 mM DTT in vials with an N_2 – purged headspace.
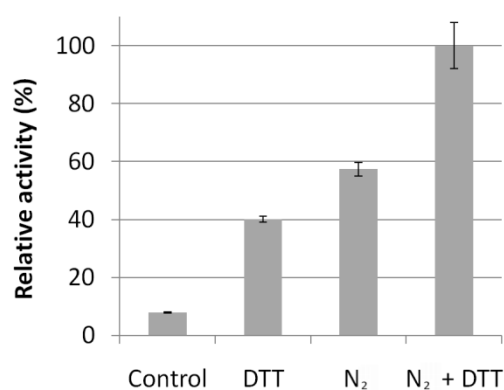
Figure 2: Relative activity of crude lysate aliquots stored at 4°C with or without the addition of 10 mM DTT, with or without an N_2 – purged headspace of the storage vial, after one week of storage. The activities are given as percentage of the highest residual activity (13.6 ± 1 U·mg^{-1}).

The enzyme was inhibited by phosphate buffer, with the activity significantly higher in 10 mM than in 50 mM sodium phosphate buffer (Figure S1). The enzyme showed the highest activity in Tris-HCl or imidazole buffer (Figure S1). Consequently, standard measurements were performed in Tris-HCl – buffer, and a set of organic buffers with overlapping pH ranges was used to determine the pH optimum. For this purpose, MES, MOPS, Tris-HCl and glycine buffers were prepared over a range from pH 5 to 10. The enzyme showed the highest activity at pH 7, with the activity decreasing significantly with a pH change both towards acidic and

basic values (Figure 3a). There was almost no remaining enzymatic activity at pH 5 and at pH 9. The temperature tolerance of the enzyme regarding its initial activity seemed to be broad. While the highest activity was measured at 25°C, with 12.8 ± 0.3 U·mg^{-1}, the initial activity at 40°C was still over 80% of the rate at 25°C (Figure 3b). The kinetic parameters were determined for the enzyme and it cleaved 3-HAA efficiently with a Kcat of 4.5 ± 0.04, which corresponded to an activity of 13.6 ± 0.01 U·mg^{-1} at Vmax, and a Km value of 21 ± 0.7 µM in 50 mM Tris-HCl buffer, pH 7.5, at 25°C.
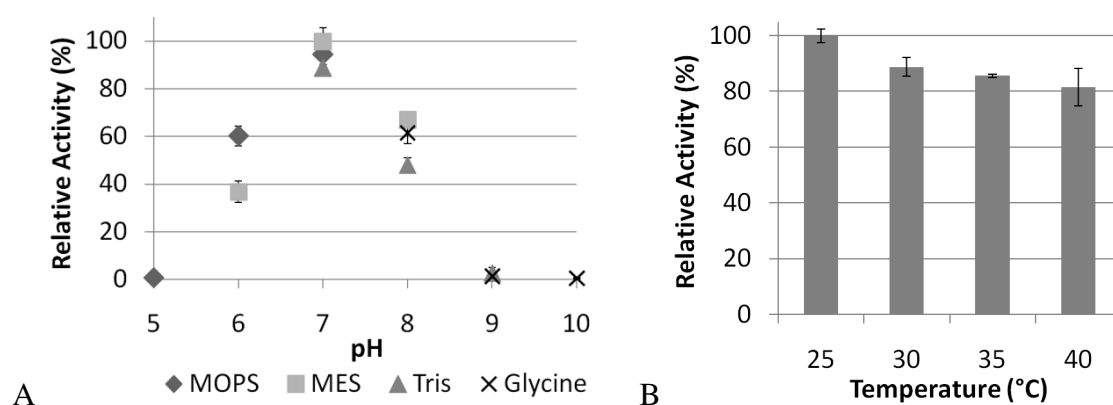
Figure 3: A: Activity of *Bp*3HAO at different pH values. 50 mM MES, MOPS, Tris-HCl and glycine buffers were used to cover a pH range from pH 5 to 10. Activities are given as percent of the highest activity, which was measured at pH 7.0 with no significant difference between MES, MOPS and Tris-HCl buffers (12.0 ± 0.7 U·mg^{-1} in MES buffer as 100%) **B:** Relative initial activity of Bp3HAO in dependence of the assay temperature. The activities are given as percent of the highest activity, which was observed at 25°C (12.8 ± 0.3 U·mg^{-1}).

3.3.3.2 Mutational analysis of *Bp*3HAO

Due to the high sequence similarity of the two enzymes, active site residues of *Bp*3HAO were identified based on the crystal structure of the highly similar 3-HAO from *C. metallidurans* (PDB ID: 1YFY). The metal-binding residues are fully conserved, as are the arginines 47 and 99 and glutamate 110 (Figure 1), which have been reported to be important for the activity of this highly similar enzyme (Zhang et al. 2005). A stick diagram of active site residues is shown in Figure 4 and kinetic data of mutants are shown in Table 1. The mutation of the positively charged residues, R47 and R99, respectively, led to a complete loss of activity, which is in agreement with the data obtained for the *C. metallidurans* enzyme (Zhang et al., 2005). Interestingly, an additional feature of these arginines seems to be necessary in addition to their positive charge, since lysines in each of the two position conferred activity much lower than that of the wild-type enzyme. A mutation of the conserved glutamate 110 (E110A), which is located in close proximity of the active site iron also resulted in almost inactive protein. Unsurprisingly, a mutation of the metal-binding glutamate 57 to alanine had no activity, even though the mutant was expressed efficiently as soluble protein (Figure S2). In contrast, the metal-binding site mutant H51A showed no detectable expression as soluble protein (Figure S2) and was expressed entirely in inclusion bodies (Figure S3), which underscores the importance of the metal-binding site for activity and folding in this class of enzymes.
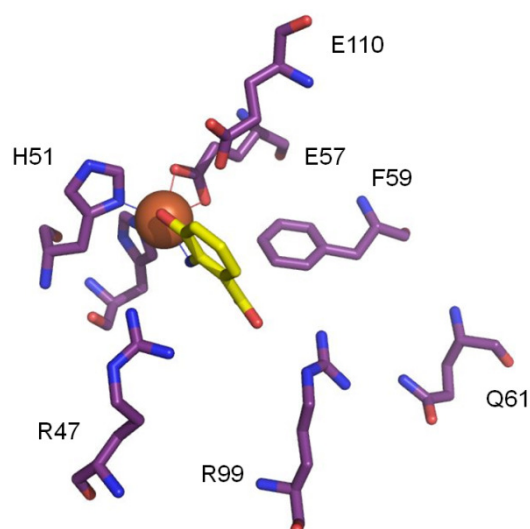
Figure 4: Stick diagram of the active site of 3-HAO from *C. metallidurans* with bound 3-hydroxyanthranilate substrate. Figure generated from PDB entry 1YFY using PyMOL.

Table 1: Kinetic data of Bp3HAO wild-type and mutants thereof, measured at 25°C in 50 mM Tris-HCl buffer, pH 7.5, 2.5 to 500 µM 3-HAA, 0.0001 – 0.5 mg·mL^{-1} total protein

Mutation	Activity (µmol· min^{-1}·mg^{1})	Km (3-HAA, µM)	Kcat (s^{-1})	Kcat/Km (s^{-1}·µM^{-1})
WT	13.6 ± 0.1	21 ± 0.7	4.5 ± 0.04	0.22
R47A	none	-	-	-
R47K	0.002	-	-	-
E57A	none	-	-	-
F59A	4.3 ± 0.07	60 ± 3	1.4 ± 0.02	0.02
F59Q	0.9 ± 0.009	89 ± 5	0.3 ± 0.003	3.5E-03
Q61A	0.1 ± 0.002	103 ± 6	0.04 ± 0.001	3.7E-04
Q61S	0.18 ± 0.002	112 ± 3	0.1 ± 0.001	5.4E-04
R99A	none	-	-	-
R99F	none	-	-	-
R99K	0.8[*]	-	-	-
E110A	0.006	-	-	-

*protein not stable under reaction conditions

In addition to the main metal-binding centre, which is clearly necessary for the activity, the bacterial 3-HAO from *C. metallidurans* as well as *Bp*3HAO contain a second (rubredoxin-like) iron centre coordinated by four cysteines (Figure 5, Zhang et al. 2005). The function of this rubredoxin-like centre is not known and it is conserved in bacterial and yeast 3-HAOs but missing in the mammalian enzyme (Li et al. 2005; Zhang et al. 2005; Đilović et al. 2009). The cysteines of this rubredoxin-like centre are also fully conserved in *Bp*3HAO (Figure 1). The cysteines were mutated in Bp3HAO, but the resulting proteins were expressed exclusively as inclusion bodies (Figure S3), which underscores the importance of the domain for proper folding of the enzyme, but unfortunately prohibits activity studies of the mutants.
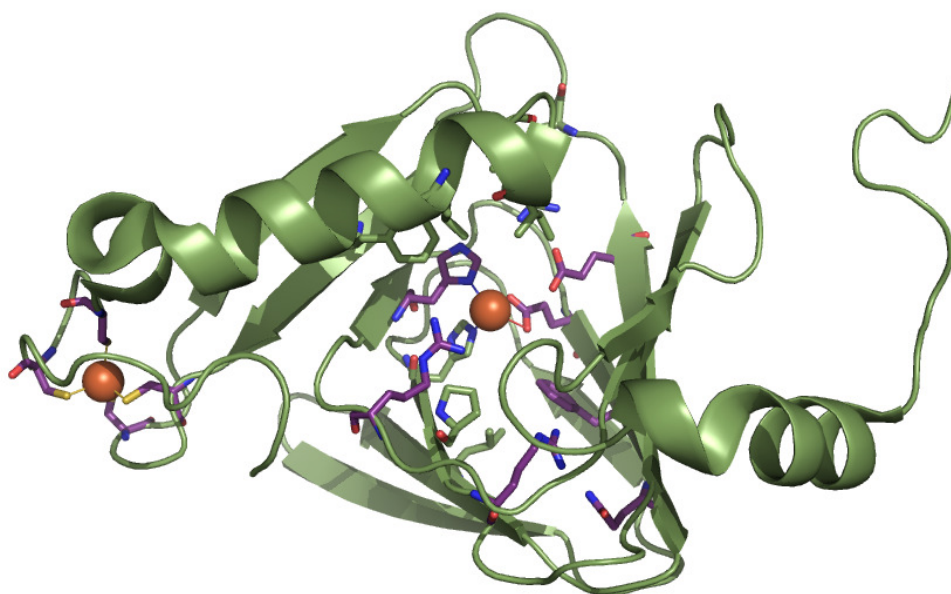
Figure 5: Illustration of the complete structure of a monomer of 3-HAO from *C. metallidurans*, showing the main active site iron (middle) and the rubredoxin-like iron centre (left). The protein backbone is shown as carton depiction in green. Iron ions are shown in magenta. Figure generated from PDB entry 1YFU using PyMOL.

3.3.3.3 Mutational analysis of *Bsp*4A3HBA23D dioxygenase

The *Bsp*4A3HBA23D enzyme was purified by others from *Bordetella* sp. 10d and a characterisation of this enzyme regarding buffer, pH and temperature as well as kinetic characterisation has been reported elsewhere (Takenaka et al., 2002). The enzyme was cloned and overexpressed in *E. coli* by the same group (Murakami et al., 2004). However, no mutagenesis study has been reported for this enzyme to date. In this work, presumed active site residues of *Bsp*4A3HBA23D were identified by analysing a homology model, which was generated based on the above mentioned 3-HAO crystal structure with bound 3-hydroxyanthranilic acid (PDB ID: 1YFY). Judging by the sequence similarity of *Bsp*4A3HBA23D to 3-HAOs (Figure 1), the metal is most likely bound by two histidines and one glutamate residue. The metal binding residues were mutated to alanine residues. All three mutants, H52A, E58A, and H97A, completely lost activity. The mutant H52A was not expressed as soluble protein at a level detectable on Coomassie-stained SDS PA-gels (data not

shown), whereas the other two mutant proteins had clearly detectable soluble expression (data not shown). The loss of activity upon mutation of metal-binding residues is an expected result and indicates that the reaction indeed takes place with participation of the metal-binding site as in 3-HAO. The mutant E111A corresponds to the *Bp*3HAO mutant E110A. This residue, and the residue corresponding to arginine 47 are completely conserved together with the metal-binding site (Figure 1), and were shown to be crucial, since both mutants were soluble but showed very little activity (Table 2; Figure S4). Interestingly, the activity of mutant F101A was abolished in spite of good soluble expression (Table 2; Figure S4). This phenylalanine may be necessary for substrate binding, since a direct catalytic involvement of this relatively inert residue would be unlikely. One additional mutant of a residue, N54A lost all activity while retaining efficient soluble expression. A number of mutants displayed an activity, which was significantly lower than that of the wild-type enzyme, but clearly measurable. These mutations thus retained activity and displayed varying effects on Km values. A full list of mutants with their activities and Km values are shown in table 2 and the expression is shown in figure S4. It is interesting to compare the effect of positively charged active site residues in *Bsp*4A3HBA23D to 3-HAOs. The positively charged residues R144 and K145 are immediately adjacent to each other in the protein, and both are positioned close to the metal-binding site, while the exact orientation is not clear from homology modelling. While in *Bp3*HAO a mutation of R99 led to a complete loss of activity, the mutants of positively charged residues in *Bsp*4A3HBA23D, R144I and K145V, both retained most of the enzymatic activity (Table 2). This is might be due to a shared function of these two positive charges, which are located adjacent to each other and may compensate the loss of each other. A combined mutant R144I/K145V lost almost all activity, with only 0.003 $U \cdot mg^{-1}$ detectable, versus ~ 12 $U \cdot mg^{-1}$ in the wild-type enzyme, which represents a reduction in activity of four orders of magnitude. In spite of the retention of significant activity in any of the single mutants, this double mutant establishes an importance of a positive charge in *Bsp*4A3HBA23D as has been observed in *Bp*3HAO.

Mutation	Activity (μmol· min^{-1}·mg^1)	Km (4A3HBA, μM)	Kcat (s^{-1})	Kcat/Km (s^{-1}·μM^{-1})
WT	11.8 ± 0.7	106 ± 16	3.9 ± 0.2	0.04
H52A	none	-	-	-
N54A	none	-	-	-
E58A	none	-	-	-
H97A	none	-	-	-
Q60A	1.6 ± 0.09	198 ± 22	0.6 ± 0.03	2.8E-03
Q60F	0.38 ± 0.05	297 ± 60	0.1 ± 0.02	4.3E-04
S62A	19.6 ± 0.54	140 ± 9	6.5 ± 0.18	0.05
S62Q	5.6 ± 1.3	> 500	1.9 ± 0.44	2.6E-03
F101A	none	-	-	-
F101R	0.001	-	-	-
R144I	10.1 ± 0.5	160 ± 16	3.4 ± 0.16	0.02
T28A	8.5 ± 0.003	112 ± 11	2.8 ± 0.1	0.03
T28N	4.7 ± 0.2	86 ± 9	1.6 ± 0.06	0.02
F40A	2.0 ± 0.26	> 500	0.7 ± 0.09	-
F40T	0.07 ± 0.008	230 ± 52	0.02 ± 0,003	1.1E-04
E46A	0.07 ± 0.006	321 ± 52	0.02 ± 0.003	6.8E-05
E46N	7.1 ± 0.4	150 ± 20	2.4 ± 0.1	0.02
E111A	0.5 ± 0.03	132 ± 20	0.18 ± 0.01	1.3E-03
K145V	16.6 ± 0.75	153 ± 16	5.5 ± 0.3	0.04
R144I/K145V	0.003	-	-	-

Table 2: Kinetic data of mutated *Bsp*4A3HBA23D variants, measured at 25°C in 50 mM Tris-HCl buffer pH 7.5, 10 – 500 μM 4A3HBA, 0.03-3.5 mg·mL^{-1} total protein

The mutant S62A is interesting because it has a significantly higher Kcat value than the wild-type enzyme. However, the Km value of this enzyme is increased compared to that of the wild-type so that the overall ratio of Kcat/Km is closer to that of the wild-type enzyme. The same is true for the mutant K145V. Phenylalanine F40 might be involved in substrate binding and positioning since both mutants F40A and F40T displayed drastically increased Km values and a significant loss of specific activity. Another phenylalanine residue, F101, appears to be even more important since the mutant F101A completely lost activity while the mutant F101R retained detectable but extremely low activity towards the native substrate (Table 2). The residues identified as important for activity thus comprise the metal-binding residues and positively charged and hydrophobic residues in close proximity of the metal-binding site. This pattern can again be rationalised based on the substrate, which contains a negatively charged carboxylic acid moiety attached to a benzene ring which can interact with positively charged amino acids and by comparing to the substrate binding mode in 3-HAO, where the hydroxy group of the substrate interacts with the active site metal while the carboxylic acid moiety in of the substrate is in direct contact with arginine 99 (Zhang et al., 2005). It can thus be said that the residues of the metal-binding site, and residues immediately adjacent to it, which comprise the site of oxygen-activation and ring-cleavage, are fully conserved between *Bp*3HAO and *Bsp*4A3HBA23D (Figure 1; Table 2). They are responsible for the activity in both enzymes, while additional residues likely confer substrate specificity.

3.3.3.4 Substrate scope

Both wild-type *Bp*3HAO and *Bsp*4A3HBA23D as well as mutant enzymes were screened with a set of 15 substrates. No activity on any of the alternative substrates was detected, while strong activity was detectable with their respective native substrates. *Bsp*4A3HBA23D has already been screened for activity on a partially overlapping set of substrates by others (Takenaka et al., 2002) and was also found to be active only on its native substrate, so the result was not surprising. All mutants, which were not completely inactive with their native substrates, were also tested with the substrates.

3.3.3.5 Converting *Bsp*4A3HBA23D into a 3-HAA dioxygenase

As has been described earlier, the native substrate of *Bp*3HAO is 3-hydroxyanthranilic acid, which is 2-amino-3-hydroxybenzoic acid, whereas the native substrate of *Bsp*4A3HBA23D is 4-amino-3-hydroxybenzoic acid. Thus, these two substrates can be viewed as 2-aminophenol derivatives differing only in the position of the carboxylic acid moiety. The carboxylic acid moiety appears to be highly important as indicated by the loss of activity in mutants which lack positively charged residues (R144I/K145V; R99A). Furthermore, in the aforementioned crystal structure of 3-HAO from *C. metallidurans* an interaction of the carboxylic acid moiety of the substrate with the active site arginine 99 was established by crystallography (Zhang et al. 2005). Upon examination of the homology model of *Bsp*4A3HBA23D aligned to the 3-HAO crystal structure, a pattern emerges of the positioning of residues close to the metal-binding site. At the position of the active site of *Bsp*4A3HBA23D towards which the substrate carboxylate is positioned according to the homology model, there is a positive charge (R144 or K145, Figure 6). In *Bp*3HAO, there is a hydrophobic residue at the corresponding position (I142). At the opposite position of the active site, there is a hydrophobic residue in *Bsp*4A3HBA23D and a positive charge in *Bp*3HAO (F101/R99).
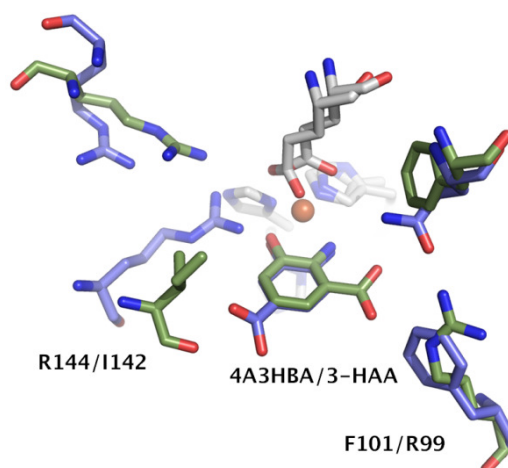
Figure 6: Crystal structure of 3-HAO with bound substrate 3-HAA (PDB entry: 1YFY; green), aligned with a homology model of *Bsp*4A3HBA23D including the substrate 4A3HBA (blue) overlaid based on the ring of 3-HAA in 1YFY. Metal-binding residues (gray) and catalytic iron (magenta), are also shown.

To further test the idea that it is the orientation of the carboxylate in the active site, which predominantly determines substrate specificity, mutations were combined. Since carboxylate orientation may be determined by active site positive charges and in the F101R mutant the second positive charge remained, this mutant was combined with mutations which exchange the positive residue R144 or the adjacent positive residue K145 for hydrophobic residues. All mutants were expressed efficiently as soluble protein, at a rate comparable to the wild-type *Bsp*4A3HBA23D (Figure S5).
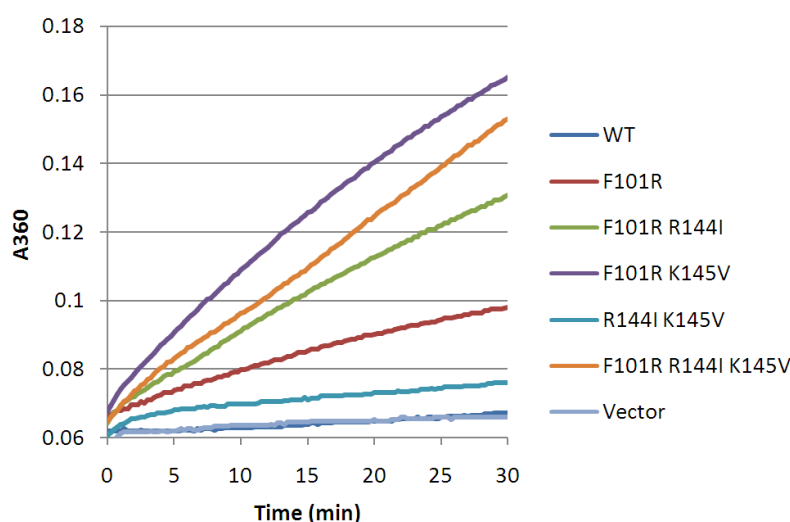
Figure 7: Cleavage of 3-hydroxyanthranilic acid by WT and mutants of *Bsp*4A3HBA23D; shown is the increase in A_{360} due to the direct cleavage product of the enzymatic reaction, as measured in 50 mM Tris-HCl buffer, pH 7.5, at 25°C, 500 µM 3-HAA, 4-7.5 mg·mL^{-1} total protein.

In spite of the low activity acquired by the mutations, the cleavage of the substrate by the active mutants was clearly visible, whereas no activity was detectable in the wild-type (Figure 7). The mutant F101R had a specific activity of $0.15 \cdot 10^{-4}$ U·mg^{-1}. The combination of this mutation with either of the mutations replacing one of two positively charged residues at the opposite side of the active site had an approximately 2- and 3-fold increased specific activity, to 0.30 and $0.45 \cdot 10^{-4}$ U·mg^{-1} in the double mutants F101R/R144I and F101R/K145V, respectively. Interestingly, the triple mutant F101R/R144I/K145V was not more active than either of the double mutants. The pattern of activity implies that the orientation of the carboxylic acid moiety of the substrate via a combination of a positive charge on one side and

hydrophobic side chains accommodating the phenyl moiety on the other side of the active site indeed confers substrate specificity.

3.3.3.6 Conversion of *Bp*3HAO to a 4A3HBA dioxygenase

Along the same line of reasoning, mutants have been produced which introduce a positively charged residue into the *Bp*3HAO scaffold at a position corresponding to positively charged residues in *Bsp*4A3HBA23D (I142R) or which remove the positively charged residue of *Bp*3HAO (R99F). A combined mutant introducing a positive charge and removing the positive charge from the opposite side of the active site (I142R/R99F) has also been constructed. A comparison of the positioning of 3-HAA in the wild-type protein and the hypothetical positioning of 4A3HBA in the double mutant is illustrated in Figure 8. All mutants expressed efficiently as soluble protein (Figure S6). No activity on any of the alternative substrates, including 4A3HBA, was detected in either *Bp*3HAO wild-type protein or any of the mutants. It should be noted that since the 4A3HBA dioxygenase assay is performed at 294 nm where protein strongly absorbs, whereas the 3-HAA dioxygenase assay measures the increase in absorbance at 360 nm, the sensitivity is limited by the amount of protein one can apply in the assay. It is thus possible that an activity on the same level with the activity displayed by the *Bsp*4A3HBA23D mutants towards 3-HAA remains undetected due to assay limitations.

A dioxygenase acting on 2-aminophenol, which differs from 3-HAA and 4A3HBA by a lack of carboxylic acid moiety have been purified from *Pseudomonas* species (Lendenmann and Spain, 1996; Takenaka et al., 1997). While this enzyme is not a cupin, the similarity of the substrates and the reaction catalysed makes this an interesting enzyme for comparison. A structure of this enzyme was solved very recently (Li et al., 2012). The enzyme contains an iron coordinated by two histidines and a carboxylate, which is a common motive for non-heme iron enzymes (Leitgeb and Nidetzky, 2008; Li et al., 2012). In addition to the metal-binding residues, a histidine and a tyrosine were identified as necessary for activity (Li et al. 2012). At least implicitly, the lack of observed positively charged residues fits well with a substrate lacking a carboxylate moiety. However, residues such as arginines and glutamates which are conserved in the cupin hydoxyaminobenozic acid dioxygenases were not reported by the authors.
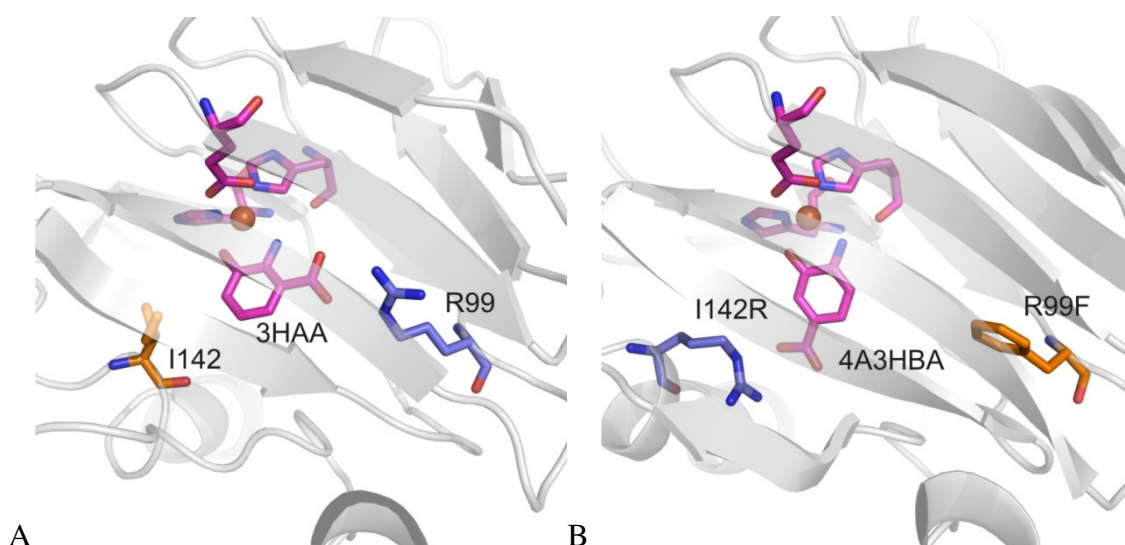
Figure 8: A: Crystal structure of 3-HAO with bound substrate 3-HAA (PDB entry: 1YFY), illustrating the positioning of R99 towards the substrate carboxylate and the positioning if I142 towards the hydrophobic side of the substrate ring. **B:** An illustration of a model of the double mutant R99F/I142R showing 4A3HBA manually positioned based on the 3-HAA ring of 1YFY, with the substrate carboxylate facing the introduced arginine residue, with the introduced hydrophobic residue on the hydrophobic side of the substrate ring.

3.3.4 CONCLUSION

Two cupin ring-cleavage dioxygenases, acting on highly similar hydroxy-amino-benzoic acid substrates, which both have high specificity for their native substrate have been characterised biochemically and through site-directed mutagenesis. In addition to the absolute importance of the metal-binding site, mutagenesis studies show the importance of additional conserved residues including a glutamate and an arginine not directly interacting with the substrate in the published substrate-bound structure of a bacterial 3-HAO (Zhang et al. 2005). Due to the nature of the substrates, it is possible that the pattern of positively charged residues which orient the substrate carboxylate and hydrophobic residues is important for substrate positioning in the active site and thus confers substrate selectivity. This hypothesis has been confirmed by the tentative conversion of a 4A3HBA dioxygenase into a weakly active but clearly demonstrable 3-HAA dioxygenase whereas no activity was detectable in the WT enzyme. The 3-HAA activity could be further improved by combined mutations, further

underscoring this approach. The work represents the first tentative generation of a 3-HAO from a 4A3HBA dioxygenase.

3.3.5 EXPERIMENTAL PROCEDURES

3.3.5.1 Cloning and heterologous expression

The sequences encoding *Bp*3HAO (gene YP_001895110, from *Burkholderia phytofirmans* PsJN, Uniprot: B2T2S5) and *Bsp*4A3HBA23D (gen BAD08206 from *Bordetella sp.*, Uniprot: Q76CA9_9BORD) were ordered as synthetic genes codon-optimised for *E. coli* (GeneArt, Life Technologies, Carlsbad, CA, USA). Sites for the restriction enzymes *Nde*I and *Hin*dIII (ThermoScientificBio, Waltham, MA, USA) were provided on the synthetic construct and were used to transfer the genes into the expression vector pET26b(+) (Novagen, Merck KGaA, Darmstadt, Germany). Site-directed mutations were generated by overlap-extension PCR. The PCR products were cloned into the vector pET28a(+) (Novagen), using *Nco*I and *Hin*dIII restriction sites for Bp3HAO mutants and *Pag*I and *Hin*dIII for *Bsp*4A3HBA23D mutants. The constructs were confirmed by sequencing (LGC Genomics, Berlin, Germany or Microsynth, Balgach, Switzerland). *E. coli* BL21-Gold(DE3) (Stratagene, La Jolla, CA, USA) cells were used for expression. All cells were grown in LB (lysogeny broth, Lennox) medium (Carl Roth Industries, Buchenau, Germany) supplemented with kanamycin sulphate (40 mg·L^{-1} final concentration, Carl Roth). Pre-cultures were diluted to an OD_{600} of ~0.1 and grown under shaking in baffled flasks at 37°C and 120 rpm until an OD_{600} of ~0.6 was reached. After cooling the cultures to 25°C, the expression was induced by addition of 0.1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG, Carl Roth). 100 µM $FeSO_4$ was added to cultures expressing *Bsp*4A3HBA23D or mutants thereof concomitantly with the induction. The induced cultures were harvested after 20 h at 25°C by centrifugation (20 min, 2800 g, 4°C). Cells were resuspended in 50 mM Tris-Cl buffer pH 7.5 with 10 mM DTT (Bp3HAO) or 10 mM NaPi pH 7.0 (4A3HBA23D) and disrupted by sonification (Branson Sonifier S-250, set to 80% duty cycle, and output control 7) two times for 3 min, cooled on ice. The crude lysates were cleared by centrifugation (1 h, 48000 g, 4°C). Cleared lysates were filtered and stored at 4°C in polypropylene vials with an N_2 - purged headspace.

3.3.5.2 Protein concentration determination and band quantification

Protein concentrations of cleared lysates were routinely determined using the Bradford protein assay (Biorad, Hercules, CA, USA) with bovine serum albumin as reference protein. The coomassie – stained SDS PA-gel bands corresponding to the overexpressed protein were quantified using a G-box HR16 device (Syngene, Cambridge UK), the Software Gene Snap v. 7.05 and Gene v. 4.00 Tools (Syngene) with the rolling disc baseline correction set at 30. The percentage of the enzyme band in total lysate bands was compared between wild-type and mutant proteins, and used together with total protein concentration to calculate specific activities. For the determination of the solubility of the enzymes, cells harvested from 1.5 mL samples of expression cultures were re-suspended in 10 µL per mg cell wet weight of BugBuster (Novagen) solution in 50 mM Tris-HCl buffer, pH 7.5, which led to cell lysis. The soluble and insoluble protein fractions were separated by centrifugation in a tabletop centrifuge (15 min, 20000 g) and the insoluble pellet containing inclusion bodies was solubilised in the same volume of 6 M urea as the original re-suspension volume. Both fractions were analysed by SDS-PAGE.

3.3.5.3 Enzyme Kinetics

All measurements were done in 96-well plates. The substrates 3-hydroxyanthranilic acid and 4-amino-3-hydroxybenzoic acid (Sigma-Aldrich, St. Louis, MO, USA) were dissolved in DMSO as 100 mM stock solutions. From these solutions all indicated concentrations were produced by dilution in 50 mM Tris-Cl buffer, pH 7.5, and for lower substrate concentrations the DMSO concentration (0.5% final in assay) was compensated. 100 µL of substrate were pipetted in each well, and 100 µL of enzyme solution diluted in the same buffer, were added to start the reactions. The photometric measurements were performed at 25°C except for the temperature studies, using a Synergy MX SMATBLD micro-plate reader (Biotek, Winooski, WT, USA), running Gen 5 software version 1.11 (Biotek). For 3-hydroxyanthranilic acid the extinction coefficient $\varepsilon_{360} = 47.5 \times 10^3$ M^{-1} cm^{-1} (Kurnasov et al., 2003) and for 4-amino-3-hydroxyanthranilic acid $\varepsilon_{294} = 7.53 \times 10^3$ M^{-1}·cm^{-1} (Takenaka et al., 2002), respectively, were used as published. The kinetic data were fitted to the Michaelis-Menten equation using the software SigmaPlot 11.0 (Systat Software) running the software module Enzyme Kinetics 1.3.

3.3.5.4 Optimisation of buffer, pH and temperature

For the optimisation of buffer conditions, 100 µL each of 100 µM 3-hydroxyanthranilic acid solutions (from 10 mM stock in DMSO) in 50 mM Na-phosphate, Na-citrate, Tris-HCl, and imidazole and 10 mM Na-phosphate buffer were each set to pH 7.0 and were dispensed in a microtiter plate, after which crude lysate of cells expressing Bp3HAO was diluted 5000 fold in each buffer separately and 100 µL were dispensed to start the reactions. The kinetic measurements were recorded at 25°C with the same setup as described above. The optimisation of pH was conducted analogously, with 50 mM MES, MOPS, Tris-HCl and glycine buffers prepared covering a pH range of 5 to 10. The optimisation of temperature was done with the same setup, whereby Tris-HCl buffer pH 7.0 was used and the spectrophotometer and enzyme-dilution buffers pre-incubated to the indicated temperatures.

3.3.5.5 Substrate screening

Alternative substrates were screened for degradation by scanning for changes in the absorption spectra between 240 and 500 nm of each reaction mixture, since all the substrates show absorption in this range, as has been published for similar substrate sets earlier (Takenaka et al., 1997; Takenaka et al., 2002). The enzymes were used as crude lysates at 50-500 µg·mL^{-1} final concentration. Substrates were diluted to a final concentration of 100 µM in 50 mM Tris-Cl pH 7.5 from 10 mM stock solutions in DMSO. The substrates 3-hydroxyanthranilic acid (2-amino-3-hydroxybenzoic acid) (97%), 4-Amino-3-hydroxybenzoic acid (97%), 3-aminosalicylic acid (97%), protocatechuic acid (97%), 2,3 dihydroxybenzoic acid (99%), gentisic acid, 2-amino-3-methylphenol (96%), 4-aminoresorcinol hydrochloride (97%), 6-amino-m-cresol (98%), 2-aminophenol (99%), 2,3-diaminobenzoic acid (95%), 3-amino-4-hydroxybenzoic acid (97%), 2-amino-3-methoxybenzoic acid, and 3,4 diaminobenzoic acid were all purchased from Aldrich (Sigma-Aldrich). The measurements were performed in 96-well plates using the plate reader described above.

3.3.6 Acknowledgements

3.3.7 Supplemental material

3.3.7.1 Supplemental Figures
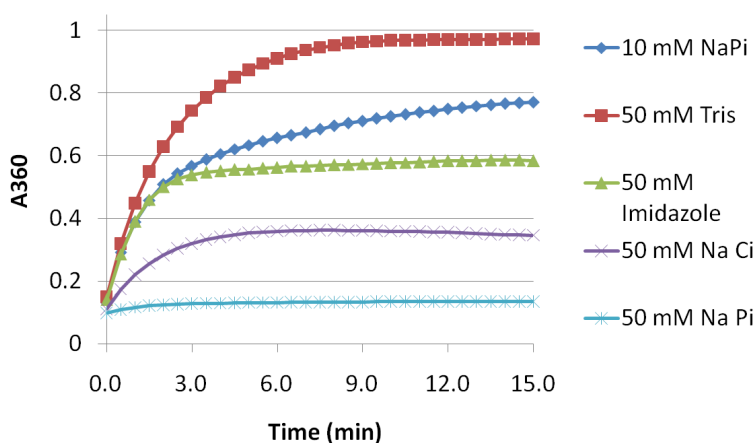
Figure S1: Activity of *Bp*3HAO in different buffers. All buffers were set to pH 7.0, Na Pi: sodium phosphate, Na Ci: sodium citrate. Conditions: Final concentration 1.7 µg·mL^{-1} crude lysate expressing *Bp*3HAO, 50 µM 3-HAA, 25°C.
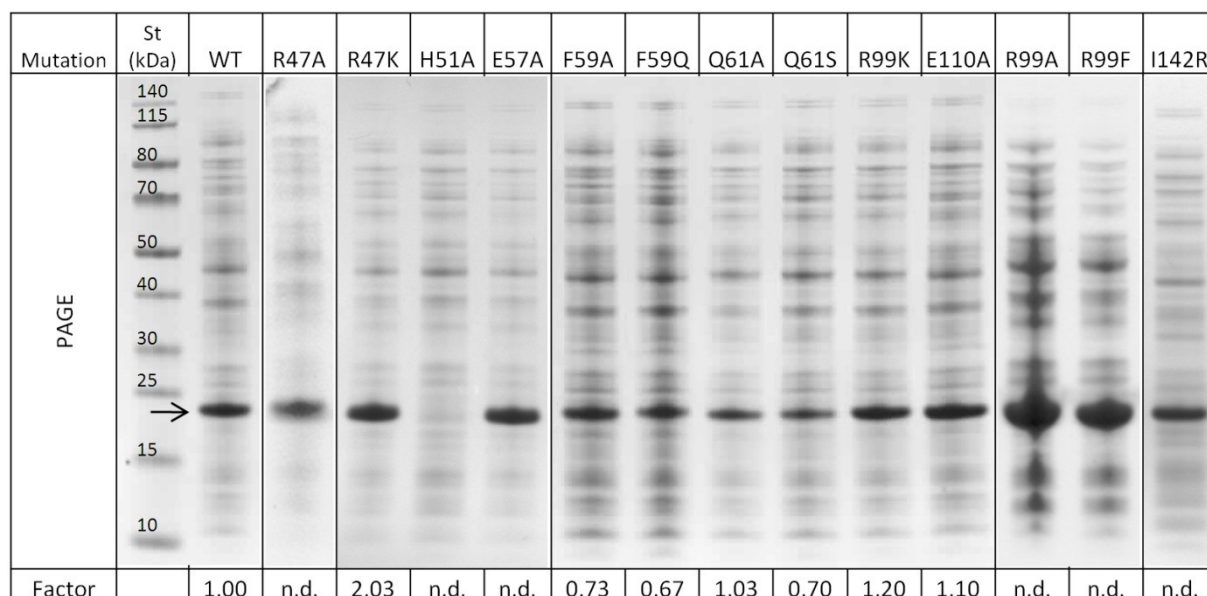
Figure S2: SDS PA-gel showing the expression of *Bp*3HAO WT and mutants. An expression factor representing the abundance of the enzyme in the crude lysate normalised to the abundance of the wild-type protein is indicated for mutants retaining detectable activity.
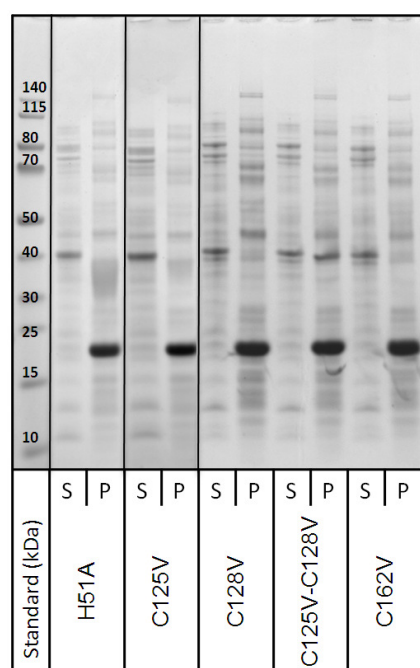
Figure S3: SDS PA-gel showing the expression of insoluble mutants of *Bp*3HAO as inclusion bodies. S: Supernatant, P: Pellet after centrifugation (15 min, 20000g) and dissolution with 6 M urea.
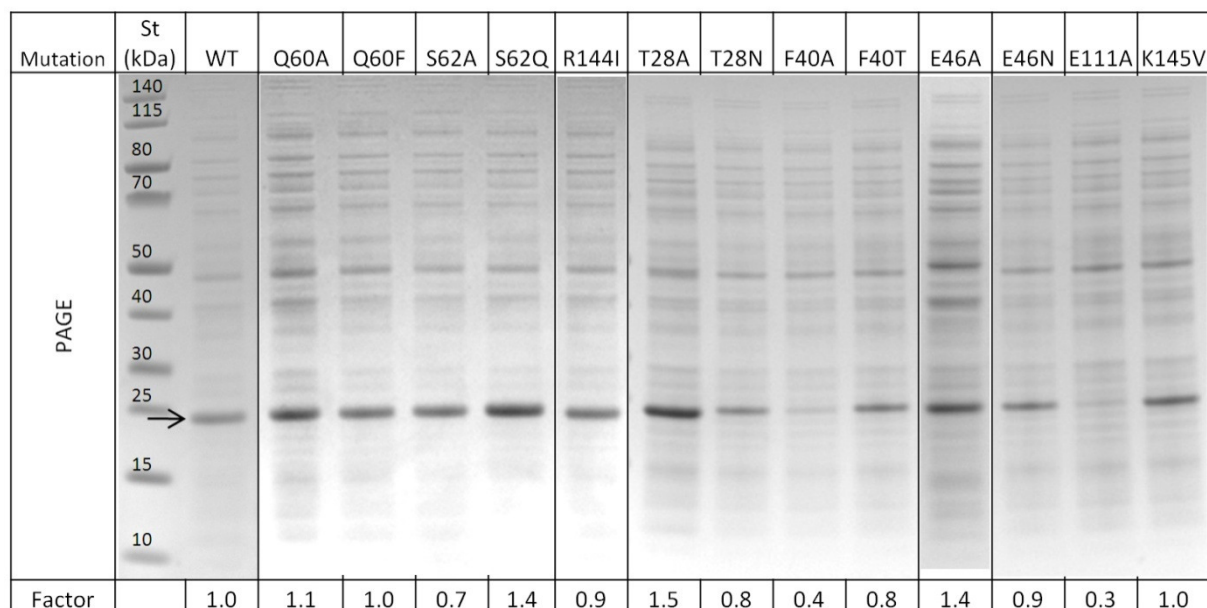
Figure S4: SDS PA-gel showing the expression of *Bsp*4A3HBA23D WT and mutants. An expression factor is indicated as above.
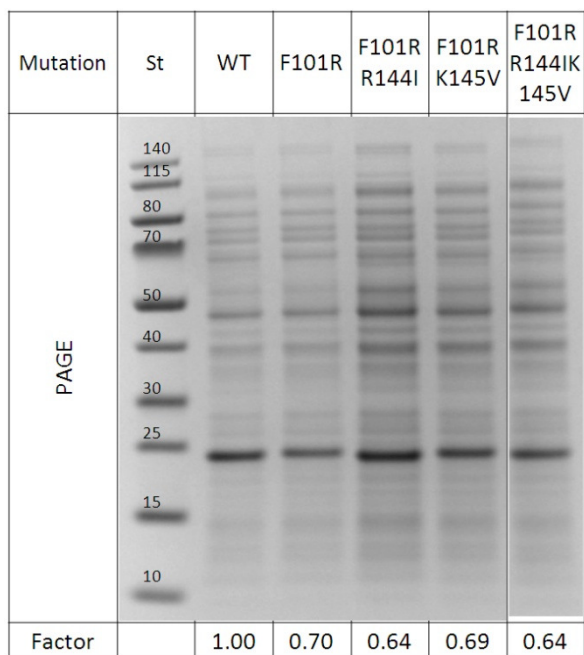
Figure S5: SDS PA-gel showing the expression of *Bsp*4A3HBA23D WT and mutants which exhibited activity on the alternative substrate 3-HAA. An expression factor is indicated as above.

Mutation	St	WT	R99F	I142R	R99F I142R
PAGE	140 115 80 70 50 40 30 25 15 10				
Factor		1.00	n.d.	n.d.	n.d.
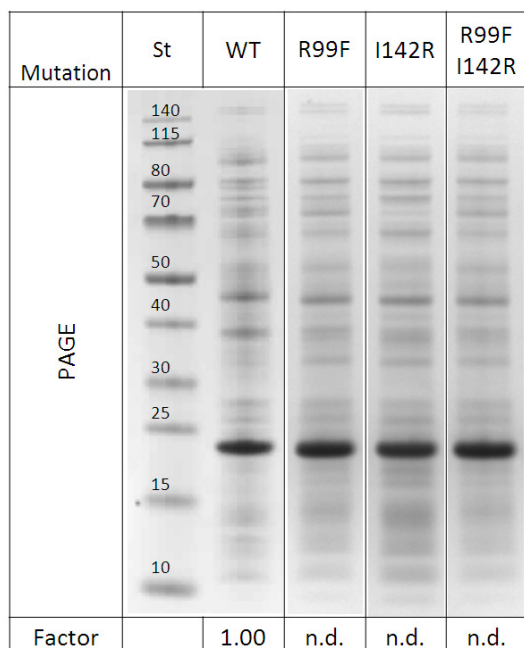
Figure S6: SDS PA-gel showing the expression of *Bp*3HAO mutants which exchange positive active-site residues for hydrophobic residues and vice-versa.

3.3.7.2 Supplemental Tables

Table S1: Cloning and mutagenesis primers used in this study are listed. Mutated sequences are shown in bold letters, restriction site sequences are underlined, start and stop codon are indicated with boxes.

Primer Name (mutation)	Sequence
*Bp*3HAO _for(*Nco*I)	AATCACC**ATG**GTGACCTATGGCAAACCGTTC
*Bp*3HAO_rev(*Hind*III)	AATCAAAGCTTTTAAACTGCTTTACCCGGATGAAC
*Bp*3HAO-H51A_for	CGTACCGATTAT**GCG**GATGATCCGCTGG
*Bp*3HAO-H51A_rev	CCAGCGGATCATC**CGC**ATAATCGGTACG
*Bp*3HAO-E57A_for	GATGATCCGCTGGAA**GCA**TTTTTTTATCAGCTG
*Bp*3HAO-E57A_rev	CAGCTGATAAAAAAA**TGC**TTCCAGCGGATCATC
*Bp*3HAO-R47A_for	GTTGGTGGTCCGAATCAT**GCT**ACCGATTATCATGATG
*Bp*3HAO-R47A_rev	CATCATGATAATCGGT**AGC**ATGATTCGGACCACCAAC
*Bp*3HAO-R47K_for	GTTGTTGGTGGTCCGAATCAT**AAA**ACCGATTATCATGATGATCC
*Bp*3HAO-R47K_rev	GGATCATCATGATAATCGGT**TTT**ATGATTCGGACCACCAACAAC
*Bp*3HAO-F59A_for	CCGCTGGAAGAATTT**GCT**TATCAGCTGCGTG
*Bp*3HAO-F59A_rev	CACGCAGCTGATA**AGC**AAATTCTTCCAGCGG
*Bp*3HAO-F59Q_for	GATCCGCTGGAAGAATTT**CAA**TATCAGCTGCGTGGTAATG
*Bp*3HAO-F59Q_rev	CATTACCACGCAGCTGATA**TTG**AAATTCTTCCAGCGGATC
*Bp*3HAO-Q61A_for	GCTGGAAGAATTTTTTTAT**GCG**CTGCGTGGTAATGCCTATC
*Bp*3HAO-Q61A_rev	GATAGGCATTACCACGCAG**CGC**ATAAAAAAATTCTTCCAGC

Primer Name (mutation)	Sequence
*Bp*3HAO-Q61S_for	GCTGGAAGAATTTTTTTATT**TCG**CTGCGTGGTAATGCCTATC
*Bp*3HAO-Q61S_rev	GATAGGCATTACCACGCAG**CGA**ATAAAAAAATTCTTCCAGC
*Bp*3HAO-R99A_for	CATGTTCGTCATAGTCCGCAG**GCT**CCGGAAGCAGGTAGCGTTTG
*Bp*3HAO-R99A_rev	CAAACGCTACCTGCTTCCGG**AGC**CTGCGGACTATGACGAACATG
*Bp*3HAO-R99K_for	CGTCATAGTCCGCAG**AAG**CCGGAAGCAGGTAG
*Bp*3HAO-R99K_rev	CTACCTGCTTCCGG**CTT**CTGCGGACTATGACG
*Bp*3HAO-R99F_for	GTCATAGTCCGCAG**TTT**CCGGAAGCAGGTAG
*Bp*3HAO-R99F_rev	CTACCTGCTTCCGG**AAA**CTGCGGACTATGAC
*Bp*3HAO-E110A_for	GTTTGTCTGGTTATT**GCG**CGTCAGCGTCCGGCAG
*Bp*3HAO-E110A_rev	CTGCCGGACGCTGACG**CGC**AATAACCAGACAAAC
*Bp*3HAO-I142R_for	GTTCAGCTGAAAAGC**CGT**GTTAGCGATCTGC
*Bp*3HAO-I142R_rev	GCAGATCGCTAAC**ACG**GCTTTTCAGCTGAAC
*Bsp*4A3HBA23D-WT _for (*Pag*I)	AATCA<u>TCATG</u>ATCATCCTGGAAAACTTCAAAATGC
*Bsp*4A3HBA23D-WT _rev (*HindIII*)	AATCA<u>AAGCTT</u>TTAATCGCGACCCGGTG
*Bsp*4A3HBA23D-H52A_for	CGTGAATATCGTAGCGAATTT**GCA**ATTAACGCCAGCTATGAAATC
*Bsp*4A3HBA23D-H52A_rev	GATTTCATAGCTGGCGTTAAT**TGC**AAATTCGCTACGATATTCACG
*Bsp*4A3HBA23D-N54A_for	CGTAGCGAATTTCATATT**GCA**GCCAGCTATGAAATCCAG
*Bsp*4A3HBA23D-N54A_rev	CTGGATTTCATAGCTGGC**TGC**AATATGAAATTCGCTACG
*Bsp*4A3HBA23D-E58A_for	CATATTAACGCCAGCTAT**GCA**ATCCAGTATAGCCTGAAAG
*Bsp*4A3HBA23D-E58A_rev	CTTTCAGGCTATACTGGAT**TGC**ATAGCTGGCGTTAATATG
*Bsp*4A3HBA23D-H97A_for	CTCCGTTTCTGCCG**GCA**AGTCCGCGTTTTG
*Bsp*4A3HBA23D-H97A_rev	CAAAACGCGGACT**TGC**CGGCAGAAACGGAG
*Bsp*4A3HBA23D-Q60A_for	CGCCAGCTATGAAATC**GCA**TATAGCCTGAAAGGTGC
*Bsp*4A3HBA23D-Q60A_rev	GCACCTTTCAGGCTATA**TGC**GATTTCATAGCTGGCG
*Bsp*4A3HBA23D-Q60F_for	CGCCAGCTATGAAATC**TTT**TATAGCCTGAAAGGTGC
*Bsp*4A3HBA23D-Q60F_rev	GCACCTTTCAGGCTATA**AAA**GATTTCATAGCTGGCG
*Bsp*4A3HBA23D-S62A_for	CAGCTATGAAATCCAGTAT**GCA**CTGAAAGGTGCACAGGATC
*Bsp*4A3HBA23D-S62A_rev	GATCCTGTGCACCTTTCAG**TGC**ATACTGGATTTCATAGCTG
*Bsp*4A3HBA23D-S62Q_for	CAGCTATGAAATCCAGTAT**CAG**CTGAAAGGTGCACAGGATC
*Bsp*4A3HBA23D-S62Q_rev	GATCCTGTGCACCTTTCAG**CTG**ATACTGGATTTCATAGCTG
*Bsp*4A3HBA23D-F101A_for	CATAGTCCGCGT**GCA**GCACCGGATAGC
*Bsp*4A3HBA23D-F101A_rev	GCTATCCGGTGC**TGC**ACGCGGACTATG
*Bsp*4A3HBA23D-F101R_for	CATAGTCCGCGT**CGT**GCACCGGATAG
*Bsp*4A3HBA23D-F101R_rev	CTATCCGGTGC**ACG**ACGCGGACTAT
*Bsp*4A3HBA23D-R144I_for	CTTCTATGTGGACGATTAT**ATC**AAAGATCCGGTTAGCCGTG
*Bsp*4A3HBA23D-R144I_rev	CACGGCTAACCGGATCTTT**GAT**ATAATCGTCCACATAGAAG
*Bsp*4A3HBA23D-T28A_for	GAAACCGGTAAACGT**GCC**CATCAGCTGTGG
*Bsp*4A3HBA23D-T28A_rev	CCACAGCTGATGGG**CAC**GTTTACCGGTTTC
*Bsp*4A3HBA23D-T28N_for	GAAACCGGTAAACGTA**AC**CATCAGCTGTGGATG
*Bsp*4A3HBA23D-T28N_rev	CATCCACAGCTGATGG**T**TACGTTTACCGGTTTC
*Bsp*4A3HBA23D-F40A_for	GATGAAACCCTGGCA**GCA**GTTGCACGTGGTCG
*Bsp*4A3HBA23D-F40A_rev	CGACCACGTGCAAC**TGC**TGCCAGGGTTTCATC
*Bsp*4A3HBA23D-F40T_for	GATGAAACCCTGGCA**ACC**GTTGCACGTGGTCG
*Bsp*4A3HBA23D-F40T_rev	CGACCACGTGCAAC**GGT**TGCCAGGGTTTCATC
*Bsp*4A3HBA23D-E46A_for	GTTGCACGTGGTCGTG**C**ATATCGTAGCGAATTTC
*Bsp*4A3HBA23D-E46A_rev	GAAATTCGCTACGATAT**G**CACGACCACGTGCAAC

Primer Name (mutation)	Sequence
*Bsp*4A3HBA23D-E46N_for	GTTGCACGTGGTCGT**AAC**TATCGTAGCGAATTTC
*Bsp*4A3HBA23D-E46N_rev	GAAATTCGCTACGATA**GTT**ACGACCACGTGCAAC
*Bsp*4A3HBA23D-E111A_for	CTTTCAGTTTATTATCG**C**ACGTGTTCGTAAACCG
*Bsp*4A3HBA23D-E111A_rev	CGGTTTACGAACACGT**G**CGATAATAAACTGAAAG
*Bsp*4A3HBA23D-K145V_for	GTGGACGATTATCGT**GT**AGATCCGGTTAGCC
*Bsp*4A3HBA23D-K145V_rev	GGCTAACCGGATCT**AC**ACGATAATCGTCCAC
*Bsp*4A3HBA23D-R144I-K145V_for	CTTCTATGTGGACGATTAT**ATCGT**AGATCCGGTTAGCCGTG
*Bsp*4A3HBA23D -R144I-K145V_rev	CACGGCTAACCGGATCT**ACGAT**ATAATCGTCCACATAGAAG

4. SUMMARISING DISCUSSION

This thesis takes a broad perspective at the cupin superfamily of proteins as a source of novel natural and engineered enzymatic activities. As has been illustrated in the introductory chapter, the functional diversity of cupins is indeed striking. Alongside non-enzymatic members, there is a large and growing repertoire of cupins with enzymatic function. Additionally, many cupins are known from high-throughput sequencing and structural genomics projects and for many if not most of these a function has not been assigned to date. This abundance forms the basis for the exploration and rationalisation of the diversity found in cupins. Starting from available literature, discoveries by previous members of our group and from database searches, a set of interesting cupins was selected for more in-depth biochemical analyses.

As part of this thesis a novel cupin hydroxynitrile lyase (HNL), termed *Gt*HNL, has been characterised in detail (Hajnal et al. 2013). Cupins with hydroxynitrile-lyase activity have been discovered earlier (Hussain et al. 2012), and *Gt*HNL has high sequence similarity to these previously-discovered cupin HNLs. However, the detailed biochemical characterisation of *Gt*HNL represents the first for this type of enzymes. Many cupins are metal-binding proteins and this enzyme was also shown to be metal-dependent. The active-site metal could be extracted and the apoprotein reconstituted with manganese, zinc or iron *in vitro*. It could further be shown that the extraction of metal does not affect overall folding or pH stability, indicating that the metal is important directly for catalytic function rather than simply folding or stability. Using these metal-exchange experiments a preference for manganese was demonstrated. Furthermore, active-site residues could be identified with site-directed mutagenesis. While mutations in metal-binding residues impact activity, and thus further support the metal-dependence of the enzyme, mutants of two histidines in the active site which are not directly involved in metal-binding also lost activity. Loss-of-function mutants were purified and the overall folding was confirmed to correspond to that of the wild-type enzyme using CD spectroscopy. While the exact reaction mechanism could not be established yet, the activity was shown to definitely depend both on metal and the protein backbone. Furthermore, the lack of any obvious structural or sequence similarity of *Gt*HNL with HNLs described so far, together with the manganese dependence, strongly indicates a unique

mechanism and underscores the contribution that this characterisation of a cupin HNL contributes to the overall understanding of hydroxynitrile lyases.

Since co-crystallisation or soaking of substrate into crystals of *Gt*HNL was not successful so far, a protein NMR approach was attempted as an alternative route towards a substrate-bound protein structure. *Gt*HNL was expressed efficiently in minimal medium, labelled with ^{15}N, and the labelled protein produced promising NMR spectra. Encouragingly, shifts in the spectra which appear strong enough not to be due to pH variations alone, have been observed when the enzyme was contacted with either the substrate mandelonitrile or the substrate analog benzoic acid. Taken together, first encouraging steps have been taken towards a substrate-bound solution structure of *Gt*HNL, and the thesis contributes multiple approaches towards the determination of the mechanism of this unusual manganese-dependent class of HNL enzymes.

In addition to the biochemical characterisation of an enzymatic function discovered previously in cupins, one part of this thesis describes the discovery of a novel alkene-cleavage activity in a cupin from *Thermotoga maritima*. No function was known previously for this protein, and no close relatives with a known function are found in BLAST searches. Despite the novelty of this enzyme, its crystal structure has been solved in a high-throughput structural genomics project (Jaroszewski et al. 2004). The alkene-cleavage reaction which is catalysed by this enzyme appears to proceed exclusively at the expense of *tert*-butyl hydroperoxide as terminal oxidant. The activity was shown to be strongly stimulated by manganese bound in the protein, as protein expressed with the addition of manganese to the medium shows increased manganese loading and increased activity. Additionally, protein expressed without manganese addition is deficient in manganese loading and activity, but can be activated by the addition of manganese salts to reaction mixtures. The enzyme has been characterised regarding its substrate scope, optimal reaction conditions, oxidant preference and tolerance of organic co-solvents. It cleaves different styrene derivatives to aldehydes or in the case of α-methyl styrenes to ketones. This is the first report of this type of activity found in cupins, and in addition to possible future applications in biocatalysis, the discovery could help in efforts to annotate the functions of some of the cupins with unknown function and high sequence similarity to this protein. It should be noted that the alkene cleavage activity of this enzyme is possibly not its natural function since *Thermotoga maritima* is an obligate anaerobe (Huber et al., 1986) and the presence of high concentrations of terminal oxidants such as peroxides or

molecular oxygen is highly unlikely in the natural source organism. However, even if the natural function of this protein might be a different one, the activity reported in this thesis is highly interesting from a biocatalysis perspective and contributes to our understanding of the diversity of reactions catalysed by cupins, both in nature and in the laboratory.

A third chapter of this thesis describes the characterisation of two dioxygenases which act on similar hydroxyaminobenzoic acids, but neither of which accepts the natural substrate of the other. The native substrate of the first is 3-hydroxyanthranilic acid (3-HAA, 2-amino-3-hydroxybenzoic acid), whereas the native substrate of the other is 4-amino-3-hydroxybenzoic acid (4A3HBA). Furthermore, the reaction catalysed by both enzymes can be described in simple terms as a dioxygenase-type ring-cleavage reaction adjacent to a hydroxy group. Within this context, the two enzymes can be used as a model system to understand the sometimes very high substrate specificity of cupin dioxygenases (Fetzner 2012). The 3-HAA dioxygenase (*Bp*3HAO), shows high sequence similarity to known 3-HAA dioxygenases and the activity was identified by a standard assay (Zhang et al. 2005). This enzyme has been characterised biochemically and the kinetic data of the wild-type protein and different mutant proteins have been determined. The 4A3HBA dioxygenase from *Bordetella sp.* (*Bsp*4A3HBA23D) has already been characterised biochemically (Takenaka et al. 2002). This thesis contributes a mutational analysis of this enzyme as well as a comparison to *Bp*3HAO and work towards the clarification of determinants of substrate specificity. In both enzymes a pattern is apparent in which a positively charged amino acid, which contributes significantly to activity, is positioned in the active site in close proximity of the carboxylic acid moiety of the substrate. At a complementary position in the active site of each of the enzymes there is a hydrophobic residue. An attempt was made to switch the substrate specificity by swapping the positively charged and hydrophobic residues in these complementary positions in each active site. Mutants which position a positive charge in *Bp*3HAO at a position where there is a positive charge in a *Bsp*4A3HBA23D homology model did not show detectable activity toward 4A3HBA. However, an equivalent approach was also attempted in *Bsp*4A3HBA23D, and mutants were generated which showed detectable activity on the substrate 3-HAA, whereas the wild-type enzyme showed no detectable activity towards this substrate whatsoever. The mutant F101R had a specific activity of $0.15 \cdot 10^{-4}$ U·mg^{-1}, which is much lower than the natural activity, but was clearly detectable photometrically as has been illustrated in the corresponding thesis chapter. When this mutation was combined with mutations which substitute positive charges at the opposing side of the active site for

hydrophobic residues, a 2- and 3-fold increase in specific activity was seen in F101R/R144I and F101R/K145V, respectively. The data thus support the hypothesis that the positioning of a positively charged residue within the active site, together with the positioning of hydrophobic residues, is an important determinant of substrate specificity in this type of dioxygenases.

Taken as a whole, this thesis investigates three quite different types of enzymatic activity found in cupins. Among the three activities, one represents a novel catalytic activity in a previously unassigned member of this large and diverse superfamily. The hydroxynitrile lyase activity and the ene-cleavage activity of cupins have been described in some detail in this thesis, and this is significant since both enzymatic activities have great potential in biocatalysis and the fine chemicals industry. The thesis thus illustrates the versatility of activities found in cupins and contributes significant research towards further understanding and utilisation of cupins as biocatalysts.

5. REFERENCES

Adams, M., and Jia, Z. (2005). Structural and Biochemical Analysis Reveal Pirins to Possess Quercetinase Activity. J. Biol. Chem. *280*, 28675–28682.

Adams, M.A., Singh, V.K., Keller, B.O., and Jia, Z. (2006). Structural and biochemical characterization of gentisate 1,2-dioxygenase from *Escherichia coli* O157:H7. Mol. Microbiol. *61*, 1469–1484.

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. *66*, 213–221.

Agarwal, G., Rajavel, M., Gopal, B., and Srinivasan, N. (2009). Structure-Based Phylogeny as a Diagnostic for Functional Characterization of Proteins with a Cupin Fold. PLoS ONE *4*, e5736.

Allard, M., Dupont, C., Muñoz Robles, V., Doucet, N., Lledós, A., Maréchal, J.-D., Urvoas, A., Mahy, J.-P., and Ricoux, R. (2012). Incorporation of Manganese Complexes into Xylanase: New Artificial Metalloenzymes for Enantioselective Epoxidation. ChemBioChem *13*, 240–251.

Altschul, S.F., Wootton, J.C., Gertz, E.M., Agarwala, R., Morgulis, A., Schaffer, A.A., and Yu, Y.-K. (2005). Protein database searches using compositionally adjusted substitution matrices. FEBS J. *272*, 5101–5109.

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 (Mosc.) *41*, 7659–7669.

Andexer, J.N., Staunig, N., Eggert, T., Kratky, C., Pohl, M., and Gruber, K. (2012). Hydroxynitrile Lyases with α/β-Hydrolase Fold: Two Enzymes with Almost Identical 3D Structures but Opposite Enantioselectivities and Different Reaction Mechanisms. ChemBioChem *13*, 1932–1939.

Andreeva, A., Howorth, D., Chandonia, J.-M., Brenner, S.E., Hubbard, T.J.P., Chothia, C., and Murzin, A.G. (2007). Data growth and its impact on the SCOP database: new developments. Nucleic Acids Res. *36*, D419–D425.

Andrews, S.C., Robinson, A.K., and Rodríguez-Quiñones, F. (2003). Bacterial iron homeostasis. FEMS Microbiol. Rev. *27*, 215–237.

Aravind, L., Anantharaman, V., Balaji, S., Babu, M.M., and Iyer, L.M. (2005). The many faces of the helix-turn-helix domain: Transcription regulation and beyond. FEMS Microbiol. Rev. *29*, 231–262.

Barney, B.M., Schaab, M.R., LoBrutto, R., and Francisco, W.A. (2004). Evidence for a new metal in a known active site: purification and characterization of an iron-containing quercetin 2,3-dioxygenase from *Bacillus subtilis*. Protein Expr. Purif. *35*, 131–141.

Bernier, F., and Berna, A. (2001). Germins and germin-like proteins: Plant do-all proteins. But what do they do exactly? Plant Physiol. Biochem. *39*, 545–554.

Berrisford, J.M., Akerboom, J., Brouns, S., Sedelnikova, S.E., Turnbull, A.P., van der Oost, J., Salmon, L., Hardré, R., Murray, I.A., Blackburn, G.M., et al. (2004). The Structures of Inhibitor Complexes of *Pyrococcus furiosus* Phosphoglucose Isomerase Provide Insights into Substrate Binding and Catalysis. J. Mol. Biol. *343*, 649–657.

Berrisford, J.M., Hounslow, A.M., Akerboom, J., Hagen, W.R., Brouns, S.J.J., van der Oost, J., Murray, I.A., Michael Blackburn, G., Waltho, J.P., Rice, D.W., et al. (2006). Evidence Supporting a cis-enediol-based Mechanism for *Pyrococcus furiosus* Phosphoglucose Isomerase. J. Mol. Biol. *358*, 1353–1366.

Bowater, L., Fairhurst, S.A., Just, V.J., and Bornemann, S. (2004). *Bacillus subtilis* YxaG is a novel Fe-containing quercetin 2,3-dioxygenase. FEBS Lett. *557*, 45–48.

Breiteneder, H., and Mills, E.N.C. (2005). Plant food allergens—structural and functional aspects of allergenicity. Biotechnol. Adv. *23*, 395–399.

Bugg, T.D., and Ramaswamy, S. (2008). Non-heme iron-dependent dioxygenases: unravelling catalytic mechanisms for complex enzymatic oxidations. Curr. Opin. Chem. Biol. *12*, 134–140.

Burrell, M.R., Just, V.J., Bowater, L., Fairhurst, S.A., Requena, L., Lawson, D.M., and Bornemann, S. (2007). Oxalate Decarboxylase and Oxalate Oxidase Activities Can Be Interchanged with a Specificity Switch of up to 282 000 by Mutating an Active Site Lid. Biochemistry (Mosc.) *46*, 12327–12336.

Calderone, V., Trabucco, M., Menin, V., Negro, A., and Zanotti, G. (2002). Cloning of human 3-hydroxyanthranilic acid dioxygenase in *Escherichia coli*: characterisation of the purified enzyme and its in vitro inhibition by Zn^{2+}. Biochim. Biophys. Acta BBA - Protein Struct. Mol. Enzym. *1596*, 283–292.

Chavez, F.A., Banerjee, A., and Sljivic, B. (2011). Modeling the Metal Binding Site in Cupin Proteins. In On Biomimetics, ISBN 978-953-307-271-5 (InTech).

Chen, J., Li, W., Wang, M., Zhu, G., Liu, D., Sun, F., Hao, N., Li, X., Rao, Z., and Zhang, X.C. (2008). Crystal structure and mutagenic analysis of GDOsp, a gentisate 1,2-dioxygenase from *Silicibacter Pomeroyi*. Protein Sci. *17*, 1362–1373.

Clissold, P.M., and Ponting, C.P. (2001). JmjC: cupin metalloenzyme-like domains in jumonji, hairless and phospholipase A2β. Trends Biochem. Sci. *26*, 7–9.

Conesa, A., Punt, P.J., and van den Hondel, C.A.M.J.J. (2002). Fungal peroxidases: molecular aspects and applications. J. Biotechnol. *93*, 143–158.

Costes, D., Rotčnkovs, G., Wehtje, E., and Adlercreutz, P. (2001). Stability and Stabilization of Hydroxynitrile Lyase in Organic Solvents. Biocatal. Biotransformation *19*, 119–130.

Crawford, R.L., Hutton, S.W., and Chapman, P.J. (1975). Purification and properties of gentisate 1,2-dioxygenase from *Moraxella osloensis*. J. Bacteriol. *121*, 794–799.

Curson, A.R.J., Sullivan, M.J., Todd, J.D., and Johnston, A.W.B. (2011). DddY, a periplasmic dimethylsulfoniopropionate lyase found in taxonomically diverse species of Proteobacteria. ISME J. *5*, 1191–1200.

Dadashipour, M., and Asano, Y. (2011). Hydroxynitrile Lyases: Insights into Biochemistry, Discovery, and Engineering. ACS Catal. *1*, 1121–1149.

Dai, Y., Wensink, P.C., and Abeles, R.H. (1999). One Protein, Two Enzymes. J. Biol. Chem. *274*, 1193–1195.

Dai, Y., Pochapsky, T.C., and Abeles, R.H. (2001). Mechanistic Studies of Two Dioxygenases in the Methionine Salvage Pathway of *Klebsiella pneumonia*. Biochemistry (Mosc.) *40*, 6379–6387.

Dias, B.B.A., Cunha, W.G., Morais, L.S., Vianna, G.R., Rech, E.L., Capdeville, G., and Aragao, F.J.L. (2006). Expression of an oxalate decarboxylase gene from *Flammulina sp.* in transgenic lettuce (*Lactuca sativa*) plants and resistance to *Sclerotinia sclerotiorum*. Plant Pathol. *55*, 187–193.

Đilović, I., Gliubich, F., Malpeli, G., Zanotti, G., and Matković-Čalogović, D. (2009). Crystal structure of bovine 3-hydroxyanthranilate 3,4-dioxygenase. Biopolymers *91*, 1189–1195.

Dominy, J.E., Simmons, C.R., Hirschberger, L.L., Hwang, J., Coloso, R.M., and Stipanuk, M.H. (2007). Discovery and Characterization of a Second Mammalian Thiol Dioxygenase, Cysteamine Dioxygenase. J. Biol. Chem. *282*, 25189–25198.

Dong, C., Major, L., Allen, A., Blankenfeldt, W., Maskell, D., and Naismith, J. (2003). High-Resolution Structures of RmlC from *Streptococcus suis* in Complex with Substrate Analogs Locate the Active Site of This Class of Enzyme. Structure *11*, 715–723.

Dong, X., Ji, R., Guo, X., Foster, S.J., Chen, H., Dong, C., Liu, Y., Hu, Q., and Liu, S. (2008). Expressing a gene encoding wheat oxalate oxidase enhances resistance to *Sclerotinia sclerotiorum* in oilseed rape (*Brassica napus*). Planta *228*, 331–340.

Dreveny, I., Kratky, C., and Gruber, K. (2009). The active site of hydroxynitrile lyase from *Prunus amygdalus*: Modeling studies provide new insights into the mechanism of cyanogenesis. Protein Sci. *11*, 292–300.

Dunwell, J.M. (1998). Cupins: a new superfamily of functionally diverse proteins that include germins and plant storage proteins. Biotechnol. Genet. Eng. Rev. *15*, 1–32.

Dunwell, J.M., and Gane, P.J. (1998). Microbial Relatives of Seed Storage Proteins: Conservation of Motifs in a Functionally Diverse Superfamily of Enzymes. J. Mol. Evol. *46*, 147–154.

Dunwell, J.M., Culham, A., Carter, C.E., Sosa-Aguirre, C.R., and Goodenough, P.W. (2001). Evolution of functional diversity in the cupin superfamily. Trends Biochem. Sci. *26*, 740–746.

Dunwell, J.M., Purvis, A., and Khuri, S. (2004). Cupins: the most functionally diverse protein superfamily?. Phytochemistry *65*, 7–17.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

Eriksson, A.E., Jones, T.A., and Liljas, A. (1988). Refined structure of human carbonic anhydrase II at 2.0 Å resolution. Proteins Struct. Funct. Genet. *4*, 274–282.

Evans, P. (2006). Scaling and assessment of data quality. Acta Crystallogr. D Biol. Crystallogr. *62*, 72–82.

Fernández-Gacio, A., Codina, A., Fastrez, J., Riant, O., and Soumillion, P. (2006). Transforming Carbonic Anhydrase into Epoxide Synthase by Metal Exchange. ChemBioChem *7*, 1013–1016.

Ferraroni, M., Steimer, L., Matera, I., Bürger, S., Scozzafava, A., Stolz, A., and Briganti, F. (2012). The generation of a 1-hydroxy-2-naphthoate 1,2-dioxygenase by single point mutations of salicylate 1,2-dioxygenase – Rational design of mutants and the crystal structures of the A85H and W104Y variants. J. Struct. Biol. *180*, 563–571.

Fesko, K., and Gruber-Khadjawi, M. (2013). Biocatalytic Methods for C-C Bond Formation. ChemCatChem *5*, 1248–1272.

Fetzner, S. (2012). Ring-cleaving dioxygenases with a cupin fold. Appl. Environ. Microbiol. *78*, 2505–2514.

Formica, J.V., and Regelson, W. (1995). Review of the biology of quercetin and related bioflavonoids. Food Chem. Toxicol. *33*, 1061–1080.

Fusetti, F., Schröter, K.H., Steiner, R.A., van Noort, P.I., Pijning, T., Rozeboom, H.J., Kalk, K.H., Egmond, M.R., and Dijkstra, B.W. (2002). Crystal Structure of the Copper-Containing Quercetin 2,3-Dioxygenase from Aspergillus japonicus. Structure *10*, 259–268.

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, J.M. Walker, ed. (Totowa, NJ: Humana Press), pp. 571–607.

Giraud, M.F., Leonard, G.A., Field, R.A., Berlind, C., and Naismith, J.H. (2000). RmlC, the third enzyme of dTDP-L-rhamnose pathway, is a new class of epimerase. Nat. Struct. Biol. *7*, 398–402.

Gopal, B., Madan, L.L., Betz, S.F., and Kossiakoff, A.A. (2005). The Crystal Structure of a Quercetin 2,3-Dioxygenase from *Bacillus subtilis* Suggests Modulation of Enzyme Activity by a Change in the Metal Ion at the Active Site(s). Biochemistry (Mosc.) *44*, 193–201.

Gorrec, F. (2009). The MORPHEUS protein crystallization screen. J. Appl. Crystallogr. *42*, 1035–1042.

Groce, S.L., and Lipscomb, J.D. (2003). Conversion of Extradiol Aromatic Ring-Cleaving Homoprotocatechuate 2,3-Dioxygenase into an Intradiol Cleaving Enzyme. J. Am. Chem. Soc. *125*, 11780–11781.

Gruber, K., Gartler, G., Krammer, B., Schwab, H. and Kratky, C. (2004). Reaction Mechanism of Hydroxynitrile Lyases of the α/β -Hydrolase Superfamily: THE THREE-DIMENSIONAL STRUCTURE OF THE TRANSIENT ENZYME-SUBSTRATE COMPLEX CERTIFIES THE CRUCIAL ROLE OF LYS236. J. Biol. Chem. *279*, 20501–20510.

Gruber, K., and Kratky, C. (2004). Biopolymers for biocatalysis: Structure and catalytic mechanism of hydroxynitrile lyases. J. Polym. Sci. Part Polym. Chem. *42*, 479–486.

Gruber-Khadjawi, M., Fechter, M.H., and Griengl, H. (2012). Cleavage and Formation of Cyanohydrins. In Enzyme Catalysis in Organic Synthesis, K. Drauz, H. Gröger, and O. May, eds. (Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA), pp. 947–990.

Guterl, J.-K., Andexer, J.N., Sehl, T., von Langermann, J., Frindi-Wosch, I, Rosenkranz, T., Fitter, J., Gruber, K., Kragl, U., Eggert, T., et al. (2009). Uneven twins: Comparison of two enantiocomplementary hydroxynitrile lyases with α/β-hydrolase fold. J. Biotechnol. *141*, 166–173.

Hajnal, I., Łyskowski, A., Hanefeld, U., Gruber, K., Schwab, H., and Steiner, K. (2013). Biochemical and structural characterization of a novel bacterial manganese-dependent hydroxynitrile lyase. FEBS J. *280*, 5815–5828.

Hansen, T., Schlichting, B., Felgendreher, M., and Schonheit, P. (2005). Cupin-Type Phosphoglucose Isomerases (Cupin-PGIs) Constitute a Novel Metal-Dependent PGI Family Representing a Convergent Line of PGI Evolution. J. Bacteriol. *187*, 1621–1631.

Harpel, M.R., and Lipscomb, J.D. (1990). Gentisate 1,2-dioxygenase from *Pseudomonas*. Substrate coordination to active site Fe^{2+} and mechanism of turnover. J. Biol. Chem. *265*, 22187–22196.

Hewitson, K.S., McNeill, L.A., Riordan, M.V., Tian, Y.M., Bullock, A.N., Welford, R.W., Elkins, J.M., Oldham, N.J., Bhattacharya, S., Gleadle, J.M., Ratcliffe, P.J., Pugh, C.W., Schofield, C.J. (2002). Hypoxia-inducible Factor (HIF) Asparagine Hydroxylase Is Identical to Factor Inhibiting HIF (FIH) and Is Related to the Cupin Structural Family. J. Biol. Chem. *277*, 26351–26355.

Hilterhaus, L., and Liese, A. (2012). Industrial Application and Processes Using Carbon-Carbon Lyases. In Enzyme Catalysis in Organic Synthesis, K. Drauz, H. Gröger, and O. May, eds. (Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA), pp. 991–1000.

Hirano, S., Morikawa, M., Takano, K., Imanaka, T., and Kanaya, S. (2007). Gentisate 1,2-Dioxygenase from Xanthobacter polyaromaticivorans 127W. Biosci. Biotechnol. Biochem. *71*, 192–199.

Holm, L., and Rosenstrom, P. (2010). Dali server: conservation mapping in 3D. Nucleic Acids Res. *38*, W545–W549.

Holt, J., and Hanefeld, U. (2009). Enantioselective Enzyme-Catalysed Synthesis of Cyanohydrins. Curr. Org. Synth. *6*, 15–37.

Horike, S., Dinca, M., Tamaki, K., and Long, J.R. (2008). Size-selective Lewis acid catalysis in a microporous metal-organic framework with exposed Mn^{2+} coordination sites. J. Am. Chem. Soc. *130*, 5854–5855.

Huber, R., Langworthy, T.A., König, H., Thomm, M., Woese, C.R., Sleytr, U.B., and Stetter, K.O. (1986). *Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C. Arch. Microbiol. *144*, 324–333.

Hussain, Z., Wiedner, R., Steiner, K., Hajek, T., Avi, M., Hecher, B., Sessitsch, A., and Schwab, H. (2012). Characterization of Two Bacterial Hydroxynitrile Lyases with High Similarity to Cupin Superfamily Proteins. Appl. Environ. Microbiol. *78*, 2053–2055.

Iwabuchi, T., and Harayama, S. (1998). Biochemical and Molecular Characterization of 1-Hydroxy-2-naphthoate Dioxygenase from *Nocardioides sp.* KP7. J. Biol. Chem. *273*, 8332–8336.

Iyer, L.M., Abhiman, S., de Souza, R.F., and Aravind, L. (2010). Origin and evolution of peptide-modifying dioxygenases and identification of the wybutosine hydroxylase/hydroperoxidase. Nucleic Acids Res. *38*, 5261–5279.

Jaroszewski, L., Schwarzenbacher, R., von Delft, F., McMullan, D., Brinen, L.S., Canaves, J.M., Dai, X., Deacon, A.M., DiDonato, M., Elsliger, M.-A., et al. (2004). Crystal structure of a novel manganese-containing cupin (TM1459) from *Thermotoga maritima* at 1.65 Å resolution. Proteins Struct. Funct. Bioinforma. *56*, 611–614.

Jeong, J.-J., Fushinobu, S., Ito, S., Jeon, B.-S., Shoun, H., and Wakagi, T. (2003). Characterization of the cupin-type phosphoglucose isomerase from the hyperthermophilic archaeon *Thermococcus litoralis*. FEBS Lett. *535*, 200–204.

Ju, T., Goldsmith, R.B., Chai, S.C., Maroney, M.J., Pochapsky, S.S., and Pochapsky, T.C. (2006). One Protein, Two Enzymes Revisited: A Structural Entropy Switch Interconverts the Two Isoforms of Acireductone Dioxygenase. J. Mol. Biol. *363*, 823–834.

Just, V.J. (2004). A Closed Conformation of *Bacillus subtilis* Oxalate Decarboxylase OxdC Provides Evidence for the True Identity of the Active Site. J. Biol. Chem. *279*, 19867–19874.

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

Kesarwani, M., Azam, M., Natarajan, K., Mehta, A., and Datta, A. (2000). Oxalate Decarboxylase from *Collybia velutipes*. MOLECULAR CLONING AND ITS

OVEREXPRESSION TO CONFER RESISTANCE TO FUNGAL INFECTION IN TRANSGENIC TOBACCO AND TOMATO. J. Biol. Chem. *275*, 7230–7238.

Koontz, W.A., and Shiman, R. (1976). Beef kidney 3-hydroxyanthranilic acid oxygenase. Purification, characterization, and analysis of the assay. J. Biol. Chem. *251*, 368–377.

Krammer, B., Rumbold, K., Tschemmernegg, M., Pöchlauer, P., and Schwab, H. (2007). A novel screening assay for hydroxynitrile lyases suitable for high-throughput screening. J. Biotechnol. *129*, 151–161.

Kucharczyk, R., Zagulski, M., Rytka, J., Herbert, C.J., (1998). The yeast gene YJR025c encodes a 3-hydroxyanthranilic acid dioxygenase and is involved in nicotinic acid biosynthesis. FEBS Lett. *424*, 127–130.

Kurnasov, O., Goral, V., Colabroy, K., Gerdes, S., Anantha, S., Osterman, A., and Begley, T.P. (2003). NAD biosynthesis: identification of the tryptophan to quinolinate pathway in bacteria. Chem. Biol. *10*, 1195–1204.

Lane, B.G. (1994). Oxalate, germin, and the extracellular matrix of higher plants. FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol. *8*, 294–301.

Lane, B.S., and Burgess, K. (2003). Metal-Catalyzed Epoxidations of Alkenes with Hydrogen Peroxide. Chem. Rev. *103*, 2457–2474.

Lane, B.G., Bernier, F., Dratewka-Kos, E., Shafai, R., Kennedy, T.D., Pyne, C., Munro, J.R., Vaughan, T., Walters, D., and Altomare, F. (1991). Homologies between members of the germin gene family in hexaploid wheat and similarities between these wheat germins and certain *Physarum* spherulins. J. Biol. Chem. *266*, 10461–10469.

Lane, B.G., Dunwell, J.M., Ray, J.A., Schmitt, M.R., and Cuming, A.C. (1993). Germin, a protein marker of early plant development, is an oxalate oxidase. J. Biol. Chem. *268*, 12239–12242.

Lara, M., Mutti, F.G., Glueck, S.M., and Kroutil, W. (2008). Biocatalytic Cleavage of Alkenes with O$_2$ and *Trametes hirsuta* G FCC 047. Eur. J. Org. Chem. *2008*, 3668–3672.

Larsson, S., Cassland, P., Jönsson, L.J., and Nilvebrant, N.-O. (2003). Treatment of Pulp and Paper Industry Process Waters with Oxalate Oxidase: Compounds Interfering with the Activity. In Applications of Enzymes to Lignocellulosics, S.D. Mansfield, and J.N. Saddler, eds. (Washington, DC: American Chemical Society), pp. 81–92.

Laskowski, R.A., Hutchinson, E.G., Michie, A.D., Wallace, A.C., Jones, M.L., and Thornton, J.M. (1997). PDBsum: a Web-based database of summaries and analyses of all PDB structures. Trends Biochem. Sci. *22*, 488–490.

Leitgeb, S., and Nidetzky, B. (2008). Structural and functional comparison of 2-His- 1-carboxylate and 3-His metallocentres in non-haem iron(II)-dependent enzymes. Biochem. Soc. Trans. *36*, 1180.

Leitgeb, S., and Nidetzky, B. (2010). Enzyme Catalytic Promiscuity: The Nonheme Fe^{2+} Center of β-Diketone-Cleaving Dioxygenase Dke1 Promotes Hydrolysis of Activated Esters. ChemBioChem *11*, 502–505.

Leitgeb, S., Straganz, G.D., and Nidetzky, B. (2009). Biochemical characterization and mutational analysis of the mononuclear non-haem Fe^{2+} site in Dke1, a cupin-type dioxygenase from *Acinetobacter johnsonii*. Biochem. J. *418*, 403.

Lendenmann, U., and Spain, J.C. (1996). 2-aminophenol 1,6-dioxygenase: a novel aromatic ring cleavage enzyme purified from *Pseudomonas pseudoalcaligenes* JS45. J. Bacteriol. *178*, 6227–6232.

Lesley, S.A., Kuhn, P., Godzik, A., Deacon, A.M., Mathews, I., Kreusch, A., Spraggon, G., Klock, H.E., McMullan, D., Shin, T., Vincent, J., Robb, A., Brinen, L.S., Miller, M.D., McPhillips, T.M., Miller, M.A., Scheibe, D., Canaves, J.M., Guda, C., Jaroszewski, L., Selby, T.L., Elsliger, M.A., Wooley, J., Taylor, S.S., Hodgson, K.O., Wilson, I.A., Schultz, P.G., Stevens, R.C. (2002). Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. Proc. Natl. Acad. Sci. *99*, 11664–11669.

Li, X., Guo, M., Fan, J., Tang, W., Wang, D., Ge, H., Rong, H., Teng, M., Niu, L., Liu, Q., Hao. Q. (2006). Crystal structure of 3-hydroxyanthranilic acid 3,4-dioxygenase from

Saccharomyces cerevisiae: A special subgroup of the type III extradiol dioxygenases. Protein Sci. *15*, 761–773.

Li, M.G., and Madappally, M.M. (1989). Rapid enzymatic determination of urinary oxalate. Clin. Chem. *35*, 2330–2333.

Li, D.-F., Zhang, J.-Y., Hou, Y.-J., Liu, L., Hu, Y., Liu, S.-J., Wang, D.-C., and Liu, W. (2012). Structures of aminophenol dioxygenase in complex with intermediate, product and inhibitor. Acta Crystallogr. D Biol. Crystallogr. *69*, 32–43.

Long, F., Vagin, A.A., Young, P., and Murshudov, G.N. (2008). BALBES: a molecular-replacement pipeline. Acta Crystallogr. D Biol. Crystallogr. *64*, 125–132.

Łyskowski, A., Steiner, K., Hajnal, I., Steinkellner, G., Schwab, H., and Gruber, K. (2012). Crystallization of a novel metal-containing cupin from Acidobacterium *sp.* and preliminary diffraction data analysis. Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun. *68*, 451–454.

Ma, Z., Jacobsen, F.E., and Giedroc, D.P. (2009). Coordination Chemistry of Bacterial Metal Transport and Sensing. Chem. Rev. *109*, 4644–4681.

Malherbe, P., Köhler, C., Da Prada, M., Lang, G., Kiefer, V., Schwarcz, R., Lahm, H.W., and Cesura, A.M. (1994). Molecular cloning and functional expression of human 3-hydroxyanthranilic-acid dioxygenase. J. Biol. Chem. *269*, 13792–13797.

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. Angew. Chem. Int. Ed. *45*, 5201–5203.

Mang, H., Gross, J., Lara, M., Goessler, C., Schoemaker, H.E., Guebitz, G.M., and Kroutil, W. (2007). Optimization of a biocatalytic single-step alkene cleavage of aryl alkenes. Tetrahedron *63*, 3350–3354.

Matera, I., Ferraroni, M., Bürger, S., Scozzafava, A., Stolz, A., and Briganti, F. (2008). Salicylate 1,2-Dioxygenase from *Pseudaminobacter salicylatoxidans*: Crystal Structure of a Peculiar Ring-cleaving Dioxygenase. J. Mol. Biol. *380*, 856–868.

McCoy, J.G., Bailey, L.J., Bitto, E., Bingman, C.A., Aceti, D.J., Fox, B.G., and Phillips, G.N.Jr. (2006). Structure and mechanism of mouse cysteine dioxygenase. Proc. Natl. Acad. Sci. *103*, 3084–3089.

McMullan, D., Schwarzenbacher, R., Jaroszewski, L., Delft, F. von, Klock, H.E., Vincent, J., Quijano, K., Abdubek, P., Ambing, E., Biorac, T., et al. (2004). Crystal structure of a novel *Thermotoga maritima* enzyme (TM1112) from the cupin family at 1.83 Å resolution. Proteins Struct. Funct. Bioinforma. *56*, 615–618.

Merkens, H., Sielker, S., Rose, K., and Fetzner, S. (2007). A new monocupin quercetinase of *Streptomyces sp.* FLA: identification and heterologous expression of the queD gene and activity of the recombinant enzyme towards different flavonols. Arch. Microbiol. *187*, 475–487.

Merkens, H., Kappl, R., Jakob, R.P., Schmid, F.X., and Fetzner, S. (2008). Quercetinase QueD of *Streptomyces sp.* FLA, a Monocupin Dioxygenase with a Preference for Nickel and Cobalt. Biochemistry (Mosc.) *47*, 12185–12196.

Mermet, J. (1991). The use of Mg as a test template for ICP-AES diagnostics. Anal. Chim. Acta *250*, 85–94.

Murakami, S., Sawami, Y., Takenaka, S., and Aoki, K. (2004). Cloning of a gene encoding 4-amino-3-hydroxybenzoate 2,3-dioxygenase from *Bordetella sp.* 10d. Biochem. Biophys. Res. Commun. *314*, 489–494.

Muraki, T., Taki, M., Hasegawa, Y., Iwaki, H., and Lau, P.C.K. (2003). Prokaryotic Homologs of the Eukaryotic 3-Hydroxyanthranilate 3,4-Dioxygenase and 2-Amino-3-Carboxymuconate-6-Semialdehyde Decarboxylase in the 2-Nitrobenzoate Degradation Pathway of *Pseudomonas fluorescens* Strain KU-7. Appl. Environ. Microbiol. *69*, 1564–1572.

Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr. D Biol. Crystallogr. *67*, 355–367.

Murzin, A.G., Brenner, S.E., Hubbard, T., and Chothia, C. (1995). SCOP: A structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. *247*, 536–540.

Mutti, F.G., Lara, M., Kroutil, M., and Kroutil, W. (2010). Ostensible Enzyme Promiscuity: Alkene Cleavage by Peroxidases. Chem. - Eur. J. *16*, 14142–14148.

Oka, T., and Simpson, F.J. (1971). Quercetinase, a dioxygenase containing copper. Biochem. Biophys. Res. Commun. *43*, 1–5.

Okrasa, K., and Kazlauskas, R.J. (2006). Manganese-Substituted Carbonic Anhydrase as a New Peroxidase. Chem. - Eur. J. *12*, 1587–1596.

Okrob, D., Paravidino, M., Orru, R.V.A., Wiechert, W., Hanefeld, U., and Pohl, M. (2011). Hydroxynitrile Lyase from *Arabidopsis thaliana*: Identification of Reaction Parameters for Enantiopure Cyanohydrin Synthesis by Pure and Immobilized Catalyst. Adv. Synth. Catal. *353*, 2399–2408.

Orii, C., Takenaka, S., Murakami, S., and Aoki, K. (2006). Metabolism of 4-Amino-3-hydroxybenzoic Acid by *Bordetella sp.* Strain 10d: A Different Modified Meta-Cleavage Pathway for 2-Aminophenols. Biosci. Biotechnol. Biochem. *70*, 2653–2661.

Outten, C.E., and O'Halloran, T.V. (2001). Femtomolar Sensitivity of Metalloregulatory Proteins Controlling Zinc Homeostasis. Science *292*, 2488–2492.

Pan, H.-Y., Whittaker, M.M., Bouveret, R., Berna, A., Bernier, F., and Whittaker, J.W. (2007). Characterization of wheat germin (oxalate oxidase) expressed by *Pichia pastoris*. Biochem. Biophys. Res. Commun. *356*, 925–929.

Pang, H., Bartlam, M., Zeng, Q., Miyatake, H., Hisano, T., Miki, K., Wong, L.L., Gao, G.F., and Rao, Z. (2003). Crystal Structure of Human Pirin: AN IRON-BINDING NUCLEAR PROTEIN AND TRANSCRIPTION COFACTOR. J. Biol. Chem. *279*, 1491–1498.

Paravidino, M., Sorgedrager, M.J., Orru, R.V.A., and Hanefeld, U. (2010). Activity and Enantioselectivity of the Hydroxynitrile Lyase *Me*HNL in Dry Organic Solvents. Chem. - Eur. J. *16*, 7596–7604.

Parkin, G. (2004). Synthetic analogues relevant to the structure and function of zinc enzymes. Chem. Rev. *104*, 699–767.

Paterson, I., Florence, G.J., Heimann, A.C., and Mackay, A.C. (2005). Stereocontrolled Total Synthesis of (-)-Aurisides A and B. Angew. Chem. Int. Ed. *44*, 1130–1133.

Pochapsky, T.C., Pochapsky, S.S., Ju, T., Mo, H., Al-Mjeni, F., and Maroney, M.J. (2002). Modeling and experiment yields the structure of acireductone dioxygenase from *Klebsiella pneumoniae*. Nat. Struct. Biol. *9*, 966–972.

Pocker, Y., and Stone, J.T. (1967). The Catalytic Versatility of Erythrocyte Carbonic Anhydrase. III. Kinetic Studies of the Enzyme-Catalyzed Hydrolysis of *p*-Nitrophenyl Acetate. Biochemistry (Mosc.) *6*, 668–678.

Pordea, A., Mathis, D., and Ward, T.R. (2009). Incorporation of biotinylated manganese-salen complexes into streptavidin: New artificial metalloenzymes for enantioselective sulfoxidation. J. Organomet. Chem. *694*, 930–936.

Purkarthofer, T., Skranc, W., Schuster, C., and Griengl, H. (2007). Potential and capabilities of hydroxynitrile lyases as biocatalysts in the chemical industry. Appl. Microbiol. Biotechnol. *76*, 309–320.

Radauer, C., Bublin, M., Wagner, S., Mari, A., and Breiteneder, H. (2008). Allergens are distributed into few protein families and possess a restricted number of biochemical functions. J. Allergy Clin. Immunol. *121*, 847–852.e7.

Rajavel, M., Mitra, A., and Gopal, B. (2009). Role of *Bacillus subtilis* BacB in the Synthesis of Bacilysin. J. Biol. Chem. *284*, 31882–31892.

Requena, L., and Bornemann, S. (1999). Barley (*Hordeum vulgare*) oxalate oxidase is a manganese-containing enzyme. Biochem. J. *343 Pt 1*, 185–190.

Ringenberg, M.R., and Ward, T.R. (2011). Merging the best of two worlds: artificial metalloenzymes for enantioselective catalysis. Chem. Commun. *47*, 8470.

Sakai, A., Xiang, D.F., Xu, C., Song, L., Yew, W.S., Raushel, F.M., and Gerlt, J.A. (2006). Evolution of Enzymatic Activities in the Enolase Superfamily: *N* -Succinylamino Acid

Racemase and a New Pathway for the Irreversible Conversion of D - to L -Amino Acids [†]. Biochemistry (Mosc.) *45*, 4455–4462.

Schaab, M.R., Barney, B.M., and Francisco, W.A. (2006). Kinetic and Spectroscopic Studies on the Quercetin 2,3-Dioxygenase from *Bacillus subtilis*. Biochemistry (Mosc.) *45*, 1009–1016.

Schwarcz, R., Okuno, E., White, R.J., Bird, E.D., and Whetsell, W.O., Jr (1988). 3-Hydroxyanthranilate oxygenase activity is increased in the brains of Huntington disease victims. Proc. Natl. Acad. Sci. U. S. A. *85*, 4079–4081.

Sharma, M., Sharma, N.N., and Bhalla, T.C. (2005). Hydroxynitrile lyases: At the interface of biology and chemistry. Enzyme Microb. Technol. *37*, 279–294.

Silvennoinen, L., Sandalova, T., and Schneider, G. (2009). The polyketide cyclase RemF from *Streptomyces resistomycificus* contains an unusual octahedral zinc binding site. FEBS Lett. *583*, 2917–2921.

Simmons, C.R., Liu, Q., Huang, Q., Hao, Q., Begley, T.P., Karplus, P.A., and Stipanuk, M.H. (2006). Crystal Structure of Mammalian Cysteine Dioxygenase: A NOVEL MONONUCLEAR IRON CENTER FOR CYSTEINE THIOL OXIDATION. J. Biol. Chem. *281*, 18723–18733.

Smith, S.J., Hadler, K.S., Schenk, G., Hanson, G.R., and Mitić, N. (2010). Manganese Metalloproteins. In Metals in Biology, G. Hanson, and L. Berliner, eds. (New York, NY: Springer New York), pp. 273–341.

Van Staalduinen, L.M., Park, C.-S., Yeom, S.-J., Adams-Cioaba, M.A., Oh, D.-K., and Jia, Z. (2010). Structure-Based Annotation of a Novel Sugar Isomerase from the Pathogenic *E. coli* O157:H7. J. Mol. Biol. *401*, 866–881.

Steiner, R.A., Kalk, K.H., and Dijkstra, B.W. (2002). Anaerobic enzyme - substrate structures provide insight into the reaction mechanism of the copper-dependent quercetin 2,3-dioxygenase. Proc. Natl. Acad. Sci. *99*, 16625–16630.

Stipanuk, M.H., Simmons, C.R., Andrew Karplus, P., and Dominy, J.E. (2010). Thiol dioxygenases: unique families of cupin proteins. Amino Acids *41*, 91–102.

Straganz, G.D., Glieder, A., Brecker, L., Ribbons, D.W., and Steiner, W. (2003). Acetylacetone-cleaving enzyme Dke1: a novel C–C-bond-cleaving enzyme from *Acinetobacter johnsonii*. Biochem. J. *369*, 573.

Stranzl, G.R., Wagner, U.G., Straganz, G.D., Steiner, W., and Kratky, C. (to be published). Crystal Structure of an Acetylacetone Dioxygenase from *Acinetobacter johnsonii*.

Suárez,, M., Ferrer, E., and Martín, M. (1996). Purification and biochemical characterization of gentisate 1,2-dioxygenase from *Klebsiella pneumoniae* M5a1. FEMS Microbiol. Lett. *143*, 89–95.

Takenaka, S., Murakami, S., Shinke, R., Hatakeyama, K., Yukawa, H., and Aoki, K. (1997). Novel Genes Encoding 2-Aminophenol 1,6-Dioxygenase from *Pseudomonas* Species AP-3 Growing on 2-Aminophenol and Catalytic Properties of the Purified Enzyme. J. Biol. Chem. *272*, 14727–14732.

Takenaka, S., Asami, T., Orii, C., Murakami, S., and Aoki, K. (2002). A novel meta-cleavage dioxygenase that cleaves a carboxyl-group-substituted 2-aminophenol. Purification and characterization of 4-amino-3-hydroxybenzoate 2,3-dioxygenase from *Bordetella sp.* strain 10d. Eur. J. Biochem. FEBS *269*, 5871–5877.

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. J. Biol. Chem. *276*, 43627–43634.

Tanner, A., and Bornemann, S. (2000). *Bacillus subtilis* YvrK Is an Acid-Induced Oxalate Decarboxylase. J. Bacteriol. *182*, 5271–5273.

The UniProt Consortium (2012). Update on activities at the Universal Protein Resource (UniProt) in 2013. Nucleic Acids Res. *41*, D43–D47.

Timpte, C. (2001). Auxin binding protein: curiouser and curiouser. Trends Plant Sci. *6*, 586–590.

Todd, J.D., Curson, A.R.J., Kirkwood, M., Sullivan, M.J., Green, R.T., and Johnston, A.W.B. (2011). DddQ, a novel, cupin-containing, dimethylsulfoniopropionate lyase in marine roseobacters and in uncultured marine bacteria: Cupin-containing DMSP lyase. Environ. Microbiol. *13*, 427–438.

Tottey, S., Waldron, K.J., Firbank, S.J., Reale, B., Bessant, C., Sato, K., Cheek, T.R., Gray, J., Banfield, M.J., Dennison, C., et al. (2008). Protein-folding location can regulate manganese-binding versus copper- or zinc-binding. Nature *455*, 1138–1142.

Tranchimand, S., Ertel, G., Gaydou, V., Gaudin, C., Tron, T., and Iacazio, G. (2008). Biochemical and molecular characterization of a quercetinase from *Penicillium olsonii*. Biochimie *90*, 781–789.

Trummler, K., Roos, J., Schwaneberg, U., Effenberger, F., Förster, S., Pfizenmaier, K., and Wajant, H. (1998). Expression of the Zn^{2+}-containing hydroxynitrile lyase from flax (*Linum usitatissimum*) in *Pichia pastoris*— utilization of the recombinant enzyme for enzymatic analysis and site-directed mutagenesis. Plant Sci. *139*, 19–27.

Tuynman, A., Spelberg, J.L., Kooter, I.M., Shoemaker, H.E., and Wever, R. (2000). Enantioselective Epoxidation and Carbon-Carbon Bond Cleavage Catalyzed by *Coprinus cinereus* Peroxidase and Myeloperoxidase. J. Biol. Chem. *275*, 3025–3030.

Uberto, R., and Moomaw, E.W. (2013). Protein Similarity Networks Reveal Relationships among Sequence, Structure, and Function within the Cupin Superfamily. PLoS ONE *8*, e74477.

Vaillancourt, F.H., Bolin, J.T., and Eltis, L.D. (2006). The Ins and Outs of Ring-Cleaving Dioxygenases. Crit. Rev. Biochem. Mol. Biol. *41*, 241–267.

Verhees, C.H., Huynen, A., Ward, D.E., Schiltz,. E., de Vos, W.M:, and van der Oost, J. (2001). The Phosphoglucose Isomerase from the Hyperthermophilic Archaeon *Pyrococcus furiosus* Is a Unique Glycolytic Enzyme That Belongs to the Cupin Superfamily. J. Biol. Chem. *276*, 40926–40932.

Waldron, K.J., and Robinson, N.J. (2009). How do bacterial cells ensure that metalloproteins get the correct metal? Nat. Rev. Microbiol. *7*, 25–35.

Whittaker, M.M., and Whittaker, J.W. (2002). Characterization of recombinant barley oxalate oxidase expressed by *Pichia pastoris*. JBIC J. Biol. Inorg. Chem. *7*, 136–145.

Winkler, M., Glieder, A., and Steiner, K. (2012). Hydroxynitrile Lyases: From Nature to Application. In Comprehensive Chirality, (Elsevier), pp. 350–371.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr. D Biol. Crystallogr. *67*, 235–242.

Winter, G (2009). xia2: an expert system for macromolecular crystallography data reduction. J. Appl. Crystallogr. *43*, 186–190.

Woo, E.-J., Marshall, J., Mauly, J., Chen, J-G., Venis, M., Napier, R.M., and Pickersgill, R.W. (2002). Crystal structure of auxin-binding protein 1 in complex with auxin. EMBO J. *21*, 2877–2885.

Woo, E.J., Dunwell, J.M., Goodenough, P.W., Marvier, A.C., and Pickersgill, R.W. (2000). Germin is a manganese containing homohexamer with oxalate oxidase and superoxide dismutase activities. Nat. Struct. Biol. *7*, 1036–1040.

Zhang, Y., Colabroy, K.L., Begley, T.P., and Ealick, S.E. (2005). Structural Studies on 3-Hydroxyanthranilate-3,4-dioxygenase: The Catalytic Mechanism of a Complex Oxidation Involved in NAD Biosynthesis. Biochemistry (Mosc.) *44*, 7632–7643.

6. APPENDIX

This appendix contains results which are relevant to the thesis but were not incorporated into the published part or manuscript drafts. One part concerns the initial experiments towards a protein NMR study of *Gt*HNL, whereby an efficient protocol for ^{15}N labelling and purification has been established and encouraging preliminary NMR data have been obtained. A second part presents the results of an investigation of possible promiscuous activities of *Gt*HNL. Also listed are two co-authored publications in peer-reviewed journals to which I have made only minor contributions and which therefore do not constitute a main part of the thesis. The first publication is a report of the crystallisation of *Gt*HNL and the second is a review article of current patents on hydroxynitrile lyases.

6.1 [15]N Labelling and preliminary protein NMR measurements of *Gt*HNL

– Towards a solution structure of substrate-bound protein –

Ivan Hajnal[1], Gabriel Wagner[2], Klaus Zangger[2], Helmut Schwab[1,3], and Kerstin Steiner[1]

[1]ACIB, Austrian Centre of Industrial Biotechnology GmbH, Petersgasse 14, 8010 Graz, Austria

[2]Institute of Chemistry, University of Graz, Heinrichstraße 28, 8010 Graz, Austria

[3]Institute of Molecular Biotechnology, TU Graz, Petersgasse 14, 8010 Graz, Austria

Author Contribution:

I contributed the preparative work encompassing expression, ^{15}N-labelling and purification of *Gt*HNL. Metal analysis was performed by Dr Helmar Wiltsche at the institute of analytical chemistry TU Graz. NMR measurements were done by Gabriel Wagner in the lab of Prof. Klaus Zangger at the institute of chemistry KFU Graz. The senior authors contributed scientific direction, fruitful discussions and critical reading of the manuscript. The manuscript was written by me.

6.1.1 INTRODUCTION

*Gt*HNL is a novel cupin type hydroxynitrile lyase (HNL) and a detailed biochemical characterisation of this enzyme constitutes a published part of this thesis (Hajnal et al. 2013). A crystal structure of the enzyme has been solved and forms a part of the same publication. The enzyme shows no structural similarity to known HNLs (Dadashipour and Asano, 2011; Winkler et al., 2012) and elucidating the exact reaction mechanism of the enzyme would be highly interesting due to its novelty. However, attempts to either co-crystallise the enzyme, or to soak crystals with the substrate mandelonitrile were not successful. As a consequence, the exact positioning of the substrate in the active site of the enzyme is not known. Attempts to dock the substrate into the active did not yield unambiguous results, possibly due to the large size of the cupin active site cavity. On the other hand, the exact orientation and binding mode of the substrate in the active site is a prerequisite for any proposal of the exact mechanism. Even though the monomer of the enzyme isrelatively small, with 14.25 kDa per monomer the tetramer is has a size of ~60 kDa. Due to the compact folding and small size of the monomer, it may be possible to use NMR to solve its structure. Initial work towards an NMR solution structure of the enzyme is presented here.

6.1.2 RESULTS AND DISCUSSION

6.1.2.1 Expression of *Gt*HNL in M9 minimal medium

In order to resolve a protein structure via NMR it is necessary to use isotopic labelling. Isotopic labelling is only possible in minimal medium, and the first experiment was aimed at determining the efficiency of overexpression of this enzyme in minimal medium as a pre-trial for the ^{15}N labelling of *Gt*HNL. The expression in minimal medium was efficient, and the expression level in the batches with or without addition of manganese showed no apparent difference (Figure 1).
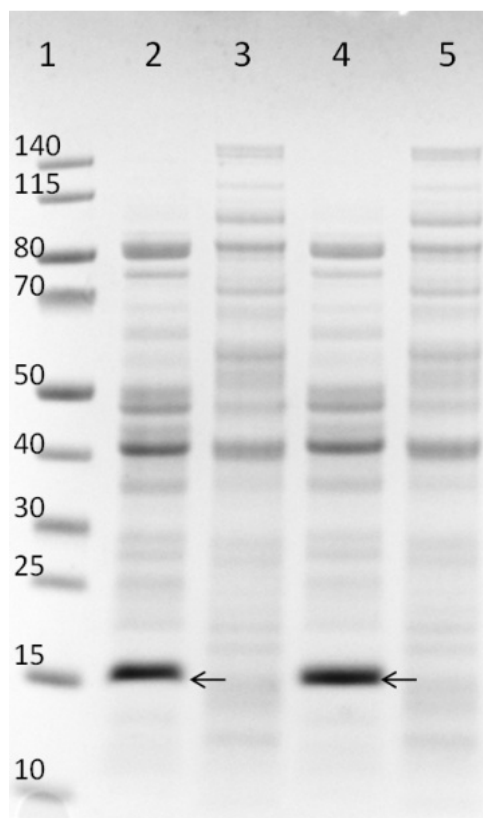
Figure 1: Expression of *Gt*HNL in M9 minimal medium **1:** Protein size standard (kDa, PageRuler, Fermentas). **2:** *Gt*HNL expressed with the addition of 100 µM MnCl₂ to the medium – supernatant. **3:** pellet fraction of **2**. 4: *Gt*HNL in M9 medium without manganese addition. **5:** pellet fraction of **4**. The bands corresponding to *Gt*HNL are indicated with arrows.

6.1.2.2 Purification of *Gt*HNL expressed in M9 minimal medium

After successful overexpression in minimal medium, purification was conducted using a protocol analogous to the protocol already published (Hajnal et al. 2013), consisting of an anion-exchange and a size-exclusion chromatography step. The anion-exchange step was conducted following the described protocol without change. However, the buffer used for the size-exclusion step in the first protocol, which is also the final storage buffer of the purified protein, was not suitable for NMR due to the presence of Tris base. Consequently, the buffer was replaced with a low-salt buffer suitable for NMR. A buffer containing only 10 mM Na-Pi and 50 mM NaCl at pH 6.0 was chosen as having a suitable pH, salt composition and concentration. The protein was stable in this buffer and eluted from the size exclusion column as a single peak without apparent aggregation. The crude lysate, ion-exchange and final size exclusion fractions of the purification procedure are shown in Figure 2.
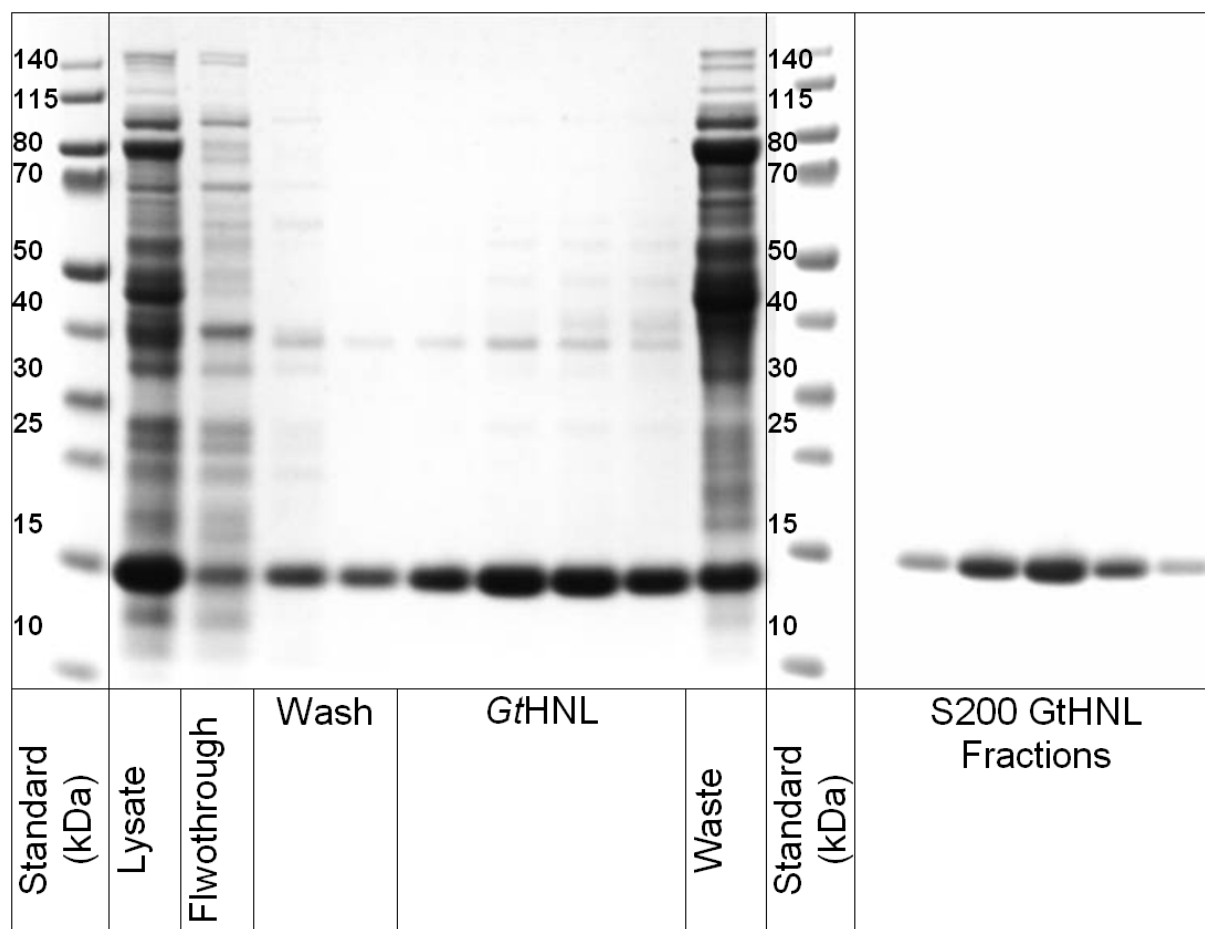
Figure 2: Purification of *Gt*HNL expressed in M9 minimal medium. Standard: Page Ruler (Fermentas); Lysate: Cleared lysate of *E. coli* cells (soluble fraction); Wash: wash fractions of anion exchange purification before salt gradient; *Gt*HNL: *Gt*HNL elution step; Waste: 1 M NaCl fraction; S200 Fractions: Size-exclusion fractions containing *Gt*HNL.

6.1.2.3 ^{15}N labelling of *Gt*HNL

For ^{15}N labelling of *Gt*HNL, the protein was expressed in M9 minimal medium made in the same way as for the initial experiments. The only difference was that the NH_4Cl stock was exchanged for the isotopically labelled $^{15}NH_4Cl$. The labelled protein was purified successfully and a preliminary NMR spectrum recorded by the group of Prof. Klaus Zangger indicated a stably folded protein (data not show).

6.1.2.4 NMR measurements of ^{15}N-labelled *Gt*HNL with mandelonitrile and benzoic acid

Since the ^{15}N-labelling and purification of *Gt*HNL was successful and the NMR spectrum was promising and indicated a stably folded protein, the second step involved a comparison of the NMR spectrum of enzyme alone and that of enzyme incubated with ligands, in this case a substrate and an inhibitor. The comparison of spectra can show shifts which might be attributed to an interaction between the protein and the ligand. In case significant shifts are visible, they may potentially be used to annotate the exact binding mode of the substrate in the active site. The protein was purified in the same way as described above. A final concentration of 100 µM $MnCl_2$ was added to the expression medium and the protein was active in the cyanolysis of mandelonitrile, indicating that it has incorporated manganese. The purified ^{15}N-labelled protein was provided to the group of Prof. Klaus Zangger for ^{15}N-Hsqc NMR. The NMR measurements were done with 8 mg mL^{-1} protein in a final volume of 556 µL in 100 mM NaPi pH 5.3. Either 17 mM of the substrate *(R)*-mandelonitrile (Km of the enzyme ~ 10 mM), or alternatively 2 mM of the substrate analog benzoic acid were used as ligands. Both ligands induced strong shifts in the ^{15}N-Hsqc NMR spectrum (Figure 3a and b). There was no change in buffer pH with mandelonitrile, and the change in pH with benzoic acid was small and is unlikely to fully explain the NMR shifts. Thus the observed shifts are most likely due to ligand binding. The shifts in spectra upon the binding of substrate are particularly interesting since the measurements were done under conditions in which the enzyme is still active. These results further imply that *Gt*HNL may be a good candidate for the annotation of the mode of substrate binding using protein NMR.
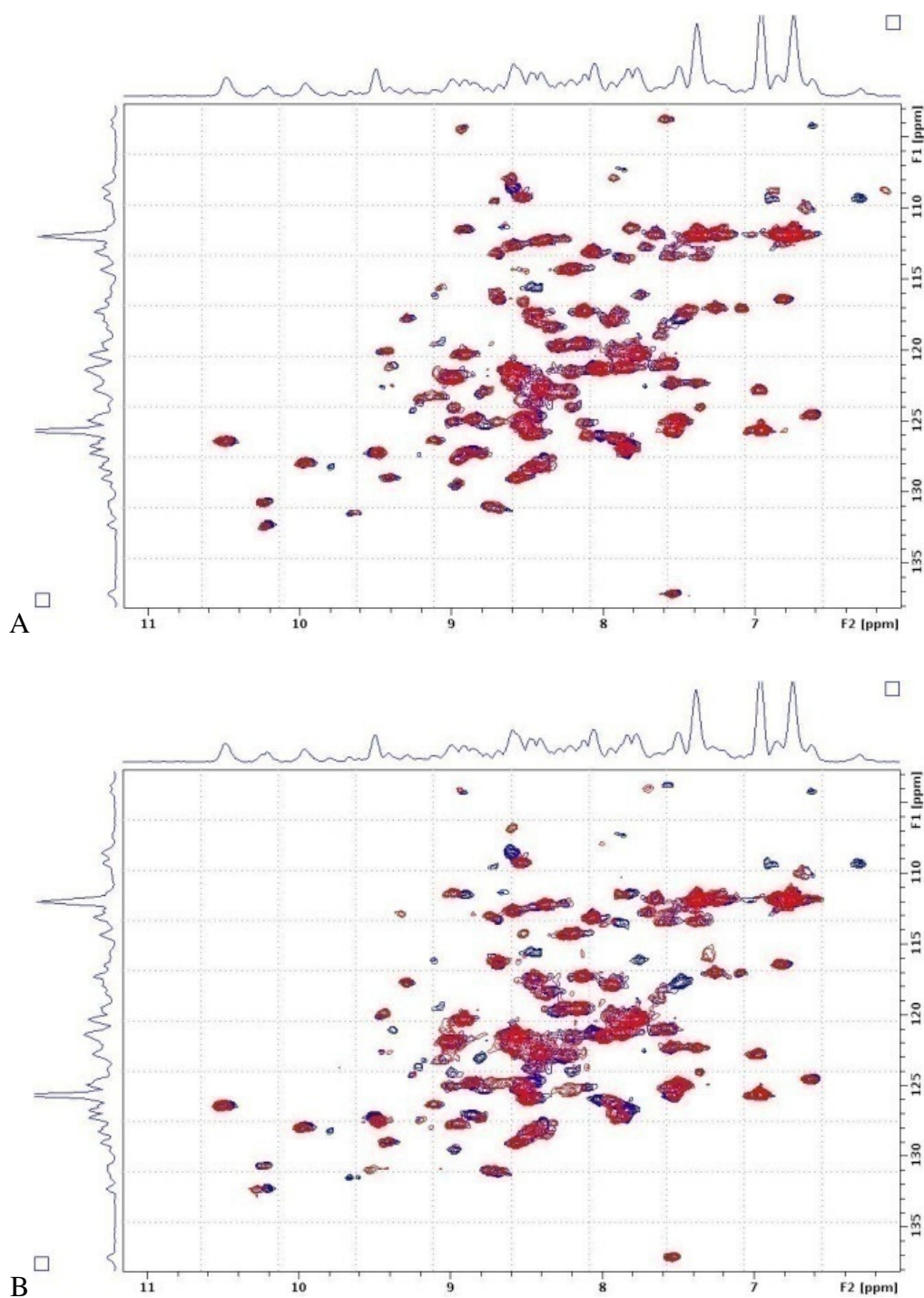
Figure 3: ^{15}N-Hsqc NMR spectra of either ^{15}N-labelled *Gt*HNL alone (blue, both images) or *Gt*HNL incubated with either 2 mM benzoic acid (A, red) or ~ 17 mM *(R)*-Mandelonitrile (B, red). The NMR spectra were recorded in 100 mM NaPi pH 5.3 with 8 mg·mL[-1] of purified ^{15}N-labelled protein.

6.1.2.5 NMR measurements of ^{15}N-labelled *Gt*HNL expressed with and without the addition of manganese

Since the substrate binding led to detectable shifts in the protein NMR spectra, and since *Gt*HNL requires manganese for activity, it was interesting to see whether the shifts with substrate are present in both the manganese-containing and the inactive manganese-depleted protein. A presence of binding-induced shifts in the metal-free protein would be an indicator that the metal is not necessary and an absence of binding-induced shifts in the metal-free protein would indicate that the metal is necessary for substrate binding. To test this idea, NMR spectra were recorded of *Gt*HNL produced with and without the addition of manganese, each with the ligands described above. Consequently, *Gt*HNL was produced in ^{15}N-labelled M9 minimal medium with and without the addition of 100 µM $MnCl_2$ and purification was performed the same way for both. The metal content was confirmed with ICP-OES and the protein expressed without manganese contained no detectable metal, whereas the protein expressed with the addition of 100 µM $MnCl_2$ showed 35% manganese loading. Preliminary data showed a difference between the spectra of the manganese loaded and metal-free protein. However, for the determination of the exact binding mode further experiments beyond the scope of this thesis will be necessary.

6.1.3 CONCLUSION

In this work it could be demonstrated that *Gt*HNL can be overexpressed efficiently in minimal medium and labelled with ^{15}N. Standard protocols for the isotopic labelling and efficient purification have been successfully adapted for this protein and can be used to further study of this enzyme. The preliminary data indicate that the protein is folded and amenable to NMR analysis. Encouragingly, strong shifts in the protein NMR spectra have been observed which are likely attributable to ligand binding, both with *(R)*–mandelonitrile and benzoic acid. Further studies will be needed to elucidate the exact binding mode of the substrate in the active site, which is a prerequisite for a mechanism proposal for this novel and unusual manganese-dependent hydroxynitrile lyase.

6.1.4 EXPERIMENTAL SECTION

6.1.4.1 Expression and ^{15}N-labelling of *Gt*HNL

For expressions in minimal medium, the same strain and construct was used as described in the chapter on the biochemical characterisation of *Gt*HNL (Hajnal et al. 2013). For expressions both with and without isotopic labelling, a standard M9 minimum medium formulation was used with 6.78 g·L^{-1} Na$_2$HPO$_4$, 3 g·L^{-1} KH$_2$PO$_4$, 0.5 g·L^{-1} NaCl, and 1 g·L^{-1} of either unlabelled or ^{15}N labelled (98 atom % ^{15}N, Sigma-Aldrich, St. Louis, MO, USA) NH$_4$Cl final concentration. NH$_4$Cl was added from a 100 g·L^{-1} stock solution and all the other salts were pre-autoclaved as a 5-fold stock. The medium was reconstituted according to the standard protocol, biotin and thiamine to a final concentration of 10 mg·L^{-1} each and FeSO$_4$ to a final concentration of 10 µM. Pre-cultures grown over night in minimal medium were diluted to an OD$_{600}$ of 0.1 in freshly reconstituted minimal medium and grown to an OD$_{600}$ of 0.6 at 37°C, under shaking at 120 rpm in baffled flasks, after which the cultures were cooled to 25°C, and expression was started by the addition of IPTG (Carl Roth GmbH, Karlsruhe, Germany) to a final concentration of 0.1 mM. The expression proceeded for 20-22 h at 25°C. Where indicated, MnCl$_2$ was added to a final concentration of 100 µM concomitantly.

6.1.4.2 Purification

Cells expressing *Gt*HNL were harvested and resuspended in Buffer A (50 mM Bis-Tris/HCl, pH 6.8, with 50 mM NaCl), lysed by sonication (Branson Sonifier S-250, set to 80% duty cycle, and output control 7) two times for 3 min under ice cooling and the lysates were cleared by centrifugation (1 h, 48000 g, 4°C). The cleared lysate was loaded to an anion-exchange column (HiTrap^{TM} Q FF 5 mL, GE Healthcare, Uppsala, Sweden), the purest fractions containing *Gt*HNL were pooled, concentrated using a centrifugal concentrator (Vivaspin 20, 10 kDa molecular-weight cut-off; Sartorius, Göttingen, Germany), and loaded either onto a Superdex 200 16/60 or a Superdex 75 HiLoad 16/600 size-exclusion column (both GE Healthcare), pre-equilibrated with 10 mM NaPi, pH 6.0, 50 mM NaCl, which was the final storage buffer.

- 135 -

6.2   Test for promiscuous activities in metal-substituted *Gt*HNL variants

Ivan Hajnal[1], Helmut Schwab[1,2], and Kerstin Steiner[1]

[1]ACIB, Austrian Centre of Industrial Biotechnology GmbH, Petersgasse 14, 8010 Graz, Austria

[2]Institute of Molecular Biotechnology, TU Graz, Petersgasse 14, 8010 Graz, Austria

Author Contribution:

I contributed the experimental data including protein purification, metal-exchange and enzymatic assays. Metal analysis was performed by Dr Helmar Wiltsche at the institute of analytical chemistry TU Graz. The manuscript was written by me. The senior authors contributed scientific direction and critical reading of the manuscript.

6.2.1 INTRODUCTION

As the introductory chapter of this thesis illustrates, cupin proteins harbour a large number of diverse enzymatic activities. What's more, very interesting promiscuous activities have been discovered in cupins, including reports that the activity can be modulated depending on the metal bound in the active site (Dai, 1999; Leitgeb and Nidetzky, 2010). Since a protocol for metal-exchange in *Gt*HNL has been established as part of this thesis, different metal-substituted variants have been investigated in search of potential promiscuous activities in *Gt*HNL. Even though many more promiscuous activities are possible for *Gt*HNL, these results are significant in their own right since they support the characterisation of *Gt*HNL as a hydroxynitrile lyase and as such are presented in this appendix.

6.2.2 Test for oxalate decarboxylase activity

As has been described in earlier chapters, *Gt*HNL was identified as a manganese-dependent cupin hydroxynitrile lyase. Interestingly, the enzyme was most active in the presence of oxalate in the buffer. On the other hand, manganese-dependent oxalate-decarboxylase activity is found in well-characterised cupins as their main function (Tanner et al., 2001). Consequently, a possible oxalate decarboxylase side-activity was investigated in *Gt*HNL. This was done using a coupled assay, which detects formic acid. As is illustrated in Figure 1, an oxalate decarboxylase activity would lead to the production of formate from oxalate, and formate in turn can be detected using formate dehydrogenase.
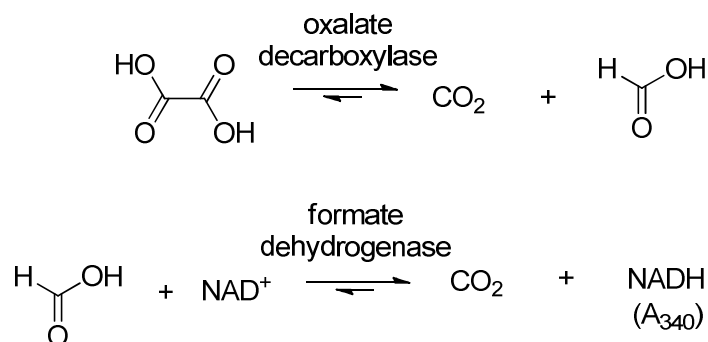

Figure 1: Principle of the coupled Oxalate-decarboxylase assay. An oxalate decarboxylase activity, if present, produces formate and CO_2 from oxalate. A formate dehydrogenase oxidises the formate to CO_2, converting NAD^+ to NADH, which is detected spectrophotometrically at 340 nm.

The assay is based on the increase in A_{340} following the reduction of NAD^+ at the expense of formate catalysed by formate dehydrogenase. No formate could be detected in the reaction mixture after 30' incubation of 0.2 $mg \cdot mL^{-1}$ *Gt*HNL in oxalate buffer under conditions optimal for the cyanolysis reaction (pH 6.0, Na-MES-Oxalate buffer, 100 mM MES and oxalic acid each). As a positive control, formic acid was added to a final concentration of 1 mM, and NAD^+ reduction was clearly visible (Figure 2). Thus, it can be largely excluded that *Gt*HNL has oxalate decarboxylase activity.
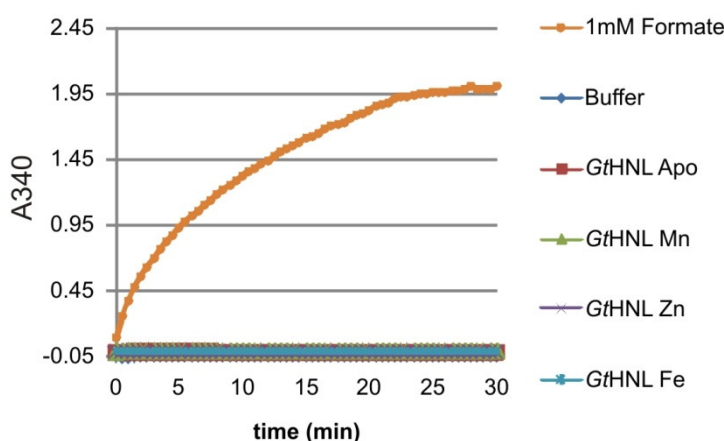

Figure 2: Test for oxalate decarboxylase activity, based on the detection of formate released from oxalate by oxalate decarboxylase, using formate dehydrogenase (see figure X). Shown are a buffer control, or 0.2 $mg \cdot mL^{-1}$ *Gt*HNL Apoprotein, Apoprotein reconstituted with manganese, zinc, or iron each, and a positive control (1 mM formate in assay buffer).

6.2.3   Test for esterase activity

No natural esterases are known so far within the cupin superfamily. However, a promiscuous esterase activity was reported in a cupin diketone dioxygenase - Dke1. The cupin is not related to *Gt*HNL, but the promiscuous esterase activity was found to be metal-dependent and strongest with zinc. (Leitgeb and Nidetzky 2012). As zinc-substituted *Gt*HNL was available, it was an interesting question whether zinc would confer esterase activity to *Gt*HNL. Consequently, aliquots of purified *Gt*HNL apoprotein reconstituted with manganese, zinc or iron were all tested for esterase activity on p-Nitrophenylacetate (p-NPA). The substrate p-NPA is a highly activated model ester and there is a detectable background reaction in buffer without any protein. While the reactions with the enzyme preparations appear to be slightly

above buffer, the difference between the apparent reaction in buffer the difference is not significant and may be an artefact due to general protein precipitation. Thus it can be said that none of the metal-reconstituted enzyme preparations had esterase activity clearly above the background at a final protein concentration of 0.2 mg·mL^{-1} (Figure 3).

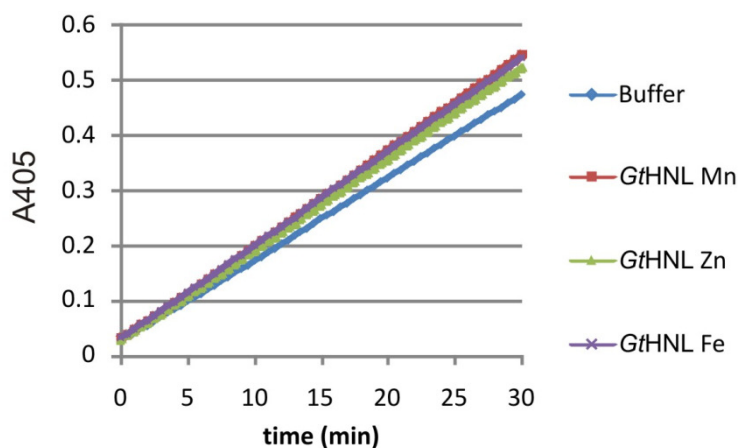**Figure 3:** Test for esterase activity with 0.2 mg·mL^{-1} of *Gt*HNL containing Mn, Zn or Iron. Conditions: 3 mM p-NPA in 20 mM Tris-Cl buffer, pH 7.5, 200 mM NaCl, 25°C.

In addition to p-nitrophenyl acetate, the well-known assay for esterase activity on naphthyl esters was chosen for comparison. The assay used was based on the detection of the liberated alpha or beta naphthol with Fastblue (4-benzoylamino-2,5-dimethoxybenzenediazonium). PLE (pig liver esterase 0.01 U/well) from commercial preparations was used as positive control. Each enzyme preparation (*Gt*HNL and metal loaded aliquots) were added to a final concentration of 1 mg·mL^{-1}. Three naphthyl ester substrates were tested: 1-naphthyl acetate, 2-naphthyl acetate and 2-naphthyl butyrate. In all cases there was no significant activity with different metal-substituted preparations of *Gt*HNL whereas the positive control PLE was clearly active (Figure 4).
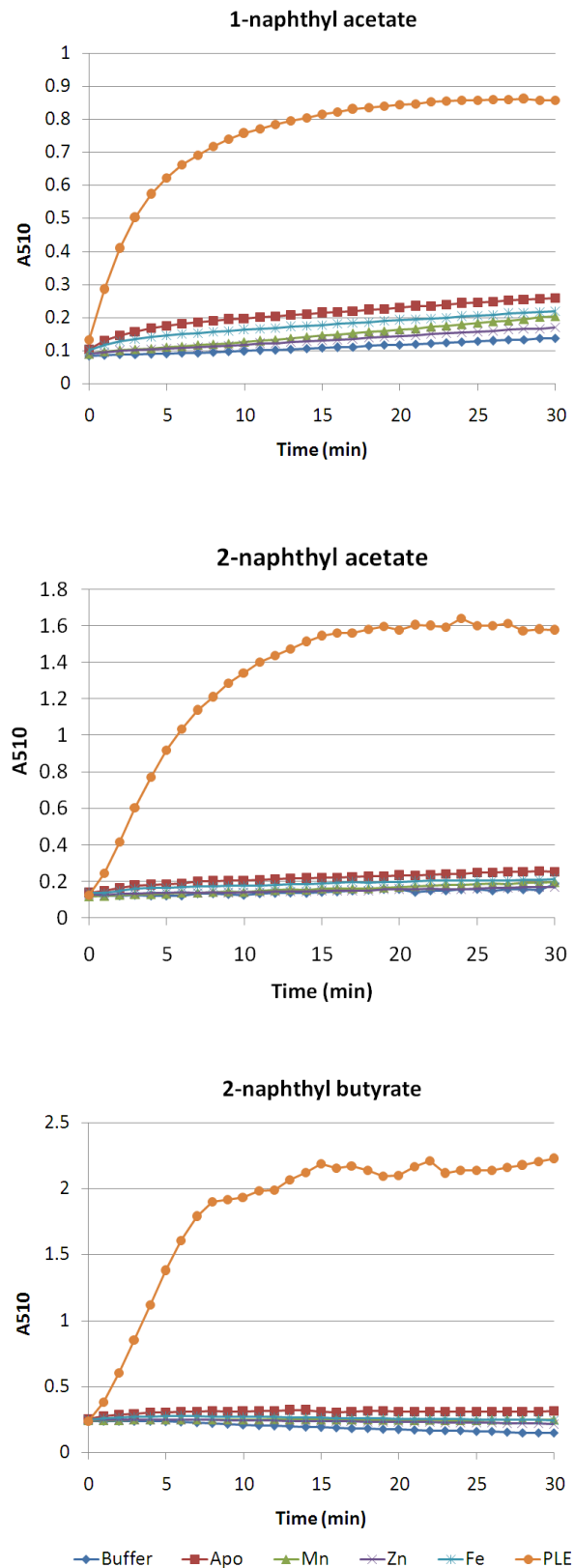
Figure 4: Test for esterase activity with 1 mg·mL^{-1} of *Gt*HNL apoprotein or protein containing Mn, Zn or Fe. PLE: pig liver esterase (0.01 U per well, positive control).

6.2.4 CONCLUSION

*Gt*HNL is a cupin hydroxynitrile lyase for which oxalate decarboxylase activity or an esterase activity with standard test substrates can be excluded. Consequently, the only known enzymatic activity for *Gt*HNL known so far is the hydroxynitrile lyase activity which has been characterised in detail in this thesis.

6.2.5 EXPERIMENTAL SECTION

6.2.5.1 Enzyme preparation and metal analysis

Apoprotein and metal-exchanged variants of *Gt*HNL were prepared following the same protocol as has been described in the published part of this thesis on the biochemical characterisation of *Gt*HNL. The metal analysis was also performed in the same way as in said publication.

6.2.5.2 Oxalate decarboxylase activity assay

Oxalic acid dihydrate, formic acid and formate dehydrogenase were purchased from Sigma (Sigma-Aldrich, St. Louis, MO, USA). Each enzyme preparation was diluted to 0.2 mg·mL^{-1} in 100 mM MES Oxalate buffer, pH 6.0, and incubated for 30 min. To 200 µL of each preparation, 10 µL 57 mM β-NAD^{+} and 10 µL formate dehydrogenase (40 U·mL^{-1} freshly diluted in cold H$_2$O) were added to start the reaction. Formic acid was added to a final concentration of 1 mM from a 100 mM stock in H$_2$O as the positive control. The photometric measurements were performed at 25°C on a Synergy MX SMATBLD microplate reader (Biotek, Winooski, WT, USA), running Gen 5 software version 1.11. The increase in NADH was followed at 340 nm for 30 min.

6.2.5.3 p-Nitrophenyl esterase activity assay

The measurements were done in microtiter plates at 25°C using the same plate reader setup as described above. Shortly, 100 µL enzyme solution (final concentration 0.2 mg·mL^{-1}) and 100 µL p-nitrophenyl acetate (final concentration 3 mM) both in 20 mM Tris-Cl buffer pH 7.5 were dispensed per well. The photometric measurements were performed on the same instrument as above. The increase in p-nitrophenol was followed at 405 nm for 30min.

6.2.5.4 Naphthyl esterase activity assay

The substrates 1-naphtyl acetate, 2-naphythyl acetate and 2-naphythyl butyrate, the positive control enzyme PLA (pig liver esterase, 19 U·mg^{-1} solid) and 4-benzylamino-2,5-dimethoxybenzenediazonium (Fastblue RR) were purchased from Sigma (Sigma-Aldrich). The photometric measurements were performed on the same instrument as above. In short, 50 µL Fastblue solution (0.077 g dissolved directly in 50 mL H$_2$O under vigorous vortexing), 50 µL of 2 mM substrate stock in 100 mM Tris-Cl, pH 7.5, 4% DMSO, and 100 µL enzyme solution (1 mg·mL^{-1} final concentration) in Tris-Cl, pH 7.5 were added, starting the reaction. The release of naphthol, which reacts with Fastblue, was followed at 510 nm for 1 h. As positive control 0.01 U of PLE were added per well in the same buffer.

6.3   Crystallization of a novel metal-containing cupin from *Acidobacterium* sp. and preliminary diffraction data analysis

Andrzej Łyskowski[1,*,‡] Kerstin Steiner[1,‡]

Ivan Hajnal[1], Georg Steinkellner[1], Helmut Schwab[1,2], and Karl Gruber[1,3]

[1]ACIB, Austrian Centre of Industrial Biotechnology GmbH, Petersgasse 14, 8010 Graz, Austria

[2]Institute of Molecular Biotechnology, TU Graz, Petersgasse 14, 8010 Graz, Austria

[3]Institute of Molecular Biosciences, University of Graz, Humboldtstraße 50, 8010 Graz, Austria

‡ These authors contributed equally to this work

Author Contribution:

I have made a minor contribution to this publication by recloning *Gt*HNL into the 6 x histidine tagged TEV-cleavable expression vector. The protein used for the crystals was produced by Dr Kerstin Steiner, and the crystallisation experiments were done by Dr Andrzej Łyskowski. These two authors also wrote the manuscript.

Crystallization of a novel metal-containing cupin from *Acidobacterium* sp. and preliminary diffraction data analysis

Andrzej Łyskowski,[a]*‡ Kerstin Steiner,[a]‡ Ivan Hajnal,[a] Georg Steinkellner,[a] Helmut Schwab[a,b] and Karl Gruber[a,c]

[a]ACIB – Austrian Centre of Industrial Biotechnology GmbH, Petersgasse 14, 8010 Graz, Austria, [b]Institute of Molecular Biotechnology, TU Graz, Petersgasse 14, 8010 Graz, Austria, and [c]Institute of Molecular Biosciences, University of Graz, Humboldtstrasse 50, 8010 Graz, Austria

‡ These authors contributed equally to this work.

Correspondence e-mail: andrzej.lyskowski@acib.at

Recombinant AciX9_0562 from *Acidobacterium* sp. MP5ACTX9 (UniProt ID E8WYN5) containing sequence motifs characteristic of the RmlC-type cupins superfamily and containing Pfam motif PF07883 has been successfully cloned, expressed and purified. AciX9_0562 crystallized in a number of conditions from the Morpheus protein crystallization screen. The best crystal diffracted to 2.7 Å resolution (space group $C222_1$; unit-cell parameters $a = 125.29$, $b = 254.63$, $c = 82.99$ Å). Structure solution was facilitated by the automated molecular-replacement pipeline *BALBES*. The initial solution was automatically rebuilt using the *PHENIX AutoBuild* wizard, with final R and R_{free} values of 0.23 and 0.26, respectively. The structure is currently undergoing manual refinement.

1. Introduction

The AciX9_0562 gene from *Acidobacterium* sp. MP5ACTX9 (Uni-Prot ID E8WYN5) encodes a 131-amino-acid protein of uncharacterized function with a molecular weight of 14.25 kDa and a calculated isoelectric point of 5.74. A *BLAST* search showed that the protein is highly conserved among many bacteria; however, all sequences with high identity are annotated as 'hypothetical proteins'. An InterPro scan (Hunter *et al.*, 2009) for sequence motifs revealed that AciX9_0562 belongs to the RmlC-type cupins superfamily (Wilson *et al.*, 2009) and also contains Pfam motif PF07883 (Finn *et al.*, 2010), a conserved cupin 2 motif (Dunwell, 1998). Currently, the cupin superfamily comprises 50 families consisting of members with diverse functions ranging from enzymatic activities such as dioxy-genases, decarboxylases, hydrolases, isomerases and epimerases to non-enzymatic functions such as auxin-binding proteins, nuclear transcription factors and seed storage proteins (Dunwell *et al.*, 2000; Agarwal *et al.*, 2009). However, no functional annotation has yet been made for AciX9_0562.

'Cupa' is the Latin term for a small barrel. All proteins that belong to the cupins adopt a barrel-like structure (Dunwell *et al.*, 2001, 2004). They can occur as monocupins or bicupins and form different oligomerization states (Dunwell *et al.*, 2004). Each cupin domain typically consists of two conserved primary-sequence motifs. Each motif corresponds to two ‚Β-strands. The motifs are separated by a less conserved region which consists of a minimum of 11 amino acids (Dunwell *et al.*, 2001).

Here, we report the successful crystallization and structure solution of AciX9_0562 from *Acidobacterium* sp. MP5ACTX9.

2. Material and methods

Plasmid construction

The synthetic gene (gene ID 322434201; see Supplementary Material[1]) encoding a putative cupin 2 conserved barrel domain protein (YP_004216413) from *Acidobacterium* sp. MP5ACTX9 was ordered codon-optimized for expression in *Escherichia coli* from GeneArt (Life Technologies). The gene was recloned into pEHisTEV

---

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: NJ5112).

vector *via Nco*I and *Hin*dIII restriction sites to introduce an N-terminal TEV-cleavable His tag (Liu & Naismith, 2009). The pEHisTEV construct was confirmed by sequencing (LGC Genomics, Berlin, Germany) and transformed into the expression host *E. coli* BL21 (DE3) Gold.

Expression and purification

E. coli BL21 (DE3) Gold cells harbouring pEHisTEV-cupin were grown in LB (lysogeny broth) medium supplemented with kanamycin (50 mg ml^{-1}) at 310 K. The expression of recombinant protein was initiated by the addition of 0.1 mM IPTG to the cultures on reaching an OD$_{600}$ of ''-0.8 and cultivation was continued at 293 K for 20 h. Protein expression and localization in cell-extract fractions was monitored by SDS–PAGE. The cells were harvested, resuspended in cold buffer *A* (20 mM sodium phosphate buffer pH 7.4 containing 0.5 M NaCl and 10 mM imidazole) and disrupted by sonication (Branson Sonifier S-250; 6 min, 80% duty cycle, output 7). The cell lysate was centrifuged for 1 h at 50 000g to remove unbroken cells and insoluble material. The cell-free lysate was filtered through a 0.45 mm syringe filter and incubated with Ni Sepharose 6 Fast Flow (GE Healthcare, Uppsala, Sweden) for 20 min. The Ni Sepharose was then transferred into an empty PD-10 column. Impurities were removed by washing with 30 mM imidazole followed by elution of the bound protein with 300 mM imidazole in buffer *A*. Fractions were analysed by SDS–PAGE, pooled and concentrated using Vivaspin 20 Centrifugal Filter Units (10 kDa molecular-weight cutoff; Sartorius) and desalted on PD-10 desalting columns (GE Healthcare) into 20 mM Tris–HCl, 200 mM NaCl pH 7. The N-terminal His tag was cleaved using 100 mg TEV protease (expressed and purified as His-tagged protein in-house) per milligram of protein in 20 mM Tris–HCl, 200 mM NaCl pH 7, 1 mM DTT, 0.5 mM EDTA at 293 K for 20 h. His tag and TEV protease were removed by reloading the protein solution onto Ni Sepharose and collecting the cleaved protein in the flowthrough. The completeness of the cleavage was analysed by SDS–PAGE. The cleaved samples were concentrated and loaded onto either a Superdex 75 10/30 column for qualitative analysis of the protein size and oligomerization state or a Superdex 200 16/60 column for preparative purification. The purified protein was again concentrated using Vivaspin 20 Centrifugal Filter Units (10 kDa molecular-weight cutoff; Sartorius). For both columns 20 mM Tris–HCl, 200 mM NaCl pH 7 was used as the running buffer. Prior to crystallization, purified and concentrated samples were stored in a cold room at 279 K and were transported on ice when necessary.

Crystallization

Purified protein samples with and without His tag were subjected to screening for crystallization conditions using the Morpheus crystallization screen (Molecular Dimensions; Gorrec, 2009) and the sitting-drop method in SWISSCI 3-Well crystallization plates. Screening drops of total volume 600 nl (1:1 screening solution to protein solution at about 10 mg ml^{-1}) were dispensed using an Oryx8 Protein Crystallization Robot (Douglas Instruments Ltd) at room temperature (293 K) and the plate was stored at 290 K between inspections. Protein variants with and without His tag crystallized rapidly in several conditions. Initial crystals were observed directly after screen setup. Two conditions that produced crystals of both
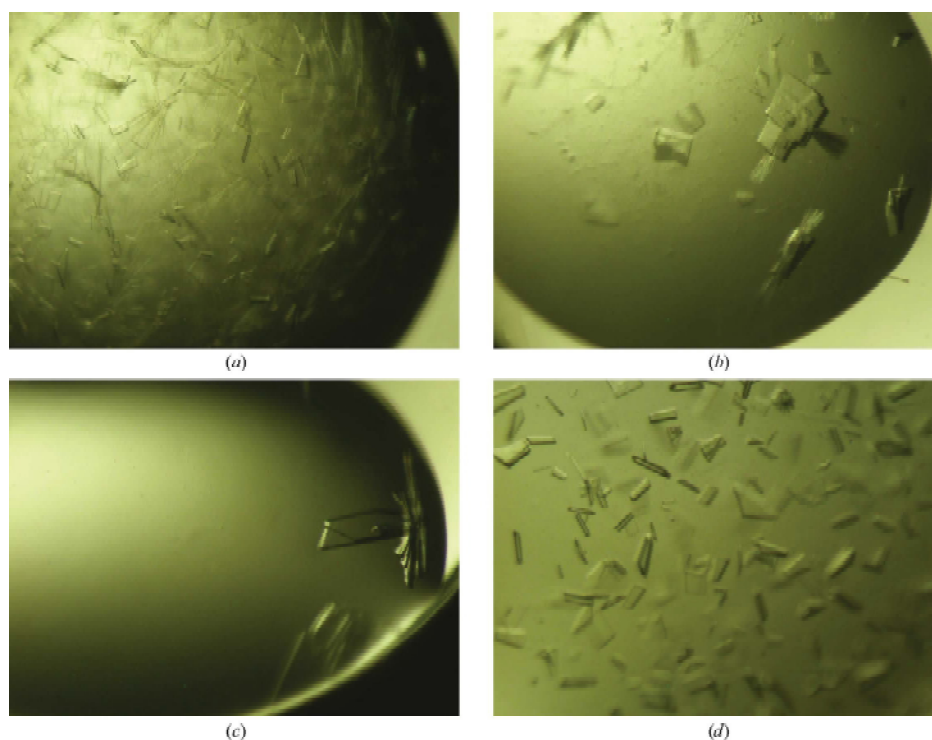

Figure 1
Crystals of AciX9_0562 from *Acidobacterium* sp. MP5ACTX9 before (*a*–*c*) and after (*d*) optimization (optimized Morpheus screen condition H02, crystal radius of between 0.05 and 0.10 mm).

Table 1
Data-processing and initial model statistics (Winn *et al.*, 2011; Evans, 2006).

Values in parentheses are for the highest resolution shell.

Data processing	
Wavelength (Å)	0.87260
Beam dimensions (mm)	8 x 5
Space group	$C222_1$
Unit-cell parameters (Å, °)	$a = 125.29$, $b = 254.63$, $c = 82.99$,
	$\alpha = \beta = \gamma = 90.0$
Resolution limits (Å)	56.21–2.68 (2.83–2.68)
R_{merge}	0.196 (0.567)
R_{meas} (all I^+ and I^-)	0.221 (0.644)
$R_{p.i.m.}$ (all I^+ and I^-)	0.082 (0.240)
Total No. of observations	270727 (37874)
No. of unique observations	37627 (5399)
hI/a(I)i	8.4 (4.1)
Completeness (%)	100.0 (100.0)
Multiplicity	7.2 (7.0)
Wilson B factor (Å^2)	40.947
Model refinement	
Matthews coefficient† (Å^3 Da^{-1})	4.96
Solvent content† (%)	75.21
Resolution range (Å)	56.2–2.7
Minimum [F_{obs}/a(F_{obs})]	1.34
Completeness for range (%)	99.97
No. of reflections	37602
Fit to data used in refinement	
R (working + test set)	0.2283
R (working set)	0.2264
R_{free}	0.2632
Free R test-set size (%)	5.03
Free R test-set count	2000

† For eight monomers in the asymmetric unit.

sample variants and that contained the visually most promising crystals were selected for further optimization: H02 [10%(*w/v*) PEG 8000, 20%(*v/v*) ethylene glycol, 0.02 *M* each of the amino acids dl-Ala, dl-Lys, dl-Ser, l-Glu and glycine, 0.1 *M* MES–imidazole buffer pH 6.5] and E05 [10%(*w/v*) PEG 20 000, 20%(*v/v*) PEG MME 550, 0.03 *M* each of the ethylene glycols diethylene glycol, penta-ethylene glycol, tetraethylene glycol and triethylene glycol, 0.1 *M* MOPS/HEPES Na buffer pH 7.5]. In both conditions initial crystals were observed as early as the day after setup and reached maximum size within a one-week period.

In order to improve the initial crystals, an optimization grid was prepared. In this grid, the original drop components were diluted with Milli-Q water: the protein sample along the long plate axis and the respective precipitant solution along the short plate axis. The resulting serial dilution grid had the following compositions in the plate corners: A01, 30% protein and 50% precipitant; A12, 50% protein and 50% precipitant (original hit condition); H01, 30% protein and 30% precipitant; H12, 50% protein and 30% precipitant. Exact drop compositions were calculated using the *XStep* program of the Oryx8 Protein Crystallization Robot. Two SWISSCI 2-Well crystallization plates, one for each of the optimized conditions, were dispensed with 1000 nl total drop volumes. Each well of the plate was occupied by the protein sample variants and was equilibrated against 1 *M* NaCl solution (Newman, 2005). The best diffracting crystals for the sample without His tag were obtained from modified condition H02 with the following composition: 33.5% protein solution and 35.5% precipitant solution (Fig. 1). The crystals were vitrified without cryoprotection and the best data-set parameters are reported.

The sample with the His tag did not produce crystals of sufficient quality for structure determination.

Structure determination

The best native data were collected on beamline ID23-2 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France at 100 K using a MAR 225 detector (Fig. 2). 180 images collected with a 1.0° oscillation range were processed using the program *MOSFLM*, imported into the *CCP*4 program suite using the program *POINT-LESS* (Winn *et al.*, 2011; Evans, 2006) and scaled using the program *SCALA* (Evans, 2006).

The data set was submitted to the *BALBES* server for automated molecular replacement at York Structural Biology Laboratory (Long *et al.*, 2008). An initial solution with a *BALBES*-modified template based on a cupin-like protein (tm1010) from *Thermotoga maritima* (PDB entry 2f4p; Joint Center for Structural Genomics, unpublished work) contained ten copies in the asymmetric unit. It was manually inspected and the number of copies was reduced to eight owing to missing electron density for two of the ten chains. The resulting structure was then rebuilt using the *PHENIX AutoBuild* wizard (v.1.7.2_869; Adams *et al.*, 2010).

Data-processing and current model statistics are presented in Table 1.

2. Results and discussion

AciX9_0562 was successfully expressed from a pET-based vector with an N-terminal TEV-cleavable His tag with high yield in the cytoplasm of *E. coli* BL21 (DE3) Gold strain (>50% of the total protein content as judged by SDS–PAGE). The His tag was cleaved from half of the sample. Both portions were purified by gel filtration and eluted as one symmetric peak from both Superdex 75 and Superdex 200 columns. Protein preparations with and without His tag were used to set up crystallization trials. Only protein without His tag produced crystals with sufficient quality for high-resolution structure determination. X-ray diffraction data from a single crystal were collected to 2.7 Å resolution on the ID23-2 beamline at the ESRF, opening the way for structural studies and analysis.
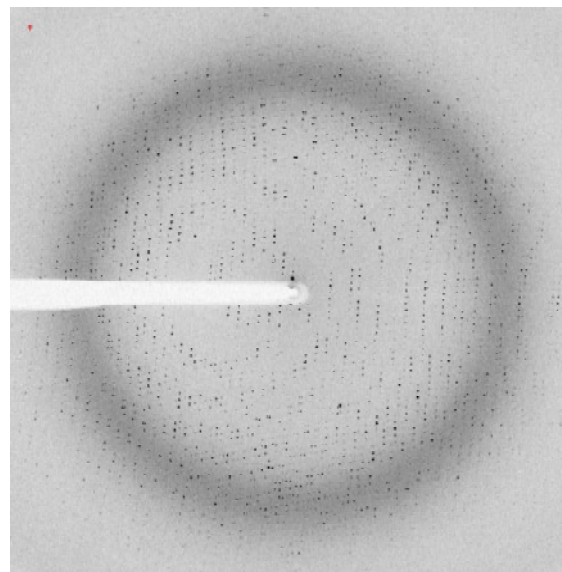

Figure 2
Diffraction image of a crystal of AciX9_0562 from *Acidobacterium* sp. MP5ACTX9 collected on the ID23-2 beamline at the ESRF, Grenoble [the image was generated with the program *ADXV* (http://www.scripps.edu/~arvai/adxv.html); the resolution at the detector edge is 2.68 Å].

Most cupins are metal-binding proteins with Mn, Fe, Ni, Cu or Zn bound in the active site. The characteristic cupin domain comprises two conserved motifs, both of which are involved in metal-ion binding. The conserved motifs comprise the following amino-acid sequences: G-X_5-H-X-H-$X_{3,4}$-E-X_6-G (motif 1) and G-X_5-P-X-G-X_2-H-X_3-N (motif 2) (Dunwell *et al.*, 2004). Two His residues and a Glu/Gln residue in motif 1 as well as the His residue in motif 2 act as ligands for binding of the active-site metal ion. At the current stage of refinement, a large unidentified electron density suitable for placement of the the hypothetical metal ion can be observed.

The biochemical and biophysical characterization as well as final and complete structure determination and refinement are currently in progress; the detailed results will be reported elsewhere.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D66, 213–221.
Agarwal, G., Rajavel, M., Gopal, B. & Srinivasan, N. (2009). *PLoS One*, 4, e5736.
Dunwell, J. M. (1998). *Biotechnol. Genet. Eng. Rev.* 15, 1–32.
Dunwell, J. M., Culham, A., Carter, C. E., Sosa-Aguirre, C. R. & Goodenough, P. W. (2001). *Trends Biochem. Sci.* 26, 740–746.
Dunwell, J. M., Khuri, S. & Gane, P. J. (2000). *Microbiol. Mol. Biol. Rev.* 64, 153–179.
Dunwell, J. M., Purvis, A. & Khuri, S. (2004). *Phytochemistry*, 65, 7–17.
Evans, P. (2006). *Acta Cryst.* D62, 72–82.
Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E. L., Eddy, S. R. & Bateman, A. (2010). *Nucleic Acids Res.* 38, D211–D222.
Gorrec, F. (2009). *J. Appl. Cryst.* 42, 1035–1042.
Hunter, S. *et al.* (2009). *Nucleic Acids Res.* 37, D211–D215.
Liu, H. & Naismith, J. H. (2009). *Protein Expr. Purif.* 63, 102–111.
Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. (2008). *Acta Cryst.* D64, 125–132.
Newman, J. (2005). *Acta Cryst.* D61, 490–493.
Wilson, D., Pethica, R., Zhou, Y., Talbot, C., Vogel, C., Madera, M., Chothia, C. & Gough, J. (2009). *Nucleic Acids Res.* 37, D380–D386.
Winn, M. D. (2011). *Acta Cryst.* D67, 235–242.

6.4   Mini-Review:

Recent Developments in Hydroxynitrile Lyases for Industrial Biotechnology

Elisa Lanfranchi,[a] Kerstin Steiner,[a] Anton Glieder,[a] Ivan Hajnal,[a] Roger A. Sheldon,[b] Sander van Pelt,[b] Margit Winkler[a,*]

[a]ACIB GmbH (Austrian Centre of Industrial Biotechnology GmbH), c/o Institute of Molecular Biotechnology, Graz University of Technology, Petersgasse 14, 8010 Graz, Austria

[b]CLEA Technologies B.V., Delftechpark 34, 2628XH Delft, The Netherlands

Author Contribution:

I have made a minor contribution to this publication consisting of the critical reading, translation and summarisation of patents written in Japanese language.

Keywords

asymmetric synthesis, biocatalysis, biotransformation, cascade reactions, C-C coupling, chiral building blocks, cyanogenesis, cyanohydrin, enantioselectivity, enzyme engineering, HNL, hydrocyanic acid, hydroxynitrile lyase, industrial process, mandelonitrile lyase, plant enzymes, oxynitrilase, stereoselective;

Glossary

HNL: hydroxynitrile lyase; HCN: hydrocyanic acid; CLEAs: cross-linked enzyme aggregates; FAD: flavin adenine dinucleotide; *ee*: enantiomeric excess; CIP: Cahn-Ingold-Prelog; WT: wild type; MTBE: methyl-*tert*-butylether

*Address correspondence to Margit Winkler, ACIB GmbH (Austrian Centre of Industrial Biotechnology GmbH), Petersgasse 14/III, A-8010 Graz, Austria; Tel: +43-316-8739333, fax: +43-316-873-9308; E-mail: margit.winkler@acib.at

6.4.1 Abstract

Hydroxynitrile lyases (HNLs) catalyze the cleavage as well as the formation of cyanohydrins. The latter reaction is valuable for the stereoselective C-C bond formation by condensation of HCN with carbonyl compounds. The resulting cyanohydrins serve as versatile building blocks for a broad range of chemical and enzymatic follow-up reactions.

A significant number of (*R*)- and (*S*)-selective HNLs are known today and the number is still increasing. HNLs not only exhibit varying substrate scope but also differ in sequence and structure. Tailor-made enzymes for large-scale manufacturing of cyanohydrins with improved yield and enantiomeric excess are very interesting targets, which is reflected in a solid number of patents. This review will complement and extend our recent review with a strong focus on applications of HNLs for the synthesis of highly functionalized, chiral compounds with newest literature, recent and current patent literature.

6.4.2 Introduction

Hydroxynitrile lyases (HNLs, EC 4.1.2.10, EC 4.1.2.11, EC 4.1.2.46 and EC 4.1.2.47; alternative names: oxynitrilase, hydroxynitrilase, mandelonitrile lyase, hydroxymandelonitrile lyase) catalyze the enantioselective cleavage and stereoselective synthesis of cyanohydrins. Cyanogenesis has been reported not only in many plant species like *Rosaceae, Linaceae, Euphorbiaceae, Clusiaceae, Olacaceae, Gramineae* and *Polypodiaceae* but also in other organisms like bacteria, fungi, lichen, arthropods or insects [1,2]. However, HNLs were also discovered in non-cyanogenic plants such as *Arabidopsis thaliana* [3].

In 1837, Wöhler and Liebig reported hydroxynitrile lyase activity for the first time as they investigated the formation of hydrocyanic acid (HCN, alternative name: prussic acid) – the cyanogenesis reaction [4]. The reverse reaction was first observed by Rosenthaler in 1908, who used emulsin – an emulsion derived from bitter almonds (*Prunus amygdalus*) – to prepare (*R*)-mandelonitrile. The first stereoselective enzyme catalyzed addition of HCN to benzaldehyde was thus discovered [5]. The enzyme responsible for this activity was an (*R*)-hydroxynitrile lyase. Since then, a number of other HNLs, both (*R*)- and (*S*)-selective were discovered. Interestingly, hydroxynitrile lyase activity is found from polypeptides with completely unrelated sequential and structural features and seems to be the result of convergent evolution [6]. HNLs from *Rosaceae* are FAD (flavin adenine dinucleotide) containing monomeric glycoproteins. Their natural substrate is mandelonitrile and this is a relatively conserved group of enzymes. In contrast, the group of FAD-independent HNLs is diverse and varies largely in physicochemical properties (primary sequence, molecular weight, oligomerization state, glycosylation, structure and natural substrate). HNLs differ in their stereospecificity and substrate range and have therefore found distinct applications for asymmetric syntheses of chiral building blocks not only in research laboratories but also at industrial level [7-9].

Chemical as well as enzyme catalyzed condensation of HCN to carbonyl compounds is reversible and the equilibrium is dependent on pH, solvent composition and the available concentrations of carbonyl compound, HCN and α-hydroxynitrile product (alternative name: cyanohydrin). The reverse reaction, the so called cyanogenesis of a hydroxynitrile, is mostly used for the discovery of new HNLs, whereas the C-C bond forming condensation is a synthetically very useful reaction: a new chiral center is formed, the carbon chain is prolonged by one carbon atom and an additional versatile functional group – the nitrile – is introduced. The cyanohydrin products serve as starting material for many enzymatic and chemical follow-up reactions [10]. For industrial applications it is often essential to immobilize the enzyme in order to optimize its performance and facilitate its recovery and reuse [11]. This can involve adsorption on solid supports such as silica or ion exchange resins or carrier-free immobilization by precipitation from aqueous buffer and subsequent cross-linking with a bifunctional reagent, such as glutaraldehyde, to afford so-called cross-linked enzyme aggregates (CLEAs). The latter have been widely applied in biocatalytic transformations [12,13], including enantioselective hydrocyanations with a variety of both (*R*) and (*S*)-selective HNLs [14]. Recently, several aspects of HNL research were reviewed [9,10,15-23]. In this article we are reviewing mainly the current patent literature and scientific literature from the last two years. For an overview, Table 1 summarizes HNL enzymes covered in current patents, including enzyme source, substrate range and selectivity.

Table 1 Hydroxynitrile lyase sources specifically mentioned in patent literature.

| Enzyme source and abbreviation	Spec.	Substrate range	Heterologous expression system	Ref
Arabidopsis thaliana (cress) *At*HNL	R	aliphatic and aromatic (substituted), heteroaromatic, α,β-unsaturated aldehydes, and methylketones	*E. coli*	[24]
Chaenomeles speciosa (flowering quince) *Cs*HNL	R	aliphatic and aromatic ketones	--	[25]
Cyclonia oblonga (quince)	--	--	--	[26]
Eriobotrya japonica (loquat) *Ej*HNL	R	aliphatic and aromatic heteroaromatic, and aldehydes	*P. pastoris*	[25]
Hevea brasiliensis (rubber tree) *Hb*HNL	S	aliphatic, aromatic, heteroaromatic, α,β-unsaturated aldehydes and methylketones	*E. coli, P. pastoris, S. cerevisiae*	[27-29]
Linum usitatissimum (flax) *Lu*HNL	R, S	R: aliphatic aldehydes and methylketones, S: aromatic (with at least on C separating) ketones	*E. coli, P. pastoris*	[30]
Malus communis (apple)	R	only aromatic aldehydes tested	--	[26]
Malus pumila (paradise apple)	--	--	--	[26]
Manihot esculenta (cassava) *Me*HNL	S	aliphatic, aromatic, heteroaromatic, α,β-unsaturated aldehydes, methylketones and aromatic ketones	*E. coli, P. pastoris, S. cerevisiae Leishmania tarentolae*	[29,31,32]
Passiflora edulis (passion fruit) *Pe*HNL	R	aromatic aldehydes, aliphatic less	--	[25]
Prunus amygdalus (almond) *Pa*HNL	R, S	aliphatic, aromatic, heteroaromatic, α,β-unsaturated aldehydes and methylketones	*P. pastoris*	[26,33]
Prunus armeniaca (apricot) *Par(s)*HNL	R	Sterically demanding aromatic aldehydes	--	[26]
Prunus avium (wild cherry)	R	aliphatic and aromatic aldehydes	--	[26]
Prunus domestica (plum)	R	aliphatic and aromatic aldehydes	--	[26]
Prunus laurocerasus (cherry Laurel)	--	--	--	[26]
Prunus lyonii (Catalina cherry)	R	Only mandelonitrile tested	--	[26]
Prunus mume (Japanese apricot) *Pm*HNL	R	aliphatic, aromatic, heteroaromatic and polycyclic aldehydes, methyl and cyclic ketones	*P. pastoris*	[26,34]
Prunus persica (peach)	R	aliphatic and aromatic aldehydes	--	[25,26]
Prunus serotina (black cherry) *Ps*HNL	R	aliphatic and aromatic aldehydes	--	[26]
Prunus sp.	--	--	--	[30]
Sorbus aucuparia	R	--	--	[25]
Sorghum bicolor (millet) *Sb*HNL	S	aromatic, heteroaromatic aldehydes and methylketones	--	[30]

6.4.3   R-Selective Hydroxynitrile lyases

The group of (*R*)-selective HNLs is known for several decades and consists of the largest number of variants. Most of these variants were characterized as preparations of plant parts or as purified enzymes (Table 1). For a long time, defatted almond flour or the soluble protein fraction obtained from it was used as biocatalyst. Interestingly, another protein fraction - without HNL activity - derived from almond flour can be used as a stabilizer for this (*R*)-HNL [38].

To allow for applications on a large scale, heterologous expression in microbial hosts is indispensable. However, because the *Prunus* HNL posed difficulties in functional recombinant expression in *E. coli*, the first recombinant (*R*)-HNL was *Lu*HNL from *Linum usitatissimum*, which has a different structure and substrate spectrum [39]. Seemingly, the expression of *Lu*HNL in *E. coli* was straight forward and was not patented; neither was the enzyme itself which was first purified in 1988 [40]. HNL isoenzymes 1 and 5 from *Prunus amygdalus* were the first (*R*)-HNLs from *Rosaceae*, that were functionally expressed in the eukaryotic expression system *Pichia pastoris* and patented [33]. Genes of other *Prunus* HNL isoenzymes have been known before [41] but initially successful recombinant expression of these glycosylated and disulfide bond containing enzymes had not been shown. In contrast to *Lu*HNL, HNLs from *Rosaceae* have a broad substrate spectrum accepting aromatic aldehydes as well as aliphatic aldehydes [42] and ketones. Today, an increasing number of (*R*)-HNL genes are known. Recently, Zhao et al. identified the coding sequence of the (*R*)-hydroxynitrile lyase from *Eriobotrya japonica* (loquat) and successfully expressed and secreted it from *Pichia pastoris* [43]. *Ej*HNL comprises 552 amino acids including a 25 amino acid signal peptide and shows 60-70% sequence identity to other HNLs from *Rosaceae* (i.e. *Prunus amygdalus*, *Prunus mume* and *Prunus serotina*). The recombinant protein showed similar catalytic properties as the natural *Ej*HNL reaching up to 98.6% *ee* in the biosynthesis of (*R*)-mandelonitrile from benzaldehyde [44].

(*R*)-2-Chloro-benzaldehyde cyanohydrin was reported by Breuer et al. to be a chiral building block for an oral antiplatelet agent [45]. This compound – and other substituted benzaldehyde cyanohydrin derivatives – were synthesized on preparative scale in good yield and enantiomeric excesses (98% and 83% for (*R*)-2-Cl-benzaldehyde cyanohydrin, respectively) using defatted almond flour, the natural source of *Pa*HNL (Scheme 1) [46]. Weis et al. used a *Pa*HNL isoenzyme 5 preparation, that was obtained as highly pure secreted enzyme by heterologous expression in *P. pastoris,* as a catalyst for (*R*)-2-Cl-benzaldehyde hydrocyanation and obtained 99.7% product yield with 90% *ee*, respectively [47]. For the same substrate, chimeric recombinant proteins based on type 1 HNL (*R*)-hydroxynitrile lyase and type 5 HNL (*R*)-hydroxynitrile lyase were patented in Japan. The combination of the high specific activity of HNL1 with the favorable pH and temperature stability of HNL5 yielded industrially versatile enzymes [48]. Additional improvement in activity and selectivity was achieved by protein engineering. The secreted *Pa*HNL5α_L1Q,A111G variant showed significantly higher specific activity for (*R*)-mandelonitrile formation compared to the secreted wild-type enzyme. Moreover, it exhibited satisfactory stability at pH 2.6 [49]. In a further example, mutein *Pa*HNL5α_N3I I108M A111G 4_E10 was cultivated on the 5 liter scale and the crude supernatant was concentrated. With the use of 16.7 mg of crude protein, 1 mol of 2-chloro-benzaldehyde was converted to 86.2% to the product in 98.5% *ee* within 4 h [26]. The above mentioned wealth of achieved improvements for the preparation of (*R*)-2-Cl-mandelonitrile was mainly related to the protein itself. However, also reaction engineering can have a valuable impact on the productivity of a process.

When 2-Cl-benzaldehyde and HCN are added continuously to a stirred mixture of *Pa*HNL in methyl-*tert*-butylether (MTBE), 41.4% of the total weight of (*R*)-mandelonitrile can be obtained in 91.5% *ee* [50]. Furthermore, *Pa*HNL, e.g. can be immobilized on porous silica gel and then used in a homogenous mixture of aqueous buffer, MTBE and HCN. The total water content influences the reaction efficiency, whereby approximately 50% of water leads to quantitative conversion of 2-Cl-benzaldehyde [51].
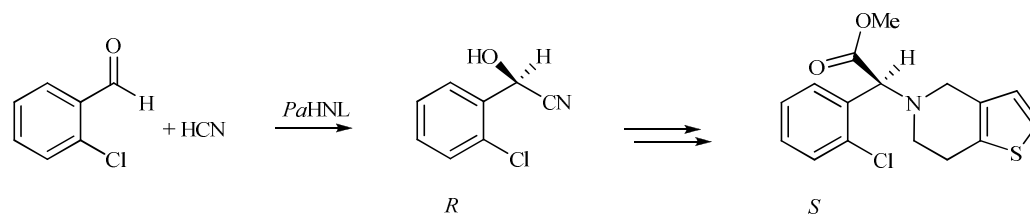
**Scheme 1.** *Pa***HNL mediated synthesis of (*R*)-2-Cl-mandelonitrile**

*Pa*HNL was also successfully immobilized as cross-linked enzyme aggregates by cross-linking with glutaraldehyde. The resulting *Pa*HNL CLEAs were highly effective catalysts for the enantioselective hydrocyanation of aldehydes under microaqueous conditions and could be recycled at least ten times without significant loss of activity [52]. Because such HNL-CLEAs perform exceptionally well in organic solvents, they can afford higher enantioselectivities than observed with the free enzymes owing to the essentially complete suppression of competing non-enzymatic hydrocyanation under these conditions. For example Roberge and coworkers obtained high enantioselectivities in the hydrocyanation of pyridine-3-carboxaldehyde (Scheme 2) [53,54]. The latter is a difficult substrate for enantioselective hydrocyanation owing to the relatively facile non-enzymatic background reaction as a result of the electron-attracting properties of the pyridine ring [55].
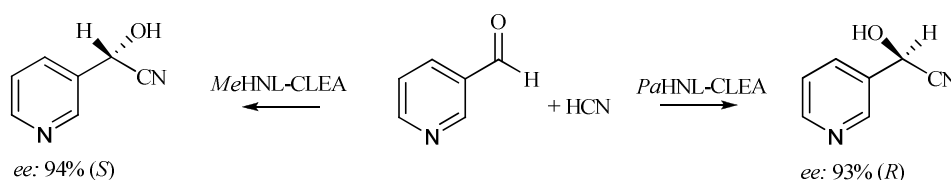
**Scheme 2. HNL-CLEAs as catalyst for the preparation of 3-pyridinecarbaldehyde cyanohydrins**

Another target molecule that was triggered with HNL technology is (*R*)-pantothenic acid (vitamin B5). Its chiral key-intermediate – (*R*)-pantolactone – is accessible by hydrocyanation of hydroxypivalaldehyde (3-hydroxy-2,2-dimethylpropanal, see Scheme 3). The two methyl groups adjacent to the aldehyde functionality imply a steric demand in the active site. Nevertheless, some HNLs formed the cyanohydrin in decent yield and *ee*. Recombinant *Lu*HNL (*Linum usitatissimum*) gave the desired product in 47% yield and 73% *ee* [56]. The special challenge with the substrate was to provide it to the enzyme in monomeric form: acidic conditions were beneficial and therefore only enzymes tolerant to low pH can be useful for this transformation on an industrial scale; native *Pa*HNL showed too low stability and afforded only 84% yield with 89% *ee* under thoroughly optimized conditions with large amounts of enzyme. Further improvement of yields and *ee*s could, however, only be achieved by enzyme engineering of the recombinant enzyme *Pa*HNL5 which is more stable at low pH than native *Pa*HNL. Following a semi-rational approach, a superior *Pa*HNL isoenzyme 5 mutant was

discovered, in which Val317 was replaced by alanine. Surprisingly, the improvement was ascribed to a reduction of stabilizing hydrophobic interactions with the less favored (S)-enantiomer and not due to changes in steric hindrance. Summarizing, this *Pa*HNL5 V317A mutein afforded the cyanohydrin intermediate with >99% conversion and 97.5% *ee* in laboratory scale experiments [57].


conv. >99%
ee 97.5%

(R)-pantolactone
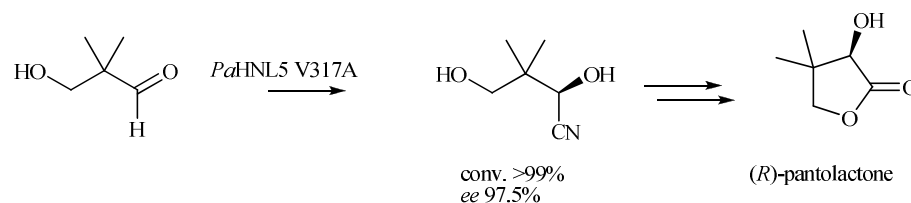
Scheme 3. Preparation of the chiral key-intermediate towards vitamin B5 synthesis

Stagonolides isolated from the fungal pathogen *Stagonospora cirsii* are nonenolides which show phytotoxic activities against some weeds and can be used as herbicides. The multi-step total synthesis of stagonolide B is started with the conversion of cheap commercially available *n*-butanal and HCN by the (R)-selective hydroxynitrile lyase from *Prunus armeniaca* (Par(s)HNL) see Scheme 4 [58].

yield: 92%
*ee*: 96%
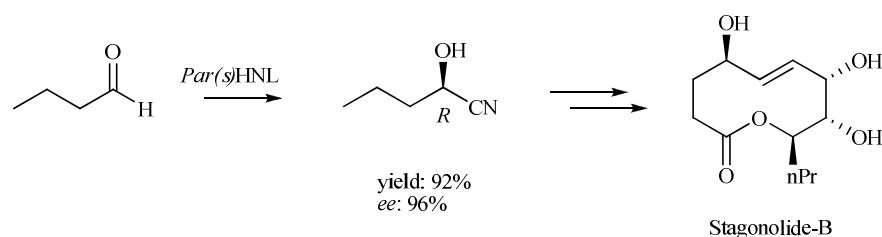
Stagonolide-B

Scheme 4. Chemoenzymatic synthesis of Stagonolide-B from *n*-butanal

Aldehydes are the classical substrates for HNL catalyzed hydrocyanation reactions; however, also ketones are transformed to their respective cyanohydrins. If the ketone substrate is not symmetric, the product represents a chiral quarternary carbon center, which is a valuable synthetic moiety that is difficult to obtain by classical chemical means. Up to now, only ketones with at least one small substituent (methyl) and cyclic ketones have been reported as HNL substrates. Notably, ketones are less reactive than aldehydes and the reaction equilibrium is often not in favor of the desired product; consequently, the reaction mixture may consist of high quantities of the substrates instead of the desired tertiary alcohol. The HNL of *Eriobotrya japonica* was recently used to convert 2-pentanone to its respective cyanohydrin which was then further hydrolyzed to give (R)-2-hydroxy-2-methylpentanoic acid in 32% overall yield and 24% *ee*. A similar yield was obtained with *Passiflora edulis* HNL, albeit the *ee* of 87% was significantly better (Scheme 5) [25].
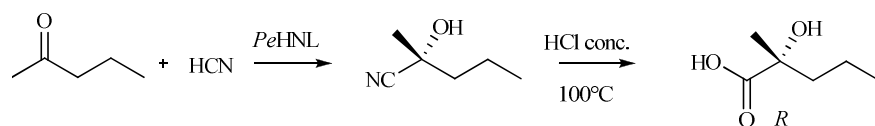
Scheme 5. From methylketone to chiral quarternary carbon center

HCN as a reagent requires special care because it is volatile and highly toxic. A well ventilated fumehood, HCN detectors and special training for the operators are prerequisites for research in this field, especially when HCN is used in neat form. To overcome the risks of HCN handling, cyanides can also be provided as acetone cyanohydrin, trimethylsilyl cyanide, and cyanoformate and the HNL catalyzed hydrocyanation reaction proceeds as transcyanation from these cyanide donors [59]. A beautiful transcyanation example is the use of ethyl cyanoformate as both an activated acyl donor as well as the source of cyanide for *Pa*HNL (Scheme 6) [60]. Despite the sophisticated concept, mixtures of both the desired carbonate and mandelonitrile were obtained [59,60].
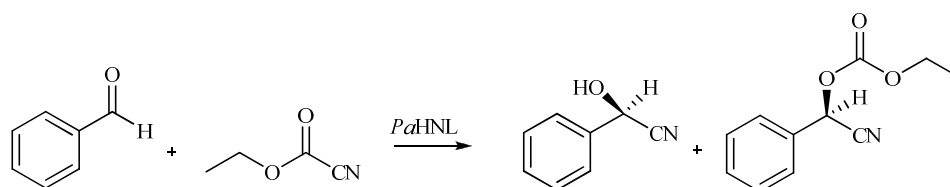

Scheme 6. Chemo-enzymatic cascade with ethyl-cyanoformate as CN source and acylation agent

Until recently, the stereopreference of HNLs was ascribed to their three dimensional structure; it was believed that α,β-hydrolase fold enzymes exclusively display (*S*)-selectivity (*vide infra*). The discovery of an (*R*)-selective HNL with high sequence similarity to (*S*)-selective enzymes from *Hevea brasiliensis* and *Manihot esculenta* came as a surprise, especially because its origin was a non-cyanogenic plant: *Arabidopsis thaliana*. Substrate range and biophysical parameters of *At*HNL were determined [3,24,61] and the structure was solved (Pdb ID: 3DQZ) [62]. Although the enzyme is highly similar to the structures of (*S*)-selective HNLs (*Hb*HNL and *Me*HNL), including the catalytic Ser-His-Asp triad in the active site, several differences in the active site were identified, which are potentially responsible for the diverging binding of the substrate and thus result in a different reaction mechanism. Based on docking studies and site-directed mutagenesis, a reaction mechanism for (*R*)-selective HNLs with α,β-hydrolase fold was proposed. In the first step, the cyanohydrin hydroxyl group is directly deprotonated by the His236 residue of the catalytic triad. The hydroxyl group also interacts with Asn12 to facilitate the deprotonation. Compared to *Hb*HNL, a Lys residue which stabilizes the negative charge of the released cyano group is missing. In contrast, the cyano group could interact with the backbone amide groups of Ala13 and Phe82 and thereby be stabilized by an oxyanion hole similar to those described in serine hydrolases. Finally, the cyanide is protonated indirectly by His236 with Ser81 as a mediator [62].

Interestingly, despite the similarity to *Hb*HNL and *Me*HNL, *At*HNL is significantly less stable at acidic pH (below pH 5.4) than the (*S*)-selective HNLs [61], which is rather deplorable because the non-enzymatic background reaction is more abundant at higher pH and thus reduces the *ee* of the product. This background

reaction is also dependent on the water content of the reaction medium. Thus, a lot of effort was recently directed to stabilize *At*HNL at low pH or, alternatively, to reduce the water content in the reaction systems. The introduction of eleven surface mutations of *At*HNL and exchange of these amino acids for the corresponding amino acids from the more stable *Me*HNL resulted in a variant, which is active at pH 4.5 (23.6 U/mg initial rate activity in the cleavage of mandelonitrile, a half-life of about 13 h at 20°C, and 70% conversion with 99.3% *ee* at pH 4.5 and 0°C in a citrate-phosphate/MTBE two-phase system) [63]. In another approach, buffer-saturated (pH 6.5) MTBE was used in the reaction system either with precipitated protein or celite-immobilized or solgel-entrapped *At*HNL [64]. While fresh solgel-entrapped *At*HNL showed the most prominent improvement in activity and *ee* in the first round of reaction, celite-immobilized *At*HNL was much more stable during recycling (in an organic solvent-resistant, fine-woven nylon mesh 'teabag') and storage. It was shown that the water-content was crucial for optimal activity. In a subsequent paper excellent enantioselectivity was achieved using *E. coli* whole cells expressing *At*HNL in a mono-phasic micro-aqueous reaction system with MTBE. Importantly, the cells were recyclable using the 'tea-bag' approach without loss of enantioselectivity, however, with decreasing conversion rates [65].

Recently, Asano and his group showed that *At*HNL also catalyzes the first (*R*)-selective HNL-catalyzed Henry reaction of aromatic aldehydes (Scheme 7). They achieved the highest enantioselectivity for benzaldehyde and MeNO$_2$ in a biphasic system at pH 7 with 50% *n*-butyl acetate (20% yield and 90% *ee*). It was shown that substituents on the aromatic ring did not influence the reactivity substantially. The best enantioselectivity was obtained with 2-naphthaldehyde, however, with low yield (7% yield, >99.9% *ee*) [66]. Aliphatic aldehydes were converted sluggishly, which was not surprising, as *At*HNL is also less active towards aliphatic aldehydes in standard hydrocyanation reactions [3].
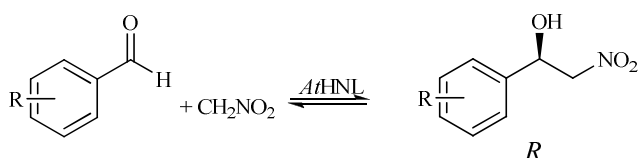
Scheme 7. *R*-selective *At*HNL catalyzed asymmetric nitroaldol condensations

6.4.4 S-Selective Hydroxynitrile lyases

In 1989, Conn et al. discovered a hydroxynitrile lyase from the rubber tree [67] that turned out to be the first HNL exhibiting (*S*)-selectivity [68]. This HNL from *Hevea brasiliensis* (*Hb*HNL) and a similar enzyme from *Manihot esculenta* (*Me*HNL) belong to the α/β-hydrolase fold-type enzyme class with the typical catalytic triad aspartic acid, serine and histidine in the active site. These two enzymes exhibit high sequence identity and show a relatively broad substrate tolerance towards aliphatic, alicyclic, unsaturated, aromatic, and heteroaromatic aldehydes and ketones. The cDNA sequence coding for the enzyme *Me*HNL was first published in 1994 [69]. The primary sequence of *Hb*HNL as well as homologous enzymes with at least 80% protein sequence identity were protected by a patent application [27], followed by the enantioselective process towards the preparation of

(*S*)-cyanohydrins [70]. From the reaction engineering viewpoint, the use of immobilized enzymes has advantages over a reaction in homogenous system: immobilized enzyme is easily removed from the reaction mixture, can be reused and facilitate product purification. *Me*HNL was, for example, immobilized on nitrocellulose and used for the preparation of an array of (*S*)-cyanohydrins derived from both aldehydes and ketones [71]. Sintered clay, silica, alumina and silica-alumina carriers with pore sizes of 10-80 nm and a surface area of more than 20 m²g⁻¹ can also be used as carriers for *Me*HNL [72]. Porous silica gel, used in a homogenous reaction system consisting of MTBE and aqueous buffer with a water content of 50%, can be re-used up to 16 times for the preparation of (*S*)-mandelonitrile without loss of activity and selectivity [51]. Notably, there is no general HNL immobilization technique, because each enzyme works optimal on different carriers and reaction conditions and each setup needs separate optimization [73].

For industrial application, it is more convenient to over-express the desired enzymes in heterologous systems than to use isolated enzyme from plants. Semba et al. claimed the expression of the *Me*HNL coding gene derived from cassava in yeasts. In particular, *Me*HNL was intracellularly expressed in *Saccharomyces cerevisiae* and *Pichia*, using vectors with the GAP or the AOX1 promoter, respectively [32]. In order to increase heterologous expression levels also in *E. coli* and thereby facilitating mutation studies, codon optimization proved useful. The codon use in plants differs from that in the prokaryotic host *E. coli* and therefore, the gene sequence of *Me*HNL was optimized: all codons which are used by *E. coli* with frequencies lower than 5% were replaced with more frequently used codons. Additional improvement of functional expression by correct folding was achieved by low temperature induction (17°C) [31,74].

In case of substrates with bulky substituents, conversions or enantiomeric excesses (*ee*s) of wild-type (*S*)-HNLs were not satisfying. One reason appeared to be the rather narrow hydrophobic substrate entrance channel to the active site. This problem was tackled by the exchange of bulky amino acid residues of *Hb*HNL and *Me*HNL with less space demanding residues such as the tryptophan in position 128 for alanine or leucine [75-77]. In addition to this rational approach, random mutagenesis revealed a range of positions which yielded improved enzyme activity for the cleavage of the standard test substrate *rac*-mandelonitrile. Also site-saturation mutagenesis at selected positions and even the insertion of an amino acid between position 128 and 129 in the *Me*HNL sequence resulted in improved variants [29]. Enzyme engineering was also performed to improve *Me*HNL's heat tolerance. The interaction of helix D3' with helix A and β-sheet 2 was increased by the introduction of specific mutations. Substitution of glycine 165 by acidic amino acids such as aspartate or glutamate in helix D3' increased hydrogen bonding with basic residues. Alternatively, hydrophobic interaction was increased by substitution of valine 173 by leucine. These variants as well as other mutants and multiple mutants in the above mentioned regions showed increased heat stability and also stability against organic solvents [28].

Histidine 103 was identified as a hotspot by random mutagenesis of wild-type *Me*HNL [78]. Specifically, substitutions to hydrophobic amino acids with aliphatic side chains (Leu, Val, Ile, Met and Cys) showed enhanced solubility during expression in *E. coli* (or a prokaryotic cell free expression system) and consequential higher activity compared to WT *Me*HNL, which is expressed mainly in inclusion bodies in *E. coli* [77,79]. However, no improvement of expression level and activity was observed when the same mutants were expressed in *P. pastoris* or *Leishmania tarentolae* or a eukaryotic cell free expression system, in which WT *Me*HNL

already expressed as soluble protein [79]. Expression level of *Me*HNL in *E. coli* was about 7-fold improved by changing the valine in position 2 for a lysine [78]. Also substitutions for Ile, Asn, Arg and Gln showed a beneficial effect on the expression level which directly translated into increased activity of cell free extracts. A combination of V2I and H103L resulted in a 10-fold improved strain as compared to the wild-type (WT). Moreover, substitution of other lysine residues at the surface of the protein by prolines increased the expression level further [78]. In the same patent application, mutants of enzymes with more than 90% sequence identity to *Hb*HNL as well as to *Me*HNL are claimed [78]. Analysis of several solvent-exposed Lys residues in *Me*HNL by site-saturation mutagenesis revealed that Pro-substitution at position 176, 199 and 224 resulted in higher solubility of the mutants. Combination of all three to a triple mutant added up to the highest activity due to an increase in solubility [80].

Recently, the first HNL, which was engineered from an esterase, was published [81]. Esterases and some HNLs belong to the α/β-hydrolase superfamily and they even share the typical conserved Asp, His and Ser catalytic triad in their catalytic site. However, the reaction mechanisms are different. Recently, two groups introduced HNL activity into esterases. Schwab et al. converted the bacterial esterase EstC from *Burkholderia gladioli* into an HNL by introduction of only one single mutation (S276K), which was sufficient to generate HNL activity and to abolish esterase activity [82]. Kazlauskas et al. switched the plant esterase SABP2 to an HNL with two point mutations (G12T and M239K). The enzyme showed strongly reduced esterase activity, but clearly detectable HNL activity with racemic mandelonitrile (20 mU/mg, $k_{cat}/K_M = 72$ min^{-1}M^{-1}). Enantioselectivity is expected to be low in cyanogenesis direction because the enzyme showed comparable activity for (*R*)- and (*S*)-mandelonitrile cleavage with 12.6 and 15.5 mU/mg, respectively. This result was confirmed in the synthesis direction, where the (*S*)-product showed 20% *ee* [81]. The sequence of the plant gene SABP2 is claimed in several patent applications not related to HNL chemistry but increased resistance of plants overexpressing such genes.

6.4.5 Cascade Reactions

The reaction products of hydroxynitrile lyases, cyanohydrins, can be hydrolyzed by nitrilases to α-hydroxycarboxylic acids and, occasionally, to amide as side product. Thus, the two enzymes can be coupled to a cascade reaction (Scheme 8). Several approaches, including engineering of the nitrilase, application of whole cells co-expressing both enzymes and the use of combi-CLEAs in combination with optimization of the reaction conditions to increase the yield of mandelic acid from benzaldehyde, have been described [10]. Similarly, a combi-CLEA consisting of a HNL and a nitrile hydratase was used for the one-pot cascade synthesis of enantiopure aliphatic α-hydroxycarboxylic amides from the corresponding aldehydes [83]. *E. coli* resting cells co-expressing *Me*HNL and an arylacetonitrilase from *Pseudomonas fluorescens* EBC191 were also used to synthesize (*S*)-atrolactic acid with 92% *ee* and atrolactamide as side product from acetophenone. Interestingly, with this coupled system a much higher conversion of acetophenone was achieved than with *Me*HNL alone [76,84], presumably because the equilibrium was shifted. The formation of amide was reduced to 4% by applying the Tyr54Val nitrilase variant [85]. In the most recent approach, whole cells of the same expression system were subsequently used in a two-phase system with water-immiscible ionic liquids [1-butyl-1-

methylpyrrolidinium-bis(trifluoromethanesulfonyl)imide (BMpl NTf2), 1-butyl-3-
methylimidazoliumbis(trifluoromethanesulfonyl)imide (BMim NTf2) or 1-butyl-3-methylimidazolium
hexafluorophosphate (BMim PF6)] [86]. Depending on the reaction conditions, 80-100% of conversion was
reached with up to 95% *ee* for (*S*)-mandelic acid and up to 99% *ee* for the (*S*)-mandeloamide. Here, amide
formation dominated mandelic acid formation.

Scheme 8 (*S*)-selective hydrocyanation in cascade with nitrile hydrolysis using unselective nitrilase.

A chemoenzymatic approach was recently applied for the preparation of the enantiomers of Venlafaxine
hydrochloride (brand name EffexorXR® marketed by Pfizer) and analogs thereof. Both enantiomers are
antidepressant agents targeting different receptors. The asymmetric synthesis started with the conversion of
cyclohexanone derivatives to their corresponding cyanohydrin by the (*S*)-selective *Hb*HNL followed by several
synthetic steps and finally a lipase catalyzed kinetic resolution [87] (Scheme 9).
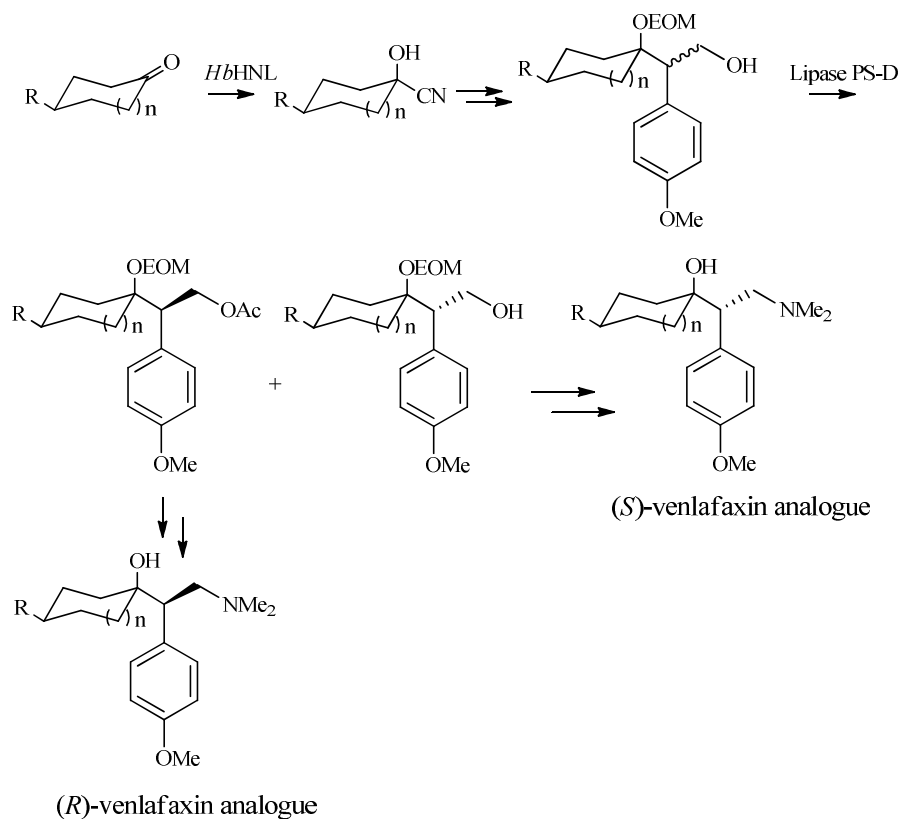
Scheme 9. Chemoenzymatic synthesis of both enantiomers of Venlafaxine hydrochloride

6.4.6 Reaction engineering

A drawback of HNL catalyzed C-C bond formation is the non-enzymatic background reaction, which can significantly compromise the enantiomeric excess (ee) of the reaction. This non-enzymatic reaction needs to be eliminated but at the same time enzyme activity must be conserved.

One way of suppression is to perform the synthesis of cyanohydrins in an acidic environment, in particular pH≤4.5, at temperatures such that the competing chemical reaction and racemization are negligible compared with the enzymatic synthesis. To overcome losses of biocatalyst activity under these conditions, the reaction can be performed at low temperature in a range from 5° to 8°C [88]. The currently most widely applied method for biotransformations with HNLs is the use of biphasic systems [73]. Organic substrates are usually soluble in the organic layer; the substrate concentration can be high without inhibiting the enzyme or decreasing its stability, due to the separation from the enzyme in the aqueous layer. Chemical side-reactions are suppressed in the near water-free organic layer and, moreover, the organic solvent can be easily removed and the enzyme in the aqueous phase can be reused. Different aldehydes and ketones can be converted to the respective (*S*)-cyanohydrins in biphasic system consisting of aqueous solution of (*S*)-HNL and a water immiscible or slightly miscible organic phase [89]. A very similar procedure was patented for aromatic (*R*)-benzaldehyde cyanohydrins [46]. Alternatively, the production of the 2-chloro-benzaldehyde and *n*-butyraldehyde cyanohydrins was very efficient in an emulsion system. After 3 hours of reaction time in continuous emulsion 100% of conversion was reached with 89.1% *ee*, compared to the traditional two-phase system which gave 98.7% conversion and 86% *ee* after 4.5 hours [90].

6.4.7 Current & Future Developments

First genes for (*S*)- and (*R*)-HNLs were identified and expressed about two decades ago and the discovery of new enzymes is still ongoing. Many international patent applications still cover the production, improved variants, new enzymes from different plant origin and applications of both (*S*)- and (*R*)-HNLs for chemical synthesis. However, besides patent protected major enzyme sources several enzymes such as *Lu*HNL, engineered esterases with HNL activity and new bacterial HNLs are targets of basic research. HNL catalysis offers exciting new opportunities for broad applications in industrial production.

6.4.8 Acknowledgements

6.4.9 Conflict of interest

The authors confirm that this article content has no conflict of interest.

6.4.10 Reference List

1. Poulton JE. Cyanogenesis in plants. Plant Physiol 1990; 94: 401-405.
2. Zagrobelny M, Bak S, Moeller B. Cyanogenesis in plants and arthropods. Phytochemistry 2008; 69: 1457-1468.
3. Andexer J, von Langermann J, Mell A et al. An R-selective hydroxynitrile lyase from *Arabidopsis thaliana* with an alpha/beta-hydrolase fold. Angew Chem Int Ed 2007; 46: 8679-8681.
4. Wöhler F, Liebig J. Ueber die Bildung des Bittermandelöls. Ann Pharm 1837; 22: 1-24.
5. Rosenthaler L. Enzyme effected asymmetrical synthesis. Biochem Zeitschrift 1908; 14: 238-253.
6. Wajant H, Forster S, Selmar D, Effenberger F, Pfizenmaier K. Purification and characterization of a novel (*R*)-mandelonitrile lyase from the fern *Phlebodium aureum*. Plant Physiol 1995; 109: 1231-1238.
7. Gruber-Khadjawi M, Fechter MH, Griengl H. Cleavage and formation of cyanohydrins. In: Drauz K, Gröger H, May O, Eds Enzyme Catalysis in Organic Synthesis. Weinheim: Wiley-VCH 2012: pp. 947-990.
8. Hilterhaus L, Liese A. Industrial application and processes using carbon–carbon lyases. In: Drauz K, Gröger H, May O, Eds Enzyme Catalysis in Organic Synthesis. Weinheim: Wiley-VCH 2012: pp. 991-1000.
9. Purkarthofer T, Skranc W, Schuster C, Griengl H. Potential and capabilities of hydroxynitrile lyases as biocatalysts in the chemical industry. Appl Microbiol Biotechnol 2007; 76: 309-320.
10. Winkler M, Glieder A, Steiner K. Hydroxynitrile lyases: from nature to application. In: Carreira EM, Yamamoto H, Eds. Comprehensive Chirality, Synthetic Methods VI - Enzymatic and Semi-Enzymatic. Amsterdam: Elsevier 2012: Vol. 7, pp. 350-371.
11. Sheldon RA. Enzyme immobilization: the quest for optimum performance. Adv Synth Catal 2007; 349: 1289-1307.
12. Sheldon RA. Cross-linked enzyme aggregates as industrial biocatalysts. Org Proc Res Devel 2011; 15: 213-223.
13. Sheldon RA. Characteristic features and biotechnological applications of cross-linked enzyme aggregates (CLEAs). Appl Microbiol Biotechnol 2011; 92: 467-477.
14. Sheldon RA. Industrial Applications of Asymmetric Synthesis using Cross-Linked Enzyme Aggregates. In: Carreira EM, Yamamoto H, Eds. Comprehensive Chirality, Synthetic Methods VI - Enzymatic and Semi-Enzymatic. Amsterdam: Elsevier 2012: Vol. 9, pp. 353-366.
15. Effenberger F, Förster S, Wajant H. Hydroxynitrile lyases in stereoselective catalysis. Curr Opin Biotechnol 2000; 11: 532-539.
16. Sharma M, Sharma NN, Bhalla TC. Hydroxynitrile lyases: at the interface of biology and chemistry. Enz Microbial Technol 2005; 37: 279-294.
17. Lu W-Y, Lin G-Q. Chiral synthesis of pharmaceutical intermediates using oxynitrilases. In: Tao J, Lin G-Q, Liese A, Eds. Biocatalysis for the Pharmaceutical Industry, Wiley VCH 2009: pp. 89-109.
18. Andexer JN, Langermann J, Kragl U, Pohl M. How to overcome limitations in biotechnological processes - examples from hydroxynitrile lyase applications. Trends Biotechnol 2009; 27: 599-607.
19. Holt J, Hanefeld U. Enantioselective enzyme-catalysed synthesis of cyanohydrins. Curr Org Synth 2009; 6: 15-37.
20. Bornscheuer UT, Kazlauskas RJ. Catalytic promiscuity in biocatalysis: using old enzymes to form new bonds and follow new pathways. Angew Chem Int Ed 2004; 43: 6032-6040.
21. North M. Synthesis and applications of non-racemic cyanohydrins. Tetrahedron-Asymm 2003; 14: 147-176.
22. Sukumaran J, Hanefeld U. Enantioselective C-C bond synthesis catalysed by enzymes. Chem Soc Rev 2005; 34: 530-542.
23. Dadashipour M, Asano Y. Hydroxynitrile lyases: insights into biochemistry, discovery, and engineering. ACS Catalysis 2011; 1: 1121-1149.
24. Eggert T, Andexer JN. (*R*)-hydroxynitrile lyase from Brassicaceae. US20100112648, 2010.
25. Asano Y. (*R*)-hydroxynitrile lyase and method for using the same. JP4746084, 2009.
26. Gaisberger R, Glieder A, Liu ZB, Pscheidt B. R-HNL random variants and their use for preparing optically pure, sterically hindered cyanohydrins. US 20100143986, 2010.
27. Hasslacher M, Schall M, Schwab H, Hayn E, Kohlwein S, Griengl H. (*S*)-hydroxynitrilelyase from *Hevea brasiliensis*. WO9703214, 1997.

28. Ichige E, Semba H, Shijuku T, Harayama S. Novel modified *S*-hydroxynitrile lyase. US20080124784, 2008.
29. Senba T, Ichige E. New highly-active modified-type S-hydroxynitrile lyase. JP4494286, 2006.
30. Effenberger F, Förster S, Wajant H. Method for the production of hydroxynitrile lyases. EP1246909, 2006.
31. Semba H, Ichige E, Mukouyama M. Method for production of *S*-hydroxynitrile lyase by use of *Escherichia coli*. US7214523, 2007.
32. Semba J, Tsuchiura-shi I. Process for producing *S*-hydroxynitrile lyase. EP1016712, 2005.
33. Schwab H, Glieder A, Kratky C, Dreveny I, Poechlauer P, Skranc W, Mayrhofer H, Wirth I, Neuhofer R, Bona R. Genes containing a DNA sequence coding for hydroxynitrile lyase, recombinant proteins derived therefrom and having hydroxynitrile lyase activity, and use thereof. US6861243, 2005.
34. Asano Y. Synthetase for optically active cyanohydrin derived from *Prunus mume* fruit and method for production of optically active substance using the same. JP2006238763, 2006.
35. Asano Y, Tamura K, Doi N et al. Screening for new hydroxynitrilases from plants. Biosci Biotechnol Biochem 2005; 69: 2349-2357.
36. Hernandez L, Luna H, Ruiz-Teran F, Vazquez A. Screening for hydroxynitrile lyase activity in crude preparations of some edible plants. J Mol Catal B-Enzym 2004; 30: 105-108.
37. Hickel A, Heinrich G, Schwab H, Griengl H. Screening for hydroxynitrile lyases in plants. Biotechnol Techniques 1997; 11: 55-58.
38. Senba T, Dobashi Y. (*R*)-hydroxynitrile lyase composition and stabilization method. JP2004113053, 2004.
39. Trummler K, Wajant H. Molecular cloning of acetone cyanohydrin lyase from flax (*Linum usitatissimum*) - definition of a novel class of hydroxynitrile lyases. J Biol Chem 1997; 272: 4770-4774.
40. Xu LL, Singh BK, Conn EE. Purification and characterization of acetone cyanohydrin lyase from *Linum usitatissimum*. Arch Biochem Biophys 1988; 263: 256-263.
41. Cheng IP, Poulton JE. Cloning of cDNA of *Prunus-serotina* (*R*)-(+)-mandelonitrile lyase and identification of a putative FAD-binding site. Plant Cell Physiol 1993; 34: 1139-1143.
42. Effenberger F, Ziegler T, Förster S. Verfahren zur Herstellung von optisch aktiven Cyanhydrinen. EP0276375, 1992.
43. Zhao GJ, Yang ZQ, Guo YH. Cloning and expression of hydroxynitrile lyase gene from *Eriobotrya japonica* in *Pichia pastoris*. J Biosci Bioeng 2011; 112: 321-325.
44. Ueatrongchit T, Komeda H, Asano Y. Parameters influencing asymmetric synthesis of (*R*)-mandelonitrile by a novel (*R*)-hydroxynitrile lyase from *Eriobotrya japonica*. J Mol Catal B-Enzym 2009; 56: 208-214.
45. Breuer M, Ditrich K, Habicher T et al. Industrial methods for the production of optically active intermediates. Angew Chem Int Ed 2004; 43: 788-824.
46. Kirschbaum B, Wilbert G, Effenberger F. Process for preparing optically active cyanohydrins and secondary products. US6653498, 2003.
47. Weis R, Poechlauer P, Bona R et al. Biocatalytic conversion of unnatural substrates by recombinant almond *R*-HNL isoenzyme 5. J Mol Catal B-Enzym 2004; 29: 211-218.
48. Dobashi Y, Senba T. Chimeric recombinant (*R*)-hydroxynitrile lyase. JP2009106167, 2009.
49. Skranc W, Glieder A, Gruber K, Weis R, Luiten R. *R*-hydroxynitrile lyases having improved substrate tolerance and the use thereof. US7572608, 2009.
50. Ogura K, Sato E, Mori H, Nogami H, Kato O, Saito T, Kokubu N, Ninomiya Y. Method for producing optically active cyanohydrin and method for producing optically active alpha-hydroxy carboxylic acid. JP2012105671, 2012.
51. Semba H, Dobashi Y. An enzyme reaction method and a method for enzymatically producing an optically active cyanohydrin. EP1160329, 2011.
52. van Langen LM, Selassa RP, van Rantwijk F, Sheldon RA. Cross-linked aggregates of (*R*)-oxynitrilase: a stable, recyclable biocatalyst for enantioselective hydrocyanation. Org Lett 2005; 7: 327-329.
53. Roberge C, Fleitz F, Pollard D, Devine P. Synthesis of optically active cyanohydrins from aromatic ketones: evidence of an increased substrate range and inverted stereoselectivity for the hydroxynitrile lyase from *Linum usitatissimum*. Tetrahedron Asymm 2007; 18: 208-214.
54. Roberge C, Fleitz F, Devine P. (*R*)- and (*S*)-cyanohydrin formation from pyridine-3-carboxaldehyde using CLEA-immobilized hydroxynitrile lyases. In: Whittall J, Sutton PJ, Eds. Practical Methods for Biocatalysis and Biotransformations. Wiley: 2010; pp. 114-118.

55. Saito T, Kubo E, Mori H. Preparation of optically-active amino alcohols and their hydrochlorides by reduction of pyridinecarbaldehyde cyanohydrins. JP2012217375, 2012.

56. Trummler K, Roos J, Schwaneberg U et al. Expression of the Zn^{2+}-containing hydroxynitrile lyase from flax (*Linum usitatissimum*) in *Pichia pastoris*-utilization of the recombinant enzyme for enzymatic analysis and site-directed mutagenesis. Plant Sci 1998; 139: 19-27.

57. Pscheidt B, Liu ZB, Gaisberger R et al. Efficient biocatalytic synthesis of (*R*)-pantolactone. Adv Synth Catal 2008; 350: 1943-1948.

58. Das T, Mahapatra T, Nanda S. Total synthesis of stagonolide B. Tetrahedron Lett 2012; 53: 1186-1189.

59. Purkarthofer T, Skranc W, Weber H et al. One-pot chemoenzymatic synthesis of protected cyanohydrins. Tetrahedron 2004; 60: 735-739.

60. Skranc W, Poechlauer P, Wirth I, Neuhofer R, Mayrhofer H. Process for preparing protected enantiomer-enriched cyanohydrins by in-situ derivatization. US6909011, 2005.

61. Guterl JK, Andexer JN, Sehl T et al. Uneven twins: comparison of two enantiocomplementary hydroxynitrile lyases with alpha/beta-hydrolase fold. J Biotechnol 2009; 141: 166-173.

62. Andexer JN, Staunig N, Eggert T et al. Hydroxynitrile lyases with alpha/beta-hydrolase fold: two enzymes with almost identical 3D structures but opposite enantioselectivities and different reaction mechanisms. ChemBioChem 2012; 13: 1932-1939.

63. Okrob D, Metzner J, Wiechert W, Gruber K, Pohl M. Tailoring a stabilized variant of hydroxynitrile lyase from *Arabidopsis thaliana*. ChemBioChem 2012; 13: 797-802.

64. Okrob E, Paravidino M, Orru RVA et al. Hydroxynitrile lyase from *Arabidopsis thaliana*: identification of reaction parameters for enantiopure cyanohydrin synthesis by pure and immobilized catalyst. Adv Synth Catal 2011; 353: 2399-2408.

65. Scholz KE, Okrob D, Kopka B et al. Synthesis of chiral cyanohydrins by recombinant *Escherichia coli* cells in a micro-aqueous reaction system. Appl Environ Microbiol 2012; 78: 5025-5027.

66. Fuhshuku K, Asano Y. Synthesis of (*R*)-beta-nitro alcohols catalyzed by R-selective hydroxynitrile lyase from *Arabidopsis thaliana* in the aqueous-organic biphasic system. J Biotechnol 2011; 153: 153-159.

67. Selmar D, Lieberei R, Biehl B, Conn EE. Alpha-hydroxynitrile lyase in *Hevea brasiliensis* and its significance for rapid cyanogenesis. Physiol Plantarum 1989; 75: 97-101.

68. Klempier N, Griengl H, Hayn M. Aliphatic (*S*)-cyanohydrins by enzyme-catalyzed synthesis. Tetrahedron Lett 1993; 34: 4769-4772.

69. Hughes J, Decarvalho JPC, Hughes MA. Purification, characterization, and cloning of alpha-hydroxynitrile lyase from cassava (*Manihot esculenta* Crantz). Arch Biochem Biophys 1994; 311: 496-502.

70. Kirchner G, Wirth I, Werenka C, Griengl H, Schmidt M. Enzymatic processes for preparing (*S*)-cyanohydrins. US6337196, 2002.

71. Effenberger F, Wajant H, Förster S, Roos J. Process for the preparation of (*S*)-cyanhydrines. EP0799894, 2004.

72. Semba H, Dobashi Y. Immobilised enzyme and method for synthesizing an optical active cyanohydrin. EP1116789, 2004.

73. Avi M, Griengl H. Biocatalysis in biphasic systems: Oxynitrilases. In: Carrea G, Riva S, Eds. Organic Synthesis with Enzymes in Non-Aqueous Media. Weinheim: Wiley-VCH 2008: pp. 211-226.

74. Semba H, Ichige E, Imanaka T, Atomi H, Aoyagi H. Efficient production of active form of recombinant cassava hydroxynitrile lyase using *Escherichia coli* in low-temperature culture. Methods Mol Biol 2010; 643: 133-144.

75. Avi M, Wiedner R, Griengl H, Schwab H. Improvement of a stereoselective biocatalytic synthesis by substrate and enzyme engineering: 2-hydroxy-(4'-oxocyclohexyl)acetonitrile as the model. Chem Eur J 2008; 14: 11415-11422.

76. Bühler H, Effenberger F, Förster S, Roos J, Wajant H. Substrate specificity of mutants of the hydroxynitrile lyase from *Manihot esculenta*. ChemBioChem 2003; 4: 211-216.

77. Lauble H, Miehlich B, Förster S et al. Structure determinants of substrate specificity of hydroxynitrile lyase from *Manihot esculenta*. Protein Sci 2002; 11: 65-71.

78. Asano Y, Akiyama T, Yu F, Sato E. Substitutional variants of hydroxynitrile lyase with high specific activity and methods of use. US8030053, 2011.

79. Dadashipour M, Fukuta Y, Asano Y. Comparative expression of wild-type and highly soluble mutant His103Leu of hydroxynitrile lyase from *Manihot esculenta* in prokaryotic and eukaryotic expression systems. Protein Expr Purif 2011; 77: 92-97.

80. Asano Y, Dadashipour M, Yamazaki M, Doi N, Komeda H. Functional expression of a plant hydroxynitrile lyase in *Escherichia coli* by directed evolution: creation and characterization of highly *in vivo* soluble mutants. Prot Eng Des Sel 2011; 24: 607-616.

81. Padhi SK, Fujii R, Legatt G et al. Switching from an esterase to a hydroxynitrile lyase mechanism requires only two amino acid substitutions. Chem Biol 2010; 17: 863-871.

82. Feichtenhofer S, Höffken WW, Schwab H. Conversion of esterase EstC from *Burkholderia gladioli* into a hydroxynitrile lyase. 2009; Poster at Enzyme Engineering XX, Groningen, The Netherlands.

83. van Pelt S, van Rantwijk F, Sheldon RA. Synthesis of aliphatic (*S*)-alpha-hydroxycarboxylic amides using a one-pot bienzymatic cascade of immobilised oxynitrilase and nitrile hydratase. Adv Synth Catal 2009; 351: 397-404.

84. von Langermann J, Mell A, Paetzold E, Daußmann T, Kragl U. A hydroxynitrile lyase in organic solvent-free systems to overcome thermodynamic limitations. Adv Synth Catal 2007; 349: 1418-1424.

85. Baum S, Williamson DS, Sewell T, Stolz A. Conversion of sterically demanding alpha,alpha-disubstituted phenylacetonitriles by the arylacetonitrilase from *Pseudomonas fluorescens* EBC191. Appl Environ Microbiol 2012; 78: 48-57.

86. Baum S, van Rantwijk F, Stolz A. Application of a recombinant *Escherichia coli* whole-cell catalyst expressing hydroxynitrile lyase and nitrilase activities in ionic liquids for the production of (*S*)-mandelic acid and (*S*)-mandeloamide. Adv Synth Catal 2012; 354: 113-122.

87. Bhuniya R, Nanda S. Asymmetric synthesis of both the enantiomers of antidepressant venlafaxine and its analogues. Tetrahedron Lett 2012; 53: 1990-1992.

88. Niedermayer U, Kragl U, Kula M-R, Wandrey C, Makryaleas K, Drauz K. Enzymatisches Verfahren zur Herstellung von optisch aktiven Cyanhydrinen. EP0326063, 1994.

89. Poechlauer P, Schmidt M, Neuhofer R, Zabelinskaja-Mackova AA, Griengl H, van den Broek C, Reintjens R, Jelle Wories H. Enzymatic process for the preparation of (*S*)-cyanohydrins. US6225095, 2001.

90. Pöchlauer P, Wirth I, Mayrhofer H, Neuhofer R. Method for the production of optically active cyanohydrins using *R*-oxynitilase. US6717006, 2004.

7. CURRICULUM VITAE

Full Name

Ivan Hajnal

Date and Place of Birth

22.02.1987 in Prijedor, Bosnia and Herzegovina, former Yugoslavia

Education

2010 - 2013	PhD thesis within the ACIB GmbH, under the supervision of Prof. Dr. Helmut Schwab, PhD program DK Molecular Enzymology, Graz University of Technology, Austria
2008- 2010	Master of Science in Molecular Microbiology "with distinction", University of Graz, Austria
2005 – 2008	Bachelor of Science in Molecular Microbiology, University of Graz, Austria
1993 – 2005	Primary and Secondary Education in Prijedor, Bosnia and Herzegovina

Research Stay

10-12/ 2011	Laboratory of Prof. U. Hanefeld, Technical University Delft, The Netherlands