Structural studies into ketosteroid dehydrogenases and S-selective \(\omega\)-transaminases
van Oosterwijk, Cornelis Christiaan

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2016

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Chapter 1

General introduction on ketosteroid dehydrogenases and ω-transaminases
Chapter 1

This thesis investigates the 3D structures of enzymes, as an aid to facilitate their optimisation for biotechnological applications. Such applications may require the enzymes to accept new substrates or deal with challenging reaction conditions. Crystal structures combined with biochemical data provide a solid base to adapt enzymes to the specific requirements of their application. Additionally, structures of proteins involved in disease may be used for drug design. Here we introduce the two main subjects of this thesis: ketosteroid dehydrogenases (KstDs) and $\omega$-transaminases ($\omega$TAs).

Ketosteroid dehydrogenases introduce a double bond into the A-ring of ketosteroids to produce unsaturated products, which are building blocks for steroid-containing pharmaceuticals. Steroids are an important class of compounds as they have found a wide range of therapeutic uses as anti-inflammatory, immunosuppressive, progestational, diuretic, anabolic and contraceptive agents. They are ranked among the most widely marketed groups of products from the pharmaceutical industry (Fernandes et al., 2003). Ketosteroid dehydrogenases are also part of the cholesterol degradation pathway of several pathogenic actinomycetes (e.g. Mycobacterium tuberculosis). Cholesterol degradation pathways have been proposed as attractive targets for inhibitor and vaccine design (Ouellet et al., 2011; van der Geize et al., 2011).

$\omega$-Transaminases are pyridoxal-5’-phosphate (PLP) dependent enzymes that are remarkably efficient for the synthesis of a wide range of industrially interesting (chiral) amines from ketones. There is currently considerable interest in the discovery of novel ($R$)- and ($S$)-selective transaminases and their optimization for new applications. Structural data for these enzymes are valuable to obtain a deeper understanding of their functioning, facilitating modifications that allow application of the enzymes.
1. General introduction on Ketosteroid dehydrogenases

Ketosteroid dehydrogenases (KstDs) are bacterial enzymes involved in the degradation pathways of steroids such as cholesterol. The enzymes use a flavin adenine dinucleotide (FAD) cofactor to introduce a double bond into a ketosteroid. In this first part of the introduction we describe steroid catabolism in bacteria, and, in particular, the cholesterol degradation pathway in actinomycetes and the 3D structures of the proteins involved in that pathway. The second part focuses on KstDs and describes both 3D structures and catalysis.

1.1 Steroid catabolism in bacteria

Steroids are organic molecules with a typical structure consisting of four fused rings named A-D (Figure 1). The seventeen-carbon atom ring system consists of three cyclohexane rings (A-C) and one cyclopentane ring (D). The four rings can be substituted on various positions. The type and number of substitutions determine the physiological function of the steroid.

![Figure 1 - The steroid core structure.](image)

Thousands of different steroid compounds occur in nature. They are prevalent in yeast, fungi, insects, plants and mammals, and have a wide range of physiological functions, mainly as signalling molecules. Steroid-based drugs are used as anti-tumor, anti-inflammatory, anti-microbial, anti-viral, anti-fungal, anti-estrogenic, anti-convulsant, or anti-allergy agents in
the treatment of a wide variety of disorders such as cancer, obesity, diabetes, rheumatoid arthritis, hypertension, asthma, eczema, inflammation and metabolic disorders (Donova & Egorova, 2012).

Natural steroid compounds can be found in e.g. animal waste products or decaying leafs. Numerous bacteria have been found to be able to degrade these steroids and to grow on them as the sole carbon source. Currently, the best characterised steroid degrading bacteria include several Rhodococcus and Mycobacterium species (van der Geize et al, 2007; Bragin et al, 2013) and Comamonas testosteroni (Horinouchi et al, 2003). These steroid-degrading bacteria and their proteins are of interest for e.g. the industrial production of bioactive steroid compounds (Fernandes et al, 2003) or the biodegradation of steroid pollutants in wastewater treatment (Yoshimoto et al, 2004; Yam et al, 2011).

Steroid-degrading bacteria may contain several different pathways for the degradation of steroid compounds. For example, Rhodococcus jostii RHA1 contains four gene clusters that putatively code for the degradation of different steroids, including pathways specific for cholesterol and cholate (Mohn et al, 2012). Indeed, the presence of cholesterol or cholate in the environment specifically induces the proteins of their respective gene clusters (van der Geize et al, 2007; Mohn et al, 2012). Nevertheless, it was shown that, besides cholate, the steroid degradation pathway in Comamonas testosteroni can also accept other input steroids, like testosterone, progesterone, epiandrosterone and dehydroepiandrosterone (Horinouchi et al, 2012). Several auxiliary proteins exist that can channel these steroids into the degradation pathway. Finally, a number of isoenzymes exist for key catalytic steps, which can take over each other’s role (Petrusma et al, 2011).

Expression levels of these isoenzymes depend on the steroids on which the bacteria are grown, suggesting a finely tuned system. Thus, steroid degradation emerges as a dynamic and carefully tuned process, in which the pathways can differ through subtle changes in the use of enzymes or the order of the degradation steps.

Steroid degradation in bacteria is well illustrated by the cholesterol degradation pathway of pathogenic actinomycetes like M. tuberculosis (van der Geize et al, 2007) and Rhodococcus equi (van der Geize et al, 2011). These organisms use cholesterol as a source of carbon and energy during the early infection stage and the dormant phase inside the host’s macrophages, where the genes involved in cholesterol degradation are specifically
General introduction on ketosteroid dehydrogenases and ω-transaminases

upregulated (Pandey & Sassetti, 2008; Yam et al, 2009). It was shown that several of the enzymes in the cholesterol degradation pathway are essential for virulence and the survival of *M. tuberculosis* inside the macrophages (Pandey & Sassetti, 2008; Rengarajan et al, 2005). Since humans lack homologs of most of these enzymes, this pathway has recently been suggested as an interesting target for inhibitor development (Ouellet et al, 2011; van der Geize et al, 2011). In addition, strains weakened by deletion of cholesterol degradation pathway enzymes may be used for vaccine development. These potential applications intensified the research effort in cholesterol catabolism, resulting in important new insights as exemplified in the next paragraph.

1.2 The cholesterol degradation pathway

For both *Rhodococcus jostii* RHA1 and *M. tuberculosis* the proteins involved in the cholesterol degradation pathway are encoded in one large gene cluster, which contains 50 or 80 genes, respectively (van der Geize et al, 2007). Although the function of each protein is not yet known, the proteins coded by the gene cluster could be categorised into three main functional groups: proteins involved in sterol uptake, enzymes involved in the elimination of the aliphatic side chain, and enzymes involved in degradation of the steroid core. More detailed, extended homology searches and biochemical characterization of several of the enzymes present in the pathway established their functions, and a cholesterol degradation route has been proposed (Figure 2) (van der Geize et al, 2007; Ouellet et al, 2011; Horinouchi et al, 2012; Garcia et al, 2012; Bragin et al, 2013).

An important step in the degradation of cholesterol is the removal of the aliphatic side chain at the C17 position in a process similar to β-oxidation. Hydroxylation of C26 by a cytochrome P450 enzyme, followed by oxidation to a carboxylate, primes the aliphatic side chain for attack by Coenzyme A (Rosloniec et al, 2009; Ouellet et al, 2010). Subsequently, the side chain is progressively shortened using a distinct acyl coenzyme A synthetase for each step (Casabon et al, 2014). Once the aliphatic side chain has been removed, a hydroxyl group or keto-group is left at the C17 position. How this latter product is exactly formed is still unclear, but it may involve a 3β,17β-hydroxysteroid-dehydrogenase (3β,17β-Hsd) (Bragin et al, 2013).
Chapter 1

Cholesterol

Steroid side chain degradation

AD

Δ1-KstD

ADD

KshAB

9OH-ADD

Non-enzymatic

Primary metabolism

3-HSA

HsaAB

3,4-DHSA

HsaC

HsaEFG

DOHNAA
General introduction on ketosteroid dehydrogenases and ω-transaminases

Figure 2 - Simplified example of the cholesterol degradation pathway in Actinobacteria (van der Geize et al, 2007; Bragin et al, 2013). For each enzyme with known X-ray structure a small picture is shown next to the step it catalyses. The enzymes shown are Δ1-KstD (PDB entry 4C3X (Rohman et al, 2013)), KshAB (PDB entry 2ZYL (Capyk et al, 2009)), HsaAB (PDB entry 3AFF (Dresen et al, 2010)), HsaC (PDB entry 2ZYQ (Yam et al, 2009)), and HsaFG (PDB entry 4JN6 (Carere et al, 2013)). The abbreviations of the compounds are AD, 4-androstene-3,17-dione; ADD, 1,4-androstadiene-3,17-dione; 9OH-ADD, 9α-hydroxy-1,4-androstadiene-3,17-dione; 3-HSA, 3-hydroxy-9,10-seconandrost-1,3,5(10)-triene-9,17-dione; 3,4-DHSA, 3,4-dihydroxy-9,10-seconandrost-1,3,5(10)-triene-9,17-dione; DOHNAA, 9,17-dioxo-1,2,3,4,10,19-hexanorandro-stan-5-oic acid.

After the removal of the aliphatic side chain and the oxidation of the C17 hydroxyl group, the C3 hydroxyl group is oxidized and the double bond is shifted from C5-C6 to C4-C5 (Figure 2). The exact order of steps and the proteins involved this process are still under investigation and may differ between organisms. Two possibilities have been proposed, involving either a cholesterol oxidase or a 3β,17β-Hsd in combination with an isomerase (Bragin et al, 2013). Intriguingly, a double knockout mutant of both cholesterol oxidase and 3β,17β-Hsd in two Mycobacterium species showed continued growth on cholesterol, indicating that, at least in some actinobacteria, other proteins may take over the role of these enzymes (Brzostek et al, 2013). The degradation of the steroid core starts with the desaturation of the steroid A ring by an FAD-dependent 3-ketosteroid-Δ1-dehydrogenase (Δ1-KstD), which introduces a double bond between the C1 and C2 atoms. Subsequently, the 3-ketosteroid-9α-hydroxylase KshAB, a two-subunit FAD Rieske oxygenase, introduces a hydroxyl group in the B ring at the C9 position. This desaturation and hydroxylation of the steroid core are key steps leading to the non-enzymatic aromatization of ring A, followed by the opening of ring B by a non-enzymatic reverse-aldol reaction. Subsequently, the A ring is hydroxylated at the C4 position by HsaAB, a two-component, flavin-dependent monooxygenase, and the A ring is opened by an extradiol dioxygenase, HsaC. The resulting product (4,9-DSHA (see Figure 2)) is then cleaved at the C5-C6 bond by a lyase, HsaD, resulting in two fragments. Whereas it remains unclear how the fragment containing the C and D core rings is further degraded, it has been found that the linear fragment originating from the A and B rings is further degraded in several
enzymatic steps (catalysed by HsaE, HsaF and HsaG) to be fed into the central metabolism (see Figure 2).

Crystal structures have been solved of most of the steroid core degradation enzymes from *Rhodococcus* and *M. tuberculosis*: Δ1-KstD (Rohman *et al*, 2013), KshA (Capyk *et al*, 2009), HsaA (Dresen *et al*, 2010), HsaC (Yam *et al*, 2009), and HsaD (Lack *et al*, 2008). The substrate-bound structure of HsaC allowed analysis of the residues responsible for substrate binding and catalysis, which inspired the screening for and design of improved inhibitors (Yam *et al*, 2009; 2011). HsaD is a C-C bond breaking hydrolase containing a serine protease-like catalytic triad. For this enzyme the structure also provided insights into the interaction of the enzyme with a range of serine protease and esterase inhibitors (Ryan *et al*, 2014). Since HsaD is essential for survival of *M. tuberculosis* in macrophages the enzyme may be a promising target for the design of new anti-tuberculosis drugs. As shown by these examples, structural information is invaluable for analysing the determinants important for binding and inhibition by potential drug lead compounds. For most of the steroid core degradation enzymes a structure is currently available, which may facilitate the development of new drugs against pathogenic *Rhodococcus* and MT.

### 1.3 Ketosteroid dehydrogenases (Δ1-KstD and Δ4-(5α)-KstD)

KstDs are flavoenzymes involved in the degradation of steroids. They catalyse the formation of a double bond in the ketosteroid A ring by the trans-diaxial elimination of two hydrogen atoms. Some of their products, such as the ketosteroids 5α-androstane-3,17-dione (1-(5α)-AD) and 1,4-androstadiene-3,17-dione (ADD), belong to the most marketed intermediates for the production of pharmaceutical steroids (Bragin *et al*, 2013). Two types of KstDs are involved in actinomycetal steroid degradation. 3-Ketosteroid-Δ1-dehydrogenase (Δ1-KstD) desaturates steroids at the C1-C2 (Δ1) position, while 3-ketosteroid-Δ4-(5α)-dehydrogenase introduces a double bond at the C4-C5 (Δ4) position (Figure 3). Although they desaturate ketosteroids at a different position, Δ1-KstD and Δ4-(5α)-KstD are structurally similar and both use FAD as a cofactor.
General introduction on ketosteroid dehydrogenases and ω-transaminases

Figure 3 - The ketosteroid dehydrogenase reactions. The substrate 1-5α-androstene-3,17-dione (5α-AD), can be converted into 5α-androstane-3,17-dione (1-(5α)-AD) or 1,4-androstadiene-3,17-dione (ADD) by the action of Δ1-KstD or Δ4-(5α)-KstD, respectively. The products of Δ1-KstD and Δ4-(5α)-KstD can serve as substrate for the other KstD, yielding 1,4-androstadiene-3,17-dione (ADD) as final product of the combined action of the KstDs. The new double bonds created by the enzymes are shown in red.

The first KstD enzymes were identified in the 1960s (Levy & Talalay, 1959a). Since then they have been found in several actinobacteria like e.g. Arthrobacter (Molnár et al, 1995), Mycobacterium (Wovcha et al, 1979), Rhodococcus (Mori et al, 1998) and Comamonas testosterone (Florin et al, 1996). However, especially since the advent of fast genome sequencing methodologies in recent years, a significant number of (putative) KstD sequences have become available. In 2012, Δ1-KstD homologues had already been identified in 100 bacterial species, while Δ4-(5α)-KstDs occurred less frequently, and were found in only 24 species (Kisiela et al, 2012). Thus, Δ1-KstDs are far more prevalent than Δ4-(5α)-KstDs. It was also found that most organisms contain several genes for Δ1-KstDs, and no or only a single Δ4-(5α)-KstD gene. The broader distribution of Δ1-KstDs is possibly related to their central role in the degradation of the steroid core in, for example, cholesterol degradation. They generate a double bond at the Δ1-
position in the A ring of the ketosteroid, which, when combined with hydroxylation by KshAB, is essential for the subsequent autocatalytic opening of the steroid B ring. In contrast, the function of the rarer Δ4-(5α)-KstDs is less clear. Δ4-(5α)-KstD is not essential for cholesterol catabolism, as the double bond at the Δ4 position can easily be generated through isomerization of the double bond at the Δ5 position. It can be envisaged that Δ4-KstDs have a function in the aromatization of the A ring of steroids that do not contain a Δ5 double bond such as the bile acid cholate (Haußmann et al, 2013), thus allowing a wider variety of steroids to enter the steroid catabolic pathways.

![Figure 4](image.png)

**Figure 4** - The isoalloxazine ring system of the FAD cofactors showing atom numbering of the main rings. The R-group represents the ribityl and the ADP parts of the FAD.

### 1.3.1 The structural properties of KstDs and the role of the FAD cofactor

Crystal structures of two KstDs are currently available, of Δ4-(5α)-KstD from *Rhodococcus jostii* RHA1 (this thesis, (van Oosterwijk et al, 2011; 2012)), and Δ1-KstD from *Rhodococcus erythropolis* SQ1 (Rohman et al, 2012; 2013). Both proteins have the p-hydroxybenzoate hydroxylase (PHBH) fold (Wierenga et al, 1979), a typical flavoenzyme fold (Mattevi, 1998). The KstDs are most closely related to the widely investigated fumarate reductases, with which KstDs share 20–24% sequence identity (Knol et al, 2008). This notable sequence identity allowed the initial identification of their fold type. Proteins with a PHBH fold have a two-domain core structure with the main domain (F-domain) binding the FAD cofactor and a smaller domain (S-domain) binding the substrate. The active site pocket is located in the interface between these domains. The FAD
cofactor binds near this interface, in a long cleft and forms part of the substrate-binding site. FAD is a versatile cofactor, which catalyses either one- or two-electron transfers. Because its special spectral properties (Ghisla et al., 1974) allow following the details of the reaction by spectroscopy, flavoenzymes have become one of the best-studied enzyme classes. The reactive part of the FAD is the isoalloxazine ring system (Figure 4); it contains the reactive N5 atom and the N1-C2=O2 locus important for stabilizing anionic FAD intermediates. While in the fully oxidized state the isoalloxazine ring is aromatic and nearly planar (Lyubimov et al., 2007; Reibenspies et al., 2000), in other oxidation states the isoalloxazine ring can adopt conformations that differ significantly from planarity. The reactivity of FAD is modulated by its environment. For instance, the presence of a positive charge near the N1-C2=O2 locus stabilizes the anionic form of the isoalloxazine ring system (Ghisla & Massey, 1989; Mattevi, 1998). A hydrogen-bonding interaction of the N5 atom with a hydrogen bond donor increases the oxidative power of the cofactor (Fraaije et al., 1999). Finally, the reactive parts of the FAD and the substrate are embedded in a solvent excluded environment. The exclusion of water enhances the strength of polar interactions, which is critical to substrate activation (Mattevi, 1998; Fraaije & Mattevi, 2000). Solvent exclusion is often achieved by closing off the active site by the substrate itself or by the movement of a loop, domain, or amino acid side chain.

1.3.2 The KstD reaction mechanism

Since the discovery of KstDs the mechanistic details of their dehydrogenation reaction have been extensively investigated. Isotope-labeling experiments showed that the dehydrogenation reaction of Δ1-KstDs involves the trans-diaxial loss of two hydrogens and a two-stage mechanism was proposed (Ringold et al., 1963). The two stages consist of abstraction of the first hydrogen by a base and the removal of the second hydrogen by the FAD cofactor. This mechanism shows many similarities to the mechanism of acyl coenzyme A dehydrogenases (Itagaki et al., 1990a).

In the case of the Δ1- and Δ4-(5α)-KstDs the dehydrogenation takes place at the C1-C2 or C4-C5 bond, respectively, which are both adjacent to the steroid C3 keto group. Enolization of the C3 keto group of the steroid is suggested to initiate the reaction. The keto-enol tautomerism is facilitated by
protonation of the keto group by an acid. As a result, the pK\textsubscript{a} of the adjacent α-carbons (C2, C4) is lowered, enabling proton abstraction from C2/C4 by a base and triggering abstraction of the hydrogen atom from C1 or C5 by the FAD cofactor, probably as a hydride ion. As in other flavoenzymes such a mechanism requires that the atom that is to be oxidised is precisely positioned next to the reactive N5 atom of the FAD (Fraaije & Mattevi, 2000).

This proposed mechanism generated considerable debate. In particular, the nature of the catalytic acid and the catalytic base remained contentious. Biochemical experiments suggested the involvement of histidine (Matsushita & Itagaki, 1992), tyrosine (Fujii \textit{et al}, 1999), and several other residues. One of the suggested tyrosine residues (Y121 in the Δ1-KstD of \textit{R. rhodochrous}) is conserved among all Δ1-KstDs and was shown to be essential for catalysis (Knol \textit{et al}, 2008), but its function remained unknown. Moreover, an early phylogenetic analysis of 42 KstD sequences did not lead to a clear identification of active site residues and the differences in the active sites between the Δ1-KstD and Δ4-(5α)-KstD enzymes. Nevertheless, these studies showed that the Δ1-KstD and Δ4-(5α)-KstD enzymes are clearly distinct (Knol \textit{et al}, 2008). Only when the first crystal structures of KstDs became available, the catalytic residues could unambiguously be identified (van Oosterwijk \textit{et al}, 2012; Rohman \textit{et al}, 2013). In both Δ1-KstD and Δ4-(5α)-KstD a tyrosine functions as the acid that promotes the keto-enol tautomerism. In Δ4-(5α)-KstD a serine residue (Ser468) is the catalytic base, but in Δ1-KstD it is a tyrosine residue (Tyr318).

For the next reaction cycle, the FAD cofactor needs to be regenerated. The nature of the electron acceptor remains unclear. Since the KstD enzymes are flavoproteins, quinones and cytochromes are likely candidates for electron transfer (Birkenmaier \textit{et al}, 2007). Molecular oxygen is a poor electron acceptor for the KstD enzymes and only limited activity was found with oxygen (Itagaki \textit{et al}, 1990b). The exact electron acceptor used in the reactions performed by KstDs is not known, although for the KstDs of \textit{Nocardia restrictus} and \textit{Corynebacterium simplex} vitamin K\textsubscript{2} has been suggested as the natural electron acceptor (Abul-Hajj, 1978).
1.4 Objectives of our ketosteroid dehydrogenase research

At the start of the research described in this thesis only limited information was available on ketosteroid dehydrogenases. A few ketosteroid dehydrogenase enzymes were identified as such, and basic biochemical data were reported. The homology with fumarate reductases, which ranges from 20 to 24% sequence identity (Knol et al., 2008), had been recognised. Based on the homology the overall fold of ketosteroid dehydrogenases was predicted to be similar to that of fumarate reductases, although the active site residues were not conserved between the two enzyme classes. Thus, the identity of the active site residues of the KstD enzymes remained unknown, as well as the details of their steroid binding mode and the factors that are responsible for the different reaction specificities of the Δ1- and Δ4-(5α)-KstDs.

Therefore, to provide further insights into the molecular details of the activity and specificity of KstDs, the Δ4-(5α)-KstD of *Rhodococcus jostii* RHA1 was crystallized (chapter 2) and its 3D structure was elucidated, both in its apo- and product-bound form (chapter 3). The product-bound structure shows how the steroid binds in the active site. The structures allow the identification of the residues involved in the dehydrogenation reaction. Surprisingly, the catalytic base residue is not a histidine or tyrosine residue as suggested by previous biochemical experiments, but a serine residue, possibly assisted by a tyrosine side chain.
2. General introduction on ω-transaminases

Transaminases are enzymes that reversibly exchange an amino group on one molecule with a carbonyl oxygen atom on another molecule. They are predominantly enzymes that depend on pyridoxal-5’-phosphate (PLP; Vitamin B₆) as a cofactor. The importance of PLP and PLP-dependent enzymes for the organisms is evident from vitamin B₆ being an essential food supplement for humans. In prokaryotes the PLP cofactor appears equally important since almost 1.5% of bacterial genomes encode putative PLP-dependent enzymes (Percudani & Peracchi, 2003). In the following we will first describe the cofactor, then rationalize the reactions catalysed by PLP-dependent enzymes and briefly describe their different fold types, and then focus on transaminases.

**Figure 5 - Chemical structure of the PLP cofactor.** The oxygen group at C4’ is the reactive part of the PLP cofactor that is involved in Schiff base formation with the catalytic lysine and substrates. The PLP cofactor can be transformed into pyridoxamine (PMP) by an enzyme and an amine donor.

2.1 The PLP cofactor

Vitamin B₆, an inactive precursor of PLP, was first discovered in the 1930s (Toney, 2011a). Its active derivatives, pyridoxal, pyridoxamine, pyridoxal-5’-phosphate (PLP) and pyridoxamine-5’-phosphate (PMP), were subsequently identified in the 1940s. After its discovery, the involvement of PLP in a wide range of enzyme-catalysed reactions was shown (Toney, 2005). All PLP-catalysed reactions follow the same basic chemical mechanism, but the enzyme environment can tune the PLP reactivity and the reaction specificity.
The chemical structure of PLP consists of a core pyridine ring with a methyl group, a phenolate oxygen function, a formyl group and a methyl-phosphate group as side chains (Figure 5). In solution, the phenolate oxygen at C3 is mostly unprotonated and negatively charged. It forms a nine-atom conjugated π-system together with the pyridine ring atoms and the formyl oxygen. This π-system allows the cofactor to act as an electron sink, stabilising negative charges.

The C4’ formyl group is the reactive site of the PLP cofactor. In the resting state of PLP enzymes it is covalently bound to the primary amine of a lysine residue as a result of a Schiff base reaction. This state is called the “internal aldimine”. In addition, the PLP cofactor is firmly anchored in the interior of the protein by up to nine hydrogen bonds to its phosphate group (Jansonius, 1998). Only in glycogen phosphorylase the phosphate group is anchored with less hydrogen bonds, as it plays a direct role in catalysis (Sprang et al, 1991). Another highly conserved interaction is a hydrogen bond of the pyridine nitrogen centre (N1) with an acidic residue. This hydrogen-bonding interaction is essential for transaminase function, as it stabilises the protonated state of N1 (Casasnovas et al, 2009).

### 2.2 Catalysis by PLP-dependent enzymes

PLP-dependent enzymes catalyse a remarkably wide range of reactions, including transamination, decarboxylation, racemization, elimination, and retro-aldol cleavage reactions (Figure 6). To rationalise how a single cofactor can catalyse such different reactions Dunathan proposed that stereo-electronic effects are the major determinants of which reaction takes place (Dunathan, 1966). In all of the above reactions the Lys-PLP bond is broken and a covalent intermediate between the C4’ of the PLP and the substrate is formed (the “external aldimine”). Subsequently, the breakdown of the external aldimine is accomplished in a way that is specific for the bond to be broken.

A protonated pyridine nitrogen centre makes the conjugated system electron withdrawing, with the strongest effect on the σ bond that is aligned parallel to the π molecular orbital system. Therefore, a bond is maximally destabilised when it is perpendicular to the plane of the pyridine ring (Kyte, 1995)). The actual degree of destabilisation depends on the cosine of angle θ (see Figure 7).
Figure 6 - Examples of reactions catalysed by the PLP cofactor (based on Toney (2011b)). Only the main reaction types are shown here. The atoms involved in the reaction are coloured according to their reaction type(s).

For example, in the first step of an α-transaminase reaction a proton is abstracted from the Cα atom. Indeed, in the external aldimine covalent complex of transaminases the Cα-H bond is oriented almost perpendicular to the plane of the PLP ring (Wybenga et al, 2012). In contrast, in amino acid decarboxylases the Cα-COOH bond takes up the perpendicular orientation (Jackson et al, 2003).

A second factor governing the reactivity of the PLP is the protonation state of the pyridine and Schiff base nitrogen atoms of the external aldimine (Toney, 2011b). For example, by favouring protonation of the pyridine
nitrogen atom, the reaction can be tuned towards transamination. Protonation of the pyridine N atom allows the formation of a quinonoid intermediate (see Par. 2.4.3, Figure 10), in which the Cα atom becomes part of the conjugated system and the negative charge formed during the reaction can be delocalized. Moreover, protonation of the pyridine N atom promotes proton transfer to the C4’ atom of PLP (Casasnovas et al, 2009), a reaction which is required to convert the quinonoid into the ketimine intermediate in transaminases (see below). Indeed, transaminases have an aspartate that ensures protonation of the pyridine N atom (Ford et al, 1980). In contrast, racemases prevent protonation of the pyridine nitrogen by having an arginine instead of an aspartate near the pyridine N1 atom (Shaw et al, 1997). As a consequence, no quinonoid intermediate is formed, and the negative charge remains on the Cα atom. This facilitates re-protonation of the Cα atom (from the other side) as required by racemases. Due to the wide range of reactions catalysed by PLP-dependent enzymes such subtle details in the active site govern which of the possible reactions is catalysed, and small changes in the protein can change its reaction specificity.

Figure 7 - The external aldimine between the PLP-cofactor and the ωTA model substrate (S)-α-MBA. The angle important for the degree of destabilization of the bond, C7 to H in this case, is shown.

2.3 PLP-dependent enzymes: folds and families

The PLP-dependent enzymes can be classified based on their amino acid sequences and their 3D-structures. After an initial classification based on sequences alone (Mehta et al, 1993), which led to many subfamilies because of undetected low homologies, Grishin et al. introduced the use of 3D structures as basis to classify PLP-dependent enzymes (Grishin et al, 1995).
Chapter 1

Figure 8 - Representative structures of the four main PLP-dependent-enzyme fold types. The enzymes of the fold types I-IV are all dimeric and their monomers are shown in green and purple. The fold type II enzymes form $\alpha_2\beta_2$ heterodimers and the $\alpha$ subunits are shown in grey.

<table>
<thead>
<tr>
<th>Fold type</th>
<th>Example</th>
<th>pdb-id</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fold type I</td>
<td>Aspartate aminotransferase</td>
<td>7AAT</td>
<td>(McPhalen et al, 1992)</td>
</tr>
<tr>
<td>Fold type II</td>
<td>Tryptophan synthase</td>
<td>1TTQ</td>
<td>(Rhee et al, 1996)</td>
</tr>
<tr>
<td>Fold type III</td>
<td>Alanine racemase</td>
<td>1SFT</td>
<td>(Shaw et al, 1997)</td>
</tr>
<tr>
<td>Fold type IV</td>
<td>D-amino acid transferase</td>
<td>1DDA</td>
<td>(Sugio et al, 1995)</td>
</tr>
</tbody>
</table>

We will use this classification, which was subsequently reviewed by (Jansonius, 1998), who suggested a structural representative for each family. Most of the hundreds of available PLP-enzyme structures can be classified
into four main structural families, named fold types I-IV and three minor fold types (V-VII) (Grishin et al., 1995; Percudani & Peracchi, 2009). The minor fold types have significantly less members and will only be discussed briefly. Table 1 gives an overview of the main fold types, including representative structures; Figure 8 shows representative 3D structures of the four fold types.

2.3.1 Fold type I
Fold type I (or the aspartate aminotransferase fold (Ford et al., 1980)) is the best characterized fold type, as the majority of the available X-ray structures of PLP-dependent enzymes belong to it (Eliot & Kirsch, 2004). The enzymes in this group catalyse the broadest range of reactions, including transamination, decarboxylation and α, β and γ elimination reactions (Percudani & Peracchi, 2003). Fold type I proteins are α/β proteins defined by a core structure consisting of two domains. The PLP-binding domain contains a mixed central β-sheet flanked by α-helices on both sides; the second, often smaller, domain comprises a small antiparallel β-sheet flanked by α-helices. The PLP cofactor binds close to the domain interface. The enzymes belonging to this fold type function as homodimers, where residues from one monomer contribute to the active site of the other monomer and vice versa. In some cases these dimers form higher order oligomers. Many fold type I enzymes can close off the active site from the solvent, for example by domain movement (Eliot & Kirsch, 2004).

2.3.2 Fold type II
Fold type II is mainly found in proteins that catalyse β-elimination reactions (Percudani & Peracchi, 2003). The first elucidated fold type II structure was that of tryptophan synthase (Hyde et al., 1988). It is a homodimer of which the monomers consist of an α and a β subunit. The α subunit consists of a (β/α)8-barrel and does not bind PLP. The β subunit binds PLP and shows the typical type II fold, consisting of two domains of approximately equal size, both containing a mixed α/β structure, with the PLP cofactor bound at the interface between the two domains. The cofactor is bound with its si-face to the solvent similar to the way it is bound in fold type I proteins. Although the
type II fold proteins share a dimeric organisation with the type I fold proteins, the active site contains residues from a single monomer only.

2.3.3 Fold type III

The fold type III enzymes have mostly racemase and decarboxylation activities (Percudani & Peracchi, 2003). The first structurally characterized protein with a type III fold was alanine racemase from \textit{Bacillus stearothermophilus} (Shaw \textit{et al}, 1997). The proteins in this fold group are active as head-to-tail-homodimers (Eliot & Kirsch, 2004), and have a classical TIM ($\beta/\alpha$)$_8$ barrel fold into which a $\beta$-strand domain is inserted. The PLP cofactor is bound at the C-terminal side of the $\beta$-barrel and, although the fold of the subunit is very different, the PLP binding mode is surprisingly similar to that of the other fold types. Besides the similar hydrogen bonds found between PLP and the protein, the PLP phosphate group is also bound close to the N-terminus of an $\alpha$-helix. Like the other fold type families, the fold type III enzymes use a lysine residue to form a Schiff base with the PLP cofactor (Couñago \textit{et al}, 2009).

2.3.4 Fold type IV

Fold type IV is the second fold type that contains enzymes with transamination activity, but also enzymes with lyase activity may have this fold type. Enzymes with this fold type invariably function as homodimers. The first structure of a representative of this fold was that of D-amino acid aminotransferase (Sugio \textit{et al}, 1995). Its active site is at the interface between the two subunits. A long loop from one subunit assists in substrate binding by the other subunit (see section 2.4.3). Each subunit has two domains, which both contain $\alpha/\beta$ and $\alpha+\beta$ elements. The small N-terminal domain consists of a four-stranded antiparallel $\beta$-sheet and two $\alpha$-helices, forming a typical “Greek key” motif. The large C-terminal domain consists of two mixed $\beta$-sheets surrounded by several $\alpha$-helices. Fold type IV enzymes are significantly smaller (~30 kDa) than the fold type I enzymes (~50 kDa). Compared to the fold type I and III enzymes the active site of fold type IV enzymes is mirrored; the active site lysine is located on the opposite side of the PLP (Sugio \textit{et al}, 1995). Moreover, all fold type IV
transaminases with known structure are (R)-selective, whereas the fold type I transaminases with currently known structure are (S)-selective.

2.3.5 Other fold types (V-VII)
Several PLP-binding proteins do not fit in the four fold types discussed above. Glycogen phosphorylase (Sprang et al, 1991) has a C-terminal PLP binding domain with a lactate dehydrogenase fold (Grishin et al, 1995), and is the first representative of fold type V. Fold types VI and VII, represented by D-lysine-5,6-aminomutase (Berkovitch et al, 2004) and lysine-2,3-aminomutase (Lepore et al, 2010), respectively, both have TIM barrel-related folds. In fold type VI PLP binds in a different position on the TIM barrel compared to the fold type III proteins, and the PLP binding lysine is located on a completely different domain. Fold type VII enzymes consist of a (β/α)_6 crescent, reminiscent of a TIM barrel lacking two β/α units. As these three fold types contain significantly fewer members than the first four and are not related to the subject of this thesis, they will not be further discussed here.

Figure 9 - The basic transaminase reaction. The keto group on the acceptor substrate is replaced by an amino group from the donor substrate alanine resulting in the formation of pyruvate and an (R)- or (S)-amine, depending on the enantioselectivity of the used transaminase. For ωTAs it is not necessary that the acceptor substrate contains a carboxylate group.

2.4 ω-Transaminases
Transaminases are classified according to the position of the amino group that is transferred. For example, an α-transaminase (αTA) performs its reaction at the α-carbon of an amino acid. On the other hand, in the case of
the ω-transaminases the use of ω in the name is less specific, as it refers to all non-α-positions (Malik et al., 2012a). Thus, an ω-transaminase (ωTA) may catalyse the transfer of any other amino group than the amino group at the α-carbon from a donor substrate to an acceptor molecule, replacing a carbonyl group on the acceptor (Figure 9). In contrast to α-transaminases, ωTAs do not necessarily require a carboxylate group in their substrates. In the following paragraphs we will discuss several aspects of ω-transaminases, which, according to the B6 database (http://bioinformatics.unipr.it/B6db), are all members of either fold type I or fold type IV (Percudani & Peracchi, 2009). We will describe the substrate-binding site, the arginine switch, their catalytic mechanism and the structural determinants governing enantioselectivity.

2.4.1 The substrate-binding sites of ω-transaminases consist of two pockets

Because all ω-transaminases act on secondary amines, they require a two-pocket substrate binding site, one for the R1 moiety and one for the R2 moiety (see Figure 9). For instance, in case of alanine as a substrate one binding pocket binds the carboxylate and the other binding pocket binds the methyl group of Ala. Originally, the two pockets were named the large (L) and the small (S) pocket (see (Shin & Kim, 2002)), where, in the case of alanine, the L pocket binds the carboxylate group and the S pocket binds the methyl group. Since in a number of transaminases the size of the carboxylate-binding pocket is smaller than the size of the amino acid side chain binding pocket (e.g. in transaminases acting on aromatic amino acids), the L/S nomenclature becomes ambiguous, and a new nomenclature was proposed (Wybenga et al., 2012). This new nomenclature is based on the proximity of the pockets to either the PLP 3’-oxygen (O pocket, previously L) or to the PLP phosphate group (P pocket, previously S). As explained above, for maximal destabilization of the scissile bond it should be oriented perpendicularly to the plane of the PLP ring. As a consequence, the R1 and R2 side chains will be approximately in the plane of the PLP ring, and will be bound in pockets near either the PLP 3’-oxygen or the PLP phosphate group. This holds for all transaminases, and thus the O- and P-pocket nomenclature is universal and independent of the fold type.

The chemical properties and shape of these pockets determine the substrate-binding mode. By providing specific binding interactions with the substrate
in both the O- and P-pocket, the enzyme ensures the formation of the correct product. A less specifically bound substrate might result in the formation of an alternative product, e.g. the wrong enantiomer. For example, reverse binding of the carboxyl and methyl groups of pyruvate (Figure 9) would lead to the formation of the opposite alanine enantiomer.

2.4.2 The arginine switch

Transaminases act on two different substrates, the amine donor and the amine acceptor, which are often chemically very different (e.g. amino acids with a hydrophobic or acidic side chain). Since they utilize the same active site for both substrates, a certain degree of adaptability is required, while substrate binding should still remain specific. Transaminases employ various strategies to achieve this dual substrate recognition (Hirotsu et al., 2005). One of these mechanisms is the “arginine switch”, which is often employed by ω-transaminases. For example, many fold type I transaminases can bind the model substrate (S)-α-methylbenzylamine (MBA) and its co-substrate alanine in the active site. To accommodate two radically different side groups, the hydrophobic phenyl group of α-MBA and the charged carboxylate of alanine, the O-pocket has two conformations, with the side chain of an arginine residue either in the “in” or in the “out” position. When the arginine is in the out position, the O-pocket is more spacious and lined with several hydrophobic residues, which is highly suitable to bind an apolar compound. When the arginine is in the in position its guanidinium group can bind the negatively charged carboxylate group of a substrate, and the space for an apolar side chain is limited. The role of the arginine switch in ω-transaminases has been affirmed by mutational studies (Steffen-Munsberg et al., 2012a).

On the other hand, in fold type IV transaminases only the in position of the active site arginine side chain has been observed, interacting with the carboxylate of a substrate analogue (Sugio et al., 1995; Peisach et al., 1998). Although this arginine is conserved in another structurally characterized fold type IV enzyme (Thomsen et al., 2014), and its mutation led to reduced activity with pyruvate as a substrate (Łyskowski et al., 2014), the absence of clear evidence for an out position questions whether this arginine does have a switching role. Moreover, the arginine is positioned in a long loop that is inserted into the active site of the other monomer in the dimer. This loop is
only weakly conserved, it varies considerably in size, and adopts diverse conformations (Łyskowski et al., 2014). Whether this conformational diversity is related to a wide range of substrate specificities or to the lack of a catalytic role of this loop is not clear at this point. Further research is needed to fully understand the role of this loop in fold type IV transaminases and to establish whether an arginine switch is operative in these proteins.

Figure 10 - The transaminase reaction. Here the half-reaction for the reversible deamination of the ωTA model substrate (S)-α-MBA to acetophenone is presented. The PLP cofactor is converted into PMP, which is used in the second half reaction, regenerating PLP.

2.4.3 The transaminase reaction mechanism

Transaminases abstract an amine group from one substrate to transfer it to a second substrate. The reversible reaction consists of two half reactions, that proceed via a similar catalytic mechanism but with a reversed order of the reaction steps. In the first half reaction the amino donor and the lysine-bound PLP (the internal aldimine) are converted to the keto product and PMP. In the second half reaction an amino acceptor is used to regenerate the covalent bond between the active site lysine and the PLP co-factor from the PMP state.

The enzyme and PLP go through various states during the reaction. The starting state of the first half reaction is the internal aldimine, while at the
end of the first half reaction the cofactor is in the PMP form. Three distinct intermediates can be observed during this half reaction, the external aldime, the quinonoid and the ketimine intermediates. In this paragraph we will discuss each state individually according to the reaction scheme shown in Figure 10, where the reversible de-amination of an aromatic model substrate is shown. In the internal aldime state (Figure 10 - 1) the active site lysine residue forms a Schiff base with the C4’ atom of the PLP. The Schiff base nitrogen atom is accessible from the active site pocket and can react with an incoming substrate. A nucleophilic attack by the substrate’s amine group on the internal aldime results in a Schiff base exchange and formation of the external aldime (Figure 10 - 2), a state that all PLP-dependent enzymes have in common. Formation of the external aldime results in an uncharged active site lysine. It can now act as a base and deprotonate the C7 carbon atom, leading to the formation of the quinonoid intermediate (Figure 10 - 3). This intermediate is usually short-lived, but it is stabilised by delocalisation of its negative charge over the π system. In this reaction step the hybridisation of the C7 atom changes from sp$^3$ to sp$^2$, making it a key step in the enantio-specificity of the protein (see section 2.4.4). In the following step the N$\zeta$ group of the lysine slightly moves to reprotoate the C4’ carbon atom of the PLP (1,3-prototropic shift), leading to the formation of the ketimine intermediate (Figure 10 - 4). Next, a water molecule activated by the active site lysine hydrolyses the ketimine intermediate. Finally, an internal rearrangement results in the formation of the aromatic keto product, leaving the amine group on the PMP cofactor. The transaminase is now ready for the second half reaction, where the second substrate undergoes the same reaction steps but in reversed order.

2.4.4 Enantioselectivity: (R)- and (S)-selective transaminases

Chiral amines are an important class of chemical building blocks, which often contribute a stereogenic center to pharmaceutically active substances, agrochemicals and other fine chemicals (Höhne & Bornscheuer, 2009). There are various ways of obtaining enantiopure amines using transaminases. For instance, chiral amines can be chemically produced as a racemic mixture, after which the unwanted enantiomer is removed using a transaminase. Such kinetic resolution has a maximum theoretical yield of 50% when assuming a fully enantioselective enzyme and no regeneration
system for the produced ketone. Alternatively, chiral amines can be directly synthesised with transaminases, which, again assuming a completely enantioselective enzyme, has a 100% yield. Although such asymmetric synthesis is more efficient, it is not yet the standard method of choice, because transaminases are rarely fully enantioselective towards the substrate of interest. Formation of the unwanted enantiomer is often a serious drawback, in particular for applications requiring an enantiopure product. In contrast, in kinetic resolution lack of enantioselectivity is not a big problem, since it only results in conversion of some of the desired enantiomer into the ketone. Although this lowers the yield, it does not affect the enantiopurity of the final product. Finally, transaminase-catalysed reactions usually favour the ketone product, thus limiting the yield of asymmetric synthesis. Kinetic resolution does not suffer from this problem; in fact, it fully exploits this property by converting the unwanted enantiomer into the ketone. Thus, although asymmetric synthesis is inherently more efficient, its application is not straightforward and requires substantial further development. In addition to improving protein stability, improved enantioselectivity and co-substrate regeneration are important challenges for the biotechnological optimization of these enzyme systems.

Currently, a large effort is devoted towards the creation of a toolbox containing a broad range of well-characterised oTAs with respect to their substrate- and enantioselectivity. From this toolbox specific enzymes can be selected for the production of chiral amines or for optimisation of these enzymes towards new substrates. Most of the transaminases currently available in this toolbox have been classified as (S)-selective enzymes (see e.g. (Koszelewski et al, 2010b; Steffen-Munsberg et al, 2012b)), on the basis of test substrates like (R)- and (S)-α-MBA, in combination with e.g. pyruvate as the amino acceptor. More extensive characterization on a much more diverse range of substrates is required for these enzymes, as well as a much deeper understanding of the molecular features responsible for their enantioselectivity. Such understanding will be key to optimizing these enzymes for specific substrates and allowing rational design strategies to improve their substrate- and enantioselectivity.

The three-dimensional organization of the transaminase active site, as e.g. visualised by protein X-ray crystallography, is an important determinant of its enantioselectivity. The two main factors governing enantioselectivity are the position of the (lysine) catalytic base and the specificity of the O- and P-
General introduction on ketosteroid dehydrogenases and \(\omega\)-transaminases

pockets. In the following two paragraphs we will discuss these two factors and compare them for the fold types I and IV transaminases.
The first factor is the position of the base with respect to the external aldime and quinonoid intermediate. In the interconversion of these two intermediates the C7 atom switches between a possibly chiral \(sp^3\) hybridised configuration and a non-chiral \(sp^2\) hybridised configuration (see Figure 10).

As the catalytic lysine side chain is the proton donor/acceptor in this step, the position of the lysine with respect to the pro-R/S side of the C7 atom in the external aldime is essential for enantioselectivity. Interestingly, the location of the catalytic lysine in the currently known fold type IV \((R)\)- and fold type I \((S)\)-selective transaminases is on opposite sides of the PLP ring system, making this difference a good indicator for the enantioselectivity of the enzymes. Most transaminases are not completely enantioselective towards all desired substrates. In addition to alternative binding modes of the substrate (see below), it can be envisaged that an alternative base at the opposite side of the intermediate may protonate/deprotonate the intermediate, resulting in the opposite enantiomer. Such an alternative base could be an amino acid side chain as e.g. observed for alanine racemase (Sun & Toney, 1999) or an activated water molecule as e.g. seen for aspartate aminotransferase (Kochhar & Christen, 1992) or the lysine-activated water molecule suggested in chapter 4. Binding of such a water molecule may be facilitated by the use of unnatural substrates. Optimisation of an alternative base may be an effective strategy for reversing the enantioselectivity of a transaminase.

The second determinant of substrate specificity and enantioselectivity is provided by the O- and P-pockets. A high specificity of the O- and P-pockets is important for the enantioselectivity of transaminases since it prevents binding of the substrate in the “opposite/wrong” orientation. In principle, an \((S)\)-selective transaminase may be engineered to catalyse \((R)\)-selective reactions by swapping the specificity of the O- and P-pockets. However, when comparing the currently available \((R)\)- and \((S)\)-selective \(\omega\)TA structures no such a swap of pocket specificities has yet been observed. Therefore, it is generally thought that swapping the specificities of the O- and the P-pockets is not a productive approach to change the enantioselectivity of a transaminase. Instead, changing the substrate-specificity of confirmed \((R)\)- or \((S)\)-selective enzymes is more often undertaken.
Enantioselectivity is a crucial factor for the industrial application of ωTAs. To enhance the use of ωTAs in industry, a much better understanding of the molecular determinants of this property is required. Structural investigations are key to improving such understanding, and may pave the way for increased application of ωTAs as outlined in the next section.

2.5 Structures of synthetically relevant ωTAs

The use of ω-transaminases in biotechnology affords significant environmental benefits compared to chemical synthesis, such as milder reaction conditions and a reduced use of organic solvents. In addition, (R)- and (S)-selective ω-transaminases facilitate the production of chiral amines, which are especially important building blocks for pharmaceutical compounds. In view of the biotechnological interest in transaminases, investigations into the structure-function relationships of transaminases have been intensified in recent years. A number of crystal structures of synthetically relevant fold type I (S)-selective ωTAs have been published recently. A first example is the structure of CV-ωTA from Chromobacterium violaceum, a transaminase with a relatively broad substrate range, which makes it attractive for use in industry (Sayer et al., 2007; Humble et al., 2012; Sayer et al., 2013). This structure was followed by the structure of PA-ωTA from P. aeruginosa (Sayer et al., 2013), an enzyme used in the synthesis of amino-alcohols (Ingram et al., 2006). These structures have provided insights into the structural rearrangements occurring upon PLP binding (Humble et al., 2012), and the molecular determinants of their amine product specificities (Sayer et al., 2013). PD-ωTA from Paracoccus denitrificans (Rausch et al., 2013) catalyses among others the conversion of the pharmaceutically important non-natural amino acid L-homoalanine from L-threonine (Park et al., 2010), but it is also active towards aromatic substrates. Its structure explained the enzyme’s strong preference for binding aromatic groups in the O-pocket. The P-pocket of PD-ωTA is unique in accepting groups larger than the typical ethyl group, and mutations based on the X-ray structure increased the activity on even bulkier substituents, up to an n-hexyl group (Park et al., 2014).

The first available structure of a fold type I β-amino acid transaminase was that of a Mesorhizobium sp. strain LUK aminotransferase (Wybenga et al., 2012). This enzyme is used for the synthesis of enantiopure β-phenylalanine,
a building block for the anticancer drug taxol (Kim et al., 2007). It can accommodate both α- and β-amino acids in its active site. Surprisingly, the binding of carboxylate and side chain is reversed for α- and β-amino acid substrates: for α−amino acids the carboxylate group binds in the O-pocket, whereas for β-amino acids it binds in the P-pocket. Although strictly taken the β-amino acid transaminases are defined as ωTAs, the differences in their active site, as well as the requirement for the presence of a carboxylate in the substrate, make them a special group of enzymes.

In addition to the structures described above, several putative ornithine-aminotransferase-like proteins, structurally characterized in structural genomics projects, were shown to be PLP-dependent and to have a fold type I fold. These enzymes were subsequently assayed against a wide range of amine donors and acceptors, showing them to have (S)-selective ω-transaminase activity with differing substrate profiles (Steffen-Munsberg et al., 2012b). This latter property makes them valuable additions to the ωTA toolbox. These results illustrate the value of structural genomics projects in providing new candidate enzymes for biotechnologically relevant reactions.

Recently, fold type IV (R)-selective transaminases were characterised that open up access to the (R)-enantiomers of industrially interesting compounds. An impressive illustration is the optimization of the (R)-selective transaminase ATA-117 for the production of the anti-diabetic sitagliptin to replace rhodium-catalysed asymmetric enamine hydrogenation (Savile et al., 2010). To use ATA-117 for the industrial process, the enzyme had to be stabilized to optimally function under the harsh reaction conditions required to keep the substrate soluble, and its active site had to be adapted to fit a completely new substrate. These goals were achieved by a combination of iterative structure-based rational design with directed evolution. Initially, a homology model was created based on an (R)-selective D-amino acid aminotransferase, which was the closest related structure at that time. Based on this model positions were selected for saturation mutagenesis, which was combined with random mutagenesis. The final ATA-117 variant contains a total of 27 mutations, and is now used in the final synthetic step of the industrial manufacturing of sitagliptin.

At the time of writing, only a limited number of structures of (R)-selective ωTAs have been solved, which is partly due to the low availability of sequences of confirmed (R)-selective ωTAs. One of the first structures of an (R)-selective ωTA was that of an Aspergillus terreus enzyme (Łyskowski et
Investigation of the active site by site-directed mutagenesis suggested a dual substrate-binding mechanism similar to that found in (S)-selective transaminases (see section 2.4.2). Other recently determined structures of (R)-selective transaminases include structures of the ωTAs from Aspergillus fumigatus (Thomsen et al, 2014) and Nectria haematococca (Sayer et al, 2014), the latter in complex with the inhibitor gabaculine. Gabaculine (5-aminocyclohexa-1,3-diene-1-carboxylic acid) is a typical suicide inhibitor, first discovered for the γ-aminobutyric acid (GABA) transaminase (Kobayashi et al, 1976). It reacts with PLP to form an external aldime, which, after deprotonation, is converted to m-carboxyphenylpyridoxamine phosphate, an extremely stable aromatic compound (Rando, 1977). In complex with the enzyme the phenyl ring of the gabaculine does not occupy the O- and P-pockets as most secondary amine substrates. Therefore, the complex cannot be used to analyse the specificities of the pockets. Instead, modelling and docking were used to suggest which residues are responsible for the enantio- and regioselectivity of these (R)-selective transaminases.

Although more than twenty structures of (R)- and (S)-selective ωTAs are available, none of them have a substrate or substrate analogue bound in their active site, with the exception of the structure of Mesorhizobium sp. strain LUK aminotransferase (Wybenga et al, 2012). The only ωTA crystal structures with a bound inhibitor contain gabaculine. Structures with better substrate analogues or actual substrates are needed to further analyse the structural determinants of enantioselectivity and substrate specificity.

2.6 Transaminase development and the role of X-ray structures herein

In view of the importance of chiral amines for biotechnology (Koszelewski et al, 2010b; Kohls et al, 2014; Tufvesson et al, 2011) interest in enantioselective transaminases has substantially increased in recent years. Transaminases allow the production of chiral amine-containing bioactive compounds, which include anti-depressants, herbicides and antibacterials (Koszelewski et al, 2008; Höhne & Bornscheuer, 2009; Koszelewski et al, 2010b). Exciting examples exist in the literature showing that the application of ωTAs for the synthesis of bioactive compounds or synthetic precursors is within reach. For instance, the total synthesis of (S)-rivastigmine, one of the most potent drugs for the treatment in the early stages of Alzheimer disease,
has been accomplished using (S)-selective ωTAs (Fuchs et al., 2010). ωTAs are also applied in the production of halogenated aromatic chiral amines, which are building blocks for potent potassium channel openers with application in the treatment of e.g. hyperparathyroidism (Malik et al., 2012b). Furthermore, an (R)-selective ωTA has been developed for application in the biocatalytic production of the anti-diabetic sitagliptin (Savile et al., 2010); the same ωTA was also used to synthesize 17-amino-steroids which are interesting intermediates for use in the synthesis of biologically active steroid derivatives (Richter et al., 2014).

From these examples it transpires that specific ω-transaminases are needed for each specific application. To make these ω-transaminases suitable for large-scale application in industry requires optimisation of their enantioselectivity for their particular substrate, as well as their adaptation to the temperature, pH, and presence of (organic) solvents during the reaction. Since such optimisation can be time-consuming, there is a need to start with ω-transaminases that are already adapted as much as possible to the appropriate substrate, conversion type and reaction conditions. Thus, discovery and characterization of new ωTAs are on-going projects worldwide. Furthermore, availability of 3D structures of the enzymes of interest would facilitate rational engineering of the enzymes to improve their enantioselectivity and other kinetic and thermodynamic properties. Especially 3D structures of ωTAs with substrates or substrate analogues would allow more accurate insights into the structural determinants of enantioselectivity and the residues involved in substrate recognition. Such insights will allow the rational mutagenesis of residues in the active site, which may permit optimization of these proteins to convert new, industrially interesting substrates.