Characterization and engineering of thermophilic aldolases

-synthesizing Nitrogen-heterocycles in biosynthetic routes-

Promotoren: Prof. Dr. John van der Oost
                               Persoonlijk hoogleraar bij het Laboratorium voor Microbiologie
                               Wageningen Universiteit

                               Prof. Dr. Willem M. de Vos
                               Hoogleraar in de microbiologie
                               Wageningen Universiteit

Promotiecommissie: Prof. Dr. Gerrit Eggink
                               Wageningen Universiteit

                               Prof. Dr. Wim Quax
                               Rijksuniversiteit Groningen

                               Dr. Willem J.H. van Berkel
                               Wageningen Universiteit

                               Dr. Theo Sonke
                               DSM, Geleen

Characterization and engineering of thermophilic aldolases

-synthesizing Nitrogen-heterocycles in biosynthetic routes-

Suzanne Wolterink-van Loo

Contents

Chapter 1

Introduction

During the last decade, a large number of enzymes has been used in biocatalytic processes by pharmaceutical and agrochemical companies. Key aspects in biocatalysis are selectivity and chirality, which are necessary to obtain a high yield of a specific product. As enzymes are generally enantio- and or regioselective, they are potentially good catalysts. Furthermore, they can catalyze reactions in a mild environment, and as such are a "green" alternative to chemical reactions. The enzymatic synthesis of agro- or pharma compounds, also known as white biotechnology, has to compete with established chemical processes. Therefore, enzymes or biocatalytic processes often need to be optimized before biotechnological application can be realized.

1.1 Classification of aldolases

In the production of fine chemicals, carbon-carbon bond formation is among the key steps. Aldolases are one of the few classes of enzymes that can catalyze this carbon-carbon bond formation. Generally aldolases have evolved to catalyze the reversible condensation of a ketone (or aldehyde) donor and an aldehyde acceptor (Fig. 1). Some aldolases are involved in degradation of metabolites (sugar degradation), whereas others are used in synthesis of a carbon backbone (amino acid synthesis). Two classes of aldolase enzymes can be distinguished based on their mechanisms. In class I aldolases, a Schiff-base intermediate is formed between a strictly conserved lysine and the donor-substrate. The activated enamine attacks the acceptor aldehyde, resulting in an enzyme-bound product that is released by hydrolysis (Dean *et al.* 2007, Fullerton *et al.* 2006). Class II aldolases have a metal cofactor (usually Zn^{2+}) that acts as a Lewis acid in activating the bound donor substrate (Fig. 2) (Dean *et al.* 2007, Hall *et al.* 2002). Both classes of aldolases are rather specific for the donor substrate, but can accept different acceptor substrates. This flexibility offers potential for application purposes, as unnatural aldehydes can replace the natural ones during condensation reactions. Another important aspect is that during aldol condensation up to two new chiral carbon atoms can be formed (Fig. 1). In combination with the stereospecificity of enzymes, this should lead to the production of pure stereoisomers (Williams *et al.* 2006).
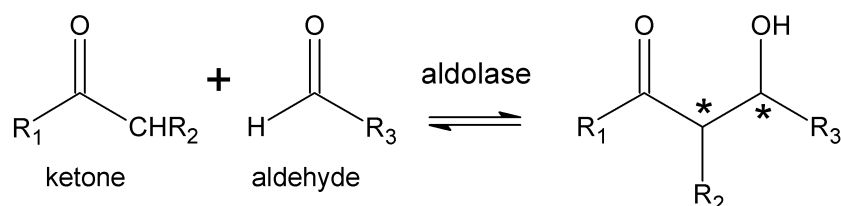
Figure 1. The aldol condensation reaction. The asterix indicates chiral carbon atoms (except when R_2 includes an H).
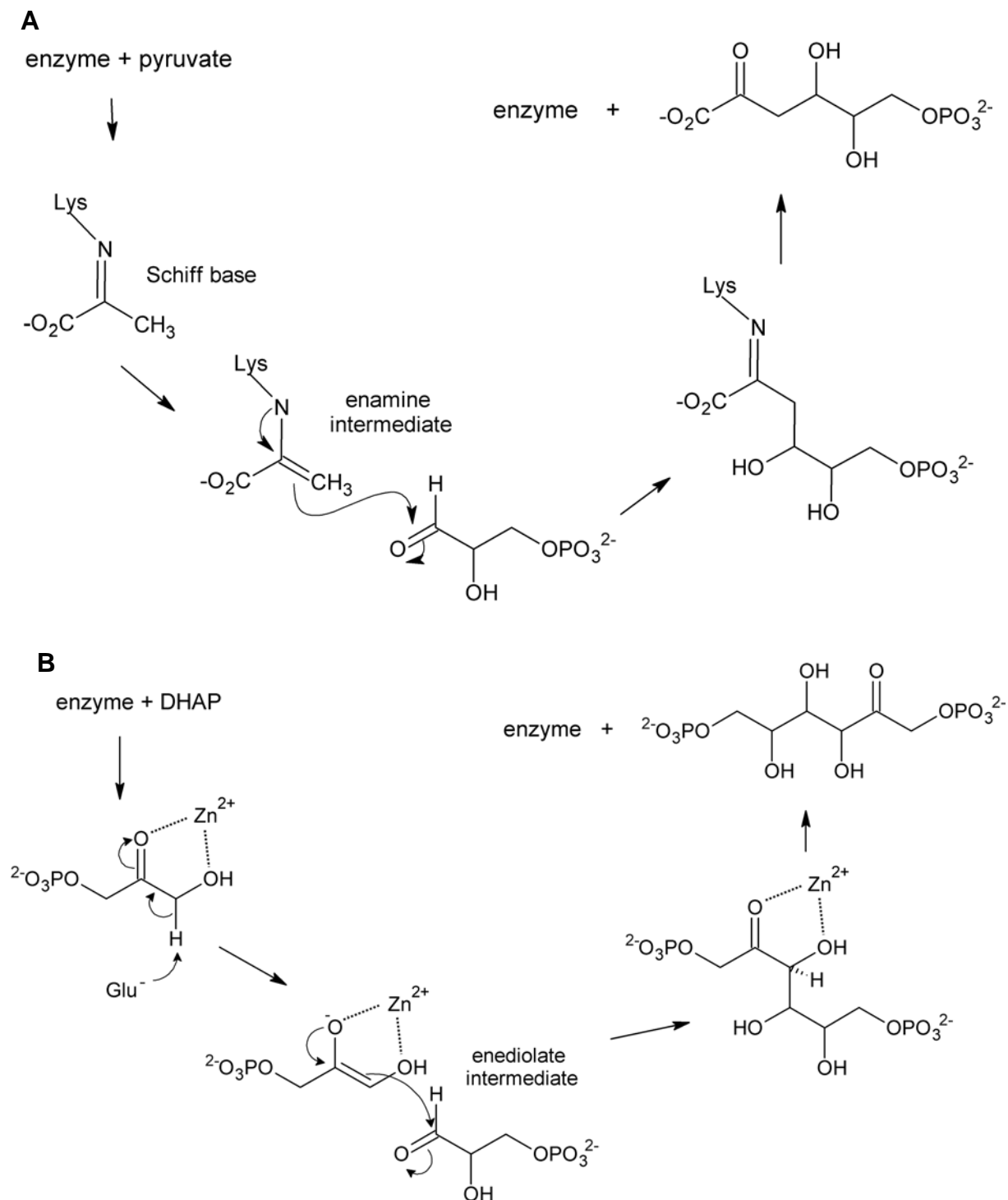
Figure 2. Examples of the two different aldolase mechanisms. **A)** Class I, demonstrated for KDPG aldolase (Fullerton *et al.* 2006). **B)** Class II, demonstrated for fructose and tagatose 1,6-bisphosphate aldolase (Hall *et al.* 2002).

Most aldolases are classified as carbon-carbon lyases, subgroup aldehyde lyases (E.C. 4.1.2.x), whereas some are present in the class of transferases (E.C. 2.1.x/2.5.x). The aldolases that have been investigated for use in the above mentioned synthesis reactions, can be divided into four main groups based on the nature of their donor substrate and include dihydroxyacetone phosphate (DHAP)-, acetaldehyde-, glycine-, and (phosphoenol)-pyruvate (PEP)-dependent aldolases (Bolt *et al.* 2008).

DHAP-dependent aldolases

Two different DHAP-dependent aldolases (fructose 1,6-bisphosphate aldolase (FBPA) (EC 4.1.2.13) and tagatose 1,6-bisphosphate aldolase (TBPA) (EC 4.1.2.40), respectively) catalyze the reversible asymmetric aldol addition of DHAP to D-glyceraldehyde 3-phosphate (D-GAP), leading to the different stereoisomers fructose 1,6-bisphosphate (FBP) and tagatose 1,6-bisphosphate (TBP) (Fig. 3). FBPA activity has been measured in species of all domains of life (eukarya, bacteria and archaea). Both class I and class II FBPAs have been characterized, e.g. some microbial FBPAs such as the class II FBPA from *Mycobacterium tuberculosis* (Ramsaywak *et al.* 2004), two archaeal class I FBPAs (Siebers *et al.* 2001) and the class I and class II FBPA in *Synechocystis* species (Nakahara *et al.* 2003). TBPA has been found as class II aldolases in bacteria, for example in *Escherichia coli* (Hall *et al.* 2002) and in several other enteric bacteria (Brinkkotter *et al.* 2002). To date, FBPA has been the most extensively studied (Samland and Sprenger 2006) and is commercially available from different resources (rabbit, spinach, *E. coli, Staphylococcus carnosus*). Rabbit muscle FBPA (RAMA) has been used in the stereoselective synthesis of monosaccharides and sugar analogs with (3*S*,4*R*) configuration (Azéma *et al.* 2000).

Two other DHAP-dependent aldolases (fuculose 1-phosphate aldolase (EC 4.1.2.17) and rhamnulose 1-phosphate aldolase (EC 4.1.2.19)) share L-lactaldehyde as acceptor aldehyde and also have different stereochemistry, leading to fuculose 1-phosphate and rhamnulose 1-phosphate (Fig. 3). The *E. coli* fuculose 1-phosphate aldolase has been used in a multi-enzyme system to catalyze carbon-carbon bond formation (Sanchez-Moreno *et al.* 2004).
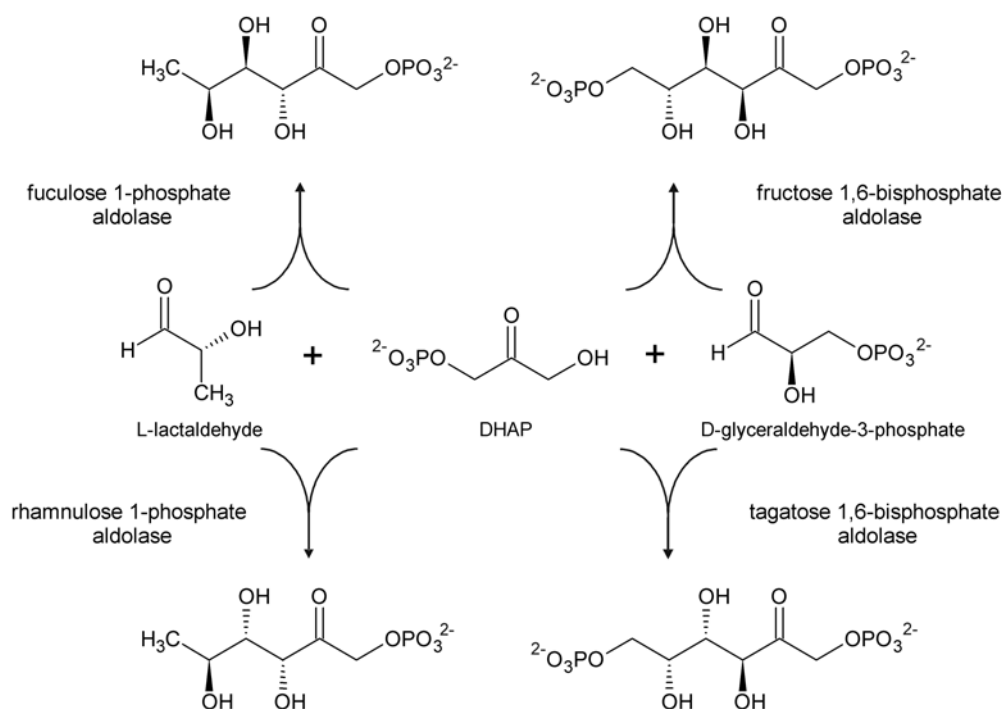
Figure 3. Reactions of four different DHAP-dependent aldolases

2-Deoxyribose-5-phosphate aldolase

In vivo 2-deoxyribose-5-phosphate aldolase (DERA, EC 4.1.2.4) catalyzes the reversible aldol reaction of acetaldehyde and D-glyceraldehyde 3-phosphate to form 2-deoxyribose-5-phosphate (Machajewski and Wong 2000) in the pentose phosphate pathway. DERA is a class I aldolase and has, with acetaldehyde as the donor, an interesting feature. The products from the DERA-catalyzed reaction are also aldehydes and are capable of becoming acceptor substrates for a second aldol condensation. Such sequential aldol reactions were first observed for DERA itself (Liu *et al.* 2004). The *E. coli* DERA has been subjected to directed evolution studies for enzymatic synthesis of a chiral precursor for statin drugs used for reducing cholesterol levels (Jennewein *et al.* 2006).

Glycine-dependent aldolases

The glycine-dependent aldolases have pyridoxal-5-phosphate as a cofactor and catalyze the reversible aldol reaction of glycine with an aldehyde acceptor to form a β-hydroxy-α-amino acid. Two classes of glycine-dependent aldolases have been found, the serine hydroxymethyltransferases (SHMT, EC 2.1.2.1) and the threonine aldolases (EC 4.1.2.5/B1), both involved in amino acid metabolism (glycine – serine – threonine) (Machajewski and Wong 2000). An L-threonine aldolase was used to prepare protected 3R,5R-dihydroxy-L-homoproline as a mimetic of idulonic acid (Miura and Kajimoto 2001).

Pyruvate and phosphoenolpyruvate-dependent aldolases

Pyruvate-dependent aldolases either have anabolic functions *in vivo* (biosynthesis) or play a role in catabolism of metabolites. When applying the right conditions they can be used to prepare similar products *in vitro*. Both class I and class II aldolases are found in this group.

A large number of different pyruvate-dependent aldolases belong to the N-acetylneuraminate lyase (NAL) subfamily, typical class I aldolases. Most of these enzymes are (homo)tetramers (Lawrence *et al.* 1997, Pearce *et al.* 2006, Theodossis *et al.* 2004), with each monomer having the $(\alpha/\beta)_8$ fold. Conservation of both their fold and their active site lysine residues indicates that they are evolutionary related. The archetype of this subfamily is N-acetylneuraminic acid (NeuAc) aldolase (EC 4.1.3.3), also named sialic acid aldolase. This enzyme is involved in aminosugar metabolism and catalyzes the reversible conversion of N-acetylneuraminate $\rightleftharpoons$ N-acetyl-D-mannosamine + pyruvate. This enzyme has shown to be remarkably flexible towards its aldehyde substrate (Takayama *et al.* 1997) and it has been used in the synthesis of sialic acid mimetics (Williams *et al.* 2005). Other members belonging to this subfamily include hydroxybenzylidenepyruvate hydrolase-aldolase (HBPHA), *trans*-2′-carboxybenzalpyruvate hydratase-aldolase (CBPHA) (Iwabuchi and Harayama 1998), D-5-keto-4-deoxyglucarate dehydratase (KDGDH), 2-keto-3-deoxyoctanoate (KDO) aldolase, dihydrodipicolinate synthase (DHDPS) and 2-keto-3-deoxygluconate aldolase (KDGA) (Barbosa *et al.* 2000, Lawrence *et al.* 1997).

Dihydrodipicolinate synthase (EC 4.2.1.52) catalyzes the conversion of pyruvate + (*S*)-aspartate-β-semialdehyde ((*S*)-ASA) $\rightleftharpoons$ 2,3-dihydrodipicolinate as part of the lysine synthesis pathway of bacteria and plants. The *Escherichia coli* DHDPS has been extensively characterized (Blickling *et al.* 1997b, Dobson *et al.* 2004b, Karsten 1997). The enzyme has a very high substrate specificity towards its aldehyde substrate, (*S*)-ASA. Furthermore a number of plant DHDPSs have been characterized (Kumpaisal *et al.* 1987, Mazelis *et al.* 1977). In plants and in some bacteria, this enzyme is inhibited by the end product of the pathway, lysine. Therefore this enzyme is also engineered to allow for the overproduction of lysine (Kennerknecht *et al.* 2003, Tsujimoto *et al.* 2006). An other reason to investigate DHDPS is because of the fact that it is a potential target of antibacterial compounds (Mitsakos *et al.* 2008). The structures of a variety of DHDPSs has been solved (*E.coli, Thermotoga maritima, Nicotiana sylvestris, Bacillus anthracis* and *Mycobacterium tuberculosis* (Blagova *et al.* 2006, Blickling *et al.* 1997a, Blickling *et al.* 1997b, Dobson *et al.* 2005, Dobson *et al.* 2004c, Kefala *et al.* 2008, Mirwaldt *et al.* 1995, Pearce *et al.* 2006).

The KDGA catalyzes the reversible conversion D,L-2-keto-3-deoxygluconate (KDG) $\rightleftharpoons$ pyruvate + D,L-glyceraldehyde (GA) (Buchanan *et al.* 1999). Also this enzyme appeared very flexible in accepting aldehyde substrates as it accepts D- and L-GA as well as GAP

(Ahmed *et al.* 2005). When condensation experiments were performed with pyruvate and D-GA, both D-KDG and D-2-keto-3-deoxygalactonate (KDGal) were identified as products (Lamble *et al.* 2005b). This rather promiscuous aldolase is part of the semi-phosphorylated Entner-Doudoroff pathway in archaea (Ahmed *et al.* 2005). The structures of *S. solfataricus* and *S. acidocaldarius* are known and have been investigated in relation to their promiscuity (Theodossis *et al.* 2004) / chapter 3).

The structurally unrelated 2-keto-3-deoxy-6-phosphogluconate aldolase (KDPGA, EC 4.1.2.21) catalyzes the reversible conversion D,L-2-keto-3-deoxy-6-phosphogluconate $\rightleftharpoons$ pyruvate + D-GAP (Shelton *et al.* 1996). This enzyme accepts short-chain, non-carbohydrate electrophilic aldehydes as substrates with a preference for phosphorylated substrates (Cotterill *et al.* 1998, Shelton *et al.* 1996). The quaternary structure of KDPGA is a trimer, based on crystal structures of *E. coli* and *Pseudomonas putida* (Allard *et al.* 2001, Bell *et al.* 2003).

1.2 Creating the perfect aldolase in the laboratory

In creating a perfect aldolase for application in biocatalysis, there are some hurdles to be taken. First of all an aldolase with some of the desired activity should be selected. With (partial or complete) genome sequences of many species being freely available today (4603 (July'08) - www.ncbi.nlm.nih.gov), it is an easy approach to search for new possible aldolases using BLAST search with a known aldolase as query sequence. Another approach is to make use of metagenomic libraries (Cowan *et al.* 2004), i.e. screening DNA fragments from environmental libraries, since there are still many locations (which hold unknown organisms) to be explored. For stability and long shelf life, enzymes from thermostable origin were explored. Another hurdle is the requirement for sufficient amounts of (active) protein. For this purpose, the genes coding for aldolases are generally heterologously expressed in a host system (see 1.2.1). Subsequently enzymes can be analyzed, and if necessary be optimized for the desired activity by changing the coding sequence in a rational or random fashion, using computational design (see 1.2.2) and laboratory evolution (see 1.2.3). A final issue is the industrial applicability of the enzyme and for efficient biocatalysis immobilization of the enzyme is required to separate product and enzyme, while it also allows for reuse of the enzyme.

1.2.1 Heterologous expression

Heterologous expression of proteins from prokaryotes is usually established in a bacterial host. Bacteria are less suitable for heterologous expression of eukaryotic proteins, as bacteria often do not have the capacity to perform the required posttranslational modifications. *E. coli* is by far the most frequently used host for protein production. It is relatively cheap and easy to

grow, and production is rather fast. Moreover multiple distinct expression systems are available.

Table 1. *E. coli* expression vectors

Vector	Promoter	Inducer	Selection marker	Copy number	Tag	Host requirements
pUC	p_{lac}*	IPTG*	Amp^R	60-180	-	*lac* repressor (LacI) on chromosome and/or plasmid
pMAL[1]	p_{tac}	IPTG	Amp^R	~20	N-MBP	*lac* repressor (LacI) on chromosome and/or plasmid
pGEX[2]	p_{tac}	IPTG	Amp^R	~20	N-GST	*lac* repressor (LacI) on chromosome and/or plasmid
pRSET[3x]	p_{T7}	IPTG*	Amp^R	>200	N-His	T7 RNAP (p_{lac}*) (e.g. BL21(DE3)
pET[4]	p_{T7}	IPTG*	Amp^R Kan^R	~40	N-His C-His N-Trx N-Nus N-Strep	T7 RNAP (p_{lac}*) (e.g. BL21(DE3)) ; *lac* repressor (lacI) on chromosome and/or plasmid
pET[4]	p_{T7}	L-arabinose**	Amp^R Kan^R	~40	N-His C-His N-Trx N-Nus N-Strep	T7 RNAP (p_{BAD}**) (e.g. BL21 AI); *ara* activator (AraC) and *lac* repressor (LacI) on chromosome and/or plasmid
pBAD[3]	p_{BAD}**	L-arabinose**	Amp^R Kan^R	~15 or >600	N-His N-MBP N-TrxA N-GST N-NusA	*ara* activator (AraC) on chromosome and/or plasmid
pQE[5]	p_{T5}	IPTG	Amp^R	~20-30	N-His C-His	*lac* repressor (lacI) on chromosome and/or plasmid (pREP4)

[1] New England Biolabs
[2] GE life sciences
[3] Invitrogen
[4] Novagen
[5] Qiagen
[x] also a set available with fluorescent tags

In Table 1, an overview is provided of some commonly used *E. coli* expression vectors. An origin of replication in the plasmid is needed to maintain sufficient copies in the cell, as well as a selection marker (usually antibiotic resistance) to prevent loss of the plasmid. For controlled expression, inducible promoters are used, for example the bacteriophage T7 promoter. Transcription of the gene is accomplished by T7 RNA polymerase, which is encoded by the genome (integration of a phage genome) of the host (λDE3). Isopropyl-β-D-thiogalactopyranoside (IPTG) can be used to induce the T7 RNA polymerase gene, which is controlled by the lacUV5 promoter. Often a cytoplasmic localization is desired, but some plasmids do have signal peptide sequences that direct the protein to the periplasm (e.g., PelB, DsbA and DsbC (available in pET) and gIII (available in pBAD)). Furthermore, a number of affinity tags were developed for ease of purification. Proteins with a poly-histidine tag (often six) can be easily purified by nickel affinity chromatography. Other tags were designed to increase solubility of the protein, for example Thioredoxin (Trx) and N-utilization substance A (NusA). Glutathione S-transferase (GST), Maltose binding protein (MBP) and streptavidin tag (Strep) can increase solubility, but can also be used for affinity chromatography using glutathione-, maltose- or biotin-matrices as column material, respectively. Commercial expression vectors are often further optimized (and published as such), for example to visualize expression by adding fluorescent or colored tags (Kaper *et al.* 2008, Park *et al.* 2003), or to increase solubility by using different fusion proteins (Cabrita *et al.* 2006, Kwon *et al.* 2005).

When archaeal proteins are expressed in *E. coli*, a problem emerges. Archaea often have low GC levels, resulting in different codon usage compared to *E. coli*. When the right tRNAs are not available in the cell, translation can stall resulting in truncated or misfolded protein (Sorensen *et al.* 2003). To allow for the heterologous expression of genes with deviating codon usage, several additional vectors were designed that contain extra copies of some *E. coli* tRNAs in order to prevent stalling of protein synthesis. Plasmids encoding those tRNAs are pRIL (Stratagene - encoding tRNA of arginine (AGA, AGG), isoleucine (AUA) and leucine (CUA) (chloramphenicol resistant)), pSJS1240 (encoding tRNA of arginine (AGA,AGG) and isoleucine (AUA) (Kim et al 1998)), pSJS1244 (encoding tRNA of arginine (AGA/AGG), isoleucine (AUA) and lysine (AAG) (Kim *et al.* 1998, Sorensen *et al.* 2003) (both spectinomycin resistant)) and pLysRARE (Novagen - encoding tRNA of arginine (AGG, AGA), isoleucine (AUA), leucine (CUA), proline (CCC) and glycine (GGA) on a chloramphenicol-resistant plasmid).

1.2.2 Rational design – optimizing enzymes

Rational design aims at specific, targeted alterations of amino acid residues, usually focusing on the catalytic site. For altering protein functions with rational design, knowledge of the protein (and gene) sequence is needed, as well as information on the mechanism. Nowadays, in the rational (re)design of enzymes, mutations are predicted by computer programs (Looger and Hellinga 2001, Saraf *et al.* 2006) for a review see Lippow and Tidor (2007). A three dimensional structure of the enzyme is required, preferably with ligands, inhibitors or intermediates present in the active site (Shao and Arnold 1996). It is not yet completely clear which determinants need to be included in the prediction of certain enzyme activity (Antikainen and Martin 2005), whereas rational design of binding has been very successfully already (Dwyer *et al.* 2003, Looger *et al.* 2003), although some progress has been made in redesign and de novo design of enzymes (Jiang *et al.* 2008, Lassila *et al.* 2005). In general mutations suggested by computational design are introduced to the gene by site-directed mutagenesis using different strategies (see below).

Site-directed mutagenesis techniques

Three techniques that are used to introduce mutations in a gene are described below and represented in Figure 4. One technique that is often used is overlap extension polymerase chain reaction (PCR) (Ho *et al.* 1989) (Fig. 4A). In this technique, two primers in the gene and two flanking primers are used. In a PCR the first half of the gene is amplified (using primer F1 and G2), incorporating a mutation at the 3' end. A second PCR (using primer F2 and G1) amplifies the 3'-end of the gene, introducing the complementary mutation. A third PCR, combining the obtained PCR-products and using flanking primers amplifies the whole gene. This method allows for multiple mutations to be build in one gene, using several PCR reactions. The original gene / plasmid should be removed by agarose gel electrophoresis and the desired products should be purified subsequently.

In the second method, a complete plasmid is amplified using PCR and two primers containing the desired mutation (Fig. 4B). The parental (methylated) DNA is destroyed by a methyl-dependent restriction enzyme (DpnI), resulting in a circular plasmid with remaining nicks. After transformation the breaks in the dsDNA are repaired by the host cell machinery. This whole-plasmid PCR method is commercialized by Stratagene (QuikChange method). This method is quite straightforward, and if restriction of the parental DNA is complete no wild types will be found. Unfortunately this technique does not work well with large plasmids (>10kb) and typically one amino acid change can be accomplished at a time (Antikainen and Martin 2005).
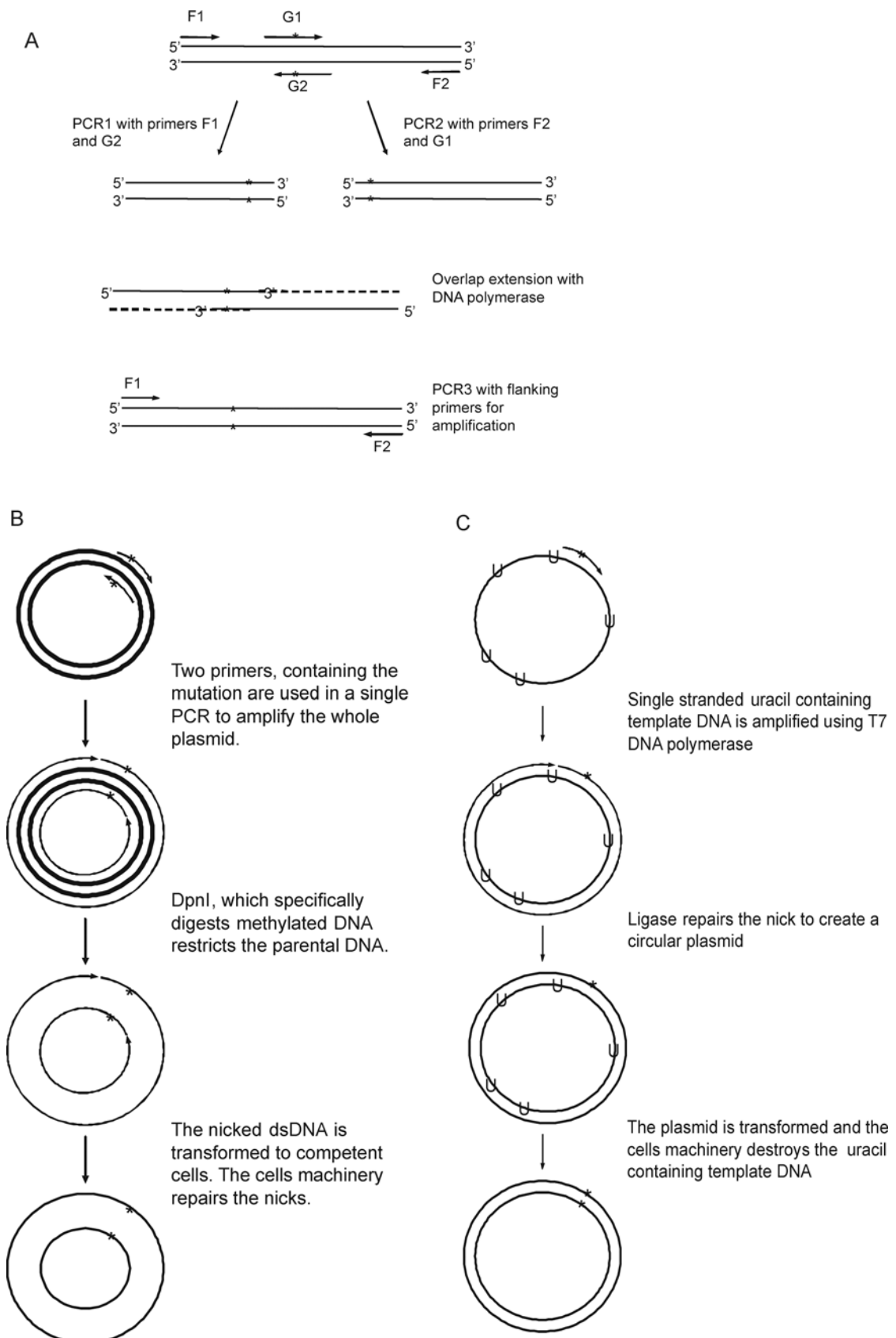
Figure 4. Site directed mutagenesis. Primer are represented by arrows, site of mutation by *. **A)** Overlap extension method (modified from (Antikainen and Martin 2005)) **B)** Whole plasmid PCR (modified from (Antikainen and Martin 2005)) **C)** Kunkel method (Kunkel *et al.* 1991)

A third method, referred to as the Kunkel method (Kunkel *et al.* 1991) starts with the synthesis of single stranded template DNA using an *E. coli* strain that incorporates uracil residues in stead of thymine. A single primer with the desired mutation(s) anneals, and DNA is subsequently extended using T7 DNA polymerase. DNA-ligase is used to create the circular plasmid. After transformation (into usual expression strains) the uracil-containing template DNA is destroyed by the cells machinery, and the complementary strand is generated by the cells DNA polymerase (Fig. 4C). This method is little more complicated, but one can introduce multiple mutations at once.

1.2.3 Laboratory evolution – optimizing enzymes

Laboratory evolution (also referred to as directed evolution), can literally be regarded as the laboratory variant of evolution. In nature, evolution generates a pool of variants by random mutations through single nucleotide alterations (point mutations) and exchange of larger DNA fragments (recombination), followed by selection of the 'fittest' variant (Jaeger *et al.* 2001).
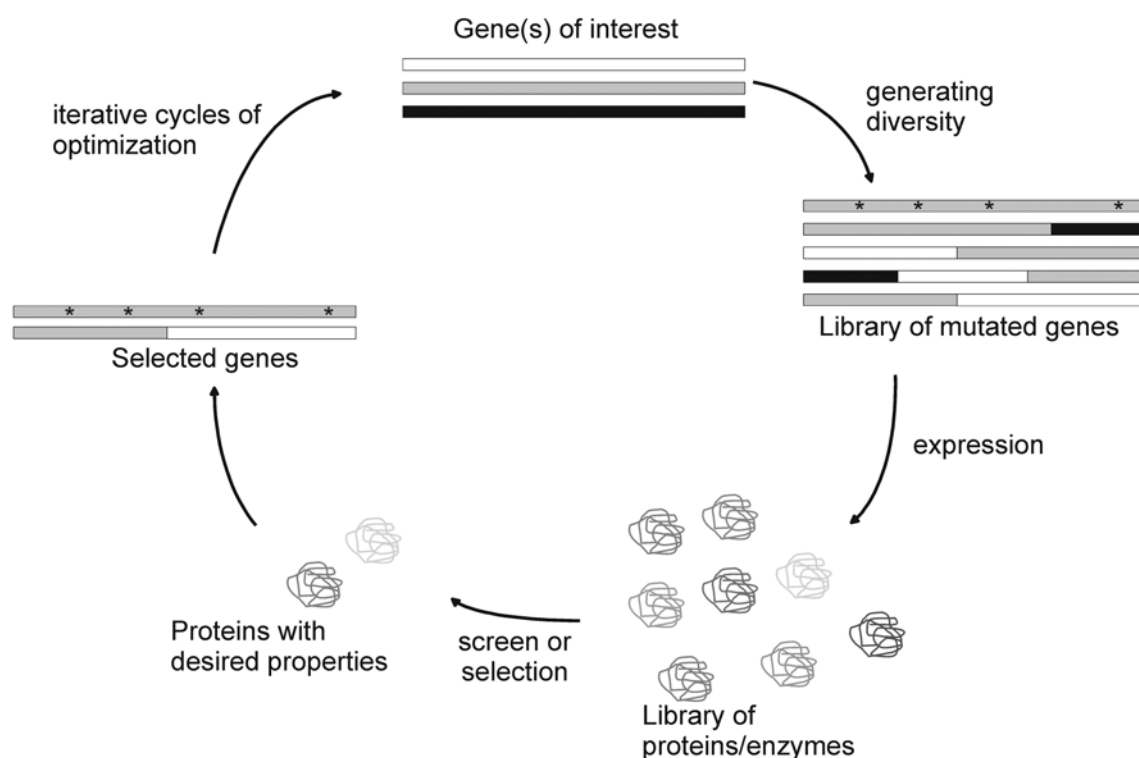
Figure 5. A schematic overview of the laboratory evolution concept. Diversity is created in a (set of) gene(s) – (with sometimes selection for fragment length). After expression in a high throughput fashion, the proteins are screened or selected for the desired properties. The whole process of creating diversity and screening / selection can be repeated several times for refinement.

By using modern molecular biology methods for mutagenesis, recombination, and selection, this process can be mimicked *in vitro*. Molecular diversity is generated by random mutagenesis (mimicking asexual evolution), homologous and heterologous DNA recombination/shuffling (resembling sexual evolution and horizontal gene transfer respectively). Next a screening or selection is used to enrich for the mutated proteins that have the desired beneficial mutations from the pool of variants (Fig. 5). In nature this pressure is imposed by the environmental conditions, in the lab it is the scientist who creates these conditions. For these optimization techniques, there is no need for a priori knowledge of the protein. However, one requires either efficient (high-throughput functionality) screening facilities, or a smart selection process. This is the bottle-neck of the laboratory evolution approach, since this has to be developed specifically for each optimization experiment (Boersma *et al.* 2007).

Creating variant-libraries

In the past decades a range of techniques has been developed to generate diversity in genes. The most common way to generate a collection (library) of genetically altered enzyme variants is error-prone PCR (epPCR) developed by Leung *et al.* (1989). In this PCR reaction, random mutations are introduced in the target gene by a DNA polymerase, that lacks proof-reading activity, therefore introducing mutations at a specific rate (Valetti and Gilardi 2004). A frequently used enzyme is *Taq* polymerase in skewed conditions, and commercially available systems such as mutazyme (Stratagene). Recently Machielsen *et al.* (2008) improved an alcohol dehydrogenase from *P. furiosus* for the production of (2S, 5S)-hexandiol at 30°C. Furthermore, Gratz and Jose (2008) introduced an upgraded version of the epPCR; protein domain library generated by overlap extension (PDGLO). In this variant method, requiring at least some insight in the encoded protein, epPCR is applied to introduce random mutations in only one part of the gene (avoiding mutation of a signal sequence), the other part is amplified by (high fidelity) PCR and both fragments are combined using an overlap extension protocol.

A breakthrough in directed evolution was made by Stemmer *et al.* with the development of DNA shuffling (Stemmer 1994a, Stemmer 1994b) as an alternative way to create diversity. This technique starts with a purified PCR product of the gene(s) to be engineered and its subsequent digestion with DNaseI to obtain random fragments. Next reassembly of those fragments (30-100 bp fragments) is achieved by a primerless PCR, followed by PCR with gene flanking primers and selection of fragments with the correct gene size. Zhao and Arnold (1997) made a high fidelity protocol to recombine related genes *in vitro*, minimizing the amount of point mutations by using other DNA-polymerases (from *Pfu*

or *Pwo* instead of the low fidelity *Taq* polymerase)*,* and digesting the plasmid DNA directly (without PCR). In general DNA shuffling can be performed with a pool of related genes (derived from nature or epPCR). This technique requires high homology of the DNA (>60% identity), because overlapping fragments have to anneal in the PCR. The relatively simple Staggered Extension Protocol (StEP) was developed by Zhao *et al.* (1998), in which the different DNA templates are mixed in one PCR-tube (with primers) and amplified using very short cycles of annealing and extension. Switching of the growing products to different templates and further extension creates shuffled genes, thus also requiring high DNA sequence homology. Li *et al.* (2007) used both DNA-shuffling and StEP on nitrite oxido-reductase for effective degradation of inorganic nitrogen.

Ostermeier *et al.* (1999) developed a method by which combinatorial gene fusion libraries are generated in a manner that is independent of DNA homology: incremental truncation for the creation of hybrid enzymes (ITCHY). Two genes can be recombined after generation of N- and C-terminal libraries of those genes by progressive truncation, using exonuclease III. Recombination by (simple) ligation results in the single cross-over library. The ITCHY protocol was also combined with shuffling (SCRATCHY, (Lutz *et al.* 2001)), to increase the amount of cross-overs to three, and thereby increasing the number of combinations possible. Another method that can create libraries of single cross-over hybrids of unrelated or distantly related proteins is Sequence Homology−Independent Protein RECombination (SHIPREC) (Sieber *et al.* 2001). Here, a fusion is made of two genes, including a restriction site. This fusion is subsequently fragmented, with the creation of blunt ends. Fragments with the correct size are ligated to create a circular DNA. This is then restricted at the initial fusion site, to create the library with a single cross-over distributed over the entire length. On the other hand, when using RACHITT (Random CHImeragenesis on Transient Templates) (Coco *et al.* 2001) one is able to create a very large number of cross-overs (14 at average) per chimeric gene, with a high percentage of active clones. This method is based one parental ssDNA-template, containing uracil, assembly of trimmed (several) parental genes (fragments of 45-200nt) and filling the gaps using the uracil containing template. As the template is degraded, this results in only chimeric genes.

Of these techniques of creating variants of a specific (set of) gene(s), epPCR and DNA shuffling are used most frequently, although protocols are often optimized by the creating scientist. Very often, epPCR is applied at first, and the best mutants are subsequently recombined by DNA-shuffling, for example to improve an alpha-amylase for activity at low pH (Jones *et al.* 2008) or enhancing thermostability of a beta-agarase (Shi *et al.* 2008). Techniques that (can) recombine distantly related genes are used less frequently and apparently do not easily give variants with desired properties.

Bridging the gap, "random design"

To explore the available sequence space as much as possible, some techniques were developed that generate diversity in a random fashion, but with rational behind the technique. In other words, methods that require some insight in structure and function relations.

Degenerate Oligonucleotide Gene Shuffling (DOGS) (Gibbs *et al.* 2001) is a procedure for gene shuffling that uses perfectly complementary degenerate primer (CDE) pairs. Based on an alignment of different genes, degenerate nested primers –with a non-degenerate core- are designed, and each fragment is amplified separately in a PCR. The non-degenerate core of individual CDE primers is generally based upon the corresponding coding sequence of one gene designated as the parental gene for shuffling. This results in the formation of chimeric fragments that retain parental sequence at the points of segment overlap. It enhances the potential sites of crossing over, and as such leads to increased levels of recombination between the parental genes. Bergquist *et al.* (2005) combined this technique with an additional technique, random drift mutagenesis (RNDM), introducing random point mutations. RNDM combines random mutagenesis with selection for retained activity.

O'Maille *et al.* (2002) developed a semi-rational protein engineering approach that uses information from the protein structure. On the basis of the tertiary structure, sites of recombination are selected. In a series of PCR reactions with hybrid oligonucleotides that code for variable connections between structural elements, libraries of chimeric genes with up to five crossovers are constructed. Also for Sequence-independent Site-directed Chimeragenesis (SISDC) (Hiraga and Arnold 2003) sequence information is necessary of the genes that are to be recombined. Nucleotide sequences of the parent genes are aligned, and consensus sequences are determined. Marker tags –containing endonuclease sites, present at 5' and 3' and a different endonuclease site in the middle– are inserted into targeted sites. Sticky ends are produced with the first endonuclease, after which mixed fragments from different parental genes are ligated in the right order. The second endonuclease is now used to remove any untreated tag. Cross-over sites were identified using the computational algorithm SCHEMA, which uses structural information to predict polypeptide elements that can be swapped among related proteins with minimal disruption. With synthetic shuffling (Ness *et al.* 2002) a high recombination frequency is obtained. It uses partially degenerate oligos and spiking oligos (to be able to include non-coded amino acids) of 15 different genes (subtilisin). In a PCR they are reassembled in complete genes, with the possibility of each amino acid to recombine independently. Another method, Assembly of Designed Oligos (ADO) (Zha *et al.* 2003), relies on sequence information to design oligonucleotides coding for the sequences of two or more parental genes with appropriate degeneracy at given positions; controlled

overlaps can be introduced in conserved regions and a desired degree of random codon mutations, insertion or deletion further ensures the exploration of sequence space while retaining a high percentage of active variants. If the homology level is too low to allow such overlap, a ligation method previously used on ssDNA with unknown terminal sequence for "Random Insertion and Deletion" (RID) mutagenesis was adapted to extend the oligos.

Using saturation mutagenesis one can change a potential interesting amino acid into every other amino acid, resulting in small libraries (32 different codon possibilities, using NNK degeneracy). Reetz *et al.* (2005) developed a combinatorial active-site saturation test (CAST). Based on tertiary and secondary structure of a protein sequence, two residues that are close to each other in a protein's 3D structure, are mutated using saturation mutagenesis to be able to find synergistic mutations. Libraries do get a lot larger with ~1000 possible combinations (32*32). Hidalgo *et al.* (2008) developed a PCR technique using long, spiked oligonucleotides, which allow randomising of one or several cassettes in any given position of a gene, One-pot Simple methodology for CAssette Randomization and Recombination (OSCARR). A mutated oligonucleotide and a flanking primer is used to perform a PCR. The PCR is finalized by introducing the other flanking primer (and additional different parental DNA for recombination) and additional rounds of PCR to amplify the whole product. This method was demonstrated using oligonucleotides, which introduce a stop-codon for the ease of screening.

Screen vs selection

After having generated the variation in proteins, a suitable assay should be applied, ideally capable of rapidly isolating valuable variants from a large mutant library. Usually, the number of protein variants that can be screened is the limit. Laboratory evolution experiments are therefore often limited by the availability of a suitable high-throughput screening or selection system (Boersma *et al.* 2007). A very important rule in laboratory evolution is that the screening (or selection) method results in a protein with the screened (selected) activity and/or selectivity: "you get what you screen for". Hence, the screening substrate should be the one of interest, or at least resemble that as close as possible. Finally, a sensitive high-throughput assay is needed, that covers the correct range.

In a screening method, the desired product of an enzymatic reaction is usually measured directly or indirectly. Screening methods often rely on the use of fluorogenic or chromogenic substrates, which are converted into spectroscopically different products (Aharoni *et al.* 2005, Otten and Quax 2005, Wahler and Reymond 2001). If a spectroscopic determination of product (or substrate) is not directly possible, even more laborious assays

need to be developed. A drawback of screening is that every variant from a pool is tested, including inactive and possibly wild type enzymes.

Selection strategies are based on the necessity of the gene(product) for its host, resulting in the exclusive survival of the desired variants; this implies that variants of no interest are never seen. As a drawback selection can be solely applied to enzymatic activity that is of major importance to the host cell. However, more general selection methods have been developed to overcome this problem. Baker *et al.* (2002) managed to link activity to transcription of an essential gene using a Yeast three-hybrid system. A heterodimeric small molecule –the substrate– bridges a DNA-binding domain–receptor fusion protein and an activation domain–receptor fusion protein, activating transcription of a downstream reporter gene (LacZ) *in vivo*. The functional enzyme separates the two functional domains and transcription of the gene is no longer activated. Adaptations to such a system are possible, aiming to couple enzyme activity to induction of an essential gene or repression of a toxic gene. Van(van Sint Fiet *et al.* 2006) used microbial chemosensors to identify better catalysts. The desired product binds to a regulator resulting in expression of a gene with clear screenable or selectable phenotype (LacZ and tet^R). By adapting the ligand-binding site of the regulator, this method can be used for different interesting products.

1.3 Aim and outline

This thesis is the result of a project within the Integration Biosynthesis Organic Synthesis (IBOS) program. Four academic groups (Rutjes-Nijmegen, Kieboom-Leiden, Franssen-Wageningen, Van der Oost-Wageningen) and an industrial group (DSM-Geleen) have aimed for an integrated approach to synthesize nitrogen heterocycles from aldehydes via biocatalytic cascades (Fig. 6).

Two different routes have been investigated. In one route, carbon-carbon bonding is accomplished by using (enantioselective) hydroxynitrile lyases (HNL). In the other route, the project described in this thesis, carbon-carbon bonding is catalyzed by aldolases. In a separate IBOS project, oxidases are explored that have the capacity to convert alcohols into (functionalized) aldehydes. Both the nitrile and aldolase route can potentially form cascades in combination with these oxidases.

The main objectives of the research of this thesis were (a) identification of potential aldolases in genome databases, (b) production and characterization of those enzymes, and (c) optimization of aldolases in a random and/or rational fashion for their use as industrial biocatalyst. In our search for enzymes we focused on thermophilic organisms as their enzymes usually have a very high stability (even in organic solvents) and possibly long shelf life.
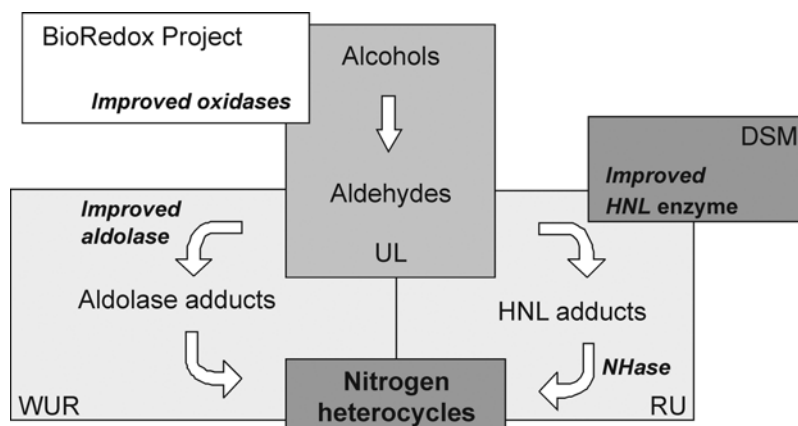
Figure 6. Schematic overview of the IBOS-project: Synthesis of nitrogen heterocycles via biocatalytic routes.

Outline

Chapter 1 gives an overview and classification of the different aldolases. Special focus is given to microbial aldolases that have potential in biotechnological applications. Furthermore several laboratory techniques are described, necessary for the characterization and optimization of enzyme functions. The different possible enzymatic routes / cascades towards (bio)synthesis of Nitrogen heterocycles are indicated.

Chapter 2 describes the production and characterization of the dihydrodipicolinate synthase of the thermophilic bacterium *Thermoanaerobacter tengcongensis*. Enzyme kinetics were determined at 60°C, as close as possible to *in vivo* conditions. A structural model of the enzyme has been generated for comparison of the active site, as well as the lysine binding site that is involved in feedback inhibition in the orthologous enzyme from *E. coli*.

Chapter 3 reports on the extensive characterization of 2-keto-3-deoxygluconate aldolases (KDGAs) originating from 3 *Sulfolobus* species. Apart from determination of enzyme kinetics, the aldehyde substrate range of the enzymes was explored. Furthermore, the 3D-structure of *Sulfolobus acidocaldarius* KDGA was solved and compared with structures of other pyruvate-dependent aldolases.

Chapter 4 presents the random optimization of the *Sulfolobus acidocaldarius* 2-keto-3-deoxygluconate aldolase. A general screening was established, based on the conversion of the pyruvate substrate. A library generated with error-prone PCR was screened for increased activity at 50°C, 40° below its optimal temperature.

Chapter 5 describes the engineering of the *Sulfolobus acidocaldarius* 2-keto-3-deoxygluconate aldolase using computational design, towards the nitrogen-containing substrate 2-azidoacetaldehyde. Furthermore, product characterization of natural and unnatural (aldol) condensation reactions has been performed, focusing on the separation of the different (stereo)isomers. This has allowed for the characterization of the stereoselectivity of SsoKDGA, SacKDGA and SacKDGA mutants.

Chapter 6 summarizes the results of the experimental chapters. A general discussion and implication of the results is given.

Chapter 2

Characterization of a thermostable dihydrodipicolinate synthase from *Thermoanaerobacter tengcongensis*

Suzanne Wolterink-van Loo, Mark Levisson, Maud C. Cabrières, Maurice C.R. Franssen and John van der Oost

Abstract

Dihydrodipicolinate synthase (DHDPS) catalyzes the first reaction of the (*S*)-lysine biosynthesis pathway in bacteria and plants. The hypothetical gene for DHDPS (*dapA*) of *Thermoanaerobacter tengcongensis* was found in a cluster containing several genes of the diaminopimelate lysine – synthesis pathway. The *dapA* gene was cloned in *Escherichia coli*, DHDPS was subsequently produced and purified to homogeneity. The *T. tengcongensis* DHDPS was found to be thermostable ($T_{0.5} = 3$ h at 90°C). The specific condensation of pyruvate and (*S*)-aspartate-β-semialdehyde was catalyzed optimally at 80°C at pH 8.0. Enzyme kinetics were determined at 60°C, as close as possible to *in vivo* conditions. The established kinetic parameters were in the same range as for example *E. coli* DHDPS. The specific activity of the *T. tengcongensis* DHDPS was relatively high even at 30°C. Like most DHDPSs known at present, the DHDPS of *T. tengcongensis* seems to be a tetramer. A structural model reveals that the active site is well conserved. The binding site of the allosteric inhibitor lysine appears not to be conserved, which agrees with the fact that the DHDPS of *T. tengcongensis* is not inhibited by lysine under physiological conditions.

Introduction

Dihydrodipicolinate synthase (DHDPS) (EC 4.2.1.52) is the key enzyme in lysine biosynthesis via the diaminopimelate pathway, which is used by bacteria (Blagova *et al.* 2006, Cremer *et al.* 1988, Shedlarski and Gilvarg 1970, Tsujimoto *et al.* 2006), some phycomycetes and higher plants (Dereppe *et al.* 1992, Frisch *et al.* 1991, Ghislain *et al.* 1995, Vauterin and Jacobs 1994). The enzyme catalyzes the condensation of (*S*)-aspartate-β-semialdehyde ((*S*)-ASA) and pyruvate to 2,3-dihydrodipicolinate (DHDP) (Yugari and Gilvarg 1965) (Fig. 1). Blickling *et al.* (1997b) reported (4*S*)-4-hydroxy-2,3,4,5-tetrahydro-(2*S*)-dipicolinic acid (HTPA) as direct product of this reaction. In all cases examined, the reaction proceeds via a ping-pong mechanism in which pyruvate binds as a Schiff-base to the amino group of a lysine residue in the active site of the enzyme. Then, (*S*)-ASA binds to enzyme-bound enamine of pyruvate, and after transamination and cyclization the product is released (Blickling *et al.* 1997b). Subsequently DHDP may be formed spontaneously under physiological conditions (Blickling *et al.* 1997b). It is generally assumed that DHDP is substrate for dihydrodipicolinate reductase (DHDPR) (Bao *et al.* 2002, Scapin *et al.* 1997) (Fig. 1).
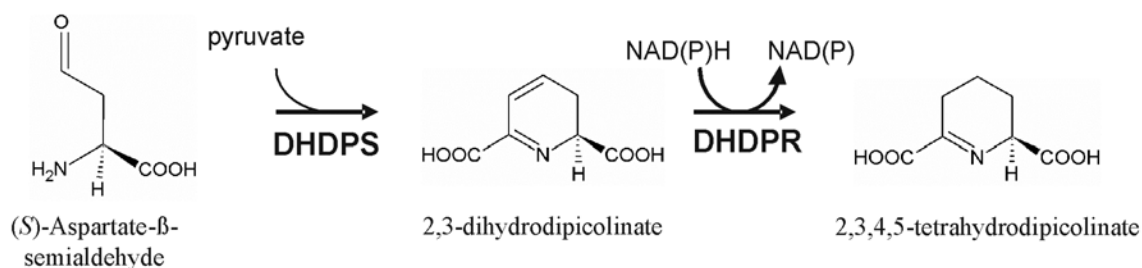
Figure 1. Reactions catalyzed by DHDPS (EC 4.2.1.52) and DHDPR (EC 1.3.1.26)
According to (Blickling *et al.* 1997b) the immediate product formed by DHDPS is (4*S*)-4-hydroxy-2,3,4,5-tetrahydro-(2*S*)-dipicolinic acid.

According to the classification of Grazi *et al.* (1962), DHDPS is a class I aldolase belonging to the Neuraminate lyase (NAL) subfamily (Choi *et al.* 2006). The amino acid sequences of DHDPSs from different organisms are well conserved. Most DHDPS-structures known to date have a homotetrameric quaternary structure (Blagova *et al.* 2006, Blickling *et al.* 1997a, Mirwaldt *et al.* 1995, Pearce *et al.* 2006), with each monomer consisting of an (α/β)$_8$ barrel and a C-terminal alpha-helical domain.

DHDPS in higher plants is generally feedback inhibited by low concentrations of (*S*)-lysine (µmolar range, K_i = 11-51 µM) (Dereppe *et al.* 1992, Frisch *et al.* 1991, Ghislain *et al.* 1995, Kumpaisal *et al.* 1987, Mazelis *et al.* 1977), while their counterparts in gram-negative bacteria are only moderately inhibited (K_i = 0.4 mM) (Karsten 1997, Laber *et al.* 1992).

Interestingly, DHDPSs from gram-positive bacteria have shown to be insensitive to (*S*)-lysine (Cahyanto *et al.* 2006, Hoganson and Stahly 1975, Stahly 1969). Recently, residues that are probably responsible for binding of the (*S*)-lysine effector in *E. coli* DHDPS have been identified (Dobson *et al.* 2005).

In this study we analyzed a gene cluster encoding enzymes that participate in the diaminopimelate pathway for lysine synthesis, present in *Thermoanaerobacter tengcongensis,* a gram-positive, thermophilic bacterium with optimal growth at 75°C (Xue *et al.* 2001). The first 2 genes of this pathway (*dapA & dapB*) of *T. tengcongensis* are cloned and expressed in *E. coli.* The *dapA* gene is coding for the DHDPS, which is characterized in detail and compared to other known DHDPSs.

Materials and methods

(*S*)-ASA synthesis

(*S*)-2-*tert*-Butoxycarbonylamino-*N*-methoxy-*N*-methyl-succinamic acid *tert*-butyl ester (**1**) (Wernic *et al.* 1989)

All glassware was dried 24 h in an oven at 140°C before use and the dichloromethane was freshly distilled. To a stirred solution of *N*-Boc-Asp-OtBu (1eq., 17 mmol, 5 g) in dry dichloromethane (175 mL) under nitrogen atmosphere and at room temperature was added triethylamine (1.1 eq., 19 mmol, 3 mL) and solid (benzotriazol-1-yloxy)tris(dimethylamino)phosphonium hexafluorophosphate (BOP.PF$_6$, 1.1 eq., 19 mmol, 8.40 g), giving a homogeneous mixture. After 5 min, solid *N,O*-dimethylhydroxylamine hydrochloride (1.1 eq., 19 mmol, 2 g) was added, followed by another portion of triethylamine (1.1 eq., 3 mL). After 10 min, the suspension was completely dissolved. The mixture was stirred at room temperature for 5 h. The reaction was followed by TLC on silica with petroleum ether:EtOAc 1:1 as eluent. The solution was then washed with HCl (1M, 3 x 50 mL), H$_2$O (1 x 50 mL), NaHCO$_3$ (1M, 2 x 50 mL). The organic layers were dried with Na$_2$SO$_4$ and the solvent was evaporated *in vacuo* (T< 30°C, 350 mmHg). A crude yellowish oil was obtained which was purified on a silica chromatography column using petroleum ether:EtOAc 6:4 as eluent. A colorless oil was obtained (5.50 g, 96 %).

^{1}H NMR (200 MHz) in DMSO: δ 1.36 (s, 18H), 2.73 (d, 2H), 3.07 (s, 3H), 3.65 (s, 3H), 4.20 (m, 1H), 7.10 (d, 1H).

(*S*)-2-*tert*-Butoxycarbonylamino-4-oxo-butyric acid *tert*-butyl ester (**2**) (Wernic *et al.* 1989)

All glassware was dried 24 h in an oven at 140°C before use and THF was freshly distilled. To a stirred solution of **1** (Fig. 2) (1 eq., 7.52 mmol, 2.5 g) in dry THF (100 mL) at -78°C and under nitrogen atmosphere was added dropwise a solution of di*iso*butylaluminium hydride

(DIBAL) in dichloromethane (1 M, 1.5 eq., 12.0 mmol, 12 mL) in about 30 min. The reaction was stirred in these conditions for 3 h and the progress was followed by TLC on silica with petroleum ether:ether 1:1 as eluent. The reaction mixture was then partitioned between ether (150 mL) and a solution of $NaHSO_4$ (0.35 M in water, 100 mL). The aqueous layer was extracted with ether (3 x 50 mL). The combined organic layers were washed successively with solutions of HCl (1M, 3 x 50 mL), $NaHCO_3$ (1M, 3 x 50 mL) and brine (3 x 100 mL) and dried using Na_2SO_4. The solvent was removed *in vacuo* (RT, 650 < P < 50 mmHg). A colorless oil was obtained which crystallized by standing overnight at room temperature. In case a mixture of product and starting material was obtained, a purification on silica chromatography column with petroleum ether/EtOAc as eluent was performed. The solid was purified by recrystallization in cold *n*-hexane (0.90 g, 44 %).

^{1}H NMR (300 MHz) in $CDCl_3$: δ 1.42 (s, 18H), 2.95 (t, 2H), 4.40 (m, 1H), 5.29 (d, 1H), 9.6 ppm (s, 1H).

^{13}C NMR (300 MHz) in $CDCl_3$: δ 28.2 (CH_3), 28.7 (CH_3), 46.8 (CH_2), 49.7 (CH), 155.7 (COOtBu), 170.3 (NHCO), 199.7 (CHO).

(*S*)-2-Amino-4,4-dihydroxy-butyric acid, trifluoroacetate salt [(*S*)-ASA hydrate, TFA salt] (Roberts *et al.* 2003)

All glassware was dried 24 h in an oven at 140ºC before use and dichloromethane was distilled and dried on molecular sieves 3Å before use. A solution of **2** (Fig. 2) (0.6 g, 2.2 mmol) was stirred in dry dichloromethane (15 mL) under nitrogen atmosphere for 15 min. Then, trifluoroacetic acid (15 mL) was added rapidly. The reaction mixture was stirred 2.5 h after addition. The solvent was evaporated *in vacuo* (rt, 200 mmHg). The yellow oily residue obtained was partitioned between ethyl acetate (60 mL) and deionized water (60 mL). An emulsion was observed which separated upon standing. After decantation, the aqueous phase was washed twice with ethyl acetate (2 x 60 mL) and the water was evaporated *in vacuo* (rt, P < 60 mmHg) but not to dryness to avoid the removal of TFA and therefore the polymerization of the product. The presence of polymer can be checked by TLC. A yellow solution was obtained (y = 98%). The concentration of the product in water was determined by ^{1}H and ^{13}C NMR with fumaric acid as internal standard. $[\alpha]^{20}_D$ = + 3.73 (c = 1.5, H_2O), lit [2] +3.33 (c = 1.5, H_2O).

^{1}H NMR (200 MHz) in D_2O: δ 2.0 (m, 2H), 3.9 (dd, 1H), 5.1 (t, 1H), 1.2 (s, presence of residual *t*-butanol).

^{13}C NMR (200 MHz) in D_2O: δ 38.7 (CH_2-C3), 52.3 (CH-C2), 90.3 (CH-C4), 119.0 (q, CF_3), 165.0 (CF_3C̲), 174.0 (COOH), 29.0 ppm (*tert*-BuOH).

Concentration determination by ^{1}H NMR:

The (S)-ASA solution (30 mg) and fumaric acid solution (200 μL, 5 g L^{-1}, 25 mg dissolved in 5 mL of D$_2$O) were put in a volumetric flask of 2 mL. This solution was used to fill the NMR tube. The NMR spectrum was determined with suppression of the water signal. After comparison of the signal intensity of both compounds, a concentration of 81% w/v was found.
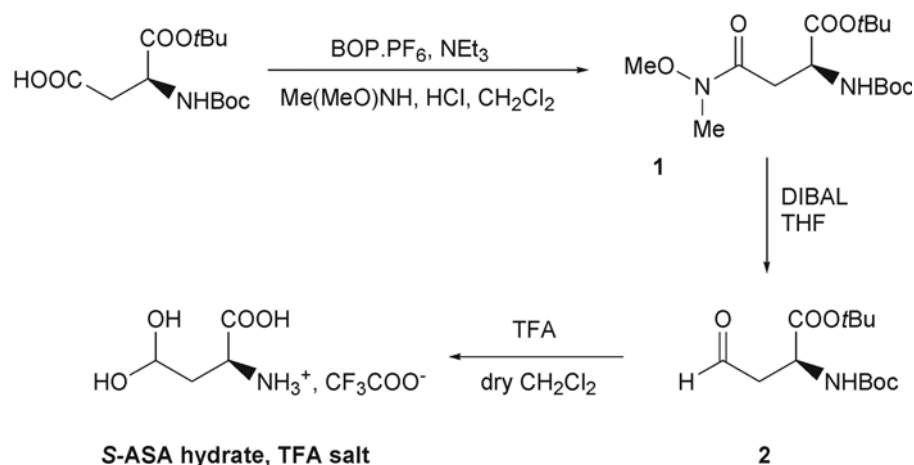
Figure 2. Synthetic scheme of (S)-ASA

Cloning

The predicted *T. tengcongensis* (DSM 15242^T, JCM11007) genes *dapA* (TTE0832; Genbank Identifier 20515826) and *dapB* (TTE0831: Genbank Identifier 20515825) were PCR-amplified from genomic DNA of *T. tengcongensis* using Pfu DNA-polymerase (Stratagene) and primer pairs BG1758-1759 (BG1758 - sense: 5'-GCGCCCATGGCACCTGTATTTAAAGGTTCATGTGTGGC-3'; BG1759 - antisense: 5'-GCGCCTCGAGATTTTCCTCCTTTAAAAGGCCCATACTGC-3') and BG1953-1954 (BG1953 - sense: 5'-CGCGCTCATGATGATAAGGCTAATAATCCAC-3'; BG1954 - antisense: 5'-CGCGCGTCGACTCATCCTTTCTTTACAAGGTCTTC-3') respectively. The *dapA* gene was cloned into a pET24d vector, as a NcoI / XhoI restriction fragment, resulting in vector pWUR194, which, after expression, results in a protein with an extra alanine at the second place and a C-terminal His-tag: LEHHHHHH. The *dapB* gene was cloned into the pET24d vector using BspHI (in NcoI) and SalI resulting in vector pWUR213, encoding the 'wild type' protein (without a His-tag). Both plasmids were transformed to *E. coli* BL21(DE3) cells.

Production and purification of enzymes

Production of *T. tengcongensis* DHDPS (TteDHDPS) was performed using 1 liter LB culture, supplemented with 50 μg/mL kanamycin, inoculated with *E. coli* BL21(DE3) cells containing the pWUR194. After 4 h of growth (37°C with shaking), the culture was induced with 1 mM

IPTG (isopropyl-β-D-thiogalactopyranoside) and growth was continued overnight at 37 °C. The cells were harvested by centrifugation (3800 x g, 10 min), resuspended in 20 mL 50 mM sodium phosphate buffer pH 7.0 – 0.5 M NaCl and lysed using a French press. After centrifugation (27,000g, 20 min) the soluble protein fraction was incubated at 70°C for 30 min. This was centrifuged (48,000g, 30 min) to obtain the heat-stable cell-free extract (HSCFE). The TteDHDPS was further purified on an ÄKTA FLPC (Amersham Pharmacia Biotech), using affinity chromatography (Nickel-chelating Superose column), and a one-step elution with 50 mM sodium phosphate buffer pH 7.5, containing 0.5 M NaCl and 0.5 M imidazole. Fractions containing Tte-DHDPS were pooled and then applied to a desalting column (HiPrep 26/10, Pharmacia Biotech), using 50 mM sodium phosphate pH 7.0, 50 mM NaCl, to remove the imidazole.

T. tengcongensis DHDPR (TteDHDPR) was produced in a similar way, using 0.15 mM IPTG for induction. Cells were resuspended in 30 mL 90 mM HEPES, 160 mM KCl, pH 7.0 – the heat incubation was shortened to 15 minutes. Tte-DHDPR was purified on an ÄKTA FLPC using anion-exchange chromatography (Q-sepharose; 20 mM Tris-HCl, pH 8.5, gradient 0-1 M NaCl, DHDPR in flowthrough). Fractions containing TteDHDPR were subsequently desalted, and further purification was performed using anion-exchange chromatography (MonoQ; 20 mM Tris/HCl pH 8.5, gradient 0-1M NaCl, elution at 0.28 M). Proteins were analyzed on a 12% SDS-PAGE after staining with Coomassie Brilliant Blue.

Enzyme assays

Pyruvate depletion assay

DHDPS activity was measured by assaying the remaining pyruvate in a standardized lactate dehydrogenase (LDH) assay in duplicate. Reactions of 1 mL volume containing 50 mM sodium phosphate buffer, pH 6, 0.5 mM pyruvate, 0.5 mM (*S*)-ASA and an appropriate amount of enzyme were incubated at 60 °C for 5-10 minutes. The reaction was stopped by adding 100 µL 12% trichloroacetic acid (TCA) and precipitated protein was removed by centrifugation (16,000g, 10 minutes). From a mixture containing 100 mM Tris/HCl pH 7.5, 0.16 mM NADH and 300µL sample the starting absorbance at 340 nm was read. After adding 3 units of LDH and incubating for 5 min, the absorbance was read again. The absorbance difference in NADH (molecular extinction coefficient 6.22 mM^{-1} cm^{-1}; (Horecker and Kornberg 1948)) was used to calculate the amount of pyruvate left in the reaction mixture.

Substrate specificity was tested as follows. Different aldehydes (20 mM) were incubated with pyruvate (10 mM) and 25 µg DHDPS in a 100 µl volume (buffered in 50 mM sodium phosphate / 50 mM sodium citrate pH 8.0) for 1 h at 60°C. Reactions were stopped by adding 10 µL 12% TCA. A 16 µL sample was used to determine the remaining pyruvate as

described above. Reactions with the aldoses as substrates were additionally analyzed with a thiobarbituric acid – assay as described previously (Wolterink-van Loo *et al.* 2007).

Coupled assay to DHDPR

DHDPS activity was measured in a continuous assay, using DHDPR, as previously described by Yugari and Gilvarg (1965). Standard incubations contained 50 mM sodium phosphate / 50 mM citrate buffer pH 8.0, 10 mM sodium pyruvate, 0.16 mM β-NADH, an excess of TteDHDPR (usually 30 µg) and an appropriate amount of TteDHDPS (0.1-1 µg) in 1 mL volume in a quartz cuvet. This mixture was heated up to 60°C for 5 min, the reaction was then initiated by adding 1 mM (*S*)-ASA. The reaction was followed by measuring the absorption decrease of β-NADH at 334 nm (molecular extinction coefficient: $6.18*10^3$ $M^{-1}cm^{-1}$ (Ziegenhorn *et al.* 1976)). Measurements were performed in duplicate or triplicate. (*S*)-ASA was stable for the time taken in the experiments up to 60°C (initial velocity taken within 30 sec). Kinetic data are expressed in U/mg, where 1 unit is equal to 1 µmol of NADH oxidized per second. When kinetic parameters for (*S*)-ASA were determined, the concentration varied between 0.1 mM and 8.0 mM and pyruvate was present in excess at 20 mM, while the kinetic parameters for pyruvate were determined by varying the concentration between 0.2 and 20 mM, using 2 mM (*S*)-ASA. Michaelis Menten kinetics were modeled using TableCurve 2D and the following equations:

$$v = V_{max} * [S] / (K_M + [S])$$
$$v = V_{max} * [S] / (K_M + [S] + ([S]^2 / K_i))$$

Where v is the enzyme activity at a particular substrate concentration ([S]) and K_i the inhibition constant of (*S*)-ASA.

Optimum pH and temperature

The assays were performed as the standard coupled assay, with replacement of the buffer with similar ionic strength to vary the pH (pH 4.0-8.0: phosphate-citrate buffer, pH 8.0-9.5: glycine). To determine the optimal temperature, the assay temperature was varied from 30 to 90°C.

Lysine inhibition

Inhibition caused by (*S*)-lysine was tested by performing the standard coupled assay and adding increasing amounts of (*S*)-lysine (0- 200 mM).

Heat stability

Aliquots of 12.6 µg/mL DHDPS in assay buffer were incubated at 90°C, removed at time intervals (up to 5 hours) and cooled rapidly to 0°C. Residual activity was measured using the standard coupled assay described above.

Alignments

Protein sequence alignments were made using MegAline from DNASTAR and manually checked using GeneDoc. A structural model was made by submitting the Tte-DHDPS protein sequence to PHYRE (Protein Homology/analogY Recognition Engine). The models based on *E. coli* and *T. maritima* DHDPS had low E-values of 4.3e-44 and 1.8e-43. The model (monomer of *T. tengcongensis* DHDPS) based on *E. coli* DHDPS was used further. PyMol was used to fit the model on the structures of *E. coli* and *T. maritima* DHDPS for comparison of active sites and possible binding site of lysine.

Results and discussion

Genomic location of the lysine synthesis genes

The diaminopimelate (DAP) pathway towards (*S*)-lysine includes nine consecutive steps, starting with (*S*)-aspartate as a precursor. The first two steps, involving an aspartate kinase and an aspartate-semialdehyde dehydrogenase, are shared with the threonine biosynthesis pathway. In *Thermotoga maritima* and *T. tengcongensis* part of these (non-characterized / hypothetical) *dap*-genes are clustered in an operon-like structure (Fig. 3). In *T. tengcongensis* the clustered genes include the aspartate-semialdehyde dehydrogenase gene (TTE0833, *asd*), the dihydrodipicolinate synthase gene (TTE0832, *dapA*), the dihydrodipicolinate reductase (TTE0831, *dapB*) and the N-succinyl transferase gene (TTE0830, *dapD*). In the corresponding cluster of *T. maritima* (TM1517-1523) three additional genes are present: *lysC*, encoding the enzyme that catalyzes the first step of the pathway and *dapF* and *lysA* (responsible for the last two steps). Additionally, related gene clusters are present in *Clostridia* species comprising *asd, dapA* and *dapB* (Fig. 3). The remaining *T. tengcongensis* enzymes in the DAP-pathway, for which (hypothetical) genes are missing in this cluster, had alternative (hypothetical) genes elsewhere on the genomes, based on either KEGG PATHWAY analysis or BLAST analysis: *lysC;* TTE1382, *dapC;* TTE2440, *dapE;* TTE1746, *dapF;* TTE1514, *lysA;* TTE0210.
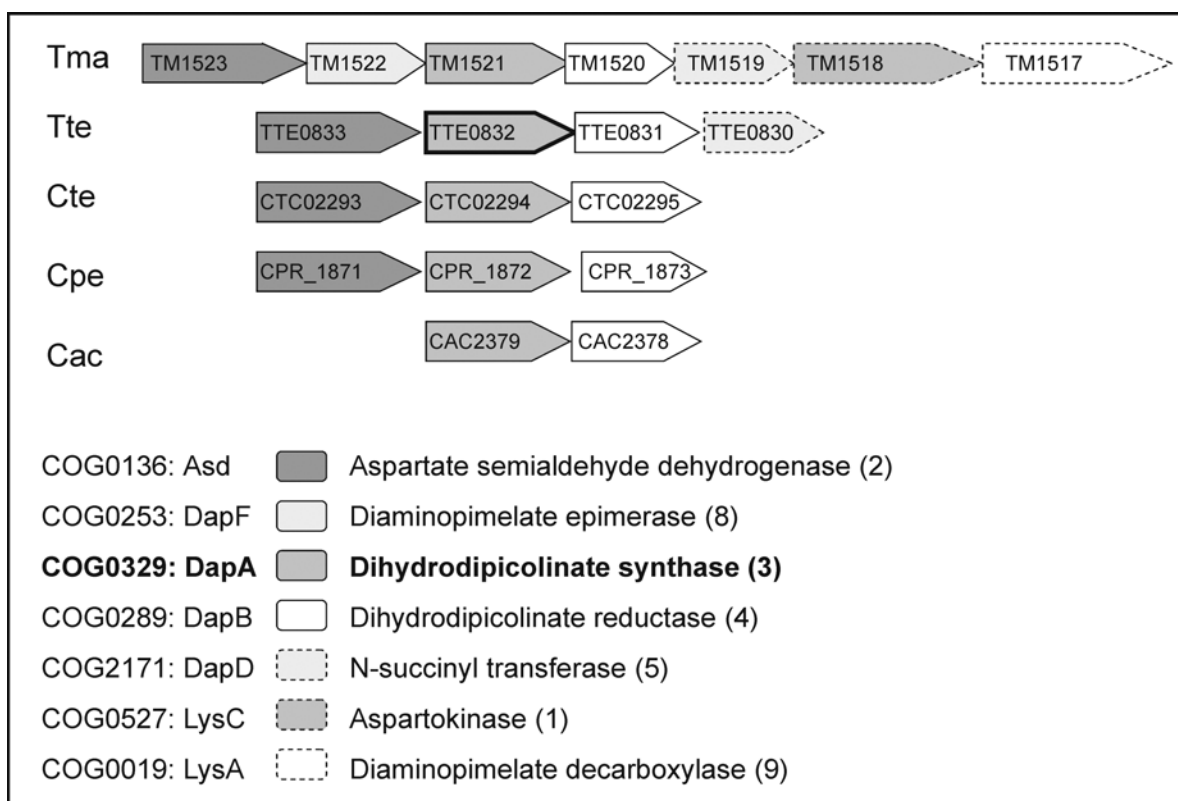
Figure 3. Gene clusters of the diaminopimelate pathway in bacteria
Comparison of clustered genes of lysine synthesis in *T. tengcongensis* (Tte) and *T. maritima* (Tma), *Clostridium tetani* (Cte), *C. perfringens* (Cpe), *C. acetobutylicum* (Cac). Genes are indicated by their systematic gene name.

Biochemical characterization

The *dapA* and *dapB* genes of the *T. tengcongensis* genome, coding for DHDPS and DHDPR, were cloned and expressed in *E. coli* and purified to homogeneity (Fig. 4a), using a heat incubation step and additional chromatography (as described in material and methods). The DHDPS activity was first analyzed with a pyruvate depletion assay to test DHDPS activity during purification.

Biochemical characterization of TteDHDPS (Table 1) was performed using a coupled assay, using an excess of Tte-DHDPR and following the oxidation of β-NADH. Optimal activity was found at pH 8.0, which is similar to that of the *E. coli* enzyme (EcoDHDPS) (Shedlarski and Gilvarg 1970) and some plant DHDP synthases (Dereppe *et al.* 1992, Kumpaisal *et al.* 1987, Tam *et al.* 2004). Furthermore, an attempt was made to measure the temperature optimum. It was necessary to keep the assay time very short due to instability of (*S*)-ASA (Coulter *et al.* 1996). Using this method the highest specific activity could be measured at 80°C (Fig. 4b). At physiological temperature (75°C), under the conditions used, (*S*)-ASA did completely disappear within 7 minutes as tested with residual DHDPS activity. Obviously the stability of (*S*)-ASA is sufficient for metabolic conversion *in vivo* due to high

turnover, and possibly because of the presence of compatible solutes. Furthermore, although no experimental evidence is yet available, the enzymes involved in (S)-lysine synthesis could be organized in functional complexes for efficient transfer of intermediates, a phenomenon called metabolic channeling (Massant *et al.* 2002).

Unlike a recent analysis of *T. maritima* DHDPS by Pearce *et al.* (2006), we decided to perform activity measurements at elevated temperature (60°C) to characterize *T. tengcongensis* DHDPS, closer to its *in vivo* conditions. During characterization, the reaction time was kept very short and the reaction was initiated by adding (S)-ASA. As such, it was possible to determine initial reaction velocities. When varying the pyruvate concentration, the apparent V_{max} was 17.8 (± 0.3) U/mg, with the K_M^{app} for pyruvate being 0.85 (± 0.05) mM, using 2 mM (S)-ASA. When the (S)-ASA concentration was varied, using 20 mM pyruvate, the apparent V_{max} was 19.0 (± 0.8) U/mg and the K_M^{app} for (S)-ASA 0.38 (± 0.05) mM. When modeled with substrate inhibition: $K_{M(S)-ASA}^{app}$ was 0.92 (± 0.13) mM, with inhibition constant, $K_{i(S)-ASA}$ = 4.7 (± 0.8) mM. These kinetic values at 60°C are somewhat higher compared to those of EcoDHDPS and TmaDHDPS, which were determined at 30°C. Most likely, this depends on the different assay temperature used here, as in previous enzyme analyses it has been shown, that the K_M of an enzyme can either increase or decrease with increasing temperature (Vieille *et al.* 1995). The different K_M values here are nevertheless still in range with the K_M values reported for other DHDPSs (Dereppe *et al.* 1992, Dobson *et al.* 2004a, Tam *et al.* 2004). Remarkably, specific activity of TteDHDPS at 30°C (3.3 U/mg) is higher than the V_{max} of EcoDHDPS (0.58 U/mg) (Dobson *et al.* 2004a) and of the recently characterized TmaDHDPS (1.01 U/mg) at 30°C (Pearce *et al.* 2006).
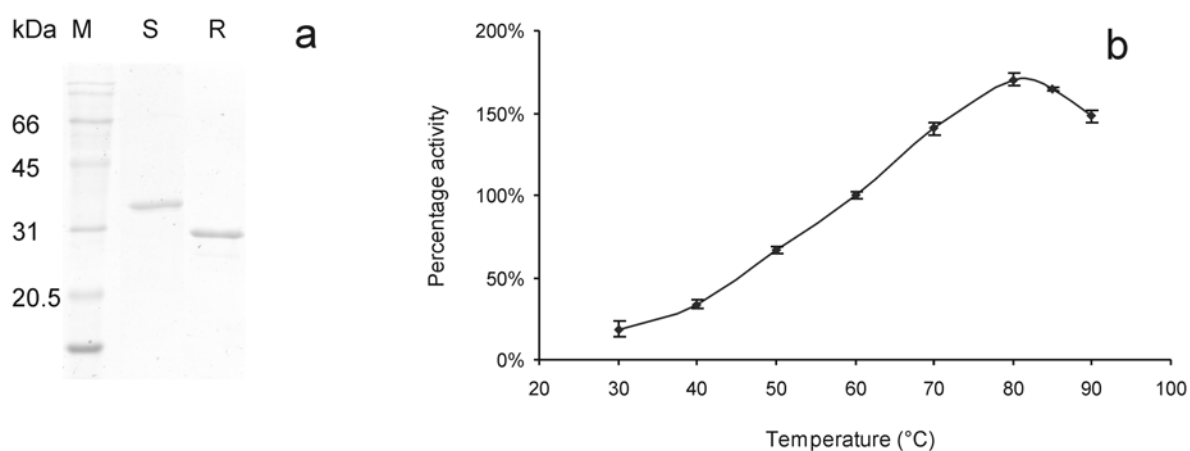
Figure 4. Expression and activity of DHDPS. **a)** SDS-PAGE of purified DHDPS and DHDPR - M; marker proteins, S; DHDPS, R: DHDPR **b)** Temperature curve of relative activity of TteDHDPS - Absolute activity at 60°C was 15.7 U/mg, which is set at 100%.

Table 1: Biochemical data for bacterial dihydrodipicolinate synthases.
One unit (U) of enzyme activity is equal to the loss of 1 $\mu mol_{NADH} s^{-1}$.

|	*Tte-DHDPS*	*Eco-DHDPS*	*Tma-DHDPS*
pH optimum	8.0	8.0	ND
Temperature optimum	85°C	30°C	ND
$V_{max(pyr)}^{app}$	17.8 (0.3) U/mg	0.58 U/mg[b]	1.01 U/mg[a]
$K_{m(pyr)}^{app}$	0.85 (0.05) mM	0.25 mM[b]	0.053 mM[a]
$V_{max(ASA)}^{app}$	19.0 (0.8) U/mg	0.58 U/mg[b]	1.01 U/mg[a]
$K_{m(ASA)}^{app}$	0.38 (0.05) mM	0.11 mM[b]	0.16 mM[a]
K_i (lysine)	180 mM	3.9/0.32 mM	>> 4mM[a]
$T_{1/2}$ 90°C	175 minutes	< 30 seconds[a]	> 400 minutes[a]

[a] Data from Pearce *et al.* (Pearce *et al.* 2006), true K_m values determined
[b] Data from Dobson *et al.* (Dobson *et al.* 2004a), true K_m values determined

Different studies of DHDPSs reported substrate inhibition by (*S*)-ASA (Karsten 1997, Kumpaisal *et al.* 1987, Stahly 1969), while others, knowing about this possible problem just avoided high concentrations of this substrate (Tam *et al.* 2004). (Dobson *et al.* 2004a) state that this inhibition problem is caused by a side product formation during (*S*)-ASA synthesis by ozonolysis; other methods (Coulter *et al.* 1996, Roberts *et al.* 2003, Tudor *et al.* 1993) showed no inhibition. Our method of (*S*)-ASA synthesis is similar to that of (Roberts *et al.* 2003) and therefore should not give any inhibition problems. However at concentrations of (*S*)-ASA higher than 2 mM, enzyme activity decreased. The latter coincided with formation of an orange color, which indicates some side reactions during incubation, probably due to the (relatively) high temperature used in our studies. It is unclear if the lower activity is caused by this orange compound or by substrate inhibition.

Substrate specificity

In order to investigate the substrate specificity of TteDHDPS a range of commercially available aldehydes were tested for their ability to replace (*S*)-ASA in the condensation reaction. None of the tested 3-carbon aldehydes (propionaldehyde and D,L-glyceraldehyde), 4-carbon aldehydes (D-erythrose, L-erythrose, D-threose, L-threose) or 5-carbon aldehydes (D-arabinose, L-arabinose, D-ribose, D-xylose) could replace (*S*)-ASA. Of the 2-carbon aldehydes (acetaldehyde, glycolaldehyde, glyoxylate), TteDHDPS only showed significant activity with glycolaldehyde at a very low rate ~0.005% relative to the reaction with (*S*)-ASA. Hence, the TteDHDPS enzyme is very specific for (*S*)-ASA, in agreement with reports on EcoDHDPS and DHDPS of wheat, which were inactive on (*R*)-ASA (Dobson *et al.* 2004a, Kumpaisal *et al.* 1987).
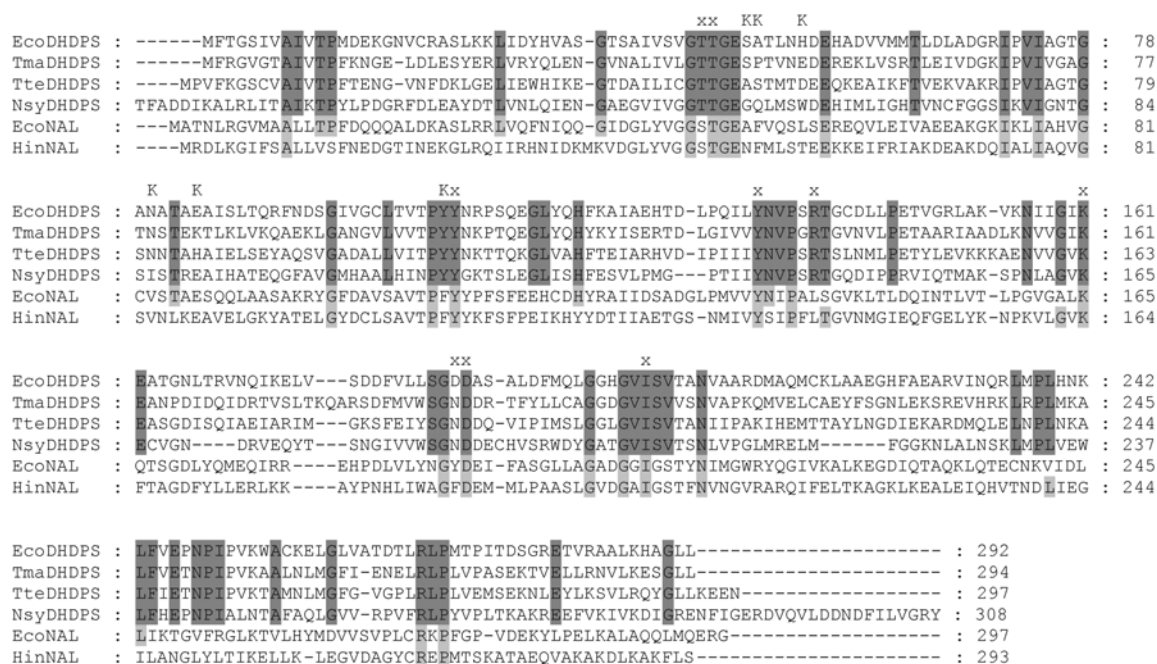
Figure 5. Alignment of TteDHDPS

Alignment of different dihydrodipicolinate synthases and Neuraminate lyases. Eco; *E. coli*, Tma; *T. maritima*, Tte; *T. tengcongensis*, Nsy; *Nicotiana sylvestris*, Hin; *Heamophilus influenzae.* Catalytic residues are shown with x. Residues involved in (*S*)-lysine binding in *E. coli* are indicated with K. Conserved residues are highlighted in grey in DHDPS sequences, and in a lighter grey in Neuraminate lyases.

Structurally conserved features

Alignment of TteDHDPS with *E. coli*, *T. maritima* and *Nicotiana sylvestris* DHDPSs and NALs from *E. coli* and *Haemophilus influenzae* showed that TteDHDPS has high homology know DHDPSs, especially with TmaDHDPS (57%-identity) and EcoDHDPS (54%-identity) (Fig 5). The active site lysine residue is present in all aligned DHDPSs and NALs as is the catalytic triad, consisting of Thr-44, Tyr-107 and Tyr-133 (*E. coli* numbering), previously shown by (Dobson *et al.* 2004c). The NALs only miss the Thr-44, which is conservatively replaced by a serine (Fig. 5). A structural model of TteDHDPS (Fig. 6a) (model of the monomer) showed that the active site residues were structurally conserved as well (Fig. 6b) within DHDPSs from *T. tengcongensis*, *T. maritima* and *E. coli*. The calculated size of the TteDHDPS monomer (including the His-tag) is 34 kDa; analysis with size exclusion chromatography revealed a dominant form of ca. 150 kDa, indicating a tetramer in solution, which is similar to oligomeric states of most DHDPSs known to date (Blickling *et al.* 1997a, Dobson *et al.* 2004c, Pearce *et al.* 2006).
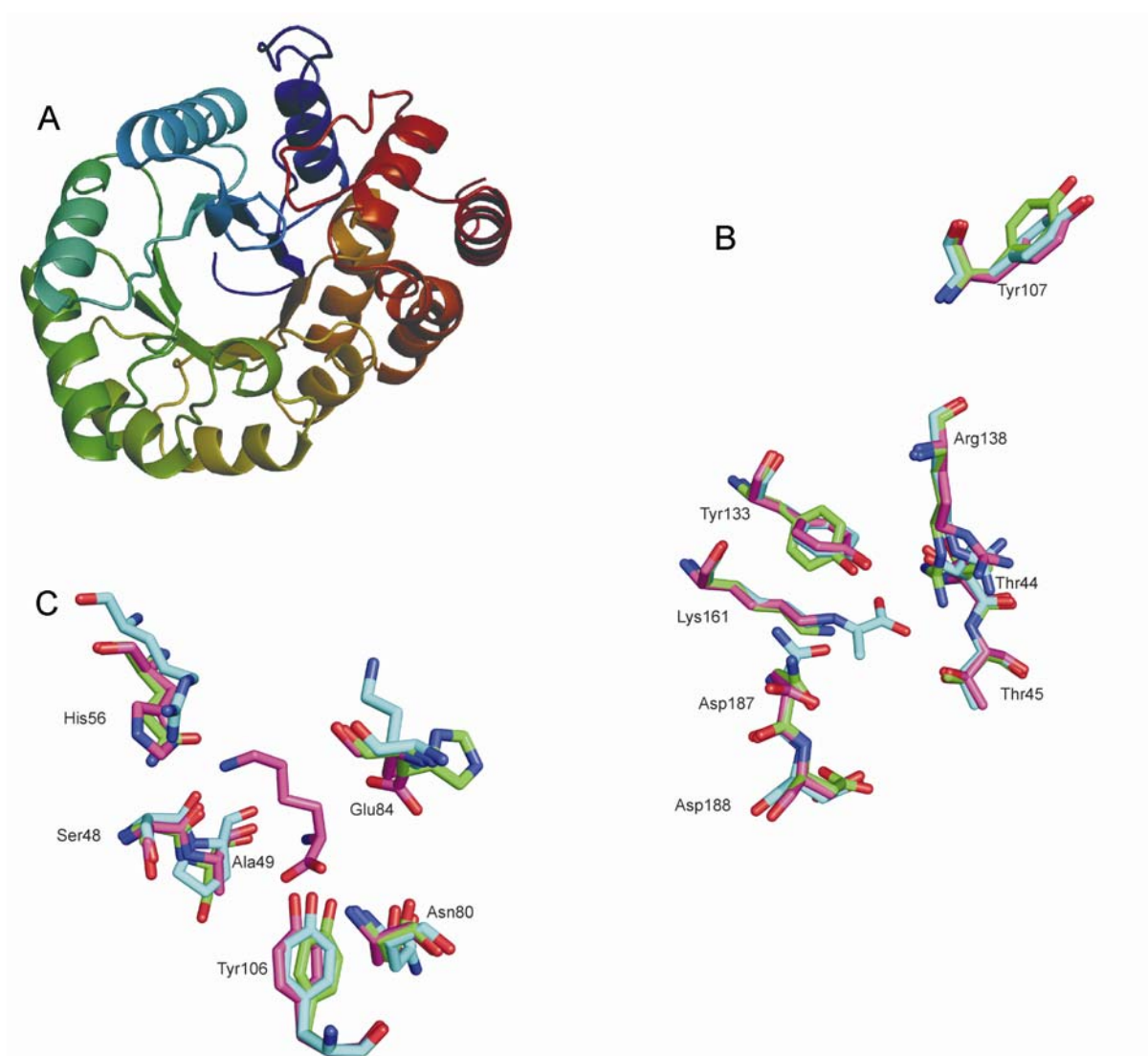
Figure 6. Structural analysis of TteDHDPS
A) Tte-DHDPS monomer, top view. **B)** Active site residues: Overlay of the active sites of TteDHDPS (model) (green), TmaDHDPS (including bound pyruvate) (cyan) and EcoDHDPS (magenta), numbering of residues as in EcoDHDPS. **C)** Site of binding of (*S*)-lysine in *E. coli*. Overlay of the (*S*)-lysine binding site of Eco-DHDPS (magenta), TmaDHDPS (cyan) and TteDHDPS (model) (green), numbering of residues as in EcoDHDPS.

Lysine inhibition

DHDPS of gram-positive bacteria is generally not inhibited by lysine (Cahyanto *et al.* 2006, Hoganson and Stahly 1975, Stahly 1969), and TteDHDPS is not an exception in that respect. When testing for possible lysine inhibition, adding lysine up to a concentration of 20 mM did not lead to detectable difference in activity. Only when extremely high concentrations of lysine were used, an apparent inhibition constant could be determined ($K_{i(lysine)}$ = 180 mM). In EcoDHDPS, which is inhibited by lysine, the mode of binding was determined by co-crystallization (Dobson *et al.* 2005). When looking at the residues involved in binding in

EcoDHDPS, only the Tyr-106 is conserved in all DHDPSs and Asn-80 is conserved in nearly all bacterial DHDPSs (including TteDHDPS). Ser-48 seems to be present in gram-negative DHDPSs, Ala-49, His-56 and Glu-84 are less conserved (Fig. 5), but are present in some gram-negative DHDPSs. This might indicate that inhibition of DHDPS by lysine in gram-negatives is similar to the mechanism in EcoDHDPS. From a structural overlay of the lysine binding site of EcoDHDPS and aligned residues in TteDHDPS and TmaDHDPS it is clear that only few residues are conserved (Fig. 6c). According to Dobson *et al.* (2005), the nitrogen of His-56 and the carbonyl oxygen of Glu-84 are within bonding distance in EcoDHDPS, which are replaced by other residues in TteDHDPS and TmaDHDPS. Those residues apparently do not bind the lysine, leading to the insensitivity of these two DHDPSs for (*S*)-lysine.

Acknowledgements

Chapter 3

Biochemical and structural exploration of the catalytic capacity of *Sulfolobus* KDG aldolases

Suzanne Wolterink-van Loo#, André van Eerde#, Marco A. J. Siemerink, Jasper Akerboom, Bauke W. Dijkstra, and John van der Oost

#Both authors contributed equally

Abstract

Aldolases are enzymes with potential applications in biosynthesis, depending on their activity, specificity and stability. In this study, the genomes of *Sulfolobus* species were screened for aldolases. Two new 2-keto-3-deoxygluconate aldolases (KDGA) (2-oxo-3-deoxygluconate) aldolase from *Sulfolobus acidocaldarius* and *Sulfolobus tokodaii* were identified, overexpressed in *Escherichia coli*, and characterized. Both enzymes were found to have biochemical properties similar to the previously characterized *S. solfataricus* KDGA, including the condensation of pyruvate and either D,L-glyceraldehyde or D,L-glyceraldehyde 3-phosphate. The crystal structure of *S. acidocaldarius* KDGA revealed the presence of a novel phosphate-binding motif that allows the formation of multiple hydrogen-bonding interactions with the acceptor substrate, and enables high activity with glyceraldehyde 3-phosphate. Activity analyses with unnatural substrates revealed that these three KDGAs readily accept aldehydes with two to four carbon atoms, and that even aldoses with five carbon atoms are accepted to some extent. Water-mediated interactions permit binding of substrates in multiple conformations in the spacious hydrophilic binding site, and correlate with the observed broad substrate specificity.

Introduction

Aldolases are among the few classes of enzymes capable of enlarging the carbon skeleton of molecules in a specific way by catalysing the condensation of a ketone donor and an aldehyde acceptor. These enzymes can be a tool for the specific production of carbon-carbon bonds in organic synthesis, since it is difficult to achieve stereochemical control with the widely used chemical aldol condensation reaction (Machajewski and Wong 2000). Aldolases have already successfully been used for the biosynthesis of a limited number of specialty compounds (Samland and Sprenger 2006). However, many of the currently used aldolases are expensive, unstable, limited in their catalytic range, and often require expensive substrates. The discovery of new aldolases or the engineering of known aldolases is therefore of continuing importance (Samland and Sprenger 2006). The availability of many sequenced genomes opens up new opportunities for identification of potentially suitable aldolases.

In Nature, many different aldolases are encountered. Some aldolases play a vital physiological role in the degradation of metabolites, for instance during glycolysis; other aldolases are employed for carbon backbone assembly, e.g. in amino acid synthesis (Gijsen *et al.* 1996). Two classes of aldolase enzymes are distinguished based on their mechanisms. Class-I aldolases form a Schiff-base intermediate with the donor substrate, whereas class-II aldolases are metal-dependent enzymes (Machajewski and Wong 2000, Takayama *et al.* 1997). Although both classes can usually accept different acceptor substrates, they are rather specific for the donor substrate. Using a different classification, synthesizing aldolases are therefore also grouped according to their donor specificity: (1) dihydroxyacetone phosphate (DHAP)-dependent aldolases; (2) 2-deoxy-ribose-5-phosphate aldolase (DERA) – acetaldehyde dependent; (3) (phosphoenol)pyruvate-dependent aldolases; and (4) glycine-dependent aldolases (Takayama *et al.* 1997). Novel aldolases belonging to these classes may be found in organisms with divergent sugar metabolic pathways or degradation pathways of xenobiotics (Samland and Sprenger 2006). The thermostability of their proteins make extremophiles such as *Sulfolobus* particularly interesting sources of new enzymes.

Sulfolobus species are thermoacidophilic archaea that typically grow optimally at high temperatures (75-85°C) and at a low pH (pH 2-4) (Brock *et al.* 1972, Grogan 1989, Suzuki *et al.* 2002). Because of their global abundance (DeLong and Pace 2001) and ease of cultivation (Grogan 1989), they have become model archaea; genetic tools have been established (Aravalli and Garrett 1997, Cannio *et al.* 1998, Stedman *et al.* 1999) and the complete genome sequences of *Sulfolobus solfataricus, S. tokodaii* and *S. acidocaldarius* have recently been unravelled (Chen *et al.* 2005, Kawarabayasi *et al.* 2001, She *et al.* 2001). In particular because of their resemblance to eukaryotic counterparts, several protein complexes from *Sulfolobus* have been selected for the analyses of fundamental processes, such as replication

(Dionne *et al.* 2003, Klimczak *et al.* 1985, Rossi *et al.* 1986), transcription, and translation (Bell and Jackson 1998, Reiter *et al.* 1988, Zillig *et al.* 1979). In addition, potential applications have been described for some of the thermostable *Sulfolobus* enzymes, including an unusual 2-keto-3-deoxygluconate aldolase (KDGA, 2-oxo-3-deoxygluconate aldolase) (Lamble *et al.* 2005b).

KDGA from *S. solfataricus* is a class-I aldolase and is a member of the N-acetylneuraminate lyase (NAL) subfamily. This NAL subfamily consists of tetrameric enzymes that are specific for pyruvate as donor substrate but use different aldehydes as acceptor substrates. The *S. solfataricus* KDGA was first characterized as an enzyme that catalyses the reversible conversion D,L-2-keto-3-deoxygluconate (KDG) $\rightleftharpoons$ pyruvate + D,L-glyceraldehyde (GA). Later it was discovered that the enzyme was not only able to synthesize KDG but that it could also synthesize D-2-keto-3-deoxygalactonate (KDGal) from pyruvate and D-GA (Lamble *et al.* 2003). Recently, Ahmed *et al.* (2005) showed that the enzyme also catalyses the interconversion D,L-2-keto-3-deoxy-6-**phospho**gluconate (KDPG) $\rightleftharpoons$ pyruvate + D,L-glyceraldehyde-3-**phosphate** (GAP) (Ahmed *et al.* 2005), suggesting that this bifunctional enzyme is part of a semi-phosphorylated Entner-Doudoroff pathway (Ahmed *et al.* 2005, Lamble *et al.* 2005b). Previously, the crystal structure of *S. solfataricus* KDGA (SsoKDGA) was solved; in agreement with the abovementioned specificity studies, substrate soaking studies with KDG and KDGal revealed promiscuous binding with multiple conformations of these substrates owing to different water-mediated interactions of the O-5 and O-6 hydroxy groups in the rather spacious binding cavity (Theodossis *et al.* 2004).

In the present study, two hypothetical KDGA genes from *S. acidocaldarius* and *S. tokodaii* were cloned, and, after expression in *Escherichia coli*, their biochemical properties were assessed, together with the previously characterized *S. solfataricus* KDGA. In addition, the crystal structure of *S. acidocaldarius* KDGA was determined, providing a molecular basis for the observed substrate specificity and stereoselectivity. Furthermore we have identified a novel phosphate-binding motif in the KDGAs that explains their high activity with phosphorylated substrates.

Materials and methods

Cloning and expression

The hypothetical KDGA genes from *S. acidocaldarius* (Saci_0225; Genbank[®] Identifier 70606067) and *S. tokodaii* (ST2479; Genbank[®] Identifier 15623602) were PCR-amplified from genomic DNA with primer pairs BG 1816 & BG 1817 and BG 1783 & 1784 (Table 1). The obtained products were restricted with NcoI and SalI and ligated into NcoI/SalI restricted pET24d (Novagen). These plasmids, designated pWUR193 (containing Saci_0225) and

pWUR192 (containing ST2479) were transformed to *E. coli* BL21(DE3) and *E. coli* BL21(DE3)/pRIL cells. The clone pWUR122, containing the *kdgA* from *S. solfataricus* (SSO3197; Genbank® Identifier 2879782), described previously (Ahmed *et al.* 2005), was also transformed in *E. coli* BL21(DE3) cells.

Table 1 Primers that are used in this study

Primer	Sequence
BG1783	CGCGCCCATGGTGTTCAAAATTTTAAGTATGGATATTG
BG1784	CGCGCGTCGACTTTACGAAACAGCTCTTTCTATTTC
BG1816	CGCGCCATGGAAATAATTTCACCTATCATTAC
BG1817	GCGCGTCGACTTAATGTACCAGTTCTTGAATCTTTC

Purification

S. acidocaldarius KDGA (SacKDGA) was produced by inoculating 2 litres LB (Luria-Bertani) medium supplemented with 50 μg/ml kanamycin with *E. coli* BL21(DE3) cells containing pWUR193. Subsequently, after 4 hours of growth at 37°C, the culture was induced with 0.1 mM IPTG (isopropyl-β-D-thiogalactopyranoside) and growth was continued overnight. The cells were harvested by centrifugation (3800*g*, 10 min.), resuspended in 50 mM HEPES buffer pH 7, 20 mM KCl, and lysed using a French press. After centrifugation (27,000*g*, 20 min.) the soluble protein fraction was incubated at 75°C for 30 min. This was centrifuged (48,000*g*, 30 min.) to obtain the heat-stable cell-free extract (HSCFE). The SacKDGA was further purified using anion-exchange chromatography (Q-sepharose; 20 mM Tris-HCl, pH 8.5, gradient 0-1 M NaCl), eluting at 0.30 M NaCl and gel filtration (Superdex™ 200 10/300 GL; 50 mM Tris/HCl, pH 7.8, 100 mM NaCl). *S. tokodaii* KDGA (StoKDGA) was produced in the same way, except that the culture – inoculated with BL21(DE3)/pRIL cells containing pWUR192 – was induced with 0.02 mM IPTG on ice and further grown at 20°C for 2 days (Q-sepharose-elution at 0.21 M NaCl). *S. solfataricus* KDGA (SsoKDGA) was also produced in a similar way as SacKDGA, but the cell pellet was resuspended in 90 mM HEPES, pH 7, 160 mM KCl and only purified by heat incubation and anion-exchange chromatography, eluting at 0.32 M NaCl. The enzyme purity was checked by SDS/PAGE (10% gels), staining the proteins with Coomassie Brilliant Blue. Protein concentrations were determined with the Bio-Rad Bradford method.

Activity measurements

KDGA activity was measured with the thiobarbituric acid (TBA) assay as described by Buchanan *et al.* (1999). In a standard aldolase assay the aldolase reaction was performed using the following conditions: 50 mM sodium phosphate buffer, pH 6, 50 mM sodium

pyruvate, 20 mM D,L-GA and an appropriate amount of enzyme (~ 1 µg) were incubated at 70°C for 10 min. The TBA-assay was then performed as described previously (Buchanan *et al.* 1999). V_{max} and K_m determinations were performed at optimal pH at 70°C.

Aldolase activities with different aldehydes were measured by assaying the remaining pyruvate in a standardized lactate dehydrogenase (LDH) assay in triplicate. Reactions of 100 µl volume containing 50 mM sodium phosphate buffer, pH 6, 10 mM pyruvate and 20 mM aldehyde and 1– 100 µg enzyme were incubated at 70°C for 10-60 minutes. The reaction was stopped by adding 6 µl 20% trichloroacetic acid (TCA) and precipitated protein was removed by centrifugation (16,000*g*, 10 min). From 984 µl of assay buffer (containing 100 mM Tris/HCl pH 7.5, 0.16 mM NADH and ~3 units LDH) the starting absorbance at 340 nm was read. After adding 16 µl sample and incubating for ~5 minutes, the absorbance was read again. The absorbance difference in NADH (molecular extinction coefficient 6.22 mM^{-1} cm^{-1}; (Horecker and Kornberg 1948)) was taken to be the pyruvate amount remaining. When aldose substrates were used, the TBA assay was also used.

Thermostability of the enzymes at 100°C was checked as follows. KDGA was diluted to 40 µg/ml in 50 mM sodium phosphate, pH 6, divided into 175 µl aliquots and closed into HPLC-vials, extra sealed with nylon. Vials were immersed in a 100°C oil bath and sampled at different time points and immediately cooled on ice. Residual activity was measured with the standard TBA assay.

Crystallization

SacKDGA was crystallized using the hanging-drop method. Hexagonal crystals of SacKDGA with typical dimensions of 0.3 x 0.3 x 0.2 mm were grown at 18°C. Drops were a mixture of 1 µl of a 19 mg/ml protein solution in water and 1 µl of reservoir solution (0.1 M HEPES, pH 7.5, 30 % (v/v) PEG 400 and 0.2 M MgCl2). Diffraction data of native crystals were collected at the ID14-1 beam line at the ESRF (European Synchrotron Radiation Facility; Grenoble, France) at 100 K using the reservoir solution as a cryoprotectant. Data were processed with XDS (Kabsch 1993) and programs from the CCP4 package (Collaborative Computational Project 1994), with different crystals being indexed either in space group P3121 or in a P6522 cell with related dimensions (Table 5). MOLREP (Murshudov *et al.* 1997) was first used to find a molecular replacement solution for the P3121 data, using the co-ordinates of one SsoKDGA subunit (PDB code 1w3i, (Theodossis *et al.* 2004)) as a search model. Afterwards, a molecular replacement solution was found in the second space group, P6522, using the preliminary model of SacKDGA as a search model. The prime-and-switch protocol of RESOLVE (Terwilliger 2004) was used for density modification and removal of model bias, after which final models were obtained by iterative cycles of refinement in Refmac5

(Murshudov *et al.* 1999) and model building in Xtalview (McRee 1999). To obtain a Schiff-base intermediate with pyruvate, a crystal was transferred to a solution of 0.1 M MES buffer, pH 6.3, 32 % (v/v) PEG 400, 0.2 M MgCl2 and 50 mM sodium pyruvate at 291 K and was flash-frozen after 20 minutes. Data were collected at the ID14-2 beamline at the ESRF and processing was as before. Coordinates have been deposited with the Protein Data Bank for the native SacKDGA structure in P3121 (accession code 2NUW), in P6522 (accession code 2NUX), and the complex with pyruvate (accession code 2NUY).

Results and discussion

Sulfolobus genomes contain putative KDGAs

The three sequenced *Sulfolobus* genomes contain several open reading frames encoding hypothetical aldolases, *i.e.* enzymes that potentially catalyze aldol-cleavage or aldol-condensation reactions (Table 2). These include DHAP-dependent fructose-1,6-bisphosphate aldolases (Siebers *et al.* 2001), pyruvate-dependent dihydrodipicolinate synthases (DHDPS) (Dobson *et al.* 2004c) and KDGAs. DHDPS and KDGAs do not require (expensive) phosphorylated substrates, and are thus potentially interesting for industrial applications. Previous studies on the KDGA from *S. solfataricus* suggested a broad substrate specificity (Lamble *et al.* 2003, Lamble *et al.* 2005b) and therefore we focussed on homologues of this enzyme. A BLAST analysis with the protein sequence of the *S. solfataricus* KDGA (SSO3197) revealed high homology with the KDGA from *Thermoproteus tenax* (TteKDGA) and hypothetical KDGAs from *S. acidocaldarius, S. tokodaii, Thermoplasma* and *Picrophilus torridus*. In particular, the active site residues (Lys-153, Tyr-129, Gly-177, Val-193, Pro-6) are conserved in these KDGAs, and also in some well-characterized enzymes of the NAL subfamily of class-I aldolases, such as the NALs from *Haemophilus influenzae* (Barbosa *et al.* 2000) and *E. coli* (Izard *et al.* 1994) and the DHDPS from *E. coli* (Dobson *et al.* 2004c) (Fig. 1). Based on this high similarity and a well-conserved gene context, in which the genes for the putative KDGA, KDG kinase and GAPN-type dehydrogenase are clustered, the hypothetical aldolases from *S. acidocaldarius* and *S. tokodaii* are most likely part of a semi-phosphorylated Entner-Doudoroff pathway as was demonstrated for SsoKDGA and TteKDGA (Ahmed *et al.* 2005). Based on that, the here analyzed homologues were expected to have a similar activity with pyruvate and D,L-GA (phosphorylated and non-phosphorylated) as the KDGA from *S. solfataricus* (Ahmed *et al.* 2005, Buchanan *et al.* 1999, Lamble *et al.* 2005b).

Table 2 Overview of possible enzymes in *Sulfolobus* species that hypothetically catalyze aldolase reactions

Enzyme	Function	Reaction
DAHP aldolase	Synthesis	Phosphoenolpyruvate + Erythrose 4-phosphate → 7-phospho-2-dehydro-3-deoxy-D-arabinoheptonate
Citrate lyase	Degradation	Citrate → oxaloacetate + acetate
DHDPS-like*	Synthesis	Pyruvate + aspartate-β-semialdehyde → dihydrodipicolinate
FBP-aldolase[1]	Degradation	*D-fructose 1,6-bisphosphate → DHAP + D-GAP*
Isocitrate lyase[2]	Degradation	Isocitrate → succinate + glyoxylate
KDGA	Degradation	KDG → pyruvate + GA

* 3 paralogs in *S. solfataricus* 1 in *S. tokodaii* and *S. acidocaldarius*
[1] In *S. solfataricus* and *S. tokodaii* (Siebers *et al.* 2001)
[2] In *S. solfataricus* and *S. acidocaldarius*

```
                *        20         *        40         *        60         *        80         *
Sac-KDGA  : -----MEIISPIITPFDKQGKVNVDALKTHAKNLLEK-GIDAIFVNGTTGLGPALSKDEKRQNLNALYDVTH---KLIFQVGSLNLNDVM : 81
Sso-KDGA  : ----MPEIITPIITPFTKDNRIDKEKLKIHAENLIRK-GIDKLFVNGTTGLGPSLSPEEKLENLKAVYDVTN---KIIFQVGGLNLDDAI : 82
Sto-KDGA  : -----MDIVTPILTPFTKEGKIDVEKLKAHAKFLIDN-GIDLLFVNGTTGLGPALSKEEKLTTLKTIYDVTN---KVIFQVGSLNINDVI : 81
Tte-KDGA  : -----MEIVAPVITTF-RGGRLDPELFANHVKNITSK-GVDVVFVAGTTGLGPALSFQEKKMLAEYCSDITD---RVIFQVGSLNLEESI : 80
Tac-KDGA  : ----MPFKIVPILTPF-RGNEIDGEKFKDHCLKLLRD-GIDMIFVAGTNGMGPALSFQEKKMLAEYCSDITD---RVIFQVGSLNLEESI : 81
Tvo-KDGA  : ----MKLKIVPVVTPF-KEGKIDGEKFIHHCMSLLSV-GVNYVFISGTTGLGPSLSINERMDLIKLSSNFAD---KAIMQVGSMNLEDSI : 81
Pto-KDGA  : ----MPLKIVPVITPF-KNNVLDMDILREHCERLLNS-GIDKVFLAGTNGLGISLDFSEKLRIINELDIDGD---KLIIQIGSLNLEKSM : 81
Eco-NAL   : MATNLRGVMAALLTPFDQQGALDKASLRRLVQFNIQQ-GIDGLYVCGGSTGEAFVQSLSEREQVLEIVAEEAKGKIKLIAHVGCVSTAESQ : 89
Hin-NAL   : -MRDLKGIFSALLVSFNEDGTINEKGLRQIIRHNIDKMKVDGLYVGGSTGENFMLSTEEKKEIFRIAKDEAKDQIALIAQVGSVNLKEAV : 89
Eco-DHDPS : ---MFTGSIVAIVTPMDEKGNVCRASLKKLIDYHVAS-GTSAIVSVGTTGESATLNHDEHADVVMMTLDLADGRIPVIAGTGANATAEAI : 86

             100        *        120        *        140        *        160        *        180
Sac-KDGA  : ELVKFSNEMDILGVSSHSPYYFPRLPEKFLAKYYEEIARISSHSLYIYNYPAATGYDIPPSI---LKSLP--VKGIKDTNQDLAHSLEYK : 166
Sso-KDGA  : RLAKLSKDFDIVGIASYAPYYPPRMSEKHLVKYFKTLCEVSPHPVYLYNYPTATGKDIDAKVAKEIGCF----TGVKDTIENIIHTLDYK : 168
Sto-KDGA  : DLVKASKDFDIVGIASYPPFYFPRLPEKFLLKYFTTIANYSPHSLYIYNYPLATGYDISAKIVYQMKDL---ITGLKDTNQDLSHSLEYK : 168
Tte-KDGA  : ALAKYAESRGAEAVASLPPYYFPRLSERQIAKYFRDLCSAVSIPVFLYNYPAAVGRDVDARAAKELGC----IRGVKDTNESLAHTLAYK : 166
Tac-KDGA  : ELARIARSMNFYAVAALPPFYFGGIPEEWIIRYYTQISEVMET--FIYNYPKFTGYDINADLARKIKKKGGNILGVKDTVSDIDHMMRYK : 169
Tvo-KDGA  : KLARMKAENGIYGISALPPYYFGGIPEEWLIKYYVEISKIIPT--FIYNYPKFTGYDINSSLAIKINKAGGNIVGVKDTIADISHMLSYR : 169
Pto-KDGA  : ELARAARAKGYFAIASVPPYYYSDLDENAIIRYFLDLSRHYKT--IVYNYPKMTGHDVNARMLKRIIDSGGDIIGVKDTVNDIMHMLEYK : 169
Eco-NAL   : QLAASAKRYGFDAVSAVTPFYYPFSFEEHCDHYRAIIDSADGLPMVVYNIPALSGVKLTLDQINTLVTLP-GVGALKQTSGDLYQMEQIR : 178
Hin-NAL   : ELGKYATELGYDCLSAVTPFYYKFSFPEIKHYYDTIIAET-GNNMIVYSIPFLTGVNMGIEQFGELYKNP-KVLGVKFTAGDFYLLERLK : 177
Eco-DHDPS : SLTQRFNDSGIVGCLTVTPYYNRPSQEGLYQHFKAIAEHTD-LPQILYNVPSRTGCDLLPETVGRLAKVK-NIIGIKEATGNLTRVNQIK : 174

                *        200        *        220        *        240        *        260        *
Sac-KDGA  : LNLFG-VKVYNGSNTLIY-YSLLSLDGVVASFTNFIPEVIVKQRDLIKQGKLDDALRLQELINRLADILRKYGSISAIYVLVNEFQGYDV : 254
Sso-KDGA  : RLNPN-MLVYSGSDMLIATVASTGLDGNVAAGSNYLPEVTVTIKKLAMERKIDEALKLQFLHDEVIEASRIFGSLSSNYVLTKYFQGYDL : 257
Sto-KDGA  : ILMPN-LKVYNGSDSLVF-YSLTSLDGSVTAASNYLPHVMKKMKEHITSGQVSKAIELQKLINKALDISRKYGQLSAIYYLVKEFLGYDV : 256
Tte-KDGA  : RYLPQ-ARVYNGSDSLVFASFAVRLDGVVASSANYLPELLAGIRDAVAAGDIERARSLQFLLDEIVESARHIGYAAAVYELVEIFQGYEA : 255
Tac-KDGA  : YLIDD-FMVFSGPSNYILHAAVNGLDGVVAAGNYATEMFVQI---L-DDTARYGRANQIIIDEIFDAASRYGQWASNYSLVRILRGYDA : 254
Tvo-KDGA  : DLGND-FIILSGPSNSIYEALRSGVDGSVAAAGNFMPKTFIELADKF-EAEASF--KLQKYINSVLSIANEYGSWSAYYSLIKILRRYSA : 255
Pto-KDGA  : NIEDD-LIVLAGPSTHIFSALRSGVDGTVSAAGNYMPNIFRYLSDHV-YSNNAF--KLQILINSVIDASAHFGIYSANYSLVKIIKNYYC : 255
Eco-NAL   : REHPD-LVLYNGYDEIFASGLLAGADGGIGSTYNIMGWRYQGIVKALKEGDIQTAQKLQTECNKVIDLLIKTGVFRGLKTVLHYMDVVSV : 267
Hin-NAL   : KAYPN-HLIWAGFDEMMLPAASLGVDGAIGSTFNVNGVRARQIFELTKAGKLAEALEIQHVTNDLIEGILANGLYLTIKELLK-LEGVDA : 265
Eco-DHDPS : ELVSDDFVLLSGDDASALDFMQLGGHGVISVTTNVAARDMAQMCKLAAEEHFAEARVINQRLMPLHNKLFVEPNPIPVKWACKELGLVAT : 264

              280        *        300
Sac-KDGA  : GYPRPPIFPLTDEEALSLKREIEPLKRKIQELVH--- : 288
Sso-KDGA  : GYPRPPIFPLDDEEERQLIKKVEGIRAKLVELKILKE : 294
Sto-KDGA  : GYPRGPIFPLEEDEVKALLSEIQPVKKELERAVS--- : 290
Tte-KDGA  : GEPRGPVYPLDPEEKAWLRAAVAKAKSQLRL------ : 286
Tac-KDGA  : GDPRPPFYPLDDHTEEKLRSEVRSIMQKYGRESKL-- : 289
Tvo-KDGA  : GDSRAPFYPLNALDEEKLRHDVFFVLPEEQRHQ---- : 288
Pto-KDGA  : GEPRPPFYPLTESDENNLKKIIKNILKDGDNV----- : 287
Eco-NAL   : PLCRKPFGP-VDEKYLPELKALAQQLMQERG------ : 297
Hin-NAL   : GYCREPMTSKATEEQLAKAKDLKAKFLS--------- : 293
Eco-DHDPS : DTLRLPMTPITDSGRETVRAALKHAGLL--------- : 292
```

Figure 1. Alignment of the three *Sulfolobus* KDG aldolases with *Thermoproteus tenax* KD(P)G aldolase, hypothetical KDG aldolases from *Thermoplasma acidophilum*, *T. volcanium* and *Picrophilus torridus*, N-acetyl neuraminate lyases from *E. coli* and *Haemophilus influenzae* and the *E. coli* dihydrodipicolinate synthase. Residues that are conserved between the KD(P)G aldolases are indicated with dark shading, while those residues that are also conserved in the hypotheticals and other aldolases are shaded lighter. At the arrow, the conserved catalytic lysine residue. Catalytic residues (▼) and residues forming the putative phosphate binding motif in SacKDGA are indicated (♦).

Stability and activity of Sulfolobus KDG aldolases

For a thorough characterization of their properties the hypothetical KDGA genes from *S. acidocaldarius* and *S. tokodaii* were cloned and expressed in *E. coli* using the pET expression system. The SsoKDGA was produced using a previously generated pET24d-derived construct (Ahmed *et al.* 2005). The three recombinant KDGAs were readily purified by heat incubation and subsequent anion-exchange chromatography. KDGA activity was assayed in the synthesis direction with the substrates pyruvate and D,L-GA producing KDG or its C4-epimer KDGal. The recombinant enzymes from *S. acidocaldarius* and *S. tokodaii* had KDGA activity in the same range as the SsoKDGA. For a thorough comparison between the three KDGAs, their pH and temperature optima, and Michaelis-Menten kinetics were determined. The biochemical properties of SacKDGA and StoKDGA are not very different from the previously studied SsoKDGA (Table 3). All three enzymes had higher activity with GAP than with GA (Table 4), similar to reports by Ahmed *et al.* (2005) and Lamble *et al.* (2005b) on SsoKDGA activity. The stability of these enzymes at high temperature was also determined. The KDGAs show a high thermostability, although the half-life time of SacKDGA at 100°C is relatively short, which may be related to the optimal growth temperature of *S. acidocaldarius*, which is lower than that of the other two species (Grogan 1989). Taken together, the biochemical experiments and the conserved gene context indicate that these enzymes are involved in a modified Entner-Doudoroff pathway, similar to what has recently been reported for the SsoKDGA (Ahmed *et al.* 2005).

Table 3 Comparison of biochemical data of *Sulfolobus* KDGAs

	S. acidocaldarius	*S. solfataricus*	*S. tokodaii*
Optimum pH	6.5	6*	5.5
Optimum temperature °C	~99	95	90
$V_{max (pyruvate)}$ (U/mg)	26.3 (± 0.9)	15.7 (± 0.3)*	17.8 (± 0.7)
$K_{m (pyruvate)}$ (mM)	1.1 (± 0.2)	1.0 (± 0.1)*	0.7 (± 0.1)
$V_{max (GA)}$ (U/mg)	33.1 (± 1.3)	17.1 (± 0.4)*	20.3 (± 0.8)
$K_{m (GA)}$ (mM)	6.3 (± 0.5)	5.2 (± 0.1)*	3.9 (± 0.4)
$T_{1/2}$ (min at 100°C)	18.9 (± 2.1)	120 (150*)	110

* Data from (Buchanan *et al.* 1999)

Table 4 Substrate specificity of *Sulfolobus* KDGAs based on the rate of condensation with pyruvate as donor substrate (units/mg)

Similar results were obtained with StoKDGA. SD values are given in parentheses

Aldehyde acceptor		SacKDGA*		SsoKDGA*	
(D,L)-glyceraldehyde		37.3	(2.0)	15.1	(0.7)
(D,L)-glyceraldehyde-3-P		93.0	(4.6)	46.5	(0.9)
glycolaldehyde		40.6	(1.0)	10.1	(0.7)
glyoxylate		16.4	(0.9)	11.6	(0.2)
acetaldehyde		0		0	
propionaldehyde		0.0055	(0.0007)	0.0058	(0.0011)
(D)-erythrose		5.7	(0.2)	1.4	(0.1)
(L)-erythrose		2.3	(0.1)	2.0	(0.2)
(D)-threose		2.5	(0.1)	1.9	(0.1)
(L)-threose		10.6	(0.5)	2.3	(0.2)
(D)-ribose		0.028	(0.003)	0.015	(0.002)
(L)-arabinose		0.025	(0.002)	0.020	(0.003)
(D)-arabinose		0.010	(0.003)	0.023	(0.002)
(D)-xylose		0.050	(0.003)	0.034	(0.005)

Crystal structure of SacKDGA

Structural information can provide detailed insights into the molecular basis of the catalytic properties of these aldolases. SacKDGA crystallized in two space groups, $P3_121$ and $P6_522$, for which diffraction data were obtained to a maximum resolution of 1.8 Å and 2.5 Å, respectively (Table 5). The structure could be determined by molecular replacement using the coordinates of previously elucidated SsoKDGA (Theodossis *et al.* 2004) (50% amino acid sequence identity) as the starting model. Initial electron density maps were of good quality and allowed the models in space groups $P3_121$ and $P6_522$ to be refined to acceptable R factors of 15.7 and 17.6 %, respectively (Table 5). Both space groups have two molecules in the asymmetric unit, with both of them containing all 288 residues of the native amino acid sequence. The total of four independently refined molecules are nearly identical with pair-wise root-mean-square differences (rmsds) of their Cα atom positions of less than 0.20 Å.

SacKDGA has a tetrameric structure (Fig. 2A) in which each subunit consists of a $(\beta/\alpha)_8$ barrel with a C-terminal helical extension (Fig. 2B). The tetramer forms a ring-like structure with a hollow core of ~ 20 Å in diameter. The active site of each subunit is located at the C-terminal end of the β-barrel (Fig. 2B) and is only accessible from the inside of the tetrameric ring. This structure is not only similar to the structure of SsoKDGA, as evidenced by the rmsd between 288 matching Cα atoms of only 1.0 Å, but it is also similar to the structures of other crystallized members of the NAL subfamily, illustrating the structural conservation within this group of pyruvate-dependent aldolases (Theodossis *et al.* 2004).

Table 5 Crystallographic data collection and refinement statistics on SacKDGA
Values in parentheses are for the highest resolution shell. Rsym = $|I - <I>|/ I$, where I is the observed intensity and $<I>$ the average intensity. R_{cryst} = hkl,work $||$Fobs$|$- k$|$Fcalc$||/$ hkl$|$Fobs$|$ x 100%, where F obs = observed structure factor and F calc = calculated structure factor. Rfree = R calculated with 5% of randomly selected data that were omitted from the refinement.

|	Native, pH 7.5	Native, pH 7.5	Pyruvate, pH 6.3
Space group	$P3_121$	$P6_522$	$P3_121$
Unit cell dimensions (Å)	$a=b$=108.2 c=171.9	$a=b$=109.5 c= 319.6	$a=b$=109.1 c=171.3
Resolution (Å)	30-1.8(1.88-1.8)	36-2.5(2.59-2.5)	50-2.5(2.65-2. 5)
Wavelength (Å)	0.934	0.934	0.933
Unique reflections	107995	36273	39118
Completeness (%)	99.9 (99.8)	90.1 (55.2)	99.6 (97.1)
Redundancy	10.9 (4.0)	33.1(19.8)	6.2 (4.2)
$I/\sigma(I)$	14.1 (2.3)	23.8 (8.4)	17.6 (5.3)
R_{sym} (%)	9.6 (55.9)[*]	15.5 (37.8)	5.3 (34.8)
Refinement			
R_{cryst} (%)	15.7	17.6	15.9
R_{free} (%)	17.9	21.7	19.1
# atoms protein	4634	4598	4616
pyruvate	-	-	10
solvent	412	95	162
Avg. B factor (Å^2) protein	35.5	52.0	52.9
pyruvate	-	-	42.6
solvent	44.2	43.6	50.8
R.m.s. deviations bonds/angles (Å/°)	0.012/1.5	0.008/1.3	0.008/1.3

[*]Value for the highest resolution shell is affected by anisotropy in diffraction
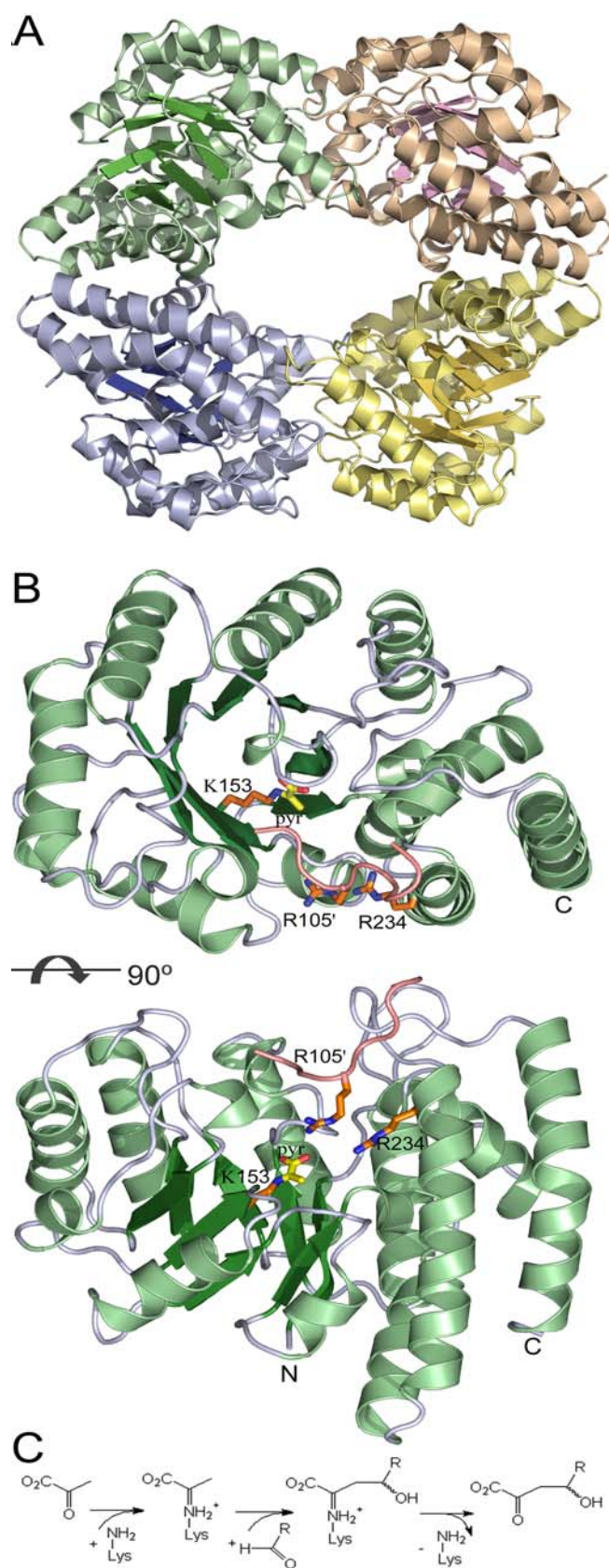
Figure 2. Overall structure of SacKDGA **(A)** SacKDGA tetramer **(B)** Orthogonal views of the SacKDGA monomer. The Schiff-base forming Lys-153 is indicated in stick presentation with pyruvate (yellow) attached. To show the layout of the substrate binding pocket, the loop of the neighbouring subunit carrying Arg-105 is also shown, together with Arg-105 and Arg-234 (orange). **(C)** General scheme of the aldol condensation reaction catalyzed by the KDG aldolases.
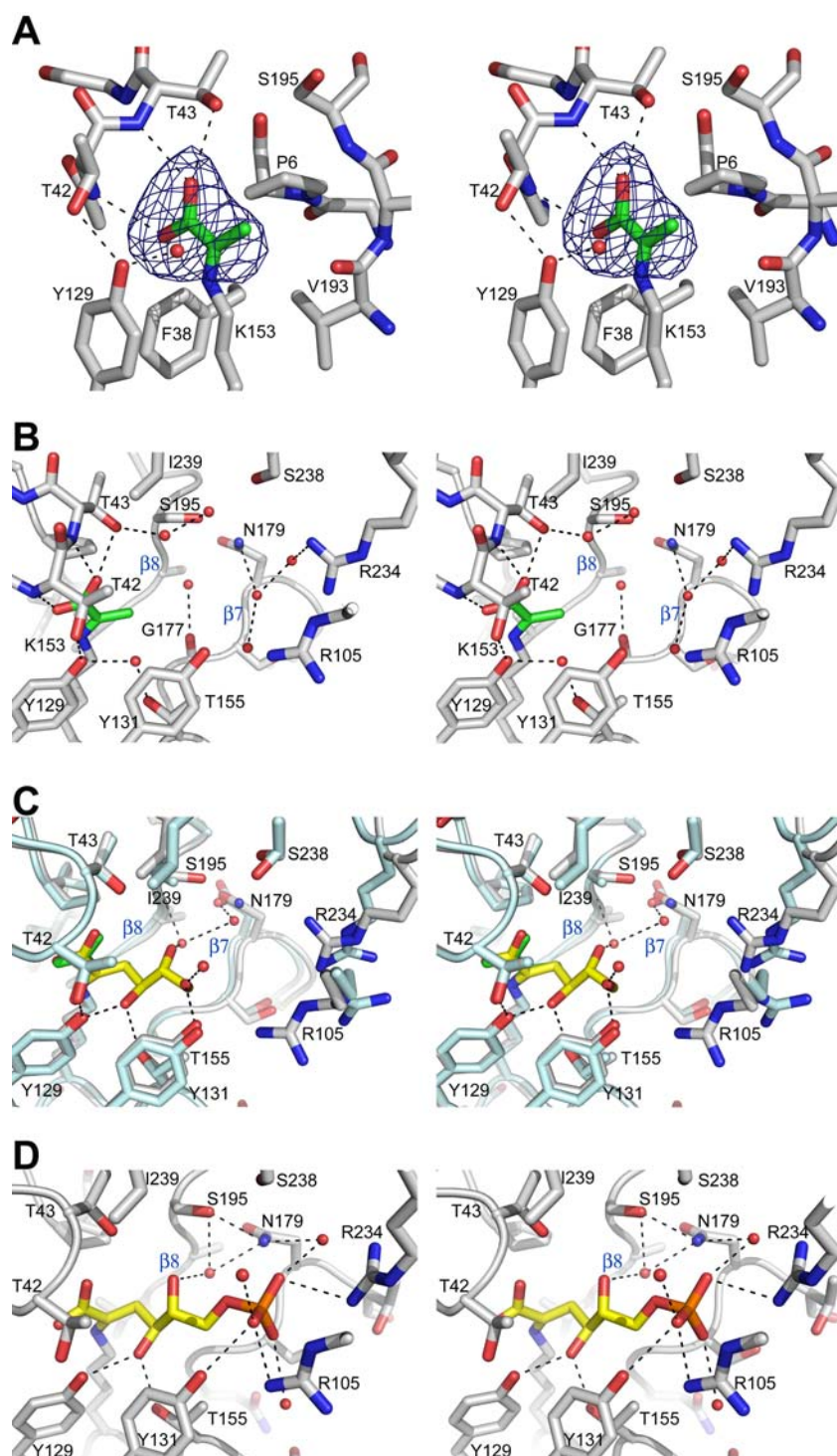
Figure 3. Stereo views of the SacKDGA active site (A) Weighted difference density map of pyruvate (green) bound to Lys-153 in the pyruvate-binding cavity of SacKDGA. Map is contoured at 4 σ. (B) The substrate binding pocket of SacKDGA with pyruvate. Strands β7 and β8 are indicated. (C) Superposition of the Pyr:SacKDGA (green:grey) and KDG:SsoKDGA complex ((Theodossis *et al.* 2004), yellow:cyan). Hydrogen bonding interactions with the pyruvate moiety have been omitted for clarity and the orientation is slightly different from B. Water molecules shown are from the KDG:SsoKDGA complex. (D) KDPG (yellow) modelled in the active site of SacKDGA on the basis of the KDG:SsoKDGA complex (Theodossis *et al.* 2004). Water molecules from the crystal structure of native SacKDGA, which would have favourable interactions with the ligand are also shown. Some hydrogen bonding interactions have been omitted for clarity and the orientation is slightly different from B and C.

The active site is among the most highly conserved regions in the protein. It consists of a pyruvate-binding cavity close to the catalytic Lys-153 (Fig. 3A), which is at the bottom of a wider hydrophilic pocket (Fig. 3B), accessible from within the tetrameric ring (Fig. 2B). The entrance to this hydrophilic pocket is somewhat constricted by the flexible side chains of two arginine residues, Arg-234 and Arg-105' from a neighbouring subunit (Fig. 2B, 3B). The SacKDGA active site residues adopt somewhat different conformations as a function of pH. Most conspicuously, at pH 7.5 the catalytic Lys-153 side chain displays some conformational variability. In its most populated conformation, its side chain points away from the pyruvate binding cavity and buries its amino-group into the interior of the protein in a hydrogen bond with the Ser-96 side chain (not shown). Since SacKDGA has its activity optimum around pH 6.5, data was also collected from a crystal that had been transferred to a cryosolution at pH 6.3 before freezing. At this pH, the lysine is in a conformation that would allow Schiff-base formation and catalysis (data not shown).

To assess the structure of the Schiff-base intermediate with pyruvate a crystal was soaked in a pyruvate-containing solution at pH 6.3, after which data were collected to 2.5 Å resolution. The well-defined pyruvate moiety is buried inside the active site and makes close interactions with backbone atoms of Thr-42, Thr-43, Val-193, Gly-177, and the ring of Pro-6, thereby ensuring the high specificity of the enzyme for pyruvate (Fig. 3A). This part of the active site and the pyruvate-binding mechanism are highly conserved with Tyr-129 participating in both substrate binding and in the substrate-assisted catalysis, by relaying a proton between the pyruvate carboxylate and the aldehyde of the acceptor (Barbosa *et al.* 2000, Theodossis *et al.* 2004).

Despite the different conformations of the Lys-153 side chain, the backbone conformations in the active site region of the SacKDGA structure are rigid and remain almost invariant upon pyruvate binding. As this rigidity was also observed for the SsoKDGA structure (Theodossis *et al.* 2004), it may correlate with thermostability. Interestingly, the active sites of NALs from mesophilic *H. influenzae* and *E. coli* are nearly identical to the KDGA active sites (Barbosa *et al.* 2000, Izard *et al.* 1994), indicating that the backbone conformation itself is not an adaptation to higher temperatures. The thermostable KDGAs appear to have stabilized this region, and therefore the protein, by rigidifying the loop that carries Tyr-129, which displays a higher flexibility in *H. influenzae* and *E. coli* NAL (Barbosa *et al.* 2000, Izard *et al.* 1994).

In contrast to the strict requirement for pyruvate, the ability of KDGA to accept both 2-keto-3-deoxy(-6-phospho)gluconate as well as 2-keto-3-deoxy(-6-phospho)galactonate, in their D- as well as their L-enantiomer conformation, already indicates that the specificity for the acceptor substrate is much more relaxed than for the donor substrate. Previously, the

structures of D-KDG and D-KDGal Schiff-base intermediates in SsoKDGA (Theodossis *et al.* 2004) have revealed the binding mode for these non-phosphorylated substrates during catalysis. The binding modes of the carbon backbone of the glyceraldehyde moieties show considerable variation across the independently determined structures and between D-KDG and D-KDGal. Still, in all cases the C-4 hydroxy group maintains a hydrogen bond with the Tyr-130 (Tyr-129 in SacKDGA) hydroxy group (Fig. 3C), which is important for catalysis. The C-5 and C-6 hydroxy groups show more flexibility and have different hydrogen bonding partners for D-KDG and D-KDGal (Theodossis *et al.* 2004). These hydrogen bonding partners are conserved in the SacKDGA structure as well as in the sequence of StoKDGA (Fig. 1, 3), which is in agreement with their comparable aldolase activity on pyruvate and glyceraldehyde. No structural data is available for the binding of the L-enantiomers of both KDG and KDGal in the KDGA active site, but the hydrophilic nature of the pocket allows their flexible binding in a similar way as the D-enantiomers. This not only provides a structural explanation for the lack of stereocontrol of the KDGAs with their natural non-phosphorylated substrates, but also gives important clues about the applicability of these enzymes with non-physiological substrates.

Sulfolobus KDGAs contain a novel phosphate-binding motif

All KDGAs show higher activity with GAP than with GA (Table 4 and (Ahmed *et al.* 2005)), which may be of physiological relevance in the semi-phosphorylated Entner-Doudoroff metabolic pathway in *Sulfolobus* spp. It was suggested recently that starvation conditions may favour the slower non-phosphorylated route, whereas otherwise a KDG kinase may phosphorylate KDG to KDPG (Lamble *et al.* 2005b). The possible binding modes of these larger phosphorylated substrates were investigated by using the SsoKDGA and SacKDGA intermediate structures (Fig. 3C) as templates. In SacKDGA, the position of a bulky and highly polar phosphate group on the C-6 position is restricted on one side by the apolar side chains of Ile-239 (replaced by Leu in SsoKDGA and StoKDGA) and Thr-42 (Fig. 3C). This side of the pocket would not be able to provide favourable interactions to the phosphate. Model building shows that minor rearrangements of the C-6 moiety would bring the phosphate closer to the entrance of the pocket and within hydrogen bonding distance of the polar side chains of Arg-234, Tyr-131, and, additionally Arg-105' of a neighbouring subunit (Fig. 3D). These three residues are absolutely conserved among the KDGAs with confirmed KDPGA activity from *Sulfolobus* spp. and *T. tenax* (Fig. 1), and form a novel phosphate binding motif. In addition, water-mediated contacts would be possible with the nearby side chains of Ser-238, Asn-179 and Ser-195 (Fig. 3D). The promiscuity of binding of KDGA, which has already been revealed for the GA moiety (Theodossis *et al.* 2004), is also present in

the bound phosphate through these water-mediated contacts and the large conformational flexibility of the Arg-105' and Arg-234 side chains (Fig. 3D). Charge complementarity between the phosphate group and the two conserved arginine residues allows efficient targeting of GAP to the active site and additional enzyme-substrate interactions may also increase substrate affinity. These additional interactions provide a structural explanation for the higher activity observed with GAP and the higher catalytic efficiency with KDPG than KDG. On the other hand, the phosphate group does not influence catalysis directly and does not significantly affect the conformation of the GA moiety (Fig. 3D), in line with the observation that the KDGAs do not absolutely require phosphorylated substrates.

The novel phosphate binding motif is only conserved among the KDGAs from *Sulfolobus* spp. and *T. tenax*. In putative KDGAs identified in other thermophilic archaea like *Thermoplasma* and *Picrophilus torridus*, the two arginine residues of the motif are not conserved (Fig. 1), neither is the genomic context similar to that of the *Sulfolobus* spp. Significantly, these organisms use a non-phosphorylative Entner-Doudoroff pathway (Budgen and Danson 1986, Reher and Schonheit 2006), in which their KDGAs do not require the dual KDGA/KDPGA functionality of KDGAs active in a semi-phosphorylative pathway. To our knowledge, the KDGAs from *Sulfolobus* spp. and *T.* tenax are the only members of the NAL subfamily identified so far that act on phosphorylated substrates and contain this motif. Therefore, the motif may be used to identify aldolases from this family with KDPGA activity.

Intriguingly, many of the enzymes of the $(\beta/\alpha)_8$ fold family that specifically act on phosphorylated substrates contain a conserved phosphate-binding site (Nagano *et al.* 2002), which differs entirely from the phosphate binding motif of the KDGAs. In these enzymes, phosphate moieties are often bound between the diverging ends of strands β7 and β8 by their backbone amide groups (Nagano *et al.* 2002). Our structural analysis reveals that this structural β7- β8 feature is not conserved in the *Sulfolobus* KDGAs. In the SacKDGA structure, a depression exists between strands β7 and β8, where two water molecules reside, hydrogen bonding to Gly-177, the side chain of Ser-178 and the main-chain amides and side chains of Asn-179 and Ser-195 (Fig. 3B). Manual docking indicates that a phosphate group cannot be accommodated at this position; the backbone amide groups in SacKDGA and SsoKDGA are not oriented properly for an optimal interaction with the phosphoryl oxygen atoms and the side chain of Asn-179 in SacKDGA would give steric hindrance (Fig. 3B,D). In addition, Asn-179 is substituted by an aspartate residue in both SsoKDGA and StoKDGA (Fig. 1), which would give unfavourable interactions with a negatively charged phosphate moiety. These structural differences lead us to conclude that this β7-β8 site is not favourable for phosphate binding and support the novel phosphate binding mechanism that is outlined above.

The phosphorylated KDPG is also a substrate for the KDPG aldolases from *E. coli* and *Pseudomonas putida*. These class I aldolases also have a $(\beta/\alpha)_8$ fold, but they do not belong to the NAL subfamily and the details of their catalytic mechanisms are different (Allard *et al.* 2001, Bell *et al.* 2003). The KDPG aldolase from *E. coli* has for example a rmsd of 2.8 Å over 171 matching Cα atoms with SacKDGA. Remarkably, these aldolases do contain a hydrophilic space between the β7 and β8 strands with the structural features of a phosphate-binding site. Significantly, in both the *E. coli* and *P. putida* KDPG aldolase crystal structures a sulphate ion is bound at this β7-β8 site (Allard *et al.* 2001, Bell *et al.* 2003). The overall similarity between sulphate and phosphate ions strongly suggests that, in contrast to the *Sulfolobus* KDGAs, the β7-β8 site functions as a binding site for the phosphate group of their substrate. Interestingly, the KDPG aldolases from *E. coli* and *P. putida* do not perform well without the phosphate group on the GA moiety. This indicates that their active site confers considerably more spatial specificity towards their phosphorylated acceptor substrate than the open hydrophilic binding site of the *Sulfolobus* KDGAs. The strict specificity of the bacterial KDPG aldolases fits their biological function, but makes them less suitable for biotechnological applications. In contrast, the unique phosphate binding motif of the archaeal KDGAs incorporates the flexibility that is necessary for promiscuous substrate binding. This flexibility may extend to, or be easily adapted to, the conversion of a wider range of acceptor substrates.

Structural requirements for acceptor substrates of KDGAs

KDGAs display substrate promiscuity towards their natural aldehyde substrates (Ahmed *et al.* 2005, Lamble *et al.* 2003, Lamble *et al.* 2005b). It was reported recently that SsoKDGA can also accept glyceraldehyde acetonide in the synthesis reaction (Lamble *et al.* 2005a). Furthermore, the structures of SacKDGA and SsoKDGA suggest that the substrate-binding site is able to accept a wide range of acceptor substrates. To identify the structural requirements for aldehydes as acceptor, the activity of SacKDGA, SsoKDGA and StoKDGA was assayed with a range of non-physiological aldehyde acceptors (Table 4, and results not shown). Like their other properties, the substrate specificities of the different *Sulfolobus* KDGAs are quite comparable and will not be discussed separately.

First, the importance of the C-6 moiety was investigated by testing the 2-carbon aldehydes glycolaldehyde and glyoxylate. The available structural information suggests that the binding site should be able to accommodate and orient these smaller substrates through hydrogen bonds with the C-5 hydroxyl group. Indeed, activity rates with these substrates were similar to or only slightly lower than that of GA (Table 4), establishing that the KDGAs do accept substrates with only two carbon atoms. The reduced activity with glyoxylate may be

the result of its resemblance to pyruvate and weak competitive inhibition of Schiff-base formation. In comparison, KDPG aldolases from *P. putida, E. coli* and *Zymomonas mobilis* (Shelton *et al.* 1996) react quite well with glyoxylate, but react with very low rates only with glycolaldehyde. This provides additional evidence that the KDGAs from *Sulfolobus* can perform well without a negative charge on the substrate and establishes that they have a catalytic range which is distinctly different from the bacterial KDPG aldolases.

After having established that the C-6 hydroxyl group is dispensable, the importance of the hydroxyl group at the C-5 position was investigated. In the SsoKDGA KDG and KDGal complexes, this hydroxyl group has hydrogen bonding interactions with either Tyr-131 or water-mediated contacts with Gly-177 and Ser-195 ((Theodossis *et al.* 2004), see also Fig. 3C). Two aldehydes were tested that do not contain such a hydroxyl group at the C-5, and this showed that acetaldehyde is not accepted and propionaldehyde only at a very low rate (<0.1% of GA) (Table 4). This suggests that the polar hydroxyl substituent at the C-5 position is indispensable for proper binding and orientation of the aldehyde in the open hydrophilic binding pocket.

Next, the ability of the KDGAs to accept aldehydes with one more carbon atom than their natural substrates was explored using the simple four-carbon aldotetroses in all possible stereoisomers. The large hydrophilic binding pocket should be able to provide favourable interactions for the additional hydroxyl groups of these sugars. Indeed, all aldotetroses tested were readily accepted in synthesis reactions at 5-25 % of the GA rate. As all stereoisomers are accepted, albeit with a substantially higher rate for the reaction with L-threose, the KDGAs show a limited enantioselectivity on these larger substrates. The relative rates are again higher than those of the different KDPG aldolases of *P. putida, E. coli* and *Z. mobilis* (Shelton *et al.* 1996), which accepted these compounds at 0-0.5% of the GAP rate only.

In contrast, the larger aldopentoses that were tested were accepted at low rates (<0.5% of GA) (Table 4). The binding site should, however, be large enough to accommodate these five-carbon acceptors. The low rates with the aldopentoses are likely due to a reduced effective concentration of the reactive open-chain conformation of these sugars, which is only ~ 0.1 % of the total aldopentose concentration at 28°C (Drew *et al.* 1998). Significantly, the activity of the cleavage of the 2-keto-3-deoxyoctonate (KDO) (5 mM) to D-arabinose and pyruvate, is much higher than its condensation, reaching 0.146 ±0.006 U/mg (SacKDGA), 0.470 ±0.040 U/mg (SsoKDGA) and 0.393 ±0.025 U/mg (StoKDGA). Activity appears to be correlated with the effective concentration of the open-chain forms of the substrates, which is probably much higher for the ketone KDO than for the aldehyde D-arabinose, suggesting that the KDGAs do not catalyze the ring-opening step of sugar substrates. Ring opening as the rate-limiting step may also play a role with the aldotetroses (Serianni *et al.* 1982), and product

cyclization may effectively limit the rate of the reverse aldol reaction (Lamble *et al.* 2005b, Midelfort *et al.* 1977).

Taken together, these results define the substrate requirements, or lack thereof, of these KDGAs. The substrate requirements are primarily restricted to the substituents of the carbon atom immediately adjacent to the aldehyde moiety, making glycolaldehyde the smallest acceptor substrate that can still be used by the KDGAs. Despite the absence of enantioselectivity in the condensation reaction, the hydroxyl group at the C-5 position of the synthesis product is indispensable for catalysis by providing a significant binding affinity of the aldehyde substrate. Our measurements do not rule out the possibility that additional hydrogen bonding interactions of other parts of an aldehyde acceptor may partially compensate for the decrease in binding affinity. The structure of the binding sites of SacKDGA and SsoKDGA combined with our activity experiments explain the general lack of specific requirements on the structure of the acceptor beyond the glycolaldehyde moiety. This suggests that these enzymes should be applicable in the biosynthesis of a broad range of compounds which carry this structural motif.

The three different *Sulfolobus* KDGAs are not the only NAL family members which accept several different (non-phosphorylated) aldehydes in the condensation reaction with pyruvate. Other members of the NAL subfamily also have broad acceptor specificity, indicating the general versatility of their binding sites. Strikingly, although their binding sites are very similar, acceptors are bound very differently in different enzymes. In NAL, for example, substrate analogues may be accommodated in a different part of the active site pocket (Barbosa *et al.* 2000). Given the similarity between active sites and substrates, it was previously suggested that the substrates sialic acid and KDG would bind in a similar way and that substrate specificity within the NAL subfamily would be modulated in a subtle way by only a few amino acid mutations (Barbosa *et al.* 2000). The structural and biochemical evidence on SacKDGA and SsoKDGA (Theodossis *et al.* 2004) now give striking illustrations of how structurally similar binding sites may bind substrates in entirely different ways. This suggests that a limited number of mutations in the active site may already lead to novel substrate specificities. This has recently been demonstrated by the creation of NAL enzymes with DHDPS activity (Joerger *et al.* 2003) or with L-KDO aldolase activity (Hsu *et al.* 2005). The thermostability, ease of production and well-characterized acceptor specificity of the *Sulfolobus* KDGAs make these enzymes highly suitable platforms for such an approach.

Acknowledgements

Chapter 4

Optimizing low-temperature activity of *Sulfolobus acidocaldarius* 2-keto-3-deoxygluconate aldolase

Suzanne Wolterink-van Loo, Marco A.J. Siemerink, Georgios Perrakis, Thijs Kaper and John van der Oost

Summary

Sulfolobus acidocaldarius 2-keto-3-deoxygluconate aldolase (SacKDGA) displays optimal activity at 95°C and is studied as a model enzyme for aldol condensation reactions. For application of SacKDGA at lower temperatures, a library of randomly generated mutants was screened for improved synthesis of 2-keto-3-deoxygluconate from pyruvate and glyceraldehyde at the sub-optimal temperature of 50°C. The single mutant SacKDGA-V193A displayed a three-fold increased activity compared to wild type SacKDGA. The K_M and k_{cat} had equally increased and as such the catalytic efficiency of SacKDGA-V193A remained unchanged. The increased specific activity at 40-60°C of this mutant appeared to be increased not only for the condensation of pyruvate with glyceraldehyde but also with some other relatively small acceptor substrates. The optimal temperature for activity of SacKDGA-V193A had decreased compared to the wild type enzyme, but the stability of the mutant was like the wild type, indicating that activity and stability were uncoupled. Val193 has Van der Waals interactions with Lys153, which covalently binds the substrate during catalysis. The mutation V193A introduced space close to this essential residue and the increased activity of the mutant presumably resulted from increased flexibility of Lys153. The increased activity of SacKDGA-V193A with unaffected stability demonstrates the potential of optimization of extremely thermostable aldolases for synthesis reactions at moderate temperatures.

Introduction

The aldol reaction leads to the formation of a new carbon-carbon bond, generating a polyhydroxylated product with at least one new chiral centre. Aldolases are enzymes that reversibly catalyze the asymmetric aldol condensation reaction and are therefore of interest for bio-organic synthesis of chiral compounds (Bolt *et al.* 2008, Dean *et al.* 2007). In contrast to traditional organic synthesis, enzymatic aldol catalysis is usually stereospecific, operates under mild reaction conditions and does not require any protection and deprotection steps. Unfortunately, applications of aldolases are often hampered by high enzyme production costs, limited protein stability, and the requirement of expensive (phosphorylated) substrates (Dean *et al.* 2007). The 2-keto-3-deoxygluconate aldolases (KDGA) found in *Sulfolobus* species are very stable enzymes with high activities on (cheap) non-phosphorylated substrates (Buchanan *et al.* 1999, Wolterink-van Loo *et al.* 2007). These class I aldolases catalyze the reversible reaction of D,L-glyceraldehyde + pyruvate $\rightleftharpoons$ D,L-2-keto-3-deoxygluconate (KDG), in which a Schiff base intermediate is formed between the substrate and the catalytic lysine residue (Lamble *et al.* 2005b). It has been shown that these enzymes can accept a wide variety of (aldehyde) substrates, and therefore have great potential for biotechnological applications (Lamble *et al.* 2005a, Lamble *et al.* 2003, Lamble *et al.* 2007, Wolterink-van Loo *et al.* 2007).

Enzymes have been optimized by natural evolution to perform optimally in a certain biological context. Using these enzymes in industrial settings (e.g. in bioreactors), generally under rather distinct conditions, often requires optimization with respect to activity, stability, selectivity and/or specificity (Reetz 2001). In the past decades several techniques, including rational design and random mutagenesis, have been developed to improve enzymes for desired applications (Sen *et al.* 2007). Enzyme catalytic rates are the result of general protein features and can be governed by long-range interactions, which are usually difficult to address by rational protein design (Kihara 2005). Therefore, the random approach of laboratory evolution seems particularly suitable for improving catalytic rates of an enzyme. Generally, diversity is created by error-prone PCR, followed by an activity screen or selection of mutants to identify the best mutants. Several examples of enzyme features that are optimized by random methods include stability (Brouns *et al.* 2005, Shao *et al.* 1998), substrate specificity (Fishman *et al.* 2004, van den Heuvel *et al.* 2004, Xu *et al.* 2003), enantioselectivity (Lipovsek *et al.* 2007, van Loo *et al.* 2004) and low-temperature catalysis (Lebbink *et al.* 2000, Machielsen *et al.* 2008, Merz *et al.* 2000, Sriprapundh *et al.* 2003). Many laboratory evolution studies of aldolases have been directed towards broadening of substrate specificity

or improvement of enantioselectivity (Bolt *et al.* 2008) and some report on increasing aldolase thermostability (Hao and Berry 2004).

In this study random mutations were introduced into the thermostable KDGA of *Sulfolobus acidocaldarius* using error-prone PCR (epPCR). A library of 1500 variants was screened for enhanced condensation activity using pyruvate and D,L-glyceraldehyde at 50°C and retention of kinetic stability. A single mutant, SacKDGA-V193A, showed enhanced activity and was analyzed in detail for its biochemical properties and stability. This is an example of optimization of extreme thermostable aldolases for application in cascade synthetic reactions with mesophilic enzymes. To our knowledge, it is the first report on increasing low-temperature activity of (hyperthermophilic) aldolases.

Materials and methods

Chemicals, strains, plasmids, and media

All chemicals used are of analytical grade and are purchased from Sigma-Aldrich (Zwijndrecht, Netherlands). A previously constructed vector pWUR193 (Wolterink-van Loo *et al.* 2007), a pET24d derivative containing the SacKDGA gene (Saci_0225), was used to produce SacKDGA. *Escherichia coli* XL1-Blue cells were used for cloning and *E. coli* BL21(DE3) cells were used for expression. The growth medium consisted of LB (1% tryptone, 0.5% yeast extract, 0.5% NaCl) supplemented with 50μg/ml kanamycin, unless stated otherwise. Growth was performed at 37°C and liquid cultures were incubated in erlenmeyer flasks in a rotary shaker at 100 rpm.

Construction of SackdgA random library

Random mutations were introduced by PCR using two DNA polymerases in parallel to generate a library with balanced mutation ratio. *Taq* polymerase which preferentially mutates As and Ts and Mutazyme, which mutates more Gs and Cs, were used. Error-prone PCR with *Taq* DNA polymerase was performed using skewed nucleotide concentrations, using the primers BG417 (5'-CTTTAAGAAGGAGATATACCATG-3') and BG1852 (5'-GGCCGCAACCTTGTCGACTTA-3') and the pWUR193 plasmid as template DNA. The 50 μl reaction mixtures contained 10 ng template DNA, 200 ng forward primer, 200 ng reverse primer, 0.2 mM dATP and dGTP, 1 mM dCTP and dTTP, 0,5 mM $MnCl_2$ and 5U REDTaq DNA polymerase (Sigma) in REDTaq reaction buffer. Additionally, the GeneMorph Random mutagenesis kit (Stratagene) was used for the epPCR with the Mutazyme DNA polymerase using the same primers and template DNA according to the manufacturer's instructions, aiming for a mutation rate of 3-7 mutations per kb.

The PCR products obtained were purified, double digested with *NcoI* and *SalI* and ligated into equally digested pET24d using T4 DNA ligase. The different ligation mixtures were transformed to *E. coli* XL-1 blue cells, plated and grown overnight. Finally a total of 1514 colonies were obtained, 1084 originated from the *Taq* DNA polymerase PCR and 430 from the Mutazyme DNA polymerase PCR. Plasmid DNA was collectively isolated from all colonies, resulting in a DNA-plasmid library. Random sequencing of the constructs revealed a mutation rate of 1 per gene on average.

Screening of random SacKDGA variants

For construction of the expression library BL21(DE3) expression cells were transformed with the plasmid library, plated on Qtray LB-agar plates (25cm x 25cm) and incubated overnight. Single colonies were picked with a colony picker (QPix, Genetix) and used to inoculate 96-wells plates (Costar 3370; Corning Inc.) with 200µl medium per well supplemented with 0.1 mM IPTG. In each plate, two wells were manually inoculated with the negative control (pET24d in BL21(DE3)) and positive control (pWUR193 in BL21(DE3)). After 16 hrs of incubation at 37°C, with shaking, the OD_{600} (Spectramax Plus384; Sopachem b.v.) was recorded. A replica plate was made by duplication of the induced plates in fresh 96-well plates containing 200 µl medium with 10% glycerol per well, incubation overnight (37°C, and shaking) and subsequent storage at -80°C for future use. The induced plates were centrifuged to pellet the cells in a Sigma 4-15C centrifuge for 30 min at 1000 x g. The pellets were stored at -20°C until further use. Cells were lysed by addition of 100 µl 25% B-PER (Pierce) in 50 mM sodiumphosphate buffer (pH 6.0) to the pellets and additional shaking at 700 rpm for 15 min at 37°C.

Aldolase reactions were performed in 2-ml deep-well plates (Greiner) containing per well 250 µl 50 mM sodiumphosphate buffer pH 6.0, 10 mM pyruvate, 20 mM D,L-glyceraldehyde, and 40 µl lysed cells. The plates were incubated for 60 min at 50°C and placed into the freezer (-20°C) afterwards to stop the reaction. Reactions were diluted three times by addition of 500 µl water. The remaining amount of pyruvate was then determined with L-lactate dehydrogenase (LDH) / NADH assay. 190 µl LDH assay-buffer (containing 250 mM Tris-HCl buffer pH 7.5, 0.2 mM NADH, 5 U/ml LDH (in excess)) was transferred to each well of a 96-wells plate. The OD_{340} was measured, and then 10 µl diluted sample of the aldolase reaction was added and incubated at 37°C for 15 minutes. An end measurement of the OD_{340} was performed to measure difference in NADH absorption and hence differences in pyruvate concentration of the samples.

DNA sequencing

Plasmids were isolated using the Qiagen plasmid isolation kit according to the manufacturer's protocol. Inserts of the isolated plasmids were sequenced by single primer extension reaction using the T7 primer set (Forward: 5′-AATACGACTCACTATAG-3', reverse: 5'-GCTAGTTATTGCTCAGCGG-3').

Production and purification of SacKDGA variants

BL21(DE3) cells, containing a (mutated) pWUR193 plasmid, were incubated in 3 ml medium and grown overnight. One liter cultures were subsequently started by transfer of 0.5 ml of overnight cell culture. After 4 hours of growth (OD_{600} ~ 1.0) IPTG was added to a final concentration of 0.1 mM and the cultures were grown further overnight. Cells were harvested by centrifuging at 3800g for 15 minutes at 4°C. Pelleted cells were resuspended in 20 ml of 50 mM HEPES/20 mM KCl pH 7.0 and subsequently lysed by passage through a French Press (three times). Centrifugation (17000g for 30 min at 4°C) yielded the cell free extract (CFE). This was incubated at 75°C for 30 min and centrifuged again (same conditions) to obtain heat stable cell free extract (HSCFE). SacKDGA was further purified on an ÄKTA FLPC using anion-exchange chromatography as described previously (Wolterink-van Loo *et al.* 2007). Fractions containing aldolase were pooled. The enzyme purity was checked by SDS-PAGE, staining the proteins with Coomassie brilliant blue. Protein concentrations were determined with the BioRad – Bradford method.

Aldol condensation activity measurements

KDGA activity was measured by detection of the condensation product, 2-keto-3-deoxygluconate, with the thiobarbituric acid (TBA) assay as described previously (Wolterink-van Loo *et al.* 2007) performing the standard assay at pH 6.0 and 50°C, using 50 mM pyruvate and 20 mM D,L-glyceraldehyde.

Aldolase activities with different aldehyde acceptors were measured by assaying the remaining pyruvate in a LDH assay in triplo. Reactions of 150 μl volume containing 50 mM sodium phosphate buffer, pH 6, 10 mM pyruvate and 20 mM aldehyde and 1.5 μg enzyme were incubated at 50°C for 20-90 minutes. The reactions were stopped by rapidly cooling the reaction mixtures on ice. From 985 μl assay buffer (containing 100 mM Tris/HCl pH 7.5, 0.16 mM NADH and ~3 units LDH) the starting absorbance at 340 nm was read. After addition of 15 μl sample and incubation of 5 minutes the absorbance was read again. The absorbance difference in NADH (molecular extinction coefficient 6.22 mM^{-1} cm^{-1} (Horecker and Kornberg 1948)) was taken to calculate the amount of unconsumed pyruvate.

Thermostability of the enzymes at 90°C was checked as follows. KDGA was diluted to 40 µg/ml in 50 mM sodium phosphate, pH 6, divided into 175 µl aliquots and closed into HPLC-vials, extra sealed with nylon. Vials were immersed in a 90°C water bath and sampled at different time points and immediately cooled on ice. Residual activity of each sample was measured with the TBA assay.

Results and discussion

Construction of a random mutant *SackdgA* library

Synthetic cascade reactions in which aldolases are involved are being performed either separately, or as one-pot reactions (Sugiyama *et al.* 2007, Yu *et al.* 2006). Obviously, combinations of different enzymes often requires establishing a compromise of the chemical and physical reaction conditions (Jennewein *et al.* 2006, Littlechild *et al.* 2007, Pollard and Woodley 2007, Straathof *et al.* 2002). When combining enzymes from thermophilic and mesophilic organisms, optimization of either one of them is required.

The KDGA from *S. acidocaldarius* (SacKDGA) is an extremely stable pyruvate-dependent aldolase, with a broad substrate specificity and high affinity for non-phosphorylated substrates (Wolterink-van Loo *et al.* 2007). A potential application of this enzyme would be a cascade with an available mesophilic alditol oxidase (Heuts *et al.* 2007), in which the latter enzyme provides the aldehyde substrate for the aldolase. Since the SacKDGA has an optimal activity at 95°C and at 37°C SacKDGA activity is only approximately 5%, we decided to improve SacKDGA activity at sub-optimal temperatures, preferentially with retention of its stability. As little is known about the molecular basis that determines activity at different temperatures, a random approach was used to optimize low-temperature activity of SacKDGA. Diversity was generated in the SacKDGA gene by using error prone PCR (epPCR) and a library with on average 1 mutation per gene was obtained. A plasmid library of 1500 clones was isolated from *E. coli* XL-1 Blue cells, which was used for transformation of *E. coli* BL21(DE3) for expression. A total of 5000 transformants were screened to ensure sufficient coverage of the genetic diversity in the *kdgA* mutant library.

Screening for increased condensation activity at 50°C of SacKDGA

The desired operating temperature for biocascade reactions in combination with mesophilic enzymes is 37°C (or lower). The incubation time necessary to detect any wild type SacKDGA activity at this temperature were in the order of hours, which was deemed incompatible with a high-throughput screening strategy. Therefore, the mutant KDGA library was screened for aldol condensation activity at 50°C, because at this temperature a small, but significant amount of activity of the wild type could be detected in a relatively short assay time (60 min).

The mutants were screened for their ability to convert D,L-glyceraldehyde and pyruvate to D,L-2-keto-3-deoxygluconate. Since protein expression levels were not taken into account, winners in the screening procedure would be more active than wild type due to increased activity or increased enzyme production levels. Initially, 162 mutants with up to 4-fold increased activities were identified. However, screening in triplicate proved most candidates to be false positives. This was attributed to the indirect screening method for measuring residual pyruvate, which accumulated more inaccuracies compared to screening for condensation product. The effect of other pyruvate consuming/producing activities in the *E. coli* extract were taken into account by subtracting 'activity' of BL21(DE3) with an empty pET24d vector. Finally 12 clones were selected for further testing and were recultivated, their plasmid DNA was isolated for sequencing, and used to re-transform BL21(DE3) to obtain the proteins for further analyses. Wild type SacKDGA and the 12 selected clones were produced on a one-liter scale. Heat stable cell free extracts (HSCFE) were prepared from the cultures, and the variants were compared with activity of the wild type using the TBA assay (measuring the amount of condensation product formed) (Fig. 1). Most clones had equal or less activity than wild type KDGA and sequencing revealed that these either contained no amino acid change or a single mutation of a solvent exposed residue (Table 1). Variants with lower activity had less protein in the HSCFE (data not shown). Variant t34D12 appeared to be on average 1.5 times more active than wild type KDGA, but did not contain any mutations. t34B9 was 1.5 times more active as well and contained mutation N168D, located at the protein surface. The most active variant was t37F3 with a 3-fold increased activity compared to wild type SacKDGA. This variant contained a single mutation of buried Val193 to alanine.

Table 1: Mutations of the twelve selected clones and their position in SacKDGA

| Clone	Mutation	Place
t32C5	M167S	Start of alfa helix, solvent exposed
m5G7	no mutation	-
t34D12	silent mutation CUU--> CUG, possible positive effect on expression	-
t37F3	V193A	Buried, close to catalytic Lys153
t34B9	N168D	In alfa helix, solvent exposed
t26D4	no mutation	-
t19H9	no mutation	-
t1A1	I121T	In alfa helix, solvent exposed
t1G1	K51R	Start of alfa helix, solvent exposed
m6H8	G171E	In a loop, solvent exposed
t34C12	R228G	In alfa helix, solvent exposed
t36F8	silent mutation AAA--> AAG	-
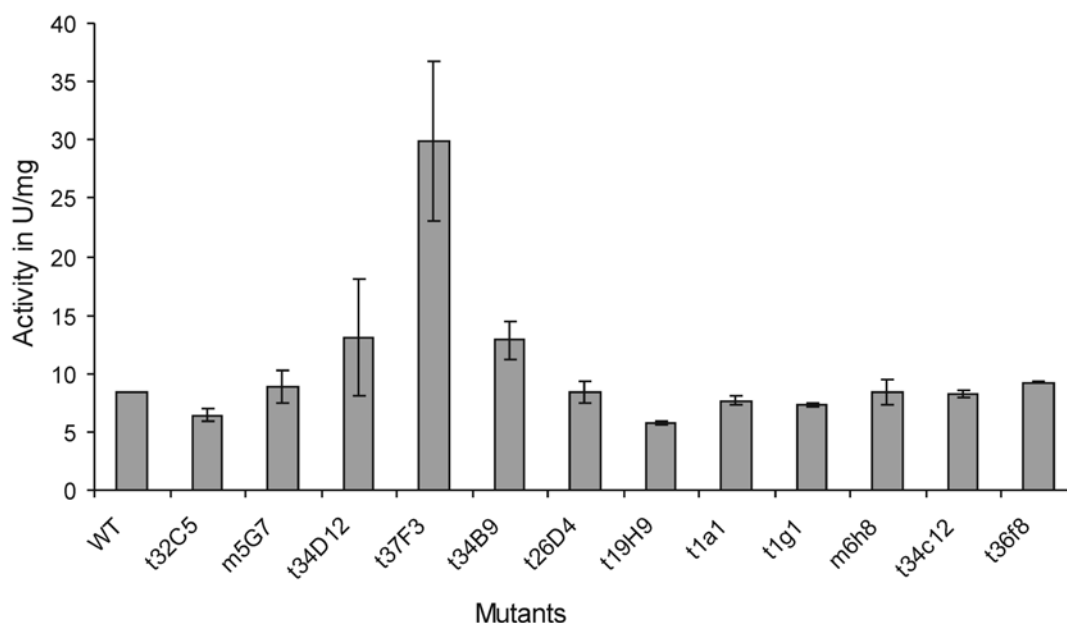
Figure 1. Activity of HSCFE of 12 selected mutants at 50°C, condensation of D,L-glyceraldehyde (20 mM) and pyruvate (50 mM).

Structural analysis of SacKDGA-V193A

Variant t37F3, containing mutation V193A, was 3-fold more active compared to wild type SacKDGA. Val193 is well conserved (but not completely) in KDGAs from hyperthermophilic organisms (Fig. 2). In wild type SacKDGA, Val193 has Van der waals interactions with the catalytic lysine (Lys153) that forms a Schiff base intermediate with the substrate (Fig. 3). Mutation V193A created space in close proximity to Lys153 and possibly allowed a higher degree of mobility of the active site compared to wild type SacKDGA. Similar mutations, creating space in close proximity to the active site, have resulted in increased activity at sub-optimal temperatures of various proteins such as *Pyrococcus furiosus* β-glucosidase (Lebbink *et al.* 2000), *Thermotoga neapolitana* xylose isomerase (Sriprapundh *et al.* 2003), and *Sulfolobus solfataricus* indoleglycerol phosphate synthase (Merz *et al.* 2000). In addition, a natural cold adapted elastase (Papaleo *et al.* 2006) has increased local flexibility (in loop regions) compared to its mesophilic counterparts.

Figure 2. Partial alignment from different pyruvate-dependent aldolases (KDGA) from Sulfolobus acidocaldarius (Sac), S. solfataricus (Sso), S. tokodaii (Sto), Thermoproteus tenax (Tte), Picrophilus torridus (Pto), Thermoplasma acidophilum (Tac), T. volcanium (Tvo), Pyrococcus furiosus (Pfu), P. horikoshii (Pho), Thermotoga maritima (Tma), E. coli (Eco) and Heamophilus influenzae (Hin). The valine residue (at the arrow) is conserved in most crenarchaeal KDGAs (Pyrobaculum aerophilum is an exception) and some euryarchaea and otherwise conservatively replaced by an isoleucine in (hypothetical, indicated by an asterisk) dihydrodipicolinate synthases (DHDPS) and Neuraminate Lyases (NAL).

```
                           ↓
SacKdgA    :  LSLDGVVASFTN  199
SsoKdgA    :  TGLDGNVAAGSN  202
StoKdgA    :  TSLDGSVTAASN  201
TteKdgA    :  VRLDGVVASSAN  200
PtoKdgA    :  SGVDGTVSAAGN  197
TacKdgA    :  NGLDGAVAAAGN  195
Tvodhdps*  :  SGVDGSVAAAGN  197
Pfudhdps*  :  LGGDGGIMACAN  209
Phodhdps*  :  LGGDGGIMACAN  208
Tmadhdps   :  AGGDGVISVVSN  212
Ecodhdps   :  LGGHGVISVTAN  209
EcoNAL     :  AGADGGIGSTYN  212
HinNAL     :  LGVDGAIGSTFN  211
```
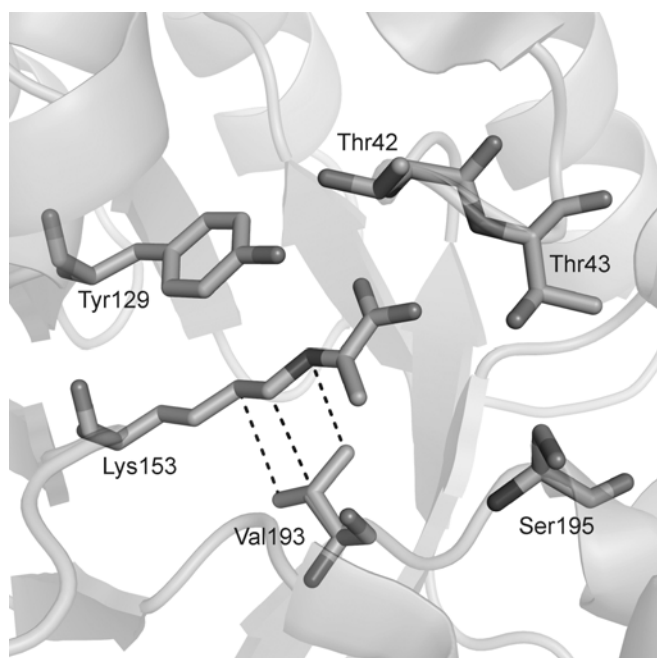

Figure 3. Close-up view of the active site of SacKDGA with pyruvate covalently bound to Lys153 (pdb: 2nuy). Van der waals interactions between the Lys153 side chain and Val193 have been indicated with dotted lines. The figure was generated using PyMol.

Figure 4. Temperature dependent activity of purified SacKDGA wild type (♦) and SacKDGA-V193A (■) Reaction conditions 10 mM pyruvate, 20 mM glyceraldehyde.
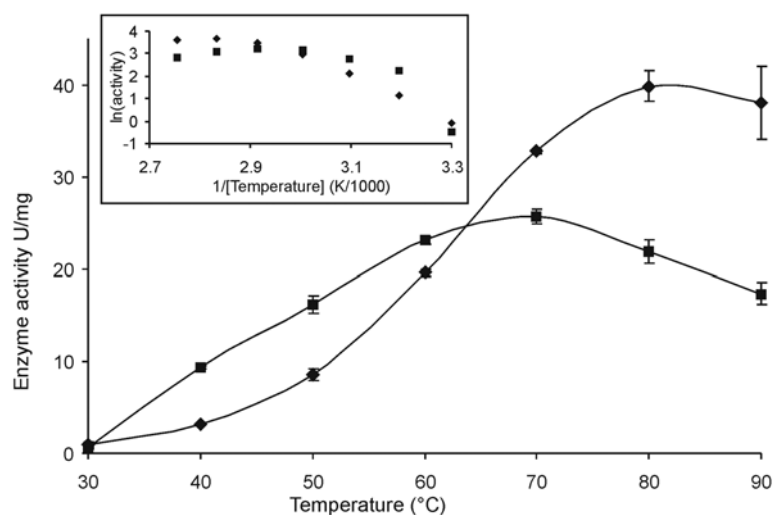
**Optimal temperature for activity and kinetic stability**

Wild type SacKDGA and variant V193A were purified to homogeneity for characterization (data not shown). To study the effect of temperature on the activity of SacKDGA-V193A and to determine whether its activity was also increased at moderate temperatures, SacKDGA-V193A and wild type SacKDGA were assayed in the temperature range of 30°C to 90°C (Fig. 4). SacKDGA-V193A was more active than wild type SacKDGA between 30°C and 60°C. The optimum temperature for activity has shifted from 90°C in wild type SacKDGA to 70°C for the mutant, similar to results obtained for a *P. furiosus* β-glucosidase CelB variant reported by Lebbink *et al.* (2000). Although at 30°C the activities of both enzymes are just detectable above background level, it was found that mutant V193A is significantly more active at the temperature range of 35 – 60°C. In addition, it is concluded that V193 contributes significantly to the activity of SacKDGA at the optimal growth temperature of *S. acidocaldarius* (80°C), which agrees well with the fact that it is relatively well conserved in homologous enzymes from hyperthermophiles; however, in some orthologs valine is replaced by isoleucine, which probably gives comparable restricted flexibility of the active site lysine residue.

An attractive feature for use of SacKDGA in biotechnological applications is its extreme thermostability. Obviously, improving the enzyme's low-temperature activity should not interfere with protein stability. Variant SacKDGA-V193A remained soluble during purification, which included a heat incubation step of 30 min 75°C, indicating that the variant had retained significant residual thermostability. Furthermore, inactivation studies at 90°C of both wild type and variant aldolase showed that they had equal activity half-lives of around 2 hours. Therefore, mutation V193A affected only the activity of SacKDGA, which demonstrates that activity and thermostability can be uncoupled for thermostable Class I aldolases. Previously, similar observations have been made for mesophilic class II fructose-bisphosphate aldolase, in which case aldolase thermostability was improved by directed evolution without reduction in catalysis at mesophilic temperatures (Hao and Berry 2004).

Table 2. Biochemical data for SacKDGA and mutant V193A, determined at 50°C

	SacKDGA[1]	SacKDGA-V193A[1]
k_{cat} (s^{-1})	4.4 (0.2)	15.7 (0.5)
$K_{M(GA)}$ mM	2.1 (0.3)	5.4 (0.5)
$k_{cat}/K_{M(GA)}$ s^{-1}mM^{-1}	2.1	2.9
k_{cat} (s^{-1})	5.5 (0.1)	15.0 (0.5)
$K_{M(pyr)}$ mM	0.8 (0)	5.4 (0.6)
$k_{cat}/K_{M(pyr)}$ s^{-1}mM^{-1}	7.0	2.8

[1] values between brackets are standard deviations.
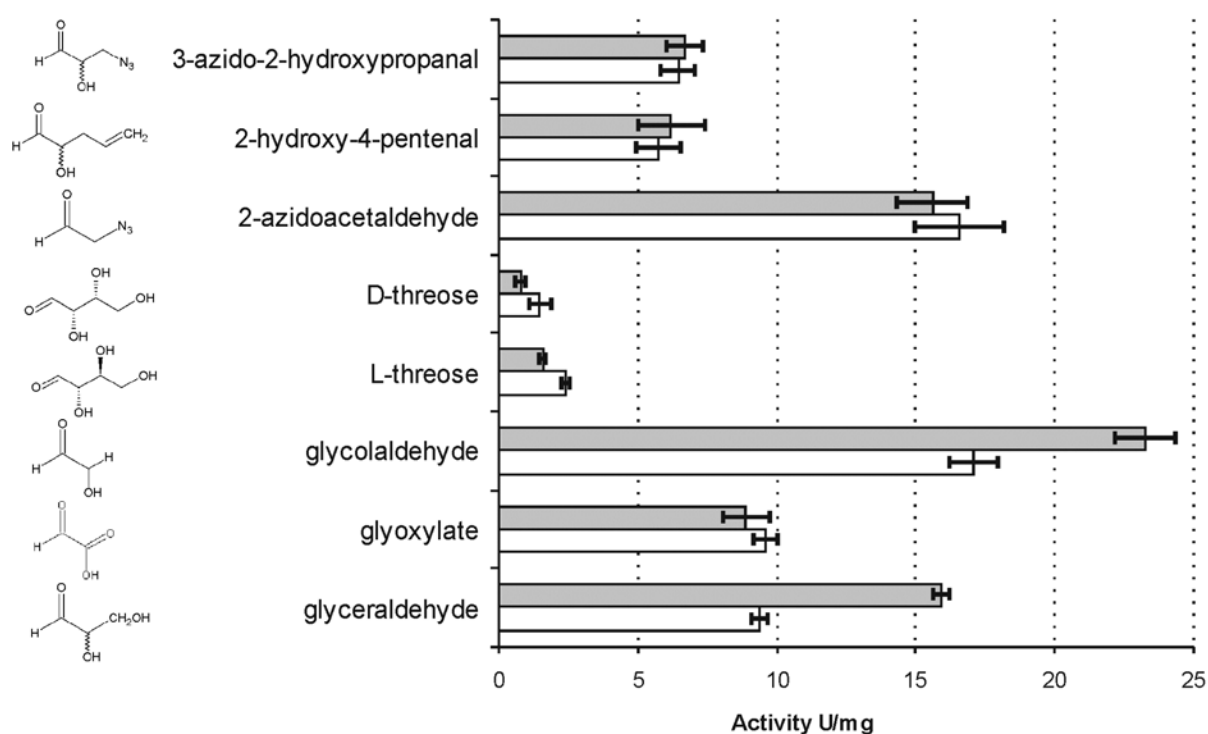
Figure 5. Activity of purified SacKDGA (white bars) and SacKDGA-V193A (grey bars) on different (unnatural) aldehyde-substrates at 50°C. Activities were determined in 50 mM sodium phosphate buffer (pH 6.0), 10 mM pyruvate and 20 mM aldehyde.

Kinetic characterization at 50°C of mutant and wild type SacKDGA

To study the effect of mutation V193A on catalysis, the kinetic reaction constants of SacKDGA-V193A and wild type aldolase for condensation of pyruvate and D,L-glyceraldehyde were determined at 50°C (Table 2). The turnover, k_{cat}, of the mutant enzyme, had increased 3 times. The K_M had increased as well, 3 times for glyceraldehyde and 5 times for pyruvate, resulting in a similar catalytic efficiency value for mutant and wild type. The kinetic constants confirm the validity of the high-throughput screen, which was performed under saturating substrate conditions.

Substrate specificity of mutant and wild type SacKDGA

S. acidocaldarius KDGA is part of the Entner-Doudoroff pathway for degradation of glucose (Selig *et al.* 1997) and probably galactose (Lamble *et al.* 2003) to pyruvate and glyceraldehyde. The enzyme displays broad substrate specificity and can accept various non-natural acceptors in the condensation reaction with pyruvate (Wolterink-van Loo *et al.* 2007). Therefore, we investigated the effect of mutation V193A on the condensation activity at 50°C using pyruvate and alternative aldehyde acceptors (Fig. 5). Wild type and variant aldolases were equally active on half of the tested acceptor substrates. Mutation V193A increased the

activity at low-temperature activity for glyceraldehyde, the acceptor used in the screening, and glycolaldehyde, but decreased the activity on D- and L-threose. Therefore, it seems that the effect of V193A on activity at low-temperatures is limited to smaller acceptor substrates and the putatively increased mobility of the active site interferes with catalysis of some larger substrates. As such, this study confirms the first law of directed evolution "You get what you screen for" (Schmidt-Dannert and Arnold 1999).

Conclusions

In this study, the extremely thermostable Class I KDGA of *S. acidocaldarius* with optimal activity at 95°C was successfully improved for catalysis at the sub-optimal temperature of 50°C to enhance its potential for use in one-pot synthesis reactions together with mesophilic enzymes. Screening of 1500 variants resulted in isolation of a single variant SacKDGA-V193A, which was three times more active than wild type SacKDGA for condensation of pyruvate and glyceraldehyde. The mutation increased the activity specifically for the screening substrates and did not affect the extreme thermostability of the enzyme. Despite the relatively small size of the generated library, the chosen laboratory evolution approach has proven to be an efficient strategy for improving the desired property into SacKDGA. To our knowledge this is the first report on improving an aldolase for activity at sub-optimal temperatures.

Chapter 5

Stereoselectivity of *Sulfolobus* KDG aldolases and engineering of substrate specificity towards azido-substituted aldehydes

Marloes Schurink*, Suzanne Wolterink-van Loo*, Thijs Kaper*, Vincent J.B. Ruigrok, Loren L. Looger, John van der Oost and Maurice C.R. Franssen

*These authors contributed equally

Abstract

The 2-keto-3-deoxygluconate aldolases (KDGAs) isolated from *Sulfolobus* species reversibly convert pyruvate and glyceraldehyde into 2-keto-3-deoxy-gluconate and -galactonate. Due to their high thermostability and activity on non-phosphorylated substrates, KDGA enzymes have potential as biocatalysts for the production of chiral building blocks for fine chemical and pharmaceutical applications. Azido-functionalized aldehydes were applied as alternative substrates in the KDGA reaction as an initial step towards the production of nitrogen heterocycles. A method for determination of the stereoselectivity of the reaction between the pyruvate donor and azido-substituted aldehyde acceptors catalyzed by KDGA was developed. *S. acidocaldarius* (Sac) KDGA displayed higher stereoselectivity than *S. solfataricus* (Sso) KDGA for all tested condensation reactions. In addition, the KDGA active site was optimized for binding of the 4(*S*)-condensation product of pyruvate and 2-azidoacetaldehyde by computational protein redesign. The predicted mutations in SacKDGA slightly changed the stereoselectivity and significantly increased the relative specificity towards 2-azidoacetaldehyde.

Introduction

Stereochemically pure nitrogen heterocycles are important building blocks for fine chemicals and pharmaceuticals. Via organic synthesis it is not trivial to synthesize optically pure compounds and often several protection and deprotection steps are needed. Nitrogen heterocycles are the basis of many pharmaceuticals such as the antihypertensive drug omapatrilat (Patel 2001), the antihistamine agent astemizole (Bräse *et al.* 2002), and antifungal agents such as anidulafungin and micafungin, which were launched recently (Renslo 2007). Enzymes are generally stereo- and/or regioselective, do not require protected substrates and catalyze reactions in a mild environment and are therefore potentially useful in chemoenzymatic synthesis routes. For example, 2-deoxy-D-ribose 5-phosphate aldolase was optimized for accepting chloroacetaldehyde and used to synthesize a precursor of statin drugs (Jennewein *et al.* 2006). It is well known that azido groups are very versatile and can be easily reduced to amino groups. So, the combined use of azido-functionalized aldehyde substrates and stereoselective aldolases may yield a suitable route towards stereochemically pure nitrogen heterocycles. Hence, the research described in this paper focuses on the stereoselectivity and substrate specificity of aldolases towards azido-substituted aldehydes.

Aldolases constitute one of the few classes of biocatalysts that nature uses to enlarge the carbon skeleton of molecules. The aldol reaction leads to the formation of a new C−C bond, generating a hydroxylated product with a new stereogenic center. Aldolases are enzymes that catalyze this asymmetric aldol reaction and are therefore interesting in bio-organic synthesis (Bolt *et al.* 2008, Dean *et al.* 2007, Samland and Sprenger 2006, Sukumaran and Hanefeld 2005). In this study, aldolases from the thermophilic archaea *Sulfolobus solfataricus* and *Sulfolobus acidocaldarius* were used (Ahmed *et al.* 2005, Buchanan *et al.* 1999, Wolterink-van Loo *et al.* 2007). Their chemical and thermal stability, long shelf life, specificity for non-phosphorylated (cheap) substrates, and their metal independent catalytic mechanism are good traits for new and useful biocatalysts. The archaeal KDGA catalyzes the reversible conversion of 2-keto-3-deoxygluconate (KDG) to pyruvate and glyceraldehyde (Ga) in a modified Entner-Doudoroff pathway. As with most pyruvate-dependent aldolases, KDGA is especially 'relaxed' towards its acceptor/aldehyde substrate (Lamble *et al.* 2007, Wolterink-van Loo *et al.* 2007), which is an attractive feature for synthetic applications. Moreover, *S. solfataricus* KDGA is characterized as a very promiscuous enzyme converting D,L-KDG as well as D,L-2-keto-3-deoxygalactonate (D,L-KDGal) (to pyruvate and D,L-Ga, with an almost complete lack of stereoselectivity in the synthesis direction (Lamble *et al.* 2003, Theodossis *et al.* 2004). Stereoselectivity, however, can be induced by using a bulky substrate analogue like glyceraldehyde acetonide (Lamble *et al.* 2005a). Alternatively, improved stereoselectivity may be achieved via mutagenesis.

Enzymes are designed to perform their functions in the cell, acting on their natural substrates. In biotechnological applications, reaction conditions and substrates are often very different from the natural situation, leading to (strongly) decreased activity (Reetz *et al.* 2001). Therefore optimization towards certain substrates (substrate specificity) and/or desired products (product selectivity) is often required for their use in biotechnology. In the past decades, several techniques have been developed to improve enzymes for their desired function, as reviewed by Sen *et al.* (2007). Two different approaches for optimization can be distinguished, namely random or rational methods. For random optimization, no a priori knowledge about the protein is needed. Rational (nowadays mostly computational) design requires extensive knowledge about the structure-function relationship of the enzyme. Using rational design and site directed mutagenesis techniques, substrate specificity (DeSantis *et al.* 2003a, Joerger *et al.* 2003) as well as stereoselectivity (Scheib *et al.* 1998) can be targeted.

KDGAs from thermophilic archaea may find their application in the production of stereochemically pure nitrogen heterocycles as building blocks for fine chemicals and pharmaceuticals. Therefore, in this paper, the acceptance of different azidoaldehyde substrates by two wild type KDGAs from *S. solfataricus* and *S. acidocaldarius* was studied. Subsequent conversion of the enzymatic products to form nitrogen heterocycles simultaneously provided a possibility to analyze the stereoselectivity of the aldolase reaction. The stereoselectivity of the KDGAs in the condensation of D-glyceraldehyde and some azido-substituted aldehydes with pyruvate was determined. Furthermore, an attempt was made to improve the enantioselective formation of 5-azido-4(*S*)-hydroxy-2-oxopentanoic acid by computational design of the KDGA active site. The 3D structures of *S. solfataricus* and *S. acidocaldarius* KDGAs have been solved recently (Theodossis *et al.* 2004, Wolterink-van Loo *et al.* 2007), with substrates and products bound in the active site, providing a good basis for a mutagenesis approach. A structural explanation for the observed difference in stereoselectivity of SsoKDGA and SacKDGA at C4 of the product is provided.

Experimental

Synthesis of azido-substituted aldehydes

2-Azidoacetaldehyde (2-Aa), 2-azidopropionaldehyde (2-Ap) and 3-azidopropionaldehyde (3-Ap) were synthesized as follows. A threefold molar excess of NaN_3 (>99% Fluka, Steinheim, Germany) was added to 25 mmol bromoacetaldehyde dimethylacetal (97% Acros Organics, Geel, Belgium), 2-chloropropionaldehyde dimethylacetal (>96% Fluka, Steinheim, Germany), or 3-bromopropionaldehyde dimethylacetal (90% Aldrich, Steinheim, Germany) in 50 mL DMSO. The solution was vigorously stirred for one day at 90°C for the bromo-substituted acetals and three days at 130°C for the chloro-substituted acetal. The reaction mixture was

allowed to cool and was transferred (while still warm) to a separatory funnel containing 50 mL H_2O. The mixture was subsequently extracted 4 times with 50 mL Et_2O. The combined Et_2O layers were washed 2 times with 100 mL H_2O and 2 times with 100 mL brine. The organic layer containing the product was dried using Na_2SO_4. After removal of the drying agent, Et_2O was carefully evaporated at 30-40°C and 300-200 mbar to obtain the oily product. Complete conversion of the reaction was confirmed by NMR analysis (Bruker 300 MHz DPX300 NMR Spectrometer) and the yield of the product after extraction was 82% for 2-azidoacetaldehyde dimethylacetal, 60% for 2-azidopropionaldehyde dimethylacetal, and nearly 100% for 3-azidopropionaldehyde dimethylacetal. ^{1}H-NMR of 2-azidoacetaldehyde dimethylacetal (300 MHz, $CDCl_3$): δ 4.44 (t, 1H, *J* = 5.29 Hz); δ 3.39 (s, 6H); δ 3.23 (d, 2H, *J* = 5.29 Hz). ^{13}C-NMR of 2-azidoacetaldehyde dimethylacetal (75 MHz, $CDCl_3$): δ 103.1 (CH); δ 54.2 (CH_3); δ 51.6 (CH_2). ^{1}H-NMR of 2-azidopropionaldehyde dimethylacetal (300 MHz, $CDCl_3$): δ 4.18 (d, 1H, *J* = 5.79 Hz); δ 3.51 (dq, 1H); δ 3.47 (s, 3H); δ 3.44 (s, 3H); δ 1.22 (d, 3H, *J* = 6.72 Hz). ^{13}C-NMR of 2- azidopropionaldehyde dimethylacetal (75 MHz, $CDCl_3$): δ 106.9 (CH); δ 58.1 (CH); δ 55.3 (CH_3); δ 55.0 (CH_3); δ 14.4 (CH_3). ^{1}H-NMR of 3-azidopropionaldehyde dimethylacetal (300 MHz, $CDCl_3$): δ 4.46 (t, 1H, *J* = 5.66 Hz); δ 3.34 (t, 2H, *J* = 6.80 Hz); δ 3.33 (s, 6H); δ 1.85 (dt, 2H, *J* = 5.70 Hz, 6.81 Hz). ^{13}C-NMR of 3-azidopropionaldehyde dimethylacetal (75 MHz, $CDCl_3$): δ 102.1 (CH); δ 53.3 (CH_3); δ 47.3 (CH_2); δ 32.1 (CH_2).

2-Azidoacetaldehyde dimethylacetal, 2-azidopropionaldehyde dimethylacetal, 3-azidopropionaldehyde dimethylacetal, and 3-azido-2-hydroxypropionaldehyde diethylacetal (the latter was a gift from the University of Amsterdam, The Netherlands) were dissolved in H_2O up to a concentration of about 0.5-1 M and hydrolyzed with 0.1-0.2 eq HCl at temperatures between 25-50°C to obtain the aldehyde. Hydrolysis was monitored by HPLC (BioRad Aminex HPX-87P column 300 mm × 7.8 mm, 1 mL/min 9 mM H_2SO_4, 20 μL inj. vol., 210 nm det.) After completion each aldehyde solution was filtrated (Acrodisc® Syringe Filters, Cornwall, UK) and stored at –20°C. The final concentration of the aldehyde substrate was determined by ^{1}H-NMR using maleic acid as a standard.

Computational design of SacKDGA mutants

Redesign of the acceptor binding site of *S. acidocaldarius* KDGA was based on the structure of SsoKDGA with a covalently bound D-KDG molecule in the active site (PDB: 1W3N, (Theodossis *et al.* 2004)). All ligand and solvent molecules were deleted from 1W3N and hydrogen atoms were added using the MolProbity server (Davis *et al.* 2007). The D-KDG ligand of 1W3N was modified in Chem3D (CambridgeSoft, Cambridge, USA) to 5-azido-

4(*S*)-hydroxy-2-oxopentanoic acid, one of the two condensation products of pyruvate and 2-azidoacetaldehyde, and superimposed on the original D-KDG ligand using KiNG (Richardson Lab, Duke University, Durham, NC, USA). The program Chameleon (L.L. Looger, unpublished results) was used for the acceptor binding site redesign. Chameleon first generated a KDGA structure in which manually selected, substrate-interacting residues were replaced by alanines. Next, high-quality amino acid rotamers are tested one by one at each position and their interactions with the surrounding protein and ligand are scored with a Charmm22 (MacKerell *et al.* 1998) potential function incorporated in Chameleon. Rotamers are placed at each amino acid position, and those with an intrinsic interaction energy above a user-set threshold are eliminated. The pairwise interaction energies are computed amongst the remaining rotamers, and the resulting intrinsic and pairwise energies are used to determine the global energy minimum via the Dead-End Elimination algorithms (Looger *et al.* 2003).

Construction of mutant *Sulfolobus* KDGAs

Single mutants of SacKDGA were made using the QuikChange mutagenesis procedure (Stratagene, La Jolla, CA, USA) with pWUR193, containing the *S. acidocaldarius kdgA* gene (Saci_0225; Genbank Identifier 70606067) in pET24d, (Wolterink-van Loo *et al.* 2007) as a template and transformed to *Escherichia coli* DH5α. For mutant 1 (Y131F), primers BG2237 and BG2238 were used (Table 1). Mutant 2 (S195M) was constructed with primers BG2239 and BG2240. Primers BG2241 and BG2242 were used for mutant 3 (S238L). Mutant constructs were verified for their mutations by sequence analysis. Double and triple mutants were made using QuikChange on the single mutants, introducing a second and third mutation. The plasmids containing the mutant KDGA genes are named pWUR341-347 (encoding mutants 1-7 respectively).

Table 1 – Primers used for construction of SacKDGA mutants.

Primer	Sequence
BG2237	5'-TCTCTCTATATATACAAC**TTC**CCTGCAGCCACTGGTTACG-3'
BG2238	5'-CGTAACCAGTGGCTGCAGG**GAA**GTTGTATATATAGAGAGA-3'
BG2239	5'-TTAGATGGAGTGGTTGCA**ATG**TTTACGAACTTCATACCTG-3'
BG2240	5'-CAGGTATGAAGTTCGTAAA**CAT**TGCAACCACTCCATCTAAAG-3'
BG2241	5'-CATACTGAGAAAATACGGT**CTG**ATCTCAGCAATTTATGTG-3'
BG2242	5'-CACATAAATTGCTGAGAT**CAG**ACCGTATTTTCTCAGTATG-3'

## Production and purification of *Sulfolobus* KDGAs

Plasmids pWUR122, containing the *S. solfataricus kdgA* gene (Ahmed *et al.* 2005), pWUR193 and pWUR341-347 were transformed to *E. coli* BL21(DE3) cells. From each transformation, a single colony was used to inoculate 3 mL of Luria-Bertani medium supplemented with 50 μg/mL kanamycin and incubated overnight at 37°C while shaking. Next, 1 mL of this preculture was used to inoculate 1 liter of Luria-Bertani medium containing 50 μg/mL kanamycin in 2 liter conical flasks and incubated in a rotary shaker at 37°C for 4 hours. The culture was then induced by adding isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.1 mM (0.3 mM for pWUR122) and growth was continued overnight. The cells were harvested by centrifugation (3,800*g*, 10 min), resuspended in 20 mL 50 mM HEPES pH 7 with 20 mM KCl, and lysed by three passages through a French press (SLM, Aminco, USA) at 110 MPa. After centrifugation (27,000*g*, 20 min) the cell free extract (CFE), containing soluble proteins, was incubated at 75°C for 20 min. This was centrifuged (27,000*g*, 20 min) to obtain the heat stable cell free extract (HSCFE). The enzyme purity was checked by SDS-PAGE (Figure 1), staining the proteins with Coomassie Brilliant Blue. The extract containing SsoKDGA was additionally incubated at 78°C for 20 min to lose more *E. coli* proteins in this step. Aldolase made up ~90% of HSCFE, except for mutant 6. For kinetic analysis, the obtained HSCFEs were purified further using anion-exchange chromatography (Q-sepharose column 46 mL, equilibrated with 3 mL/min 20 mM Tris-HCl pH 8.5, proteins eluted with 0-1 M NaCl gradient). Enzymes elute at approximately 0.30 M NaCl. Protein concentrations were determined with the Bradford method (BioRad, Hercules, CA, USA) and were found to be 11.2 ± 0.07 mg/mL for SsoKDGA and 5.50 ± 0.35 mg/mL for SacKDGA.
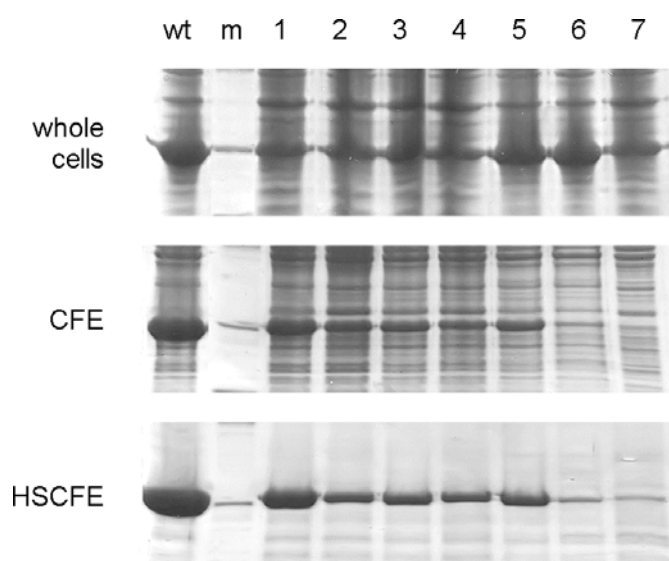
Figure 1. SDS-PAGE of whole cells, cell free extract (CFE) and heat stable cell free extract (HSCFE) of SacKDGA wild type (wt) and its mutants 1-7 along with a molecular weight marker (m).

Conversion of the KDGA reaction followed by HPLC

HSCFEs or purified enzymes of different KDGAs were prepared as described above. One unit of KDGA, with respect to the reaction between pyruvate and 2-Aa at 50°C as determined by the LDH assay (Wolterink-van Loo *et al.* 2007), per mL reaction mixture, was added to a mixture of substrates pyruvate and aldehyde in buffer at 50°C, to initiate the reaction. After addition of the enzyme, the initial concentrations of the substrates were 100 mM pyruvate (>99% Sigma, Steinheim, Germany) and 100 mM aldehyde in 100 mM NaPi pH 6.0. Because K_M values for pyruvate and Ga are far lower than 100 mM (Wolterink-van Loo *et al.* 2007), the reaction rate is expected not to be substrate-limited under these conditions. The reactions were incubated at 50°C, while gently shaken for several hours to reach equilibrium. For kinetic experiments, samples of 0.5 mL were drawn at different points in time to follow the reaction. Each sample was frozen in liquid nitrogen to stop the reaction immediately, and stored at –20°C. The enzyme was removed by filtrating the samples at 13,000g for 1.5 hours at 4°C using centrifugal filters with a molecular weight cut-off of 10 kDa (Millipore Microcon®, Bedford, USA). A small amount of each sample was diluted 6 times prior to HPLC analysis (BioRad Aminex HPX-87P column 300 mm × 7.8 mm, 1 mL/min 9 mM H_2SO_4, 20 μL inj. vol., 210 nm det.) and a calibration series was used to quantify the unreacted pyruvate. The remainder of each sample was stored at –20°C for further analysis.
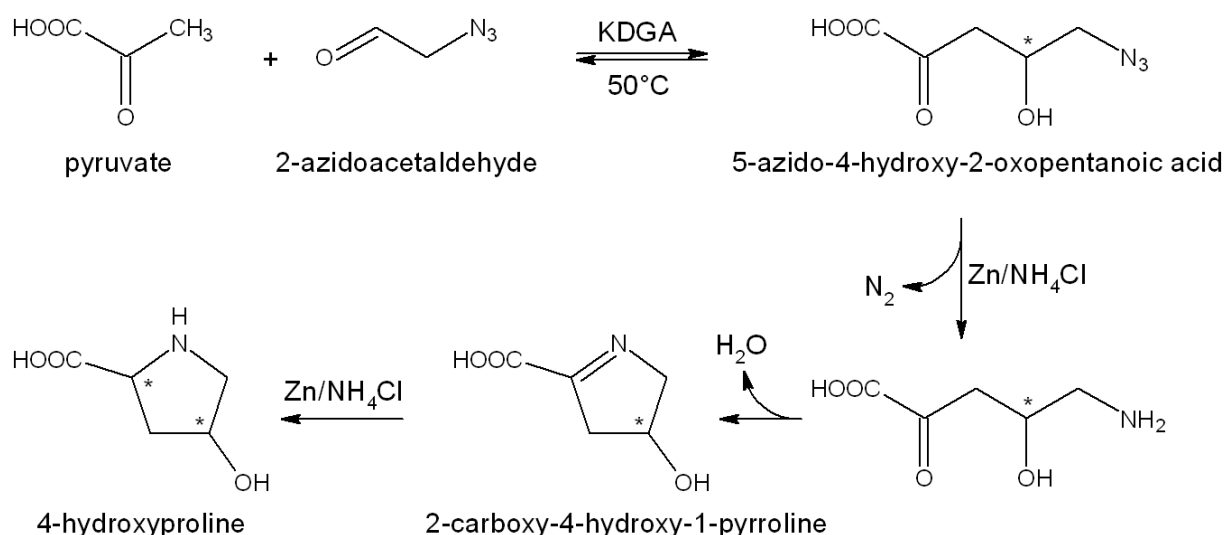
Figure 2. KDGA-catalyzed aldol condensation illustrated for the reaction between pyruvate and 2-azidoacetaldehyde and the subsequent reduction, after termination of the enzymatic reaction, of the azido group using Zn and NH_4Cl to form 4-hydroxyproline. Similar reduction products are formed when 2-Ap and 3-Ap are used as a substrate.

Stereoselectivity of the KDGA reaction determined via GC-MS

Since direct resolution of the product enantiomers using an enantioselective HPLC column (Phenomenex Chirex 3011 column 250 mm x 4.6 mm or ChromTech Chiral-AGP column 100 mm x 4.0 mm) was unsuccessful, an alternative approach was set up. After termination of the enzymatic reaction, the azido group of the product of the KDGA reaction was reduced using Zn/NH$_4$Cl (Lin *et al.* 2002) to obtain, in case 2-Aa was used as a substrate, eventually 4-hydroxyproline (Figure 2). The advantage of this approach is the formation of a natural compound of which 3 out of 4 stereoisomers are commercially available. A suspension of the reducing agent containing 44 mM Zn powder and 78 mM NH$_4$Cl in H$_2$O was freshly prepared. To 200 µL of filtrated KDGA reaction mixture, 1 mL of this reducing reagent was added. The mixture was incubated overnight at 50°C while vigorously shaken to avoid aggregation of Zn. Then the Zn was precipitated by a short centrifugation step and the supernatant was analyzed by HPLC (BioRad Aminex HPX-87P column 300 mm × 7.8 mm, 1 mL/min 9 mM H$_2$SO$_4$, 20 µL inj. vol., 210 nm det.) along with a calibration series to quantify the remaining pyruvate in each sample. HPLC analysis showed that the reduction of 5-azido-4-hydroxy-2-oxopentanoic acid (*rt* = 5.92 min) and the remaining 2-Aa (*rt* = 8.75 min) was nearly 100%, which allows proper quantification of the unreacted pyruvate (*rt* = 6.07 min) in case 2-Aa was used as a substrate. The presence of 2-carboxy-4-hydroxy-1-pyrroline (*rt* = 26.4 min) as an intermediate was confirmed based on its UV absorbance of the C=N bond (λ_{max} = 265 nm) and molecular mass (*m/z* = 112 = 129 [M] + 1 − 18) determined via LC-MS analysis (BioRad Aminex HPX-87P column 300 mm × 7.8 mm, 1 mL/min 0.5 mL/L TFA, 20 µL inj. vol., source voltage 5 eV +ESI, Thermo Finnigan LCQ Classic, San Jose, USA). Although this imine was not completely reduced to 4-hydroxyproline, both enantiomers will be reduced equally and an *ee* value for the KDGA reaction can be derived from the relative abundance of each stereoisomer of 4-hydroxyproline. To that end, the remainder of the supernatant was lyophilized to remove H$_2$O completely. Derivatization of 4-hydroxyproline present in the sample was performed according to a standard procedure for GC resolution of amino acid enantiomers (Gerard *et al.* 1997). The residue, or a small amount of *trans*-4-hydroxy-L-proline, *cis*-4-hydroxy-D-proline, *cis*-4-hydroxy-L-proline (>99% Sigma-Aldrich, Steinheim, Germany), or an equimolar mixture of these standards, was dissolved in 300 µL *i*-PrOH saturated with HCl (*g*). The vial was sealed tightly and esterification of the carboxyl group was accomplished at 110°C during 45 min. Then the sample was reduced to dryness using a N$_2$ (*g*) flow. The residue was dissolved in 300 µL CH$_2$Cl$_2$ and 100 µL pentafluoropropionic anhydride (Supelco, Bellefonte, USA) was added. The vial was sealed and heated to 110°C for 15 min. Excess of reagent was removed under a flow of N$_2$ (*g*) and the derivatized sample was dissolved in 100 µL CH$_2$Cl$_2$ for GC-MS analysis (Alltech Heliflex

Chirasil-Val column 25 m × 0.25 mm, 1 mL/min, 3 μL inj. vol., 75°C isotherm). Mass spectrum of 4-hydroxyproline derivative m/z: 319 [M]$^+$ (0), 301 (<1), 276 (1), 260 (2), 214 (100), 119 (15), 67 (14), 43 (10). Analysis of derivatized 4-hydroxyproline stereoisomers was performed in single ion mode (SIM) detection of the most abundant ion at m/z 214 and for the 4-hydroxy-5-methylproline and 2-carboxy-4-hydroxypiperidine derivatives, both resulting from the azidopropionaldehyde substrates, SIM detection was set at m/z 228. From the abundance of each 4-hydroxyproline stereoisomer the ratio between the enzymatic products 5-azido-4(S)-hydroxy-2-oxopentanoic acid and 5-azido-4(R)-hydroxy-2-oxopentanoic acid was calculated via $S/R = (trans\text{-}D + cis\text{-}L)/(trans\text{-}L + cis\text{-}D)$ with a corresponding $ee = |(S - R)/(S + R)| \times 100\%$. The stereoselectivity of KDGA reactions using other azidoaldehyde substrates was computed in a similar manner.

^{1}H-NMR analysis of the KDGA reaction

Alternatively to HPLC, the KDGA reaction was monitored directly by ^{1}H-NMR (Bruker 400 MHz DPX400 NMR Spectrometer). Apart from 10% D_2O in the reaction mixture, all reaction conditions were the same. The rate of conversion was not affected by the presence of 10% D_2O. The concentration of total product of each KDGA reaction was deduced from the ratio between signals of the methyl protons (δ 2.51 (s, 3H)) of pyruvate and the protons at C3 (δ 3.13 (m, 2H)) of the product (Figure 6A). Besides its direct nature, another advantage of this approach is that not only the conversion but also the diastereomeric excess can be calculated instantly. When 2-Ap was used as a substrate, the concentration of one diastereomer of 5-azido-4-hydroxy-2-oxoheptanoic acid (DS1) was determined from the isolated signal of a single proton at C5 (δ 3.80 (m, 1H)) in relation to the total of signals of this proton and corresponding protons of DS2 (δ 3.73 (m, 1H)) and 2-Ap (δ 3.69 (m, 1H)). Similarly, the concentration of the other diastereomer of 5-azido-4-hydroxy-2-oxoheptanoic acid (DS2) was determined independently, from the signal of its methyl protons (δ 1.44 (d, 3H)) in relation to the total of methyl protons from DS1 (δ 1.40 (d, 3H)), DS2, and 2-Ap (δ 1.38 (d, 3H)). Addition of the concentrations of DS1 and DS2 equals the concentration of total product.
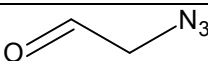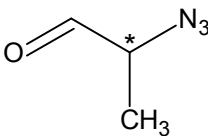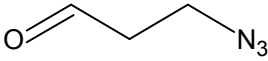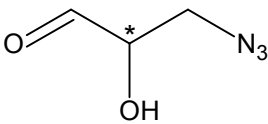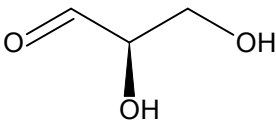
Results and discussion

Substrate specificity of *Sulfolobus* KDGAs towards azido-containing aldehydes

To broaden the scope of KDGAs for biosynthetic applications, reactions with different azido-functionalized aldehydes were studied. The azido-substituted aldehydes were readily accepted as substrates by KDGAs from *Sulfolobus solfataricus* and *Sulfolobus acidocaldarius*. This activity was not surprising as KDGAs are reported to have a broad substrate specificity

(Lamble *et al.* 2007, Wolterink-van Loo *et al.* 2007). The specific activity of *S. acidocaldarius* KDGA (SacKDGA) is generally higher than that of *S. solfataricus* KDGA (SsoKDGA), especially considering the azido-substituted aldehyde substrates (Table 2). For 2-Aa, the simplest azido-substituted aldehyde used in this study, a significant higher activity was observed for both wild type KDGAs in comparison to their natural substrate D-Ga. It was demonstrated previously that each KDGA shows a similar activity towards glycolaldehyde and DL-Ga (Wolterink-van Loo *et al.* 2007), suggesting that the $-N_3$ group of 2-Aa is more suitably accomodated in the KDGA active site than the OH-group of glycolaldehyde.

Table 2. Specific Activity (A) and stereoselectivity of SsoKDGA and SacKDGA using azido-substituted aldehyde substrates. Enantiomeric excess (*ee*) or diastereomeric excess (*de*), with respect to a 4(*S*)-configuration of the enzymatically formed product, were determined after 1 hour of reaction.

Aldehyde acceptor		SsoKDGA		SacKDGA	
		A (U/mg)	ee or de (%)	A (U/mg)	ee or de (%)
2-Aa		8.77 ± 0.1	48.7 ± 4.1	54.1 ± 0.7	85.1 ± 3.4
2-Ap		$1.35 \pm 0.4^{[c]}$	$77.6 \pm 0.5^{[a]}$ $89.4 \pm 3.1^{[b]}$	$21.7 \pm 4.9^{[c]}$	$97.3 \pm 0.3^{[a]}$ $90.4 \pm 4.6^{[b]}$
3-Ap		5.97 ± 0.3	41.5 ± 28	24.9 ± 1.6	51.3 ± 3.6
3-Ahp		1.06 ± 0.5	n.d.	7.70 ± 0.7	n.d.
D-Ga		4.98 ± 0.4	$16.6 \pm 1.0^{[d]}$	13.6 ± 2.4	$64.7 \pm 4.6^{[d]}$

[a] Determined for enantiomer E1 of 2-Ap
[b] Determined for enantiomer E2 of 2-Ap
[c] Total conversion of enantiomers E1 and E2 of 2-Ap
[d] Determined by HPLC

The azido group is a very polar group that can be described by resonance structures of which an important one is $R-N=N^+=N^-$. The partial negative charge at the terminal nitrogen may have favorable interactions with R237 and R106 in SsoKDGA and R234 and R105 in SacKDGA, similar to the phosphate group in 2-keto-3-deoxy-6-phosphogluconate (Wolterink-van Loo *et al.* 2007). Introduction of a methyl group adjacent to the $-N_3$ function lowered the activity of KDGA (Table 2), as was observed for the reaction with racemic 2-Ap. This is most likely due to an increased steric hindrance at C2 of the aldehyde substrate. Elongation of the carbon chain of 2-Aa into 3-Ap also resulted in a lower activity for both wild type KDGAs. Probably the $-N_3$ function of 3-Ap tends to stick out more to the outside of the hydrophilic binding pocket rendering less possible interactions that are favorable for binding. Nevertheless, 3-Ap is a good substrate and is converted by SsoKDGA and SacKDGA with a rate comparable to that of D-glyceraldehyde. Introduction of a hydroxyl group at C2 of 3-Ap yielding 3-azido-2-hydroxypropionaldehyde (3-Ahp), lowered the KDGA reaction rate. This was unexpected since the hydroxy-group at C2 (i.e. C5 of KDGA product) seemed to be of importance for substrate binding (Wolterink-van Loo *et al.* 2007). The favorable interactions of this hydroxy-group within the active site are either insignificant in presence of the more polar azido group or are counteracted by an accompanied increase in steric hindrance.

Enantioselectivity of *Sulfolobus* KDGA reactions with 2-Aa

The KDGA-catalyzed condensation of pyruvate with 2-Aa or 3-Ap yields enantiomeric products instead of diastereomers as occurs with D-Ga. Direct resolution and establishment of the absolute configuration of enantiomers is usually a difficult task in organic chemistry. In view of the production of enantiopure nitrogen heterocycles, the azido group of the enzymatically formed products was reduced to a primary amine that subsequently cyclizes to form a nitrogen-containing heterocyclic compound (Figure 2). Chiral resolution of such amino acid enantiomers allows determination of the stereoselectivity of the KDGA-catalyzed condensation between pyruvate and azidoaldehydes. In case of 2-Aa, the heterocyclic product is 4-hydroxyproline. Three out of four stereo-isomers of 4-hydroxyproline are commercially available, enabling accurate determination of the stereochemical course of the reaction. The relative abundance of 4-hydroxyproline stereoisomers was used to determine the absolute configuration of the major product formed in the KDGA reaction (Figure 3A). Surprisingly, both wild type enzymes, especially SacKDGA, show a selectivity towards the product enantiomer with the *S*-configuration at C4, i.e. 5-azido-4(*S*)-hydroxy-2-oxopentanoic acid. When the enantiomeric excess (*ee*) was determined as a function of reaction time (Figure 4), a maximum *ee* of 59% for SsoKDGA and even 87% for SacKDGA was observed (Table 4).

This was unexpected, since a lack of stereocontrol for the natural reaction between pyruvate and D-GA catalyzed by SsoKDGA was reported (Lamble *et al.* 2003, Theodossis *et al.* 2004). Although an increased stereoselectivity of SsoKDGA has been observed previously by the use of alternative aldehydes such as glyceraldehyde acetonide (Lamble *et al.* 2005a) and L-threose (Lamble *et al.* 2007), SacKDGA seems to be more stereoselective than SsoKDGA and displays stereocontrol in the natural reaction as well (Table 2). SacKDGA shows a diastereomeric excess (*de*) of 65% for D-KDG, while SsoKDGA shows only a slight preference for D-KDG with a *de* of 17%, which is in agreement with literature data (Lamble *et al.* 2003, Lamble *et al.* 2007).

**Table 3**. Specific activity of SacKDGA mutants on D-Ga and 2-Aa in U/mg.

Construct	Mutations	D-Ga activity	2-Aa activity	2-Aa/D-Ga
pWUR122[a]	SsoKDGA	4.50	16.3	3.6
pWUR193[a]	SacKDGA	8.27	51.0	6.2
pWUR341[b]	1. Y131F	9.09	50.5	5.6
pWUR342[b]	2. S195M	0.36	6.21	17
pWUR343[b]	3. S238L	3.43	22.3	6.5
pWUR344[b]	4. Y131F + S195M	0	7.95	∞
pWUR345[b]	5. Y131F + S238L	2.15	12.0	5.6
pWUR346[b]	6. S195M + S238L	0.17	2.84	17
pWUR347[a]	7. Y131F + S195M + S238L	0	11.1	∞

[a] Purified enzymes
[b] HSCFE

Table 4. Maximum conversion and enantioselectivity of different KDGA-catalyzed aldol condensations between pyruvate and 2-Aa.

KDGA	c_{max} (mM or %)	ee_{max} (%)	$v_0(S)/v_0(R)$
Sulfolobus solfataricus			
Wild type	73 ± 1.8	59	17.0
Sulfolobus acidocaldarius			
Wild type	67 ± 1.1	87	80.2
Mutant 1	76 ± 2.2	89	117.0
Mutant 7	63 ± 0.7	89	22.4
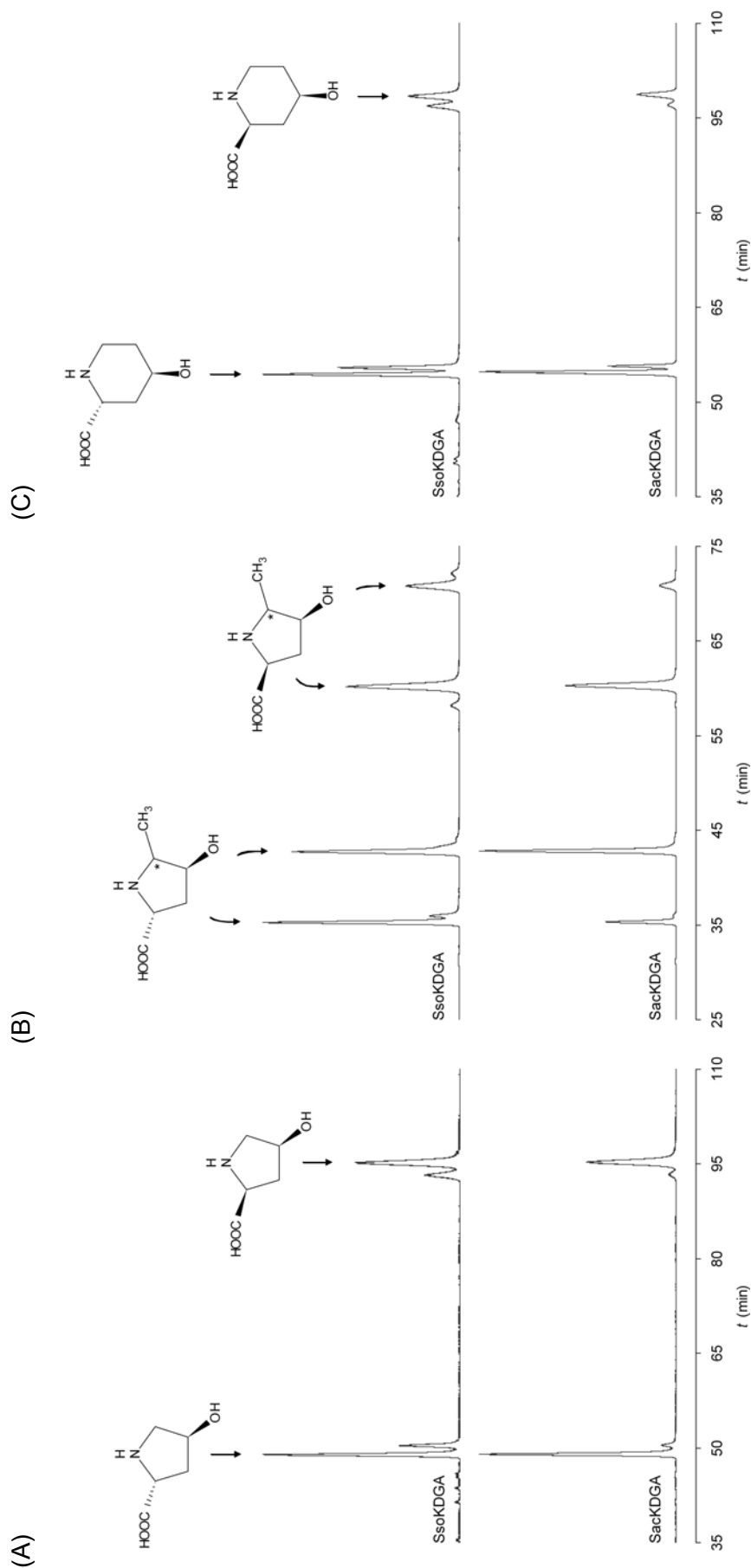
Figure 3. GC-MS analysis of nitrogen heterocycles as a tool to determine the stereoselectivity of KDGAs using different azidoaldehyde substrates. The upper chromatograms originate from the SsoKDGA reaction after 1 hour and the lower chromatograms from the SacKDGA reaction after 1 hour. Chromatograms show chiral resolution of identified stereoisomers of 4-hydroxyproline (A), 8 stereoisomers of 4-hydroxy-5-methylproline (B), and 4 stereoisomers of 2-carboxy-4-hydroxypiperidine (C) after reduction of the enzymatically formed product. For reasons of clarity, only the tentative structures of nitrogen heterocycles originating from enzymatic products with a 4(S)-configuration are indicated. Smaller, neighbouring peaks or shoulders are enantiomers of the assigned larger peaks.

Figure 4 – Enantioselective formation of 5-azido-4-hydroxy-2-oxopentanoic acid from pyruvate and 2-Aa using different KDGAs. (A) Concentration of total product (●), 5-azido-4(S)-hydroxy-2-oxopentanoic acid (◐) and 5-azido-4(R)-hydroxy-2-oxopentanoic acid (○) as a function of reaction time. (B) Enantiomeric excess with respect to the S-enantiomer of the product as a function of reaction time, for SsoKDGA wild type (▲), SacKDGA wild type (■), SacKDGA mutant 1 (◨), and SacKDGA mutant 7 (□). (C) Schematic presentation of the equilibrium between between the substrates pyruvate and 2-Aa and both product enantiomers, catalyzed by KDGA.

KDGA-mediated racemization of 5-azido-4(*S*)-hydroxy-2-oxopentanoic acid

The *ee* of the 5-azido-4(*S*)-hydroxy-2-oxopentanoic acid product reaches a maximum after some time but then slowly decreases to zero (Figure 4B). Theoretically, one would expect an *ee* of 100% at time zero of the reaction, but this was not observed. This may be explained by the detection limit of the method at low product concentrations at the beginning of the reaction. The racemization process is explained as follows (Figure 4C). KDGA shows enantioselectivity towards the 4(*S*)-product and, therefore, at the beginning of the reaction 5-azido-4(*S*)-hydroxy-2-oxopentanoic acid is formed at a high rate (i.e. $k_S > k_R$). The rate of the reverse reaction 4(*S*)-product $\rightarrow$ pyruvate + 2-Aa is negligible, because the product concentration is still too low. After several minutes a kinetically controlled maximum *ee* of nearly 90% is reached, and now a sufficient amount of 4(*S*)-product has been formed for the reverse reaction to play a role. In the meanwhile some 4(*R*)-product is produced and since the formation of this 4(*R*)-product is relatively slow, the reverse reaction 4(*R*)-product $\rightarrow$ pyruvate + 2-Aa is slow as well. Eventually, after several hours, a thermodynamic equilibrium is reached yielding a 1 : 1 mixture of 4(*S*)- and 4(*R*)-product. A maximum conversion (total product in equilibrium) between 60-80% was obtained under the reaction conditions used. In absence of KDGA, the *ee* of the product remains constant (not shown). Therefore, racemization only takes place via the KDGA-catalyzed equilibrium. It remains to be determined if such enzyme-mediated racemization eventually, after longer periods of incubation, occurs with the reaction of SacKDGA and its natural substrates pyruvate and D-Ga as well.
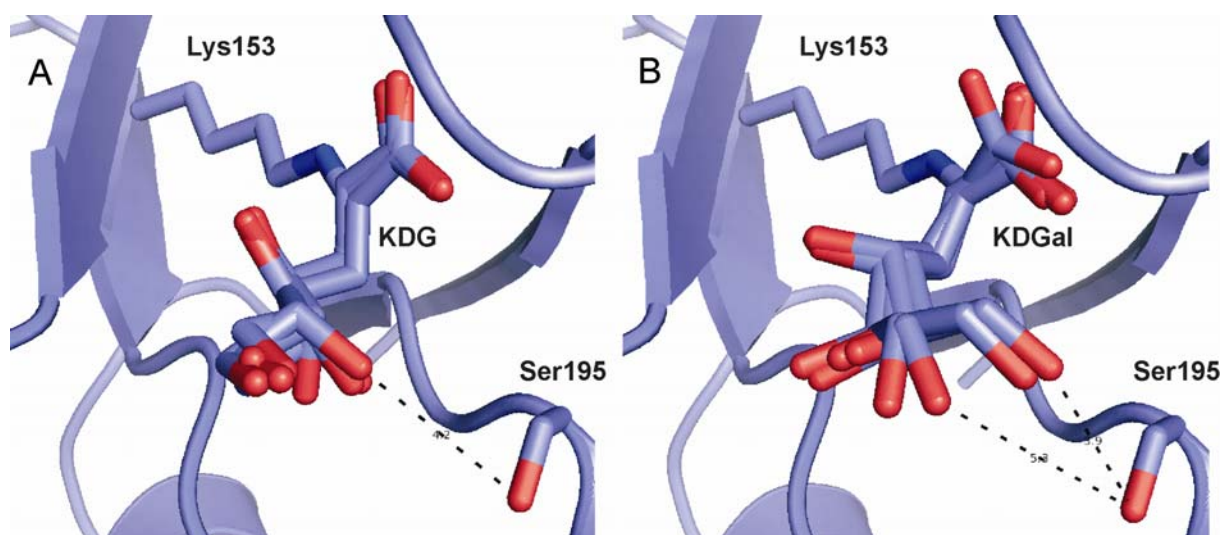
Figure 5. Superposition of 4 D-KDG molecules from individual SsoKDGA monomers (PDB: 1W3N) (A) and 4 D-KDGal molecules from SsoKDGA (PDB: 1W3T) (B) on chain A of SacKDGA (PDB: 2NUX). D-KDG, D-KDGal, catalytic residue L153 and S195 are shown in stick representation. Dashed lines indicate distances between C5–OH of D-KDG or C6–OH of D-KDGal and S195–OH. Pictures were generated using PyMol (DeLano 2002).

Structural basis for the diastereoselectivity of SacKDGA in the natural reaction

3D structures of SsoKDGA (Theodossis *et al.* 2004) and SacKDGA (Wolterink-van Loo *et al.* 2007) were determined previously in complex with various ligands, which enabled searching for features that could explain the difference in stereoselectivity between the two aldolases. Superposition of the available monomers in the deposited structures showed that the side chains of amino acid residues at the acceptor binding site adopt virtually identical conformations in the two enzymes. The active sites are identical, except for SsoKDGA A198, which is S195 in SacKDGA. The ability of SacKDGA to form an extra hydrogen bond with the substrate through the hydroxyl group of the serine side chain may explain why SacKDGA is more stereoselective than SsoKDGA. SsoKDGA has been crystallized in complex with both D-KDG and D-KDGal (Theodossis *et al.* 2004). The asymmetric unit of the crystal lattice contains the four subunits of the homotetramer. The D-KDG molecules in each of the monomers adopt similar conformations (Figure 5A). The C5–OH of D-KDG is about 4Å from SacKDGA S195, when SsoKDGA and SacKDGA are superimposed. D-KDG and S195 could interact directly when the side chain of S195 adopts a different conformation or possibly through a water molecule. The conformations of the D-KDGal molecules are less uniform than the D-KDG molecules and the C6–OHs of only two of the four molecules come within 4 Å of S195 (Figure 5B). In SsoKDGA the conformational differences of the D-KDGal molecules would not necessarily have an effect on binding. In SacKDGA, however, conformations with the C5–OH in proximity to S195 are possibly bound more tightly than in SsoKDGA. This could explain SacKDGA's preference for the formation of D-KDG over D-KDGal.

Stereoselectivity of *Sulfolobus* KDGA reactions with azidopropionaldehydes

To determine the stereoselectivity of the KDGA-catalyzed condensation between pyruvate and the azidopropionaldehyde substrates, some assumptions had to be made. First of all, it is assumed that the major stereoisomer produced by KDGA has a 4(*S*)-configuration, regardless the aldehyde acceptor used. So far, this stereopreference has been confirmed for glyceraldehyde acetonide (Lamble *et al.* 2005a), L-threose (and to a lesser extent D-Ga) (Lamble *et al.* 2007), and also 2-Aa in this study (Table 2). In addition, a significant preference of KDGAs for a 4(*R*)-configuration in the condensation product has never been reported. Secondly, as was observed for the 4-hydroxyproline stereoisomers (Figure 3A), it is assumed that the *trans*-isomers of the nitrogen heterocycles, considering the relative orientation of the carboxyl and hydroxyl group on the ring, elute before the *cis*-isomers from the GC column.

Pyruvate and 3-Ap were converted to 6-azido-4-hydroxy-2-oxoheptanoic acid, which eventually, after reduction with Zn/NH_4Cl, led to the formation of 2-carboxy-4-hydroxypiperidine. The 4 stereoisomers of this compound showed a similar GC-MS elution pattern as 4-hydroxyproline (Figure 3C). Based on this and the assumptions described above, structural assignments of the 2-carboxy-4-hydroxypiperidine peaks were made analogously to the peaks of 4-hydroxyproline. The enantioselectivity of the enzymatic reaction using 3-Ap was decreased compared to 2-Aa (Table 2). The stereoselectivity of the KDGA reactions with 3-Ahp as a substrate could not be determined via the present method, due to cyclization of the product isomers into furanose forms.

When studying the condensation of pyruvate and 2-Ap, assignment of the 4-hydroxy-5-methylproline peaks was complicated by the presence of an additional stereocenter. Both enantiomers, referred to as E1 and E2, of racemic 2-Ap were converted by KDGA to form 4 stereoisomers of 5-azido-4-hydroxy-2-oxoheptanoic acid and these eventually resulted in 8 stereoisomers of 4-hydroxy-5-methylproline (Figure 3B). The 4 largest peaks in the GC-MS chromatogram most probably originate from 5-azido-4(*S*)-hydroxy-2-oxoheptanoic acid, 2 originating from enantiomer E1 and 2 originating from enantiomer E2 of 2-Ap. The smaller, accompanying peaks are their mirror images and, hence, originate from 5-azido-4(*R*)-hydroxy-2-oxoheptanoic acid. By comparison of the chromatograms of SsoKDGA and SacKDGA (Figure 3B), it can be concluded that the two inner, largest peaks originate from one enantiomer E1 of 2-Ap and that the two outer peaks must come from the other enantiomer E2 of 2-Ap. The ratio of conversion after 1 hour between E1 and E2 for SsoKDGA appeared to be (E1/E2) = 1.5 ± 0.2 and for SacKDGA (E1/E2) = 4.1 ± 0.2, indicating a higher substrate enantioselectivity for SacKDGA. With the assignment of the 4-hydroxy-5-methylproline peaks, a *de* for each enantiomer of 2-Ap was calculated (Table 2). For 2-Ap, the stereoselectivity of the KDGA reaction was significantly increased compared to 2-Aa, probably due to the increased steric hindrance next to the aldehyde function. For SsoKDGA, the maximum *de*'s for the two enantiomers E1 and E2 of 2-Ap were 78% and 89%, respectively. For SacKDGA these were even higher, namely 97% and 90%.

(A)

(B)
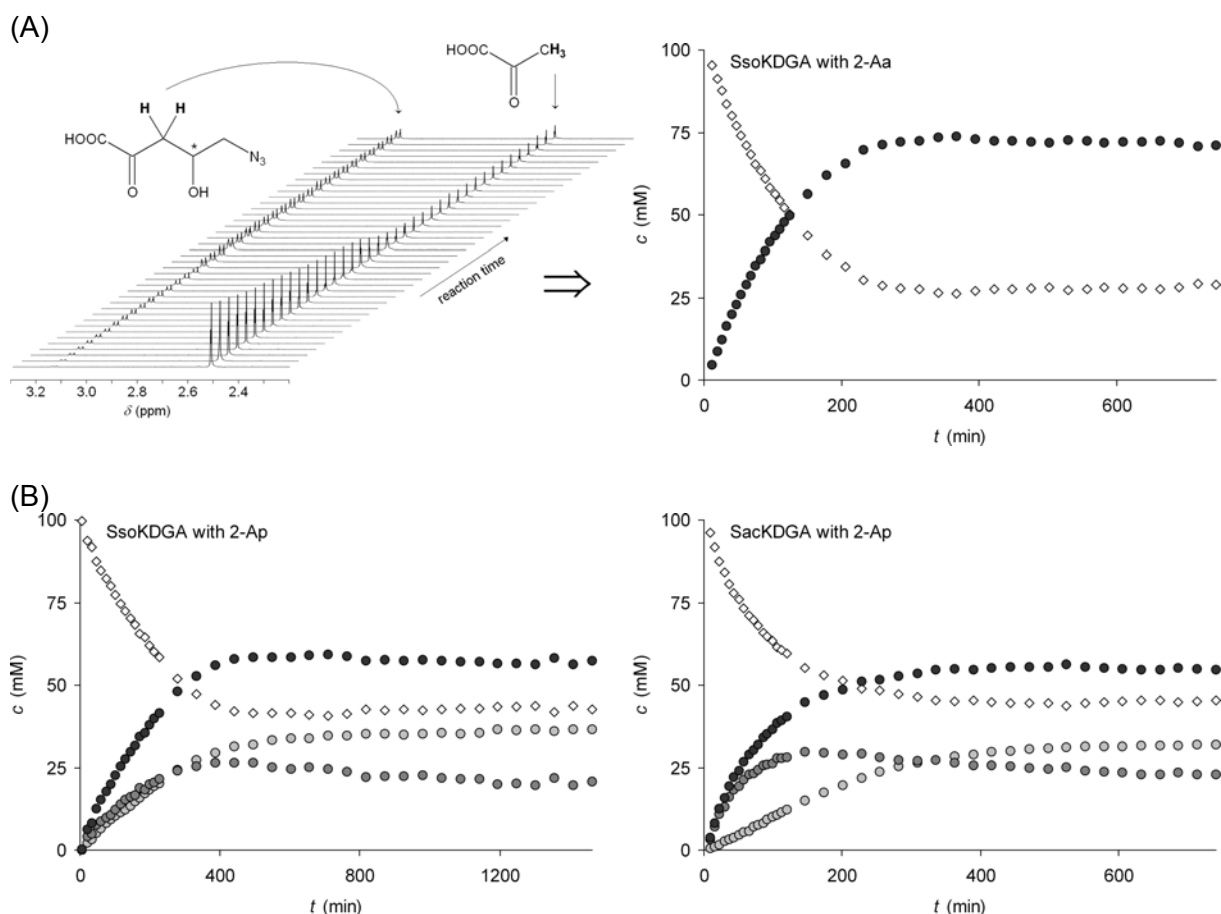
Figure 6. Enzymatic activity of SsoKDGA and SacKDGA, using 2-Aa (A) or 2-Ap (B) as a substrate, monitored directly by ^{1}H-NMR. Concentration of pyruvate (◇), total product (●), DS1 (◉), and DS2 (○) are given as a function of reaction time. Specific activities determined by NMR are 8.68 U/mg (c_{max} = 72 ± 0.7%) for SsoKDGA with 2-Aa, 4.68 U/mg (c_{max} = 58 ± 0.9%) for SsoKDGA with 2-Ap, and 27.5 U/mg (c_{max} = 55 ± 0.5%) for SacKDGA with 2-Ap.

^{1}H-NMR analysis of SsoKDGA and SacKDGA reactions with 2-Ap

In order to get more insight in the difference between SsoKDGA and SacKDGA catalyzed conversion of pyruvate and 2-Ap, the reactions were followed by NMR (Figure 6). To validate this approach, the KDGA reaction using 2-Aa was analyzed prior to the reaction with 2-Ap. For the conversion of 2-Aa by SsoKDGA, the same results regarding specific activity and position of equilibrium was obtained as with HPLC and GC-MS (compare Tables 2 and 4 with Figure 6). A major advantage of this NMR approach is the fact that the enzymatic reaction can be directly monitored without subsequent derivatization. Other advantages are the short experimental time and the more accurate results due to its direct nature. A disadvantage of NMR is that the enantioselectivity of the KDGA reaction cannot be determined, since it is not possible to distinguish between enantiomers. It is, however, possible to distinguish between diastereomers and the difference in the diastereoselectivity of

SsoKDGA and SacKDGA using 2-Ap as a substrate is nicely demonstrated via NMR (Figure 6). Initially, as described above, the majority of the product formed has a 4(*S*)-configuration. Therefore, the substrate specificity, expressed as v_0(DS1)/v_0(DS2), wherein v_0(DS1) is the initial speed of formation of the 4(*S*)-product of E1 and v_0(DS2) of the 4(*S*)-product of E2 of 2-Ap, is 1.4 for SsoKDGA and 5.3 for SacKDGA. These numbers agree well with the E1/E2 conversion ratios, 1.5 and 4.1 respectively, determined by GC-MS after 1 hour of reaction. These results suggest that SacKDGA is not only more stereoselective towards the 4(*S*)-product, but also represents a higher substrate specificity whilst SsoKDGA hardly distinguishes between both enantiomers of 2-Ap. Although it is the preferred product formed in the KDGA reaction, especially when using SacKDGA, DS1 appears to be thermodynamically less stable than DS2, since it is present in a lower concentration in equilibrium. Despite this knowledge, the absolute stereochemistry of the major product of the KDGA reaction between pyruvate and 2-Ap remains undetermined.
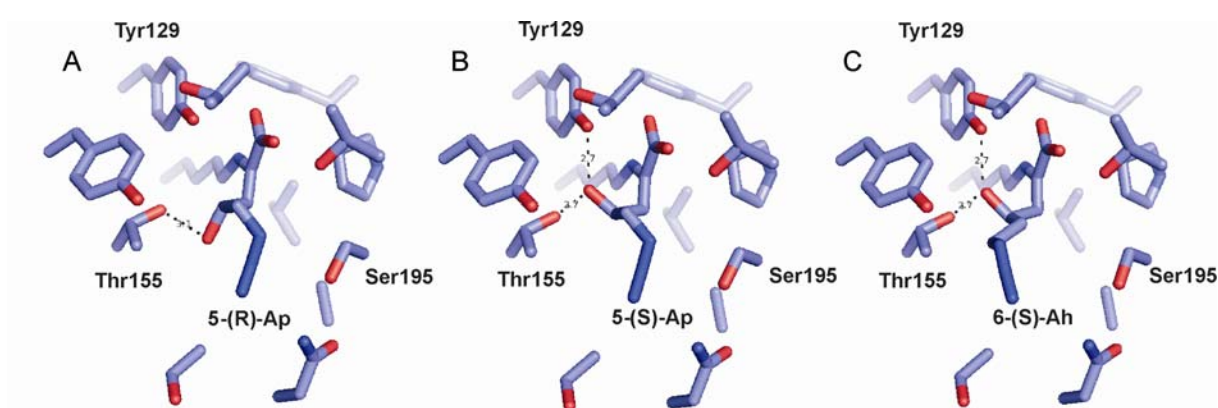
Figure 7. Condensation products of pyruvate and 2-Aa, 5-azido-4(*R*)-hydroxy-2-oxopentanoic acid (A) and 5-azido-4(*S*)-hydroxy-2-oxopentanoic acid (B), and the preferred condensation product of pyruvate and 3-Ap, 6-azido-4(*S*)-hydroxy-2-oxoheptanoic acid (C), modeled in the active site of SacKDGA complexed with pyruvate (PDB: 2NUY). Residues within 4 Å of the product are shown in stick representation. Dashed lines indicate the putative hydrogen bonds between the C4–OH of 5-azido-4-hydroxy-2-oxopentanoic acid and T155 and Y129. The acceptor-derived moieties were built and energy minimized using MOE software (Chemical Computing Group Inc, Montreal, Canada). Pictures were generated using PyMol (DeLano 2002).

Structural basis for the stereoselectivity of KDGAs with azido-substituted aldehydes

Both SsoKDGA and SacKDGA are more stereoselective for the condensation with 2-Aa than with D-Ga. The two possible products of the condensation of pyruvate and 2-Aa were modeled in the active site of SacKDGA, which resulted in very similar conformations for these enantiomers. The C4–OH of 5-azido-4(*R*)-hydroxy-2-oxopentanoic acid is within

hydrogen-bonding distance of the hydroxy-group of T155, while that of 5-azido-4(*S*)-hydroxy-2-oxopentanoic acid is within hydrogen-bonding distance of both the T155 hydroxy-group and that of Y129 (Figure 7). In contrast to D-KDG and D-KDGal, the acceptor moiety of the product does not appear to form more hydrogen bonds that could account for an equal stabilization of both product stereoisomers (Theodossis *et al.* 2004). Thus, the 4(*S*)-enantiomer is the preferred product of this condensation, because its C4–OH has more hydrogen bond interactions with the enzyme than that of the 4(*R*)-enantiomer. The distance between the side chain hydroxy-group of Ser195 in SacKDGA and the azido group of the product is close enough (3 Å) for a hydrogen bond, which is an explanation for the higher stereoselectivity of SacKDGA than SsoKDGA in the condensation of pyruvate and 2-Aa.

The product with 2-Ap is very similar to that with 2-Aa and the methyl groups of either diastereomer can be accomodated with minimal changes in conformation. The methyl groups aid in orientation of the acceptor, resulting in an increased stereoselectivity for both aldolases. The product of the coupling of pyruvate with 3-Ap is longer than that with 2-Aa, but does not have more interactions (Figure 7C), resulting in weaker binding of the ligand. The distance between the side chain hydroxy-group of Ser195 in SacKDGA and the azido group of the 3-Ap product is too large (>4 Å) for a hydrogen bond. Thus, weaker binding of the acceptor and inability to form a hydrogen bond with S195 explain the reduced activity and enantioselectivity observed for the condensation of pyruvate with 3-Ap compared to that with 2-Aa.

Optimization of SacKDGA towards 5-azido-4(*S*)-hydroxy-2-oxopentanoic acid

SacKDGA is relatively enantioselective in the condensation of pyruvate and 2-Aa (ee_{max} = 87%). However, the possibility to improve this enantioselectivity even further through protein engineering of SacKDGA was explored. Laboratory evolution of enzymes has been successfully applied for improving aldolase substrate specificity and stereoselectivity (Bolt *et al.* 2008). Successful strategies for laboratory evolution of enzymes depend on the efficiency of an assay to screen the numerous random variants in a typical enzyme library. The analysis of the reaction products of the coupling of pyruvate and 2-Aa does not fall in that category since no high throughput screening method is available. When detailed structure-function information is available, engineering of substrate specificity can be approached by *in silico* optimization of the catalysis-compatible interactions between the desired ligand and protein. This approach was chosen and resulted in a set of 7 variants, that are analysed with the available (elaborate) activity assay to determine the products in the KDGA-catalyzed condensation of pyruvate and 2-Aa.

The kinetically favored product for the KDGA-catalyzed condensation of pyruvate and 2-Aa is 5-azido-4(*S*)-hydroxy-2-oxopentanoic acid, which was modeled into the active site of KDGA. SsoKDGA was used for the modeling and design procedure, because structures with condensation products bound in the active site were available (PDB: 1W3N and 1W3T). The bound D-KDG molecule was replaced by 5-azido-4(*S*)-hydroxy-2-oxopentanoic acid *in silico*. By visual inspection, Y132, A198, and S241 were found to be close to the new acceptor molecule and selected to be targeted in the protein-redesign procedure. By selecting target residues that were in close proximity of the newly introduced acceptor part of the product, but that did not interact directly with the catalytic residue, we expected to minimize the effects of the introduced mutations on the catalytic machinery of the enzyme. The highest ranked designs (see Experimental section) had the substitutions Y132F, A198M, and S241L (Figure 8). As the structures of SsoKDGA and SacKDGA are virtually identical, especially around the conserved active site, the mutations were made in SacKDGA, as this enzyme has a higher specific activity and selectivity (Table 2). The corresponding substitutions Y131F, S195M, and S238L were introduced into SacKDGA and all possible combinations were constructed resulting in 3 single mutants, 3 double mutants and 1 triple SacKDGA mutant (Table 3).
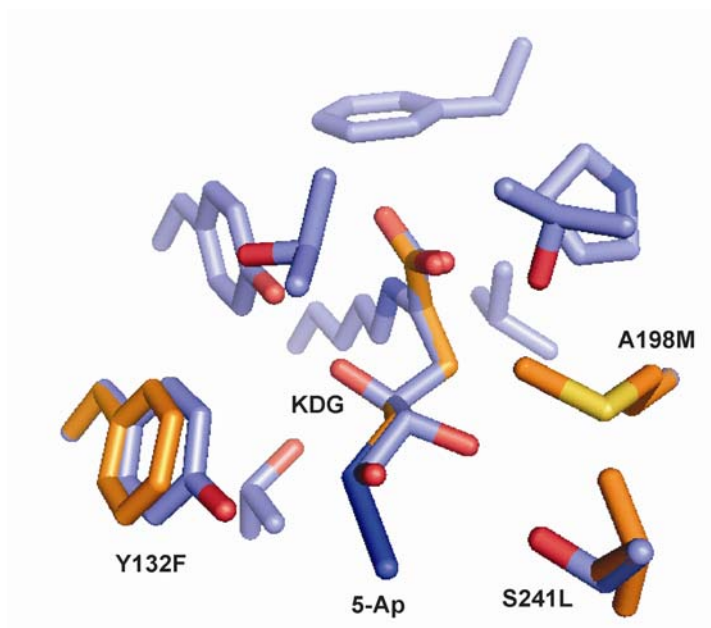
Figure 8. Superposition of SsoKDGA wild type complexed with D-KDG (PDB: 1W3N) (light gray) with highest ranked design with 5-azido-4(*S*)-hydroxy-2-oxopentanoic acid (dark grey). Pictures were generated using PyMol (DeLano 2002).

Activity and enantioselectivity of the SacKDGA mutants

The SacKDGA mutants were expressed and purified together with the wild type enzyme. The different stages of purification were monitored by SDS-PAGE (Figure 1). Mutants 6 and 7 were initially produced in inclusion bodies. Both mutants have in common that the serine residues at positions 195 and 238 are replaced by more apolar residues. The residues are located in the active site and not involved in inter-subunit contacts. It seems therefore that the apolar residues at these positions may affect protein folding, which was supported by the fact that omission of IPTG from the culture medium, resulting in a reduced expression rate, increased the amount of soluble enzyme (mutant 7, and to a lesser extent mutant 6).

The specific activity of the SacKDGA mutants for condensation of pyruvate and D-Ga or 2-Aa (Table 3), was compared to the wild type enzyme. The specific activity in the condensation of pyruvate and 2-Aa of all mutants are significantly reduced, with the exception of Y131F (mutant 1). Major differences in activity between the variants were observed for the condensation of pyruvate with D-Ga. Single mutation Y131F had no effect on either condensation reaction. This is surprising, since the hydroxy-group of Y131 (Y132 in SsoKDGA) is involved in positioning of hydroxyls at C5/C6 of D-KDGal/D-KDG and the disruption of the interaction by the Y131F substitution was expected to affect the condensation with D-GA (Theodossis *et al.* 2004). Single mutation S195M (mutant 2) reduced the D-GA activity nearly 25-fold. In SsoKDGA, the backbone –NH of A198, the residue corresponding to S195 in SacKDGA, interacts with D-KDG through a water molecule. The lower activity of mutant 2 can be explained by introduced steric hindrance, since the larger methionine side chain can interfere with the D-KDG molecule directly. S195M reduced the aldolase activity with 2-Aa about 8-fold. Since the 2-Aa molecule is smaller than D-Ga, diffusion of 2-Aa and its product are less hindered in the active site of mutant 2. Single mutation S238L (mutant 3) reduced both condensations with D-Ga and 2-Aa slightly over 2-fold. S238 is located more to the outside of the binding pocket than S195 and, consequently, the introduction of a large hydrophobic side chain at this position has a relatively small effect on both condensation reactions. The combination of Y131F + S195M eliminated all D-Ga activity in mutants 4 and 7, while the addition of S238L slightly increased the 2-Aa activity in the triple mutant 7.

SacKDGA wild type and the design variants along with SsoKDGA wild type were analyzed for their enantioselectivity in the coupling of pyruvate and 2-Aa via GC-MS analysis of the cyclized reaction products (Figure 2) as described above. Although the SacKDGA variants were designed for optimized binding to improve the production of 5-azido-4(*S*)-hydroxy-2-oxopentanoic acid over 5-azido-4(*R*)-hydroxy-2-oxopentanoic acid, the enantioselectivity proved to be lower than, or equal to the corresponding wild type. For this

reason only two of them, the most promising single mutant 1 and the triple mutant 7, were studied in more detail. For that, the *ee* of these two mutants was determined as a function of reaction time and compared to the two wild type enzymes (Figure 4). It appears that the enantioselectivity of the mutants was not significantly increased compared to the wild type enzyme. Racemization of the products occured more readily for mutant 7 compared to SacKDGA wild type. As explained above, racemization of the product mixture is caused by the reversibility of the enzymatic reaction (Figure 4C). For mutant 1 a slight delay in this process of racemization was observed indicating that the initial, relative rate of formation of 4(R)-product was decreased compared to the 4(S)-product (Table 4). At the start of the reaction, SacKDGA wild type produces about 80 times more 4(S)- than 4(R)-product and mutant 1 produces over 100 times more 4(S)- than 4(R)-product. This indicates a small increase in enantioselectivity for mutant 1 compared to the SacKDGA wild type. During the course of the reaction, however, the *ee* decreases again due to the reaction equilibrium.

In the design procedure the KDGA active site was optimized for binding of the 2-Aa derived moiety of the 5-azido-4(S)-hydroxy-2-oxopentanoic acid condensation product, without taking into account conformational changes during catalysis or diffusion of the reaction products. As such, the reduction in activity observed for the S195M and S238L mutants, which had significantly larger, hydrophobic side chains into their active sites, might be partly due to decreased diffusability of pyruvate into or product out of the active site. Thus the mutations that were predicted in the redesign procedure reduced the overall activity of SacKDGA, but increased the relative substrate specificity (2-Aa/D-Ga) of the aldolase for the condensation of pyruvate with 2-Aa. Introduction of one predicted mutation Y131F (mutant 1) slightly increased the enantioselectivity of SacKDGA for the condensation of pyruvate with 2-Aa. When the three predicted mutations were combined (mutant 7), the enantioselectivity was not increased but the enzyme changed into a specific 5-azido-4(S)-hydroxy-2-oxopentanoic acid aldolase.

Conclusions

In this study a comparison was made of the stereoselectivities of *Sulfolobus solfataricus* and S. *acidocaldarius* KDGA, in the condensation of pyruvate with D-GA and several azidoaldehydes. Both enzymes showed a significantly higher activity towards the reaction with pyruvate and 2-azidoacetaldehyde, compared to the natural reaction. Both aldolases show a preference for the 4(S)-enantiomer of 5-azido-4-hydroxy-2-oxopentanoic acid. Interestingly, SacKDGA appears to be more stereoselective than SsoKDGA in the condensations tested. Even in the natural reaction, where SsoKDGA shows no stereopreference, a stereoselectivity towards D-KDG was observed. This difference in

stereoselectivity of the natural reaction may be attributed to the presence of S195 in the active site of SacKDGA, instead of A198 in SsoKDGA. Additional hydrogen bonding via S195 in SacKDGA may affect the delicate balance of interactions described between D-KGA/D-KDGal and the SsoKDGA active site (Theodossis *et al.* 2004). Modeling of both enantiomers of 5-azido-4-hydroxy-2-oxopentanoic acid into the KDGA active site showed that the OH-group of the 4(*S*)-product can form hydrogen bonds with T155/T157 and Y129/Y130 in the active site of SacKDGA/SsoKDGA, while the 4(*R*)-product only forms only one hydrogen bond with T155/T157.

Mutations were predicted and introduced in order to get more insight in the stereochemical course of the KDGA-catalyzed reaction between pyruvate and 2-Aa. The effect of the three amino acid mutations within the active site of KDGA is an increase in all *k* values (Figure 4C), resulting in a more rapid increase and subsequent decrease of the 4(*S*)-product and a more rapid increase of the 4(*R*)-product. Although more insight in catalysis was obtained, the desired effect of increasing the stereoselectivity of KDGA towards 5-azido-4(*S*)-hydroxy-2-oxopentanoic seemed much harder. Only mutant 1 has a small increase of $v_0(S)/v_0(R)$ and seems to racemize slower. The maximum *ee* was not affected significantly. To further optimize the system towards production of the 4(*S*)-enantiomer of 5-azido-4-hydroxy-2-oxopentanoic acid, k_S/k_R should be increased drastically

Azidoaldehydes are easily accepted as substrates by SsoKDGA and SacKDGA, enabling the use of these enzymes for the preparation of nitrogen heterocycles, which are synthetically useful building blocks en route to pharmaceutically relevant molecules or combinatorial chemistry scaffolds. The enzymes react stereoselectively, but the enantiomeric or diastereomeric excess of the KDGA catalyzed reaction between pyruvate and azidoaldehydes depends on the progress of the reaction and, hence, the concentration of enzyme activity (U/mL) in the reaction and the reaction time. Despite of the high stereoselectivity, the reversibility of the enzymatic reaction may hamper application in continuous processes for the production of enantiopure building blocks. Considering the enzymes and substrates described in this paper, the selectivity is high, but not high enough for an industrially relevant process. Applicability of these enzymes in industrial production processes for nitrogen heterocycles depends on whether mutants can be obtained that give a higher stereoselectivity for the reaction or whether other unnatural nitrogen-containing substrates can be found which are converted with higher stereoselectivity. Furthermore, shortening the reaction time, for example by enzyme immobilization combined with a high flow rate, will give high stereoselectivities. It is to be expected that the spectrum of compounds being converted by pyruvate aldolases will further expand.

Acknowledgements

Chapter 6

Summary and concluding remarks

This thesis describes the results obtained in a project within the Integration Biosynthesis Organic Synthesis (IBOS) program entitled: Synthesizing nitrogen heterocycles from aldehydes via biocatalytic cascades. Interesting enzymes that can perform (part of) these cascades are aldolases. Numerous potential aldolases were found in thermophilic micro-organisms. Some of them were characterized, with respect to their substrate specificity and subsequently subjected to optimization for their use as biocatalysts.

Enzymatic synthesis of Nitrogen heterocycles

Nitrogen heterocycles are organic compounds that contain a nitrogen atom in a ring-structure. A well-known example is the simplest the 6-membered N-heterocycle piperidine, that contains one nitrogen atom (Fig. 1A). Stereochemically different stereoisomers can be created when at least one carbon atom contains substituents, for example 2-carboxy-4(S)-hydroxypiperidine (Fig. 1B). They are common structural elements in nature, often as part of more complex molecules with important biological functions. Well known examples include the bases of DNA, RNA and nicotinamide adenine dinucleotide (NAD), the side chains of the amino acids proline and tryptophan, and more complex ring structures as in flavin and riboflavin co-factors.
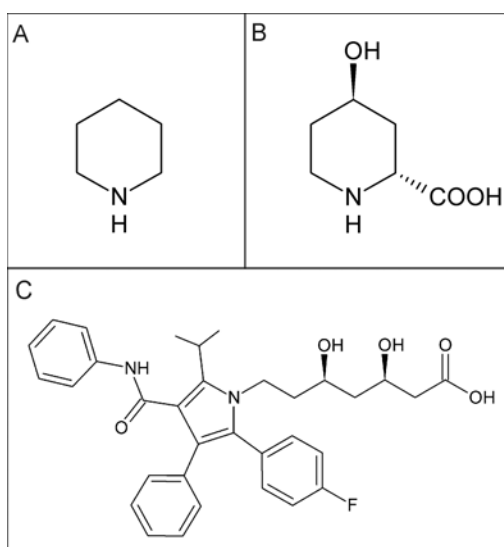
Figure 1. Nitrogen heterocycles
A Piperidine
B 2-carboxy-4(S)-hydroxypiperidine
C Atorvastatin

N-heterocycles are the basis of many pharmaceuticals such as cephalosporins (Castro *et al.* 2006), the antihypertensive drug Omapatrilat (Patel 2001), and the antihistamine drug Astemizole (Brase *et al.* 2002). To obtain stereochemically pure substituted N-heterocycles needed for these drugs, chiral precursor molecules are necessary. On the other hand, direct isolation of substituted N-heterocycles from raw materials (e.g. plant waste) or micro-organisms is a relatively cheap method. For example, isolation of L-proline, *trans*-4-hydroxyl-

L-proline or L-tryptophan from collagen is a well established production method. In addition, mutants of *Brevibacterium flavum* produce L-proline up to 40g/L (Tsuchida *et al.* 1986) and metabolically engineered *Corynebacterium glutamicum* produces up to 43 g/L of tryptophan. Moreover, a wide variety of bioconversions has been developed to prepare functionalized N-heterocycles. Many microbial conversions are performed by hydroxylases (hydroxylating α position) and nitrilases, introducing functionalized groups to substituted pyridines (Petersen and Kiener 1999). Jennewein *et al.* (2006) used 2-deoxy-D-ribose 5-phosphate aldolase (DERA) and variants thereof to synthesize a precursor of statins, (3R,5S)-6-chloro-2,4,6-trideoxyhexapyranoside. Similarly Liu *et al.* (2004) used the sequential aldol condensation catalyzed by DERA mutant Ser238Asp, and synthesized a key chiral component (tert-butyl[(4R,6R)-6-aminoethyl-2,2-dimethyl-1,3-dioxn-4-yl]acetate) of atorvastatin (Fig. 1C) using the azidosubstrate 3-azidopropionaldehyde. Other aldolases that have the potential to accept nitrogen-containing aldehydes can be used for the biosynthesis of (a range of) N-heterocycles.

Using aldolase-based biocatalysis in combination with relatively simple follow-up chemistry (**chapter 5**)**,** we were able to produce the nitrogen heterocycles 4-hydroxyproline, 4-hydroxy-5-methylproline and 2-carboxy-4-hydroxypiperidine. This has been a proof-of-concept for IBOS, the gained knowledge of which may be exploited in future applications.

Exploration of pyruvate aldolases

A wide variety of aldolases is able to catalyze an aldol-condensation reaction. Many aldolases require a phosphorylated aldehyde substrate; these substrates are relatively expensive, and since phosphate is often unwanted in the product it needs to be removed from the product. As our focus was directed to robust enzymes that convert relatively cheap substrates, we initially screened thermophile genome sequences for genes encoding a sub-class of pyruvate aldolases, belonging to COG0329, Neuraminate lyases / dihydrodipicolinate synthases, that generally accept non-phosphorylated substrates. A simple BLAST search using as query some well-characterized pyruvate aldolases that belong to COG0329 has resulted in a long list of hypothetical aldolases from (hyper)thermophilic micro-organisms (Table 1). On the basis of their resemblance to characterized 2-keto-3-deoxygluconate aldolase (Buchanan *et al.* 1999) or dihydrodipicolinate synthase (Laber *et al.* 1992) and their accessibility, 8 genes have been selected for analysis in the course of the research described in this Thesis (Table 1 – printed bolt, Fig. 2A – in boxes). Unfortunately, attempts to produce 3 aldolase candidates (Sso2274, Sso3035 and Sso3072) were unsuccessful despite the application of different vector and fusion systems (reviewed in **Chapter 1**). A close look at the sequences (Alignment of the sequences of Table 1 can be found in the appendix) reveals that the conserved motif present in

pyruvate aldolases (GTTG in 2-keto-3-deoxygluconate aldolase (KDGA), (Theodossis *et al.* 2004)) is not conserved in those hypotheticals, whereas other 'important' features such as the active site lysine (K153) and G177 (in *S. acidocaldarius*) are present. Although it is most likely that these genes encode aldolases, functionality and substrate specificity remains to be established.
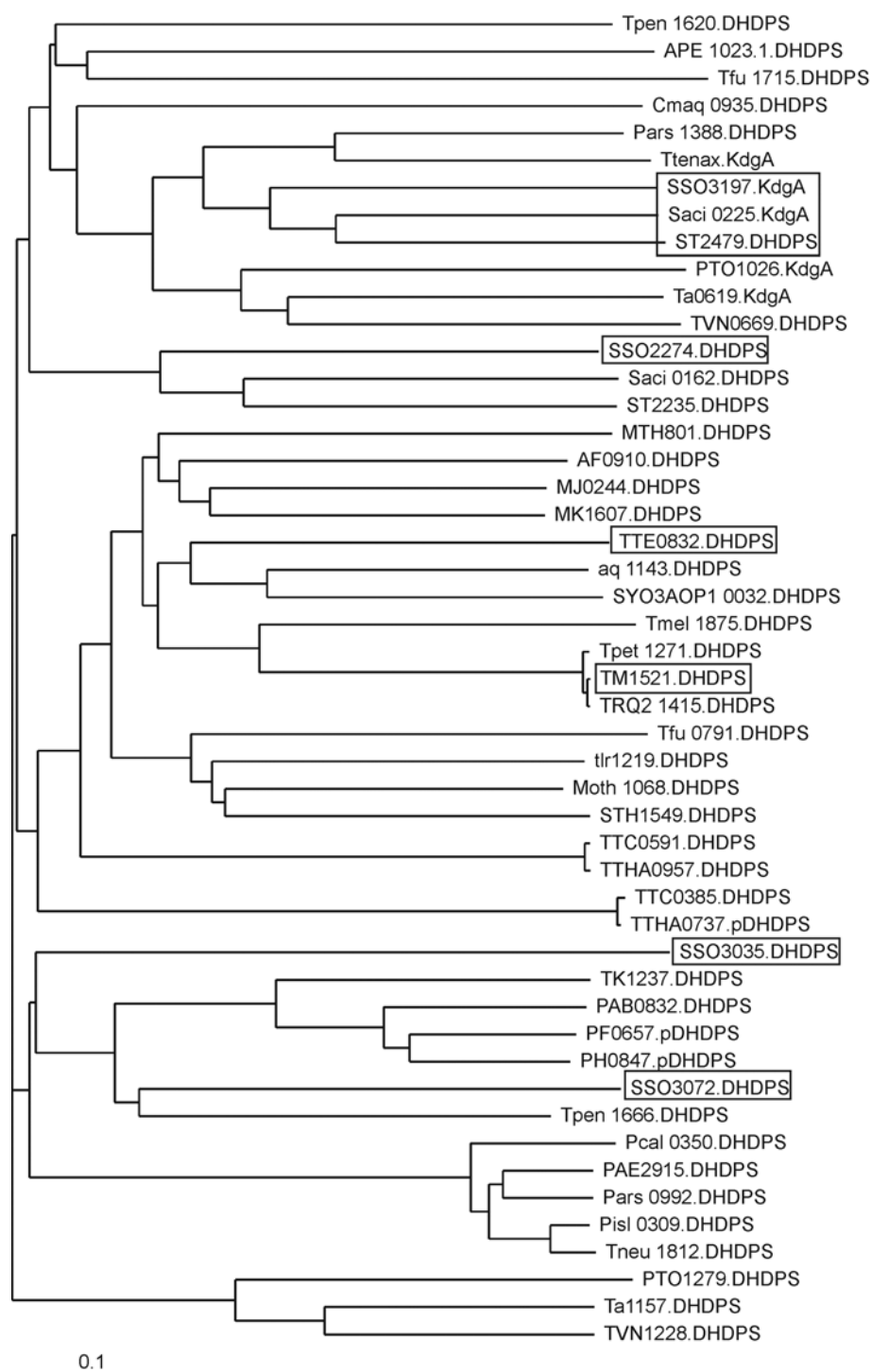
Figure 2. A phylogenetic tree of pyruvate-dependent aldolases belonging to COG0329, Neuraminate lyases / dihydrodipicolinate synthases in thermophilic organisms

Table 1. Hypothetical pyruvate aldolases belonging to COG0329, Dihydrodipicolinate synthase/N-acetylneuraminate lyase (BLAST_P, E-value < 1e-9) in thermophilic organisms. (LBMGE genome query page)

	Gene	(Hypothetical) function	Remarks
1	aq_1143	DHDPS	
2	Moth_1068	DHDPS	Cluster with genes of lysine path
3	SYO3AOP1_0032	DHDPS	
4	STH1549	DHDPS	Cluster with genes of lysine path
5	Tfu_0791	DHDPS	
6	Tfu_1715	DHDPS	
7	tlr1219	DHDPS	
8	TM1521	DHDPS	Characterized (Pearce et al. 2006)
9	Tmel_1875	DHDPS	Cluster similar to T. maritima
10	Tpet_1271	DHDPS	Cluster similar to T. maritima
11	TRQ2_1415	DHDPS	Cluster similar to T. maritima
12	TTC0385	DHDPS	
13	TTC0591	DHDPS	
14	TTE0832	DHDPS	Cluster similar to T. maritima – chapter 2
15	TTHA0737	DHDPS	
16	TTHA0957	DHDPS	
17	APE_1023.1	DHDPS	
18	Cmaq_0935	DHDPS	
19	PAE2915	DHDPS	
20	Pars_0992	DHDPS	
21	Pars_1388	DHDPS	
22	Pcal_0350	DHDPS	
23	Pisl_0309	DHDPS	
24	Saci_0162	DHDPS	
25	Saci_0225	KDGA	Characterized (chapter 3)
26	Sso2274	DHDPS	
27	Sso3035	DHDPS	
28	Sso3072	DHDPS	
29	Sso3197	KdgA	Characterized (Buchanan et al. 1999)– chapter 3)
30	St2235	DHDPS	
31	St2479	KDGA (annotated as DHDPS)	Characterized (chapter 3)
32	Tneu_1812	DHDPS	
33	Tpen_1620	DHDPS	
34	Tpen_1666	DHDPS	
35	Ttenax.KdgA[1]	KDGA	Characterized (Ahmed et al. 2005)
36	AF0910	DHDPS	Reductase dapB present downstream
37	MJ0244	DHDPS	
38	MK1607	DHDPS	
39	MTH801	DHDPS	Cluster similar to T. maritima – characterized(chapter 2)
40	PAB0832	DHDPS	
41	PF0657	DHDPS	
42	PH0847	DHDPS	
43	PTO1026	KDGA	
44	PTO1279	DHDPS	
45	Ta0619	KDGA	
46	Ta1157	DHDPS	
47	TK1237	DHDPS	
48	TVN0669	DHDPS	
49	TVN1228	DHDPS	

1 – 16 hits in Bacteria, 17 – 35 hits in Crenarchaea, 36 – 49 hits in Euryarchaea. APE – *Aeropyrum pernix*, AF – *Archaeoglobus fulgidus,* aq – *Aquifex aeolicus*, Cmaq – *Caldivirga maquilingensis*, MJ – *Methanococcus jannashii*, MK – *Methanopyrus kandleri*, MTH – *Methanothermobacter thermautotrophicus*, Moth – *Moorella thermoacetica* ATCC 39073, PTO – *Picrophilus torridus*, PAE – *Pyrobaculum aerophilum*, Pars – *Pyrobaculum arsenaticum*, Pcal – *Pyrobaculum calidifontis*, Pisl – *Pyrobaculum islandicum*, PAB – *Pyrococcus Abysii*, PF – *Pyrococcus furiosus*, PH – *Pyrococcus horikoshii*, Saci – *Sulfolobus acidocaldarius*, Sso – *Sulfolobus solfataricus*, St – *Sulfolobus tokodaii*, SYO - *Sulfurihydrogenibium* sp. YO3AOP1, STH – *Symbiobacterium thermophilum*, Tfu – *Thermobifida fusca* YX, TK – *Thermococcus kodakaraensis*, Tneu – *Thermoproteus neutrophilus*, Tpen - *Thermofilum pendens*, Ta – *Thermoplasma acidophilum*, TVN – *Thermoplasma volcanium*, Tmel – *Thermosipho melanesiensis* BI429, trl – *Thermosynechococcus elongates*, TM – *Thermotoga maritima*, Tpet – *Thermotoga petrophila*, TRQ2 - *Thermotoga* sp. RQ2, TTC – *Thermus thermophilus* (HB27 Chromosome), TTHA – *Thermus thermophilus* (HB8 Chromosome).

[1] From reference, not available in LBMGE database.

Most hypotheticals are indicated as dihydrodipicolinate synthase (DHDPS). Two bacterial DHDPSs were cloned and expressed in *E. coli* during this research (TM1521 and TTE0832). Soluble enzymes were produced (although this was much easier for TTE0832) and they were able to condensate pyruvate and (*S*)-aspartate-β-semi-aldehyde ((*S*)-ASA), a key step in the lysine biosynthesis pathway. This was in good agreement with the observation that the genes encoding these enzymes cluster with lysine-synthesis genes on the genomes of both species (as well as in the ones indicated in Table 1). *T. tengcongensis* DHDPS was characterized in detail (**Chapter 2**). As expected, it appeared to be a stable enzyme and its catalytic properties were quite similar to that of *T. maritima* and *E. coli* DHDPS (Dobson *et al.* 2004b, Pearce *et al.* 2006). DHDPS is generally very specific for the aldehyde substrate, (*S*)-ASA, which is unusual for a pyruvate dependent aldolase (Samland and Sprenger 2006). A single arginine (R138 in *E. coli* DHDPS) seemed responsible for coordinating the carboxyl from (*S*)-ASA (Blickling *et al.* 1997b). Substitution of this arginine residue made DHDPS function as a neuraminate lyase (Joerger *et al.* 2003).

Apart from the previously studied *Sulfolobus solfataricus* 2-keto-3-deoxygluconate aldolase (SsoKDGA), *S. acidocaldarius* and *S. tokodaii* 2-keto-3-deoxygluconate aldolase s (SacKDGA and StoKDGA) were characterized in **Chapter 3**. Biochemical properties of the three enzymes appeared to be very similar. All three enzymes proved to have a remarkable broad substrate specificity regarding the aldehyde acceptor: aldehydes with 2-5 carbon atoms were accepted. In **Chapter 5** it has been described that SacKDGA and SsoKDGA readily accept azido-substituted aldehydes as well. Condensation rates for the simplest aldehyde 2-azidoacetaldehyde (with pyruvate) was even higher than the rate of the natural reaction (pyruvate with D-glyceraldehyde) at 50°C. The 3-dimensional structure of KDGA proved that water-mediated interactions permit binding of substrates in multiple conformations (**chapter 3** – (Theodossis *et al.* 2004)) in the spacious hydrophilic binding site, which correlates with the observed broad substrate specificity. The crystal structure of *S. acidocaldarius* KDGA

revealed the presence of a novel phosphate binding motif that allows the formation of multiple hydrogen bonding interactions with the acceptor substrate, and enables high activity with glyceraldehyde-3-phosphate. A similar binding motif was recently found in the *T. tenax* KD(P)GA structure (Pauluhn *et al.* 2008).

Laboratory evolution of (pyruvate) aldolases

One of the crucial aspects of high-throughput laboratory evolution experiments is the development of a screening assay. To be able to generalize the assay for further use, a screening assay was developed that can be used in aldol condensation reactions of pyruvate aldolases in general. Therefore, determination of the remaining pyruvate, using an enzymatic assay with NADH and L-lactate dehydrogenase was used. Obvious drawback of this assay are that the amount of remaining substrate is detected, and that no information is obtained about the product formed. Fong *et al.* (2000) used a similar assay procedure in directed evolution of an *E. coli* KDPG aldolase. In the latter study, error-prone PCR and subsequent gene shuffling had been performed on *E. coli* KDPG aldolase, after which the generated library was screened for acceptance of the phosphate-free substrate variant KDG. The extracts were tested for activity on KDG and produced pyruvate, which could be measured using NADH and L-lactate dehydrogenase in a continuous assay. Furthermore, they screened for the synthetic reaction for variants accepting L-glyceraldehyde (which is not accepted by the wild type enzyme), thus changing selectivity aiming for the synthesis of D- and L-sugars. Hsu *et al.* (2005) also used the same continuous assay based on the production of pyruvate. They performed laboratory evolution on *E. coli* N-acetyl neuramic acid lyase. After 5 rounds of error-prone PCR they obtained an efficient L-3-deoxy-manno-2-octulosonic acid (KDO) aldolase.

In **Chapter 4** *S. acidocaldarius* KDGA aldolase mutants (clone library of ~1500 random mutants) were screened for an increased activity at 50°C (40° below the enzymes optimum temperature). After further selection (heat stability), one mutant appeared to have significant higher activity (3 times) and was selected for further analysis. This higher activity could indeed be detected with several other, relatively small aldehydes, using high substrate concentrations; sequence analysis revealed a single mutation V193A. A close look at the active site revealed that catalytic lysine has Van der Waals interactions with the mutated residue V193. Mutation of this residue to alanine presumably increased local flexibility of K153, explaining both higher activity at lower temperatures and conservation of its stability. Apparently this higher flexibility interfered with larger substrates.

A similar screening method was used with 2-azidoacetaldehyde as a substrate, at an assay temperature of 37°C –using a newly made library of random *S. acidocaldarius* KDGA

mutants–. Unfortunately no mutants could be selected with significant higher activity in - condensating pyruvate and 2-azidoacetaldehyde- compared to *S. acidocaldarius* KDGA.

Engineering by rational / computational design:

Several aldolases have been altered using rational design. DeSantis *et al.* (2003b) made 5 mutants of *E. coli* 2-deoxyribose-5-phosphate aldolase (DERA, EC 4.1.2.4) to improve activity towards the unnatural non-phosphorylated substrate, D-2-deoxyribose. This resulted in a mutant S238D that exhibited a 2.5-fold improvement over the wild type enzyme in the retroaldol reaction of 2-deoxyribose. Seebeck and Hilvert (2003) used an alanine racemase (a pyridoxal 5'-phosphate-dependent enzyme) and converted it to an aldolase upon replacement of Tyr265. This Y265A mutant is highly stereoselective for (2R,3S)-phenylserine, a D-amino acid, and does not process its enantiomer. This aldolase may be used in the synthesis of β-hydroxy-α-amino acids (threonine aldolases as well, as they can perform the same reaction).

With the availability of the 3-dimensional structures of *S. solfataricus* (Theodossis *et al.* 2004) and *S. acidocaldarius* KDGA (**Chapter 3**) (Wolterink-van Loo *et al.* 2007), attempts have been made to use a rational design approach to improve SacKDGA towards (stereoselective) synthesis of 5-azido-4(*S*)-hydroxy-2-oxopentanoic acid by modeling this exact product in the enzymes active site (**Chapter 5**). Based on the outcome of the 'chameleon' algorithm, three residues (Y131F, S195M, S238L) in the vicinity of the active site were mutated separately and in combination. We succeeded in changing the relative selectivity of the natural substrate glyceraldehyde towards 2-azidoacetaldehyde. Specific activity of most mutants was much lower than of the wild type, probably due to the introduction of larger residues in the active site, causing steric hindrance. Unfortunately the stereospecificity had not increased. Apparently the mutations optimized for 5-azido-4(*S*)-hydroxy-2-oxopentanoic acid did not block activity towards synthesis of 5-azido-4(*R*)-hydroxy-2-oxopentanoic acid. Cheriyan *et al.* (2007) were (similarly) able to improve the selectivity for unnatural substrates for a KDPGA, by introducing a single mutation (S184L) near the active site.

Based on the triple mutant of SacKDGA (Y131F, S195M and S238L) saturation mutagenesis was performed on positions: T42, T43, F131, M195, L238 and I239 (Fig. 3), choosing the residues that apparently interact with the substrate/product by visual inspection. These libraries were screened for activity with azidoacetaldehyde and pyruvate in condensation reaction. No mutants with enhanced activity compared to wild type were seen, although some mutants were selected that exhibited significantly higher (specific) activity compared to the triple mutant. Apparently due to the decrease in activity (because of a major reduction in produced soluble protein) of the triple mutant, it seemed impossible to select

mutants with improvement in comparison with the wild type. Changing F131 to serine did have a positive effect on activity compared to the triple mutant, but activity was still much lower than wild type. None of the selected mutants had higher stereospecificity compared to wild type or triple mutant SacKDGA.
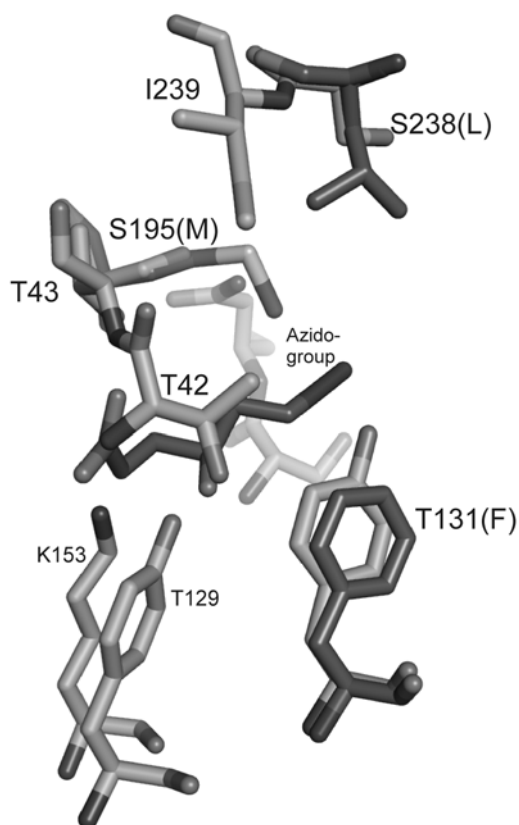
Figure 3. Active site residues in *S. acidocaldarius* KDGA. In grey, residues that are targeted in saturation mutagenesis experiments. In dark grey 5-azido-4(*S*)-hydroxy-2-oxopentanoic acid and residues of the triple mutant (residues between brackets) .

Aldolases in (bio)cascades

A nice example of a possible enzymatic cascade reaction was developed by Sanchez-Moreno *et al.* (2004). Three different enzymes: dihydroxykinase, acetate kinase and fuculose-1-phosphate aldolase (Chapter 1, Fig. 3) and catalytic amount of ATP were used. The cascade started with the (non-phophorylated) substrate dihydroxyacetone or other aldehyde substrates, yielding different amounts of product. Although starting with a cheap non-phosphorylated substrate, the end-product does contain phosphate, which is probably unwanted in the final product, and therefore still has to be removed.

Another one-pot cascade reaction was performed by Siebum *et al.* (2006) (as part of the IBOS-project). An alcohol oxidase was combined with DERA. The aldehyde synthesized could be used by the aldolase, and even lead to higher yield, as the aldehyde could not compete with the alcohol for the active site of the oxidase.

A possible biocascade reaction using the KDGA was combining this enzyme with an alditol oxidase (Heuts *et al.* 2007). Similarly to the reactions tested by Siebum *et al.*, an alcohol should be converted by the alditol oxidase to an aldehyde, which, together with pyruvate, would be substrate for the aldolase. 4-hydroxypentenol was used as a substrate, but unfortunately the desired product has never been detected (Eric van Hellemond, personal communication). The temperature and pH range of both enzymes have a narrow overlap, which was a major drawback for an effective combination of the two enzymes. Random optimization of the aldolase towards the specific optimal conditions of the oxidase (or vise versa) would be an option to functionalize the cascade reaction.

Future perspectives

The goal of the research was to synthesize nitrogen heterocycles in a combined biological-chemical approach. The experiments in **Chapter 5** showed we indeed could make the nitrogen heterocycles by combining an aldolase and follow-up chemistry. The intrinsic broad substrate specificity was exploited by using the unnatural azido-containing aldehydes. The azide group could easily be used in the subsequent cyclization of the product.

Moreover, N-heterocycles were synthesized by Ran *et al.* (Ran *et al.* 2004, Ran and Frost 2007). They used the 2-keto-3-deoxy-6-phosphogalactonate aldolase (EC 4.1.2.21) of *E. coli* and changed its activity towards the condensation reaction: pyruvate + D-erythrose-4-phosphate → 3-deoxy-D-arabino-heptulosonic acid 7-phosphate (DAHP), replacing phosphoenolpyruvate synthase and DAHP synthase, in a strain lacking all isozymes of DAHP synthase. Finally this would lead to production of shikimic acid, which is a starting material in the synthesis of Tamiflu.

The biochemical and structural characterization of KDGA and DHDPS provided knowledge necessary for further research. The knowledge obtained from enzyme 3-D structures combined with the effect of certain mutations on activity, specificity and stability will certainly be useful for further rational design approaches of the KDGA. On the other hand, it has been demonstrated in **Chapter 4**, that without any structural knowledge of the aldolase we can obtain an enzyme with properties we select for. The only, but very important prerequisite is a good screening assay for the selection of the exact desired property.

Furthermore, studying the structurally closely related pyruvate aldolases in more detail (combining sequence analysis and experimental data) reveals more specific targets for rational design. For example, (Joerger *et al.* 2003) were able to change the structurally closely related neuraminate lyase into a functional DHDPS by introducing an active site arginine at the corresponding amino acid position in *E. coli* neuraminate lyase. Based on the assumption that the arginine is a strict requirement of a DHDPS, only the bacterial hypothetical DHDPSs from

Table 1, and AF0910, MJ0244, MK1607 and MTH801 will probably be true DHDPSs, which is confirmed by the clustering in the phylogentic tree (Fig 2.). Hence, in order to broaden the substrate specificity of DHDPS, this arginine might be targeted and changed into a smaller residue. The arginine is conservatively replaced by a leucine in *E. coli* neuraminate lyases (Izard *et al.* 1994) and by an alanine (A134 in Sac) in the KDGAs, which both have specificity for an other aldehyde substrate.

An issue not directly discussed in this research is the real biotechnological application of the aldolases. In **Chapter 5** we showed the enzymes had high stereoselectivity, but application is complicated because of the reversibility of the enzymatic reaction, leading to racemization of the product. To overcome this problem, one could aim for the mutation of the enzyme to completely block the reaction towards the 'R'-product. On the other hand, developping a system in which the desired (S) product is selectively removed by either binding or further conversion, would be a good alternative.

Nederlandse samenvatting

Enzymatische synthese van heterocyclische stikstofverbindingen

Heterocyclische stikstofverbindingen zijn organische stoffen die een ringstructuur hebben met daarin tenminste één stikstofatoom. Piperidine (Fig. 1A – Summary and concluding remarks) is zo'n heterocyclische stikstofverbinding met 6 atomen die een ring vormen, waarvan één stikstof. Wanneer een koolstofatoom 4 verschillende groepen heeft (een zogeheten asymmetrisch koolstofatoom), zijn er 2 verschillende stereo-isomeren mogelijk (Fig. 1B – Summary and concluding remarks). Deze stereo-isomeren zijn elkaars spiegelbeeld (en heten enantiomeren): ze zijn ruimtelijk niet hetzelfde, chemisch wel. Als er meerdere asymmetrische koolstofatomen in een molecuul voorkomen, zijn er nog meer verschillende stereo-isomeren mogelijk die chemisch wel en ruimtelijk niet hetzelfde zijn, maar ook niet perse elkaars spiegelbeeld zijn (diastereomeren). Heterocyclische stikstofverbindingen komen in de natuur veel voor in bekende biologische stoffen zoals DNA, RNA en 'zijketens' van de aminozuren proline en tryptofaan. Heterocyclische stikstofverbindingen komen ook veel voor in medicijnen, waar de stereochemie erg belangrijk is. Vaak is het ene stereo-isomeer werkzaam en het andere niet (en soms zelfs toxisch). Het kunnen maken van één specifiek stereo-isomeer (en dus geen mengsel) is dan belangrijk.

In het onderzoek beschreven in dit proefschrift zijn we op zoek gegaan naar enzymen waarmee (stereoselectief) heterocyclische stikstofverbindingen gemaakt kunnen worden. Bij synthese met enzymen (biosynthese) zijn doorgaans minder milieubelastende condities (bijv. organische oplosmiddelen) nodig dan bij de traditionele organische synthese van heterocyclische stikstofverbindingen. Bovendien zijn enzymen vaak van nature stereoselectief. Aldolases zijn enzymen die een aldehyde (acceptor) en keton (donor) aan elkaar kunnen koppelen (Fig. 1A). Met de juiste stikstofhoudende substraten (stoffen waaruit de enzymen verbindingen kunnen maken) kunnen dan met behulp van additionele chemische synthese heterocyclische stikstofverbindingen gemaakt worden. Om stabiele enzymen te verkrijgen zijn in dit onderzoek enzymen uit thermofiele micro-organismen gehaald. De substraten zijn idealiter – met het oog op een industriële toepassing – zo eenvoudig en goedkoop mogelijk. Verder kan de biosynthese nog uitgebreid worden door verschillende enzymen samen te voegen, waarbij het product van het 1ᵉ enzym als substraat kan dienen voor het 2ᵉ enzym (biocascade-reactie). De in dit onderzoek gebruikte acceptor-substraten om heterocyclische stikstofverbindingen te maken, komen niet in de cel voor en zijn dus 'vreemd / onnatuurlijk' voor het enzym. Aangezien het enzym aangepast is aan het natuurlijke substraat is het soms nodig het te veranderen (genetisch te modificeren), zodat de katalyse van het onnatuurlijke substraat beter verloopt.
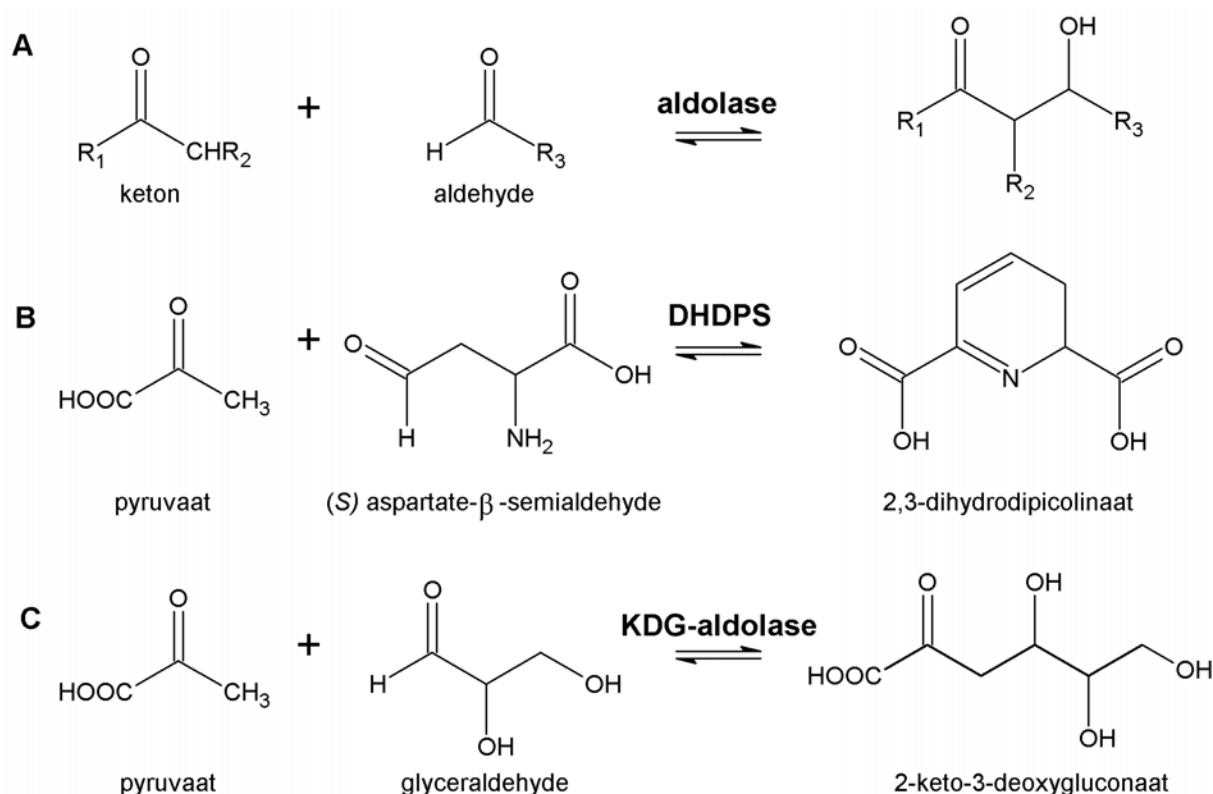
Figuur 1. (1A) Het aan elkaar koppelen van een aldehyde en keton (1B) Reactie gekatalyseerd door dihydrodipicolinaatsynthase (1C) Reactie gekatalyseerd door 2-keto-3-deoxygluconaataldolase

Aldolases

Er zijn veel verschillende aldolases en de meeste gebruiken één specifiek donormolecuul en kunnen daaraan verschillende aldehyden koppelen (één per reactie). In dit onderzoek zijn aldolases onderzocht die met pyruvaat als donor werken. Twee enzymen, dihydrodipicolinaatsynthase (DHDPS) en 2-keto-3-deoxygluconaataldolase (KDGA) kregen specifieke aandacht. DHDPS katalyseert in de cel de reactie van pyruvaat met (*S*)-aspartaat-β-semi-aldehyde en vormt daarbij 2,3-dihydrodipicolinaat (DHDP) (Fig. 1B). Deze stap is onderdeel van de *in vivo* syntheseroute van het essentiële aminozuur lysine. KDGA katalyseert de evenwichtsreactie van 2-keto-3-deoxygluconaat met pyruvaat en glyceraldehyde (Fig. 1C), als onderdeel van de Entner-Doudoroff afbraakroute (om energie uit suikers te verkrijgen).

Met behulp van de aminozuurvolgorde van bekende aldolases (DHDPS uit *Escherichia coli* en KDGA uit *Sulfolobus solfataricus*) zijn homologe potentiële aldolase-enzymen gevonden in thermofiele micro-organismen. Deze staan in een fylogenetisch boom

(Fig. 2, Summary and concluding remarks), waarbij aldolases die veel op elkaar lijken dicht bij elkaar staan. De omlijnde enzymen zijn verder onderzocht.

Productie en activiteit van aldolases

De aldolases zijn geproduceerd met behulp van expressiesystemen in de bacterie *Escherichia coli*, waarvoor de coderende DNA-sequentie wordt gebruikt. De productie van Sso2274, Sso3035 en Sso3072 is nooit gelukt, ondanks het gebruik van verschillende expressiesystemen (o.a. met een ander (stabiel) eiwit eraan vastgekoppeld). Soms werd er geen eiwit geproduceerd en als er wel eiwit werd geproduceerd, was dit niet oplosbaar, hetgeen essentieel is voor het meten van activiteit.

De productie van twee potentiële bacteriële dihydrodipicolinaatsynthases, TM1521 (TmaDHDPS) en TTE0832 (TteDHDPS), leverde wel oplosbare enzymen op. Beide vertoonden activiteit met de natuurlijke substraten pyruvaat en (*S*)-aspartaat-β-semi-aldehyde ((*S*)-ASA). Het TteDHDPS is uitgebreid onderzocht (hoofdstuk 2). De biochemische eigenschappen komen goed overeen met het DHDPS uit *E. coli*. TteDHDPS bleek heel stabiel maar was ook erg specifiek voor het aldehyde-acceptorsubstraat (*S*) ASA en is dus in dit onderzoek verder niet gebruikt voor biosynthese.

Naast het inmiddels bekende *S. solfataricus* KDGA (SSO3197), zijn ook de KDGA's uit de archaea *S. acidocaldarius* (Saci_0225) en *S. tokodaii* (ST2479) geproduceerd in *E. coli*. Deze drie enzymen zijn gekarakteriseerd (hoofdstuk 3). Niet onverwacht bleken ook deze enzymen erg stabiel (na minuten koken vertoonden ze nog steeds activiteit) en 'normale' activiteitstesten werden dan ook uitgevoerd bij 70°C. De verschillen tussen de enzymen zijn erg klein, dus zal er verder geen onderscheid gemaakt worden tussen de verschillende KDGA's. Het KDGA kan reacties katalyseren van pyruvaat met verschillende aldehydes, variërend van 2 tot 5 koolstofatomen in de keten, waar het natuurlijke substraat glyceraldehyde er 3 heeft. Het KDGA accepteert zelfs aldehydesubstraten waarin een azido-groep (N_3) voorkomt (hoofdstuk 5). SacKDGA (hoofdstuk 2) en SsoKDGA (Theodossis *et al.* 2004) zijn gekristalliseerd, waarmee de 3-dimensionale structuur van het eiwit ontrafeld kon worden. Dit leverde informatie op over de mogelijke bindingen die de substraten in en met het enzym aangaan. Het ruime maar stugge 'actieve centrum' verklaart de brede aldehydesubstraat-range (hoofdstuk 3). Verder bleek dit enzym niet 100% stereospecifiek en synthetiseerde het met het natuurlijke substraat glyceraldehyde naast 2-keto-3-deoxygluconaat ook 2-keto-3-deoxygalactonaat (dit zijn diastereomeren van elkaar) (hoofdstuk 5 - Lamble *et al.* 2005b). Hierbij bleken kleine verschillen in eiwitvolgorde verschillen in specificiteit op te leveren (het SacKDGA is specifieker dan SsoKDGA).

Aanpassen van enzymen in het laboratorium

Er zijn twee methodes gebruikt om SacKDGA te verbeteren. Eén techniek is de zogenaamde random-methode. Hiervoor heb je geen achtergrondkennis van het enzym nodig. De andere techniek, rationeel ontwerpen, vergt het nodige inzicht in de reactiemechanismen en de betrokken aminozuurresiduen.

Random optimalisatie

Met behulp van error-prone PCR werd het voor het aldolase-enzym coderende gen (*SackdgA*) vermenigvuldigd. Hierbij worden met opzet fouten in het gen geïntroduceerd, waardoor er mutaties ontstaan die kunnen resulteren in het inbouwen van een ander aminozuur in het enzym. Zo'n mutatie kan een positief effect hebben op de gewenste activiteit, maar ook een neutraal of een negatief effect. Daarom is er een goede screeningstest nodig om de (~1500) enzymvarianten te kunnen testen. De test die in hoofdstuk 4 is gebruikt, meet na een vooraf bepaalde reactietijd de overgebleven hoeveelheid pyruvaat. Er werd getest op de koppeling van pyruvaat en glyceraldehyde bij 50°C (40° onder de optimale temperatuur van het enzym). Een enzym dat bij deze (of nog lagere) temperatuur veel actiever is dan het wildtype SacKDGA zou mogelijk in een biocascade met een enzym van 'koude' origine gebruikt kunnen worden. Na verschillende, steeds nauwkeurigere testen, bleef er één enzym over dat een driemaal hogere activiteit had dan het wildtype SacKDGA. De DNA-volgorde (en daarmee ook de aminozuurvolgorde) van deze mutant is bepaald. Daaruit bleek dat deze mutant slechts één mutatie bezit die de nieuwe eigenschappen veroorzaakt. Want hoewel het enzym sneller werkt bij een lagere temperatuur is het net zo stabiel als het wildtype SacKDGA. Met dezelfde screeningstest konden andere aldehyden op activiteit worden getest, maar in deze tests zijn geen nieuwe, betere mutanten gevonden.

Rationeel ontwerpen

Tegenwoordig wordt een rationeel ontwerp vaak voorspeld met behulp van computerprogramma's. Hiervoor moeten de 3-dimensionale structuur en de interacties met het substraat bekend zijn. Zo'n computerprogramma kan dan optimale interacties (binding) berekenen voor een nieuw substraat of (desgewenst) product. Het is echter niet eenvoudig er voor te zorgen dat er dan ook activiteit optreedt. In hoofdstuk 5 is met behulp van zo'n computerprogramma het 'S' product (dwz het stereo-isomeer met S-configuratie) van de reactie van 2-azidoacetaldehyde met pyruvaat (5-azido-4(*S*)-hydroxy-2-oxopentaanzuur) (Zie Fig. 2 in hoofdstuk 5) in het actieve centrum van SacKDGA gezet. Volgens voorspelling zouden drie mutaties nodig zijn om dit specifieke stofje te binden, namelijk: Y131F, S195M en S238L. Deze mutaties zijn afzonderlijk en in combinatie aangebracht in SacKDGA en

getest op activiteit. De drievoudige mutant was vele malen beter in het koppelen van pyruvaat met 2-AA in vergelijking tot het koppelen van pyruvaat met glyceraldehyde (natuurlijke reactie), maar absoluut gezien is de reactiesnelheid niet verhoogd. De verhouding tussen 5-azido-4(S)-hydroxy-2-oxopentaanzuur en 5-azido-4(R)-hydroxy-2-oxopentaanzuur bleef ook onveranderd. Kennelijk zijn de mutaties die nodig zijn om 5-azido-4(S)-hydroxy-2-oxopentaanzuur goed te binden ook goed om 5-azido-4(R)-hydroxy-2-oxopentaanzuur te binden. Om deze verhouding en daarmee de stereospecificiteit van het enzym te verbeteren zou een model nodig zijn waarin berekend wordt welke mutaties 5-azido-4(S)-hydroxy-2-oxopentaanzuur binding positief beïnvloeden en tegelijkertijd 5-azido-4(R)-hydroxy-2-oxopentaanzuur binding negatief beïnvloeden.

(Bio)-synthese van heterocyclische stikstofverbindingen

In hoofdstuk 5 zijn diverse reacties uitgevoerd van pyruvaat met azido-substraten. De producten ervan zijn gereduceerd, waardoor uiteindelijk heterocyclische stikstofverbindingen ontstaan. Op deze manier zijn verschillende heterocyclische stikstofverbindingen geproduceerd. Het enzym had een voorkeur voor het katalyseren van de reactie naar het S-product. Het hoogste gemeten percentage van één stereo-isomeer is ca. 90%.

Appendix

```
                                                    GTTG
Ta1157.DHD . : ------------MYKGIVTPMITPMGQNG-EIDYRATEILIDNLADF-GVDGLFPMGSTCLFPMFSTDEKKFLGFVRDHSKKIEVYAGVGSSS--TQESVELSKYTEDIGIKV :  98
TVN1228.DH . : ------------MYNGIVTPMLTPNRAG-EIDYDATRLLIQDLKGY-GVSGLFPMGSTCLFPMFSLEERRKFLQFVSENRNLTVYAGVGSSS--TQESIELSKFTEDVGIEV :  98
PTO1279.DH . : ---------------MITPLDAHG-NIDYNATNILIKYLEGI-NVDYLFPMGSTCVFPYFTLKERKDFLKFVRENSKK-PIMAGVGSSS--INEVNELMKFSMDIGIEA :  89
PF0657.pDH . : ----------MEGVIVPLVTPFKEDY-SIDFEALEWHIEFLEEK-GVHGIFTNSTTCEFTSLSLEERKILAEKGRELTSKAYYLVGTGSTN--TFEVIELTKHAKEIGADG :  97
PH0847.pDH . : ----------MEGVIVPLVTPFREDY-SIDFEALEWHIRFLEEK-GVHGIFVNSTTCEFTSLNTDERKMLAEKGREITSGM-YLVGTGSTS--TLEVIELSRHAEDIGADG :  96
PAB0832.DH . : ----------MRGVIVPIVTPFKEDY-SIDVPALEEHLDYLQKV-GVHGIFINATTCEFTSLSKEERRFPLAEKGRELVTSAFYLVGTASSN--TLEVIELTKHAQDIGADY :  96
TK1237.DHD . : ----------MRLEGVVPLITPLEDY-SIDKEGLKWLVSYLSEN-GVNGIFVNSSVCEFASLTLDEMKLVTKIALDSRKSYSTLIFAGASTNFTEETIKLAKDFKDMGVDA :  97
SSO3072.DH . : ------MSARFYGVISPFITPFREDL-SLDREAVAMLARYQAEK-GVHGIFPNSTCEVFHLSREEAVEVTRLVLEAVGG-KVWVIPGISANYTEDSVALGRTFKDLGVDG : 101
Tpen 1666. . : -----------MKLEGVVATVMPFAKD-GVNYEGLRLLLSKIVEA-GYHGVFPTSSTCEVTKLLTPEERVKVMEVAKEVAGG-KALVVAGTGDHVSTIDIARKYKDVGVDA : 102
PAE2915.DH . : -----------MKLEGVIAATVTPFTKD-GVNYEGLRIILSRIVEA-GYQGVFPTSSTCEVTKLLTPEERVKVMEVAKEVAGG-KALVIAGTGDHLSTIDMVRRYKDVGVDV :  99
Pars 0992. . : ----------MSMRIEGVIVATVTPFTKD-GVNYEGLRALLSKIVSE-GYQGVFPTSSTCEVTKLITFEERVKVAMEVAKEVAGG-RALVVAGTGTGDHLSTIEIARRYRDVGVDA :  99
PisI 0309. . : ----------MRIEBGVVATVTPFAKD-GVNFEALRALLAKVAAD-GYH-LFPTSSTCEVTKLITVEERVKVMEVAKEVAGG-SVYVVAGTGDHVSTIDIIRRYRDVGVDF : 101
Tneu 1812. . : -------MDVLSPMKLEGVVAATVTPFSKD-GVNFEALRALLAKVAAD-GYH-LFPTSSTCESPTLSHEEHHKVIEKVDVVNG-RVQVIAGAGSNCTEEAIELSVFAEDVGADA :  99
Pcal 0350. . : ----------MFKGVYPAIITPFKN--KEVDFDGLEENINFLIEN-GVSGIVAVGTCBSSTLSHAEHRRVIEIVVDEVNG-KVPVIAGAGSNSTREAALELSTYAEDVGADA : 104
MJ0244.DHD . : ------MGFQIEGVIPALITPFTDDLKGINEEGLRENVSRLLEA-GVHGVVPAGTCBSSTLSHAEHRRVIEIVVDEVNG-KVPVIAGAGSNSTREAALELSTYAEDVGADA :  98
MK1607.DHD . : ------------MFEGVIPAMVTPFKED-FSVDYEGLAALENHIKE-GTDAILICGTTCEAATLSYEEHIDVVRYVAETS---KLPVIGGAGSNSTREAAIWLAKEVKAGAEA : 103
AF0910.DHD . : ------------MKIEGTVVAMVTPFTED-DVVDEAGLRENINYLIEN-GVDGLLVAGTCBSATITHEEQRRMIDILVDEVNG-RVRTVAGAGSNSSREAMGLVEYAEDAGADA :  96
MTH801.DHD . : ------------MFQGSIVALITPFKEG--EVDYEALGNLIEFHVDN-GTDAILVCGTCBSPTLTFEEHEKVIEFAVKRAAG-RIKVIAGTGNATHEAVHLTAHAKEVGADG : 100
aq 1143.DH . : ------------MFYGSVVALITPFKDG--TLDRQGLKRLIEFHIEN-GTDAIVVAGTCBSPTLFEEHELLIELAVGYSNK-RIPIIAGTGANSTHEAITLTKFAEKVGADA :  98
SYO3AOP1_0 . : ------------MPVFKGSCVAIVTPFTEN--GVNFDKLGELIEWHIKE-GTDAILICGTCBASTMTDEEQKEAIKFTVSKVAK-RIPVIAGTGSNNTAHAIELSEYAQSVGADA :  98
TTE0832.DH . : ------------MFRGVGTAIVTPFKNG--ELDLESYERLVRYQLEN-GVNALIVLGTTCBSPTVNEDEREKIVSRTLEIVDG-KIPVIVGAGTNSTEKTLKLVKQAEKIGANG : 100
TM1521.DHD . : ------------MFRGVGTAIVTPFKNG--ELDLESYERLVRYQLEN-GVNALIVLGTTCBSPTVNEDEREKIVSRTLEIVDG-KIPVIVGAGTNSTEKTLKLVKQAEKLGANG :  98
TRQ2 1415. . : ------------MFRGVGTAIVTPFKNG--ELDLESYERLVRYQLEN-GVNALIVLGTTCBSPTVNEDEREKIVSRTLEIVDG-KIPVIVGAGTNSTEKTLKLVKQAEKLGANG :  98
Tpet 1271. . : -----------MFSGIGTAIITPFKNG--EIDYHALKNFLETQRM--VDAIILLGTTCBAPNISIEERDKLIFVREYFPD-K-PLIVGVGTNSSHHTMELVKNAEKNKADA :  98
Tmel 1875. . : ------------MQWGRILTAMVTPFTAD-GKLDLDGARRLAAYLVDH-GSDGLVVAGTCRAAVIAGTGTNSTAASIELSREAEALGVDG :  95
Moth 1068. . : ------------MPTFGRLLTAMVTPMTPD-GAVDYQRAGELAKHLVAA-GSEGLVVSGTCQVSVIAGTGSNNTAESIRFSREAARTGVHG : 100
STH1549.DH . : ------------MTDFGRVITAMITPFTAD-GAIAYDVAAKLAQHLVAN-GSDGIVVCGTCKAKIIAGTGSNSTREAIDATAKAAELGLDG : 101
tlr1219.DH . : ------MVGSTTPNAPFGQMLTAMITPMLDN-GEVDYDGVARLATYLVDEQRNDGLIVNGTCBASATTSDBERKERILRTVIDAVGD-RATIVAGAGSNDTRHSIELARTAERAGADG : 101
Tfu_0791.D . : ------------MFRGSIPPLPTPFRRG--RLDEEALRRLVERVVQG-GSHGVSVGGTTCBPGTQTLBERKRAIEVVLDQVAG-RVPVIPGTGALRLEETLELTRFAKEAGADG : 109
TTC0591.DH . : ------------MFRGSIPPLPTPFRRG--RLDEEALRRLVERVVQG-GSHGVSVGGTTCBPGTQTLBERKRAIEVVLDQVAG-RVPVIPGTGALRLEETLELTRFAKEAGADG :  98
TTHA0957.D . : -----------MILPPIPTPFDRE-GRLDEEAFRELAQALEPL--VDGLLVYGSNCBGVHLTPEER-----ARGLRALRP-RKPFLVGLMEETLPQAEGALLEAKAAGAMA :  98
TTC0385.DH . : -----------MILPPIPTPFDRE-GRLDEEAFRELAQALEPL--VDGLLVYGSNCBGVHLTPEER-----ARGLRALRP-RKPFLVGLMEETLPQAEGALLEAKAAGAMA :  91
TTHA0737.p . : ----------MEGVIPAIVTPFKEN-EDIDFEALNLFLIFKRN-GVKWIFALGSTCBFNMLSLSEKQEFIRKLREITKL--KIVINVSENSLKNTLSLGKLAFDVGMEG :  91
Saci 0162. . : ----------MKGVIPAIVTPFKEN-EELDLDSLETYIQFLKRN-GIEHFFVLGSTCBFNMLSFEEKTTFLKKLADITKN--GIVN-VTENSTYNALKLAKIVVDLGFEE :  96
ST2235.DHD . : ----------MKENISALVTPYDDK-ENINTEALQQLLDFLIKK-GIKDFWVLGTSCBFNMLSQDERILVVSKIREMVKG--KIYVGVNENSTRNSLILAKKYYDLGVDY :  95
SSO2274.DH . : ---------MDIIAPIITPFR--SGKVDVETFINHAKNIMKK-GVDALFVAGTCRKPFLVGLMEETLPQAEGALLEAKAAGAMA :  96
Pars 1388. . : MSHPLIPFRANFLRAPRVLSMEIVAPVITFR--GGRLDFANHVKNITSK-GVDVVFAGTCBGPALSLQEKMELTDAATSAARR---VIVQVASLNADEAIALAKYAESRGADA :  93
Ttenax.Kdg . : ----------MEIISPIITPFDK-QGKVNVDALKTHAKNLLEK-GIDAIFVNGTTCBGPALSKDEKRQNLNALYDVTHK---LIFQVGSLNLNDVMELVKFSNEMDILG : 113
Saci 0225. . : ----------MFKILSMDIVTPILTPFTK-EGKIDVEKLKAHAKFLIDN-GIDLLFVNGTTCBGPALSKEEKLTTLKTIYDVTNK----VIFQVGSLNINDVIDLVRASKDFDIVG :  94
ST2479.DHD . : ----------MPEIITPIITPFTK-DNRIDEGKFKDHCLKLLRD-GIDMLFVAGTNCBGPALSPEEKLENLKAVYDVTNK--IIFQVGGLNLDDAIRLAKLSKDFDIVG : 100
SSO3197.Kd . : ----------MPFKIVPILTPFR--GNEIDGEKFKDHELLIRK-GVDMKLFVAGTNCBMGPALSFEEKLMLAEYCSDITDR--VIFQVGSLNLEESIELARIARSMNFYA :  95
Ta0619.Kdg . : ----------MKLKIVPVVTPK--EGKIDGEKFIHHCMSLLSV-GVNYVFISGTTCBGPSLSINERMDLIKLSSNFADK---AIMQVGSMNLEDSIKLARMAKENGIYG :  94
TVN0669.DH . : ----------MPLKIVPVITPFK--NNVLDMDILREHCERLLNS-GIDKVFLAGTNCBGISLDFSEKLRIINELDIDGDK---LIIQISLNLEKSMELARARARAKGYFA :  94
PTO1026.Kd . : ----------MPVTRLYGIIPPLITPFDTKG-NLSVDRLSGIVEFTKPY--VHGYYLCGTYCBGVLLNLEETEKRAVIKTVAEASGN--KPIVVQVGSLDWDTVAGTVKLAEKYGASA :  94
Cmaq 0935. . : ----------MAKPKVIVPMVTPLRGG--SVDKNAASWLASRLSES-GVDGVFVPGTTCBGLPMRTDQRKKVAERVAELVGG-EKLIVHVGSADTETAVELAKHAQDIGAHA :  96
Tpen 1620. . : ----------MDGVFLTNVTPFTKT-LELDLEGLKRIVEFAQNS-GVDGVFVPGTTCBEWYLLDPHERAEAFFAVASSAIA-GLRLVAGVSGYRPEESYRLAEEAAGHGAHA : 102
APE 1023.1 . : ----------MFEAQLITALPTPFTAD-ERLDTSGLATLTTWIRDS-GVDGVFAAGTTCBFSSLTVEEKVRVDHVRRALSK--GEIIAGASTSFLETGELCKKYKEIGVDK :  99
SSO3035.DH . : ----------MFEAQLITALPTPFTAD-ERLDTSGLATLTTWIRDS-GVDGVFAAGTTCBFTALDDDERLEVLRVALEVFGP--DRTYAHVGAATSRQAERLARAAVAGGARK :  97
Tfu_1715.D . :                                                                                                                          :  99
```

```
                                                                                      K                          G
Ta1157.DHD  : RVLMPT-YYIKPDEDWMYRHFSTVISAA-SNDLFIYNIPQLSGSWISESLIEKLTREFSNVKGIKDSSGDMRFFS-RIIRHKNEK---FDIFQGQDDLLFLSLSIGASGGVCGLSNIS-P : 211
TVN1228.DH  : RVLMPT-YYIKPDEEWMYKHFSEVISAS-SNDLFIYNIPQLAGSWVPENVIKKLKSEFSQVKGIKDSSGDMRFFS-RIVTLVSKD---FSVFQGQDDLLLTSMALGANGGVCGLSNIS-G : 211
PTO1279.DH  : AVLMPP-YYIKLNQEAIYHYYKEILSSN-DMDLLIYNIPQFT-NKIDPETVKNLKSEFSSVKGVKDSSADIRGFM-EMLSLSDDD---FAVFQGQDDLLFTSLELGASGGVCGTTNFS-D : 201
PF0657.pDH  : TVIVSP-YYCKLKEDAIFKHFSIVAERA-DIPIILYAIPSCANP-INVEIVRKLALEHSNIIGIKASVDSLTYLG-ELI-EVKEERKDFKVFTGLDQYFLPNLILGGDGGIMACANFAPE : 212
PH0847.pDH  : VVIVSP-YYCRLKEDAIFKHFSMVAEKV-DIPIILYAIPSCANP-ISLEVVRKLALEYSNIVGVKASVDSLTYLS-GLI-EIKEERKDFRVFTGLDQYFLSTLLLGGDGGIMACANFAPE : 211
PAB0832.DH  : IVIVAP-YYCKLKDEELKHFSMVAERV-DIPIIVYAIPSCANP-VPVDIIRKVSLEYSNIIGVKASVDSLTYLQ-ELI-EVKEERKDFRVFTGLDQHFLNTLILGGDGGIMACANFAPE : 211
TK1237.DHD  : AVIAPP-YYCPLTEEALFRHYSMIAEKT-DIPIILYNIPACANS-LSVPLVKRLRITIEYPSIAGIKATLDSVNYIR-DIILDVKGERKDFRVFTGLDQHFLNTLILGGDGGIMACANFAPE : 213
SSO3072.DH  : IVVMAP-YFFKVREKELLDHFSMIAEKV-DLPIIVYNIPLFTGIDIPISVYKMLVSQHSNIIGTKVTLDSLIFFK-KLISEIREIRKDFSILTGFDEYLLPLLYMGGNGGVMGLANAIPK : 218
Tpen 1666.  : AVVTPP-YFFKVSPERLKVHFSTILEKV-DLPIIVYNIPATTGINIPVGLYLELAKEHSNLAGAKATVESFTYFR-QLVQVVKAERKDFPAVLTGLDDLLLPVIMMGGDGGIMALANAAPQ : 219
PAE2915.DH  : ILITPP-YYIQYDWAALYAFYKKVLDKV-DIPVILYTIPLATGYNVPVEVFEMVAGEYSQVVGVKDSSGDFRYHL-DLIHLLGKR---LSVLQGLDMLFVPSLVMGAHGGILAGPNFLGR : 213
Pars 0992.  : VLITPP-YYIQYDWAAIYAFYKKVLDKT-DVPVILYTIPLATGYNIPVEVFELVANEYSQVVGVKDSSGDFRYHL-DLIYLLGRR---LSVLQGLDMLFVPSLIMGAHGGILAGPNFLGK : 213
Pisl 0309.  : LLITPP-YYIQYDWAAIYAFYKKVLDKV-DMPTVLYTIPLATGYNIPVEVFELVANEYSQVVGVKDSSGDFRYHL-DLIHLLGKR---LSVLQGLDLLFVPSLLMGAQGGVLAGPNFLGR : 215
Tneu 1812.  : LLITPP-YYIQYDWAAVYAFYKRVLDKV-DIPIILYTIPLATGYNIPVEVFELVANEYSQVVGVKDSSGDFRYHL-DLIHLLGKR---LSVLQGLDLLFVPSLLMGAHGGVLAGPNFLGR : 213
Pcal 0350.  : VLITPP-YYIQYDWAAIYAFYKKVLDKV-DVPVILYTIPLAVGYNIPVEVFELVANEYSQVVAVKDSSGDFRYHL-DLIHLLGKR---LSIILQGLDLLFVPSIVMGAHGGILAGPNFLGK : 218
MJ0244.DHD  : VLSITP-YYNKPTQEGLRKHFGKVAESI-NLPIVLYNVPSRTAVNLEPKTVKLLAEEYSNISAVKEANPNLSQV---SELIHDA-KITVLSGNDELTLPIIALGGKGVISVVANIVPK : 210
MK1607.DHD  : ILSVVP-YYNKPPQEGLFIHFSKIAEAV-ECPIILYNVPSRTGCALEPETAAKLAEEYSHIVGYKEASGDLDVV---QRFIEETPDDFILLSGVDELILPIIAVGGVGVISVTANVAPE : 217
AF0910.DHD  : AMLVTP-YYNKPAEGLYQHYKAVASEV-SIPIIVYNVPSRTGINTTPELVRRIAEID-NIFGIKEASGNLKQI---SEIIRTTPDDFVLLSGDDFLTLPILCLGKGVISVAANVAPH : 209
MTH801.DHD  : ALVITP-YYNKPQPHGLIEHYTMLEEAA-DIPLIIYNVPSRTGTDIDVDTVAELAKLD-GIIGIKEASPDLDKVSMLRSRLMDLGLDDFTVLSGNDNLITLPMISMGAEGVISVAANVDPA : 217
aq 1143.DH  : ALVVVP-YYNKPTQRGLYEHFKTVAQEV-DIPIIIYNIPSRTCVEISVDTMFKLASECENIVASKESTPNMDRIS-EIVKRLGES---FSVLSGDDSLTLPMMALGAKGVISVANNVMPR : 212
SYO3AOP1_0  : SLQVVP-YYNKPTQEGIYQHFKAIAEET-NIPLILYNIPSRTGVDMLPETFARLYGDFPNVIGIKEATGNVARVS-ETISLTNED---VLILSGDDALTLPMMAVGAKGVIANNIIPK : 212
TTE0832.DH  : LLIVTP-YYNKTTQKGLVAHFTEIARHV-DIPIIIYNVPSRTSLNMLPETYLEVKKAENVVGVKEASGDISQIA-EIARIMGKS---FEIYSGNDDQVIPIMSLGGLGVISVTANIIPA : 214
TM1521.DHD  : VLVVTP-YYNKPTQEGLYQHYKYISERT-DLGIVVYNVPGRTGVNVLPETAARIAADLKNVGIKEANPDIDQID-RTVSLTKQARSDFMVWSGNDDRTFYLLCAGGDGVISVSNVAPK : 215
TRQ2 1415.  : VLVVTP-YYNKPTQEGLYQHYKYISERT-DLGIVVYNVPGRTGVNVLPETAARIAADLKNVGIKEANPDIDQID-RMVSLTKQARSDFMVWSGNDDRTFYLLCAGGDGVISVSNVAPK : 215
Tpet_1271.  : VLVVTP-YYNKPTQIGLYQHYKYISERT-DLGIVVYNVPGRTGVNVLPETAARIAADLKNVGIKEANPDIDQID-RTVSLTKQARSDFMVWSGNDDRTFYLLCAGGDGVISVSNVAPK : 215
Tmel_1875.  : LLVVTP-YYNKPTQIGLYHYYISEHT-DLEIIIYNVPGRTGVNIAPETVRLLAASDCKNITALKEANSSFDQIN-KVLHLKPET---FKVFSGNDDISFQFLASGGNGVISVASNVIPN : 209
Moth 1068.  : LMLVVP-YYNRPSQEGLYQHFKAIAAAT-TLPIILYNIPSRTGRNMDAATTLRIA-EIKNIKAVKEASGDLDQAT---AILRQAPADFIVYSGDDSLITLPMAVGGYGIISVVAHVAGE : 213
STH1549.DH  : LLLVTP-YYNKPPQEGLYRHFRAVABAV-DLPCILYNVPSRTSVNMLPATTLRLARDVPNIVGIKECA-DVGQLA---DILSGAPEGFRVWSGDDANLLPYLTVGAYGIISVASHVVGP : 214
tlr1219.DH  : ALLVVP-YYNKPPQEGIYAHFQAIAKAVPDFPLMLYNIPGRTGQNLLPETVIRLA-EYPNIVAFKEASGSLDQAS---TLRAALPPTFRIYAGDDSLTLPLLAVGGYGVVSVASHLVGL : 215
Tfu_0791.D  : LLLVTP-YYNRPPQEGLRHFTAIADAT-GLPIMLYDIPGRTGTPIDSETLVRLA-EHPRIVANKDAKDDLGASS---WVMSRT--DLAYYSGDMLNLPLLSIGAAGFVVGHVVGS : 220
TTC0591.DH  : AMVIVP-YYVKPNQEGLYRYFAEVARAVPDFPLLIYNIPGRAGVEIAPKTVGRLRRDFPNIVGLKHSSKDLEYLS---HLFLEAGRDFLVFCGLESLTLPMMSLGAVGTIAATANWLPK : 213
TTHA0957.D  : AMVIVP-YYVKPNQEGLYRYFAEVARTVPDFPLLIYNIPGRAGVEIAPKTVGRLRRDFPNIVGLKHSSKDLEYLS---HLFLEAGRDFLVFCGLESLTLPMMSLGAVGTIAATANWLPK : 213
TTC0385.DH  : LLATPPRYHHGSLGAGLLRYYEALAEKM---PLFLYHVPQNTKVDLPLEAVEALA-PHPNVLGIKDSSGDLSRIA---FYQARLREFRVYTGHAPTFLGALALGAEGGILAAANLAPR : 202
TTHA0737.p  : LLATPPRYHHGSLGAGLLRYYEALAEKM---PLFLYHVPQNTKVDLPLEAVEALA-PHPNVLGIKDSSGDLSRIA---FYQARLQEFRVYTGHAPTFLGALALGAEGGILAAANLAPR : 202
Saci 0162.  : IALIPP-IYHKPSESGLISYF-EEMS-KLDLPLYYAYP--DKVPITYETISKLVKQD-IIDGIKITTDNLPLFK-KYLG-LKEIKNDLKILIGSDTLFVYSLIEGGNGCISAVANVAPE : 208
ST2235.DHD  : LASLPP-LYHKPSEKGIVKYF-EDLS-RLGIPLNMVYYPQ-NSYQIAENIIEKLVEEG-IIQGMKYTTNDLIVSFK-NLIS-LKEINKEFSLLIGNDELFLESYLAGGDLVSAVANIAPE : 208
SSO2274.DH  : IFSVPP-IYHKPSEKGLISYF-SELRNAIDLPLFIYNIPSFTGYNVPLHIIEKLAREY-VLDGMKYSTTDFVSFL-KYLKALKGVNKNFKVFIGEDRMLSALILDTDGAVSGISNLVPE : 212
Pars 1388.  : VASVPPYYFPGLSQROVAKYFRELCA-ATSLPVYLYNYPAAVGRNVDAEMAKAIG----CLKGVKDTNPDLAHTL---AYKRLMPEIKAYKGSDSLVMASFAVGLDGVIVASANYIPD : 203
Ttenax.Kdg  : VASLPPYYFPRLSERQIAKYFRDLCS-AVSIPVYLYNYPAAVGRDVDARAAAKEIG----CIRGVKDTNESLAHTL---AYKRYLPQARVYNGSDSLVPASFAVRLDGVVASANVAPE : 223
Saci 0225.  : VSSHSPYYFPRLPEKFLAKYYEEIAR-ISSHSLYIYNYPAATGYDIPPSILKSLP----VKGIKDTNQDLAHSL---EYKLNLPGVKVYNGSNTLIYYSLLS-LDGVVASFTNFIPE : 202
ST2479.DHD  : IASYAPYYYPRMSEKHLVKYFKTLCE-VSPHPVYLYNYPATGKDLSHSL---LITGLKDTNQDLSHSL---EYKILMPNLKVYNGSDSLVFYSLTS-LDGSVTAASNYLPH : 210
SSO3197.Kd  : IASVAPYYYPRMSEKHLVKYFKTLCE-VSPHPVYLYNYPATGKDIDAKVAKEIG----CFTGVKDTIENIIHTL---DYKRLNPNMLVYSGSDMLIATVASTGLDGNVAAGSNYLPE : 205
Ta0619.Kdg  : VAALPPFYYGGIPEEWLIRYYTQLSE-VMET--FIYNYPKFTGYDINADLARKIKKKGGNILGVKDTVSDIDHMM---RYKYLIDDFMVFSGPSNYLHAAVNGLDGAVAAAGNYATE : 206
TVN0669.DH  : ISALPPYYFGGIPEEWLIKYYVEISK-IIPT--FIYNYPKFTGYDINSSLAIKINKAGGNIIGVKDTIADISHML---SYRDLGNDFIILSGPSNSIYBALRSGVDGSVAAAGNFMPK : 206
PTO1026.Kd  : IASVPPYYYSDLDENAIIRYFLDLSR-HYKT--IVYNYPKMTGHDVNARMLKRIIDSGGDIIGVKDTVNDIMHML---EYKNIEDDLIVLAGPSTHIFSALRSGVDGTVSAAGNYMPN : 206
Cmaq 0935.  : VASIVP-IYYRPDYETIRRYFMKLNQLT-KLPTFIYNIPGNAGFNVTPDIVARLIKDGVRITGVIKEVG--VAGIKDSSGDLVQALM---QFIEMGLEVFNGSDHMIAPSVIVGAKGCISALSNSITN : 207
Tpen 1620.  : VASIVP-IYYRPDYETIRRYFMKLNQLT-KLPTFIYNIPGNAGFNVTPDIVARLIKDGVRITGVIKEVG--VAGIKDSSFDVQAFI---DYKVTLGEDFDVVVGTEALMLPTYVLGARAFIPGMSNYAPE : 214
APE 1023.1  : VASVPP-FYYHHSETHILNFFRELVDSV-SIPVYANNPPRVGYPITPELALKLKEVG--VAGIKATVTDAAYIA-GLAAVKERNPEFEVLAGGLEMLVHTSLSGGDVVDAYSNVAPK : 215
SSO3035.DH  : IMVLPP-YYLKGNEKGLMEYFKKVAELC-DSTMVLYNNPSSTGVDLSPQLIKRLNDMIPSFRVLKEARYNVVEFK---EVRLVGDKIDVIDGLEERALLGLVLGAKGFTTSMGSFSSL : 211
Tfu_1715.D  : LAAITP-FYLPAPEDAVLRYFERIVAAAGDAHVFAYLFQARSTTPATPPLLRRLADVG---VRGVKISGESDATVA---AFLAEAPHGFTFFSGNDLSYAELVRSGGTGVVSGVSSVFPK : 212
```

```
Ta1157.DHD : YITNLYHEFSAGNLEKARKIQIDEVNPLMYAIN----------EATFPAGYYAFYKMNGIKGG-YRAPMVEPTTDQKKKIDQELTKIPKKQ----------------------------------- : 292
TVN1228.DH : SIVDLYKAFQSGDLKKSREIQLNKVNPLMYAVN----------APTFPSGYYYAFYKKHNINGG-YRSPMVQPSKEHSRVIDEATSQQI------------------------------------- : 289
PTO1279.DH : GIVRLYHEYKN-NREMALKIEKNDVIPLMKKLG----------KYQFPNAYYEYFYKKNNINGG-YRPPMYRVGIEI--------------------------------------------------- : 266
PF0657.pDH : IHLEIWNAFKEKNFRRAIELAR-ELVRITKIYK----------IASSFASAVKLAMVAKGFPIRPVLRPPYVIDGEEVFNEIREIVEWRQMLRKESL---------------------------- : 298
PH0847.pDH : IHLEVWNRAFKKGNLEKAMNSAR-KLVKITKIYS---------IASSFASAVKLAMVARGFPIKPILRPPYMMDGEEVFKKIKEIVS---------------------------------------- : 287
PAB0832.DH : IHLQIWNSFKRRNFEEAIKLSR-VLGEISRIYN----------VASSFASAVKLAMIAKGFPIKPVLRPPHVIDGEEVFNEIKGILRSLENVKP---------------------------------- : 294
TK1237.DHD : LHLRLYKAFNEKRFEEAFEYSR-KLAKLSKVYD----------IASSFGSAIKLAMRVRGFSIKPVLRPPYTMDGKEVEBKVRALLLEVL------------------------------------- : 292
Tpen_1666. : LHLKVYESWKNGNLSDANKYWR-NVLNLTDIYD----------YCNSYTASIKLLLKVLNMPIKNVVRPPLS--ICEEESKIRGMIKDLSSFLTNSK------------------------------ : 302
PAE2915.DH : IHREVDAYRSGDLKRALEAWH-KLLRLVRVYD----------YATSFPTSVKTLLKVMGAPVKPYARTPLTPETREVEEKIAQIARELGLKI-------------------------------------- : 301
Pars_0992. : ITLEQYLVKEGKIAEAVALHN-KLMPLWRFMGGCGLVGKLGGKWPTLYKLATQLVHGIDMGPPREPLPPVDDKDRRELEKILKELGLI-------------------------------------- : 301
Pisl_0309. : ITLEQYRLVKEGKIAEAVALHN-KLMPLWRFMGGCGLVGKLGGKWPTLYKVATQIVRGIDMGPPREPLPPIDDRDRKELEKLLKDLGLI---------------------------------------- : 303
Tneu_1812. : ITLEQYRLVKEGKIAEAVALHN-KLMPLWRFMGGCGLVGKLGGKWPTLYKVATQMVRGIDMGPPREPLPPIDDRDRKELERLLKELGLI---------------------------------------- : 301
Pcal_0350. : LTLRQYQLVKEGKVAEAVELHN-RLMPLWRFMGGCGLVGKLGGKWPTLYKVAEMVHGIDMGPPREPLPPIDDKDRELERLLRELGIVKP---------------------------------------- : 308
MJ0244.DHD : EFVEMVNYALEGDFEKAREIHY-KLFPLMKAMFIE----------TNPIPVKTALNMMGRPAG-ELRLPLCEMSEEHKKILENVLKDLGLI---------------------------------- : 289
MK1607.DHD : LMVEMYEAWKSGDVERARELHY-ELLPLHRALFTE----------TNPIPVKAAVELVGMASS-PPRPPLKEAREDTKELLRRELKKLGLLPEGG--------------------------------- : 300
AF0910.DHD : LMKEMYEAFVEGNIERAREMHH-RLTPLFDVLFID----------TNPIPVKKALQLMGLAAG-KPRLPLVELSEEKTQKVKEVLKSLELIS------------------------------------- : 289
MTH801.DHD : RMSRLVNEALSGDFESAMKTHY-ELYSLMKVLFIE---------SNPVPVKAALNMMGKARG-S----CEDAPGTPAGCKP---------------------------------------------- : 283
aq_1143.DH : EVKELIRAALEGDFRRAREIHY-YLHDLFKVLFIE----------TNPIPVKTACWMLGMCEK-EFRLPLTEMSPENENKLREVLKKYNLPLKN----------------------------------- : 294
SYO3AOP1_0 : EISQMCKLALEGKFEEAKKIHN-QYWDLFKVLFIE----------TNPIPIKTAAYLMGLIESLERLPLYTMSKSNEEKLKEVLRKHSLIK--------------------------------------- : 293
TTE0832.DH : KIHEMTTAYLNGDIEKARDMQL-ELNPLNKALFIE----------TNPIPVKTAMNLMGFGVG-PLRLPIVEMSEKNLEYLKSVLRQYGLLKEEN------------------------------------ : 297
TM1521.DHD : QMVELCAEYFSGNLEKSREVHR-KLRPLNKALFVE----------TNPIPVKAALNLMGFIEN-ELRLPIVPASEKTVELLRNVLKESGLL--------------------------------------- : 294
TRQ2_1415. : QMVELCAEYFSGNLEKSREVHR-KLRPLMKALFVE----------TNPIPVKAALNLMGFIEN-ELRLPIVPASEKTVELLRNVLKESGLL--------------------------------------- : 294
Tpet_1271. : QMVELCAEYFSGNLEKSREVHR-KLRPLMKALFAE----------TNPIPVKAALNLMGFIEN-ELRLPLVPASEKTMELLRNVLKESGLL--------------------------------------- : 294
Tmel_1875. : QMVEMYKNIISGNISNARKIFY-TYYPLFKALFIE----------TNPIPVKQALNIMGLIEN-ELRLPIYPANKETKNILEKILKECKII---------------------------------------- : 288
Moth_1068. : KMQAMVRAFTAGDVQGAAALHR-ELFPLFKALFIT----------SNPVPVKEALQMLGLPAG-PVRLPLVGATREEKEKIAAALKETGLL---------------------------------------- : 292
STH1549.DH : QMRELIDAFLAGDTARAAAWHR-RLLPVFRGLFAV----------TNPILVKAALRLTGFPVG-PVRLPLVDATEQEEALREVLAEAGVL----------------------------------------- : 293
tlr1219.DH : RIQEMIQAFVQGDTAKATAIHC-QLLPLFKALFVT----------TNPIPIKAALSLQGWSVG-EPRLPLTSASDAVISQLKAVLDDLGLLKS-------------------------------------- : 296
Tfu_0791.D : ELHDMIDAYRAGDVARALDIHR-RLIPVYRGMFRT----------QGVITTKAVLAMFGLPAG-VVRAPLLDASPELKELLREDLAMAGVKGPTGLASAHEDAASGREAAERLTEGTA : 325
TTC0591.DH : EVALLCEKALAGDYQGARELHF-YLLEANEAIFWD----------TNPIPLKTVLSWMGLLEK-EWRPPLGPTTPEVEERLRRMAERYGLLPKEKEAA------------------------------- : 299
TTHA0957.D : EVALLCEKALAGDYQGARELHF-HLLEANEAIFWD----------TNPIPLKTVLSWMGLLEK-EWRPPLGPTTPEVEERLRRMAERYGLLPKEKEAA------------------------------- : 299
TTC0385.DH : AYRALLDHFREGRLAEAQELQK-KLFPLG-DLLAK----------GGVPLLKQALRHLGLPAG--YPRPPYPAESPLWERFLPVLEGLKEEGWVL-------------------------------------- : 283
TTHA0737.p : AYRALLDHFREGRLAEAQELQK-KLFPLG-DLLAK----------GGVPLLKQALRHLGLPAG--YPRPPYPAESPLWERFLPVLEGLKEEGWVL-------------------------------------- : 283
Saci_0162. : LMMRAYQGVREGKLNESLEIQG-MISKISDAIMSG----------DFPSGVKVALRYRGVSVG-SVRRPLKESIEVN-ARIYSVLKELGM---------------------------------------- : 285
ST2235.DHD : LVVKEYELLKGNLTEALKIQK-MIEKINRLLLIG----------DYPSTLKILKYRGIYVG-KVRSPLQESVDGN-SLIYGILKELGL---------------------------------------- : 285
SSO2274.DH : LVTNLFLEFDRGNIQRAIEIQR-ILNKLVDDVVSLG----------DYPSGIKIGLRYRGINVG-SVRKPLLEDSRAE-GEIYNVLKELGI------------------------------------------ : 289
Pars_1388. : VLVKIRDAVLKGDVEKAKRLQFLLINEILEVATPLG----------YNSSVYELVRIFQGYDPG-EPRPPIYPLESEERRKLAKAVSRIKETLNKL------------------------------------ : 287
Ttenax.Kdg : LLAGIRDAVAAGDIERARSLQFLLDEIVSARHIG----------YAAAVYELVEIFQGYEAG-EPGPVYPLDPEEKAWIRAAVAKAKSQLRI--------------------------------------- : 306
Saci_0225. : VIVKQRDLIKQGKLDDALRLQELINRLADLIKYG----------SISAIYLVNEFQGYDVG-YPRPPIFPLTDEEALSLKREIEPLKRKIQELVH------------------------------------ : 288
ST2479.DHD : VMKKMKEHITSGQVSKAIELQKLIINKALDISRKYG----------QLSAIYYLVKEFLGYDVG-YPRGPIFPLEEDEVKALLSEIQPVKKEIERAVS------------------------------- : 296
SSO3197.Kd : VTVTIKKLAMERKIDEALKLQFLHDEVIEASRIFG----------SLSSNYVLTKYFQGYDLGYPRPPIFPLDDEEERQLIKKVEGIRAKLVELKILKE------------------------------ : 294
Ta0619.Kdg : MFVQILDDTAR---YGRAN-QIIIDEIFDAASRYG----------QWASNYSLVRILRGYDAG-DPRPPFYPLDDHTEEKLRSEVRSIMQKYGRESKL------------------------------- : 289
TVN0669.DH : TFIELADKFEA---EASFKLQKYINSVLSEVMEYG----------SWSAYYSLIKILRRYSAG-DSRAPFYPLNALDEEKLRHDVFFVLPEEQRHQ---------------------------------- : 288
PTO1026.Kd : IFRYLSDHVYS---NNAFKLQILINSVIDASAHFG----------IYSANYSLVKIIKNYYCG-EPRPPFYPLTESDENNLKKIIKNILKDGDNV------------------------------------- : 287
Cmaq_0935. : LVIDTYRRAKDGDVNNALRTQALVTKVRDTANKYP----------RPAVHYSLVKLLKYDFGGVKEPLVRGLTREEEESLASDLSKLGFRVRVSGV-------------------------------- : 293
Tpen_1620. : IVFKLFKALENRDFENAAKIQYKINKVRRQVQRLG----------PTIPLVYLALRGVDAG-FPRKPFLPAPHEVQEILKSYIEELLEGVET---------------------------------------- : 297
APE_1023.1 : LAVELAESLEKGSVERVAGLQG-RVLKLSRILGKR----------SLPAVLKAVLHVLGAPITGEVRKPLKPLTQPEANATANALCTSYRDYLLPGLEC-------------------------------- : 303
SSO3035.DH : PLMIYESYEKKRNIDTAIRLYD-LLLEYRSFIKLP----------LTMSHLAKLGAWLRGLLISSYATRPPLPNLDDVVSEDVRKKMIEKINQIISFEKNTIE--------------------------- : 302
Tfu_1715.D : PFIALRDALRTGDAARAASAQPLVERAVAVKAGN----------IAHLKAALALQGLPAG-PPRVSCESVSDADLAELRATIAVAA------------------------------------------- : 288
```

References

Aharoni, A, Griffiths, AD and Tawfik, DS. (2005). High-throughput screens and selections of enzyme-encoding genes. *Curr Opin Chem Biol* **9**:210-216.

Ahmed, H, Ettema, TJ, Tjaden, B, Geerling, AC, van der Oost, J and Siebers, B. (2005). The semi-phosphorylative Entner-Doudoroff pathway in hyperthermophilic archaea: a re-evaluation. *Biochem J* **390**:529-540.

Allard, J, Grochulski, P and Sygusch, J. (2001). Covalent intermediate trapped in 2-keto-3-deoxy-6-phosphogluconate (KDPG) aldolase structure at 1.95-A resolution. *Proc Natl Acad Sci U S A* **98**:3679-3684.

Antikainen, NM and Martin, SF. (2005). Altering protein specificity: techniques and applications. *Bioorg Med Chem* **13**:2701-2716.

Aravalli, RN and Garrett, RA. (1997). Shuttle vectors for hyperthermophilic archaea. *Extremophiles* **1**:183-191.

Azéma, L, Bringaud, F, Blonski, C and Perie, J. (2000). Chemical and enzymatic synthesis of fructose analogues as probes for import studies by the hexose transporter in parasites. *Bioorg Med Chem* **8**:717-722.

Baker, K, Bleczinski, C, Lin, H, Salazar-Jimenez, G, Sengupta, D, Krane, S and Cornish, VW. (2002). Chemical complementation: a reaction-independent genetic assay for enzyme catalysis. *Proc Natl Acad Sci U S A* **99**:16537-16542.

Bao, Q, Tian, Y, Li, W and other authors. (2002). A complete sequence of the *T. tengcongensis* genome. *Genome Res* **12**:689-700.

Barbosa, JA, Smith, BJ, DeGori, R and other authors. (2000). Active site modulation in the N-acetylneuraminate lyase sub-family as revealed by the structure of the inhibitor-complexed *Haemophilus influenzae* enzyme. *J Mol Biol* **303**:405-421.

Bell, BJ, Watanabe, L, Rios-Steiner, JL, Tulinsky, A, Lebioda, L and Arni, RK. (2003). Structure of 2-keto-3-deoxy-6-phosphogluconate (KDPG) aldolase from *Pseudomonas putida*. *Acta Cryst D* **59**:1454-1458.

Bell, SD and Jackson, SP. (1998). Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features. *Trends Microbiol* **6**:222-228.

Bergquist, PL, Reeves, RA and Gibbs, MD. (2005). Degenerate oligonucleotide gene shuffling (DOGS) and random drift mutagenesis (RNDM): two complementary techniques for enzyme evolution. *Biomol Eng* **22**:63-72.

Blagova, E, Levdikov, V, Milioti, N, Fogg, MJ, Kalliomaa, AK, Brannigan, JA, Wilson, KS and Wilkinson, AJ. (2006). Crystal structure of dihydrodipicolinate synthase (BA3935) from *Bacillus anthracis* at 1.94 A resolution. *Proteins* **62**:297-301.

Blickling, S, Beisel, HG, Bozic, D, Knablein, J, Laber, B and Huber, R. (1997a). Structure of dihydrodipicolinate synthase of *Nicotiana sylvestris* reveals novel quaternary structure. *J Mol Biol* **274**:608-621.

Blickling, S, Renner, C, Laber, B, Pohlenz, HD, Holak, TA and Huber, R. (1997b). Reaction mechanism of *Escherichia coli* dihydrodipicolinate synthase investigated by X-ray crystallography and NMR spectroscopy. *Biochemistry (Mosc)* **36**:24-33.

Boersma, YL, Droge, MJ and Quax, WJ. (2007). Selection strategies for improved biocatalysts. *FEBS J* **274**:2181-2195.

References

Bolt, A, Berry, A and Nelson, A. (2008). Directed evolution of aldolases for exploitation in synthetic organic chemistry. *Arch Biochem Biophys.*

Brase, S, Gil, C and Knepper, K. (2002). The recent impact of solid-phase synthesis on medicinally relevant benzoannelated nitrogen heterocycles. *Bioorg Med Chem* **10**:2415-2437.

Bräse, S, Gil, C and Knepper, K. (2002). The recent impact of solid-phase synthesis on medicinally relevant benzoannelated nitrogen heterocycles. *Bioorg Med Chem* **10**:2415-2437.

Brinkkotter, A, Shakeri-Garakani, A and Lengeler, JW. (2002). Two class II D-tagatose-bisphosphate aldolases from enteric bacteria. *Arch Microbiol* **177**:410-419.

Brock, TD, Brock, KM, Belly, RT and Weiss, RL. (1972). *Sulfolobus*: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. *Arch Mikrobiol* **84**:54-68.

Brouns, SJ, Wu, H, Akerboom, J, Turnbull, AP, de Vos, WM and van der Oost, J. (2005). Engineering a selectable marker for hyperthermophiles. *J Biol Chem* **280**:11422-11431.

Buchanan, CL, Connaris, H, Danson, MJ, Reeve, CD and Hough, DW. (1999). An extremely thermostable aldolase from *Sulfolobus solfataricus* with specificity for non-phosphorylated substrates. *Biochem J* **343**:563-570.

Budgen, M and Danson, MJ. (1986). Metabolism of sugar via a modified Entner-doudoroff pathway in the thermoacidophilic archeabacterium *Thermoplasma acidophilum. FEBS Lett* **196**:207-210.

Cabrita, LD, Dai, W and Bottomley, SP. (2006). A family of E. coli expression vectors for laboratory scale and high throughput soluble protein production. *BMC Biotechnol* **6**:12.

Cahyanto, MN, Kawasaki, H, Nagashio, M, Fujiyama, K and Seki, T. (2006). Regulation of aspartokinase, aspartate semialdehyde dehydrogenase, dihydrodipicolinate synthase and dihydrodipicolinate reductase in *Lactobacillus plantarum. Microbiology* **152**:105-112.

Cannio, R, Contursi, P, Rossi, M and Bartolucci, S. (1998). An autonomously replicating transforming vector for *Sulfolobus solfataricus. J Bacteriol* **180**:3237-3240.

Castro, A, Castano, T, Encinas, A, Porcal, W and Gil, C. (2006). Advances in the synthesis and recent therapeutic applications of 1,2,4-thiadiazole heterocycles. *Bioorg Med Chem* **14**:1644-1652.

Chen, L, Brugger, K, Skovgaard, M and other authors. (2005). The Genome of *Sulfolobus acidocaldarius*, a Model Organism of the Crenarchaeota. *J Bacteriol* **187**:4992-4999.

Cheriyan, M, Toone, EJ and Fierke, CA. (2007). Mutagenesis of the phosphate-binding pocket of KDPG aldolase enhances selectivity for hydrophobic substrates. *Protein Sci* **16**:2368-2377.

Choi, KH, Lai, V, Foster, CE, Morris, AJ, Tolan, DR and Allen, KN. (2006). New superfamily members identified for Schiff-base enzymes based on verification of catalytically essential residues. *Biochemistry (Mosc)* **45**:8546-8555.

Coco, WM, Levinson, WE, Crist, MJ, Hektor, HJ, Darzins, A, Pienkos, PT, Squires, CH and Monticello, DJ. (2001). DNA shuffling method for generating highly recombined genes and evolved enzymes. *Nat Biotechnol* **19**:354-359.

Collaborative Computational Project, N. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**:760-763.

Cotterill, IC, Shelton, MC, Machemer, DEW, Henderson, DP and Toone, EJ. (1998). Effect of phosphorylation on the reaction rate of unnatural electrophiles with 2-keto-3-deoxy-6-phosphogluconate aldolase. *J. Chem. Soc., Perkin Trans.* **1**:1335-1341.

Coulter, CV, Gerrard, JA, Kraunsoe, JAE, Moore, DJ and Pratt, AJ. (1996). (*S*)-Aspartate Semi-Aldehyde: Synthetic and Structural Studies. *Tetrahedron* **52**:7127-7136.

Cowan, DA, Arslanoglu, A, Burton, SG, Baker, GC, Cameron, RA, Smith, JJ and Meyer, Q. (2004). Metagenomics, gene discovery and the ideal biocatalyst. *Biochem Soc Trans* **32**:298-302.

Cremer, J, Treptow, C, Eggeling, L and Sahm, H. (1988). Regulation of enzymes of lysine biosynthesis in *Corynebacterium glutamicum*. *J Gen Microbiol* **134**:3221-3229.

Davis, IW, Leaver-Fay, A, Chen, VB and other authors. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**:375-383.

Dean, SM, Greenberg, WA and Wong, CH. (2007). Recent advances in aldolase-catalyzed asymmetric synthesis. *Advanced Synthesis & Catalysis* **349**:1308-1320.

DeLano, WL (2002).The PyMOL molecular graphics system. Palo Alto, CA, USA: DeLano Scientific.

DeLong, EF and Pace, NR. (2001). Environmental diversity of bacteria and archaea. *Syst Biol* **50**:470-478.

Dereppe, C, Bold, G, Ghisalba, O, Ebert, E and Schar, HP. (1992). Purification and Characterization of Dihydrodipicolinate Synthase from Pea. *Plant Physiol* **98**:813-821.

DeSantis, G, Liu, J, Clark, DP, Heine, A, Wilson, IA and Wong, C-H. (2003a). Structure-based mutagenesis approaches toward expanding the substrate specificity of D-2-deoxyribose-5-phosphate aldolase. *Bioorg Med Chem* **11**:43-52.

DeSantis, G, Liu, J, Clark, DP, Heine, A, Wilson, IA and Wong, CH. (2003b). Structure-based mutagenesis approaches toward expanding the substrate specificity of D-2-deoxyribose-5-phosphate aldolase. *Bioorg Med Chem* **11**:43-52.

Dionne, I, Robinson, NP, McGeoch, AT, Marsh, VL, Reddish, A and Bell, SD. (2003). DNA replication in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Biochem Soc Trans* **31**:674-676.

Dobson, RC, Gerrard, JA and Pearce, FG. (2004a). Dihydrodipicolinate synthase is not inhibited by its substrate, (S)-aspartate beta-semialdehyde. *Biochem J* **377**:757-762.

Dobson, RC, Griffin, MD, Jameson, GB and Gerrard, JA. (2005). The crystal structures of native and (S)-lysine-bound dihydrodipicolinate synthase from *Escherichia coli* with improved resolution show new features of biological significance. *Acta Crystallogr D Biol Crystallogr* **61**:1116-1124.

Dobson, RC, Griffin, MD, Roberts, SJ and Gerrard, JA. (2004b). Dihydrodipicolinate synthase (DHDPS) from *Escherichia coli* displays partial mixed inhibition with respect to its first substrate, pyruvate. *Biochimie* **86**:311-315.

Dobson, RC, Valegard, K and Gerrard, JA. (2004c). The crystal structure of three site-directed mutants of *Escherichia coli* dihydrodipicolinate synthase: further evidence for a catalytic triad. *J Mol Biol* **338**:329-339.

Drew, KN, Zajicek, J, Bondo, G, Bose, B and Serianni, AS. (1998). ^{13}C-labeled aldopentoses: detection and quantitation of cyclic and acyclic forms by heteronuclear 1D and 2D NMR spectroscopy. *Carbohydr Res* **307**:199-209.

Dwyer, MA, Looger, LL and Hellinga, HW. (2003). Computational design of a Zn2+ receptor that controls bacterial gene expression. *Proc Natl Acad Sci U S A* **100**:11255-11260.

Fishman, A, Tao, Y, Bentley, WE and Wood, TK. (2004). Protein engineering of toluene 4-monooxygenase of Pseudomonas mendocina KR1 for synthesizing 4-nitrocatechol from nitrobenzene. *Biotechnol Bioeng* **87**:779-790.

Fong, S, Machajewski, TD, Mak, CC and Wong, C. (2000). Directed evolution of D-2-keto-3-deoxy-6-phosphogluconate aldolase to new variants for the efficient synthesis of D- and L-sugars. *Chem Biol* **7**:873-883.

Frisch, DA, Gengenbach, BG, Tommey, AM, Sellner, JM, Somers, DA and Myers, DE. (1991). Isolation and Characterization of Dihydrodipicolinate Synthase from Maize. *Plant Physiol* **96**:444-452.

References

Fullerton, SW, Griffiths, JS, Merkel, AB, Cheriyan, M, Wymer, NJ, Hutchins, MJ, Fierke, CA, Toone, EJ and Naismith, JH. (2006). Mechanism of the Class I KDPG aldolase. *Bioorg Med Chem* **14**:3002-3010.

Gerard, J, Lloyd, R, Barsby, T, Haden, P, Kelly, MT and Andersen, RJ. (1997). Massetolides A-H, antimycobacterial cyclic depsipeptides produced by two pseudomonads isolated from marine habitats. *J Nat Prod* **60**:223-229.

Ghislain, M, Frankard, V and Jacobs, M. (1995). A dinucleotide mutation in dihydrodipicolinate synthase of *Nicotiana sylvestris* leads to lysine overproduction. *Plant J* **8**:733-743.

Gibbs, MD, Nevalainen, KM and Bergquist, PL. (2001). Degenerate oligonucleotide gene shuffling (DOGS): a method for enhancing the frequency of recombination with family shuffling. *Gene* **271**:13-20.

Gijsen, HJ, Qiao, L, Fitz, W and Wong, CH. (1996). Recent Advances in the Chemoenzymatic Synthesis of Carbohydrates and Carbohydrate Mimetics. *Chem Rev* **96**:443-474.

Gratz, A and Jose, J. (2008). Protein domain library generation by overlap extension (PDLGO): a tool for enzyme engineering. *Anal Biochem* **378**:171-176.

Grazi, E, Rowley, PT, Cheng, T, Tchola, O and Horecker, BL. (1962). The mechanism of action of aldolases. III. Schiff base formation with lysine. *Biochem Biophys Res Commun* **9**:38-43.

Grogan, DW. (1989). Phenotypic characterization of the archaebacterial genus *Sulfolobus*: comparison of five wild-type strains. *J Bacteriol* **171**:6710-6719.

Hall, DR, Bond, CS, Leonard, GA, Watt, CI, Berry, A and Hunter, WN. (2002). Structure of tagatose-1,6-bisphosphate aldolase. Insight into chiral discrimination, mechanism, and specificity of class II aldolases. *J Biol Chem* **277**:22018-22024.

Hao, J and Berry, A. (2004). A thermostable variant of fructose bisphosphate aldolase constructed by directed evolution also shows increased stability in organic solvents. *Protein Eng Des Sel* **17**:689-697.

Heuts, DP, van Hellemond, EW, Janssen, DB and Fraaije, MW. (2007). Discovery, characterization, and kinetic analysis of an alditol oxidase from *Streptomyces coelicolor*. *J Biol Chem* **282**:20283-20291.

Hidalgo, A, Schliessmann, A, Molina, R, Hermoso, J and Bornscheuer, UT. (2008). A one-pot, simple methodology for cassette randomisation and recombination for focused directed evolution. *Protein Eng Des Sel*.

Hiraga, K and Arnold, FH. (2003). General method for sequence-independent site-directed chimeragenesis. *J Mol Biol* **330**:287-296.

Ho, SN, Hunt, HD, Horton, RM, Pullen, JK and Pease, LR. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**:51-59.

Hoganson, DA and Stahly, DP. (1975). Regulation of dihydrodipicolinate synthase during growth and sporulation of *Bacillus cereus*. *J Bacteriol* **124**:1344-1350.

Horecker, BL and Kornberg, A. (1948). The extinction coefficients of the reduced band of pyridin nucleotides. *J Biol Chem* **175**:385-390.

Hsu, CC, Hong, Z, Wada, M, Franke, D and Wong, CH. (2005). Directed evolution of D-sialic acid aldolase to L-3-deoxy-manno-2-octulosonic acid (L-KDO) aldolase. *Proc Natl Acad Sci U S A* **102**:9122-9126.

Iwabuchi, T and Harayama, S. (1998). Biochemical and genetic characterization of trans-2'-carboxybenzalpyruvate hydratase-aldolase from a phenanthrene-degrading Nocardioides strain. *J Bacteriol* **180**:945-949.

Izard, T, Lawrence, MC, Malby, RL, Lilley, GG and Colman, PM. (1994). The three-dimensional structure of N-acetylneuraminate lyase from *Escherichia coli*. *Structure* **2**:361-369.

Jaeger, KE, Eggert, T, Eipper, A and Reetz, MT. (2001). Directed evolution and the creation of enantioselective biocatalysts. *Appl Microbiol Biotechnol* **55**:519-530.

Jennewein, S, Schurmann, M, Wolberg, M, Hilker, I, Luiten, R, Wubbolts, M and Mink, D. (2006). Directed evolution of an industrial biocatalyst: 2-deoxy-D-ribose 5-phosphate aldolase. *Biotechnol J* **1**:537-548.

Jiang, L, Althoff, EA, Clemente, FR and other authors. (2008). De novo computational design of retro-aldol enzymes. *Science* **319**:1387-1391.

Joerger, AC, Mayer, S and Fersht, AR. (2003). Mimicking natural evolution in vitro: an N-acetylneuraminate lyase mutant with an increased dihydrodipicolinate synthase activity. *Proc Natl Acad Sci U S A* **100**:5694-5699.

Jones, A, Lamsa, M, Frandsen, TP, Spendler, T, Harris, P, Sloma, A, Xu, F, Nielsen, JB and Cherry, JR. (2008). Directed evolution of a maltogenic alpha-amylase from Bacillus sp. TS-25. *J Biotechnol* **134**:325-333.

Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Cryst* **26**:795-800.

Kaper, T, Lager, I, Looger, LL, Chermak, D and Frommer, WB. (2008). Fluorescence resonance energy transfer sensors for quantitative monitoring of pentose and disaccharide accumulation in bacteria. *Biotechnol Biofuels* **1**:11.

Karsten, WE. (1997). Dihydrodipicolinate synthase from *Escherichia coli*: pH dependent changes in the kinetic mechanism and kinetic mechanism of allosteric inhibition by L-lysine. *Biochemistry (Mosc)* **36**:1730-1739.

Kawarabayasi, Y, Hino, Y, Horikawa, H and other authors. (2001). Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* strain7. *DNA Res* **8**:123-140.

Kefala, G, Evans, GL, Griffin, MD, Devenish, SR, Pearce, FG, Perugini, MA, Gerrard, JA, Weiss, MS and Dobson, RC. (2008). Crystal structure and kinetic study of dihydrodipicolinate synthase from Mycobacterium tuberculosis. *Biochem J* **411**:351-360.

Kennerknecht, N, Peters-Wendisch, P, Eggeling, L and Sahm, H. (2003). Metabolic Engineering: Entwicklung vor Bakterienstämmen für die Lysinproduktion. *BIOspektrum* **9**:582-585.

Kihara, D. (2005). The effect of long-range interactions on the secondary structure formation of proteins. *Protein Sci* **14**:1955-1963.

Kim, R, Sandler, SJ, Goldman, S, Yokota, H, Clark, AJ and Kim, SH. (1998). Overexpression of archaeal proteins in Escherichia coli. *Biotechnology Letters* **20**:207-210.

Klimczak, LJ, Grummt, F and Burger, KJ. (1985). Purification and characterization of DNA polymerase from the archaebacterium *Sulfolobus acidocaldarius*. *Nucleic Acids Res* **13**:5269-5282.

Kumpaisal, R, Hashimoto, T and Yamada, Y. (1987). Purification and Characterization of Dihydrodipicolinate Synthase from Wheat Suspension Cultures. *Plant Physiol* **85**:145-151.

Kunkel, TA, Bebenek, K and McClary, J. (1991). Efficient Site-Directed Mutagenesis Using Uracil-Containing DNA. *Bacterial Genetic Systems edn*: Academic Press.

Kwon, SY, Choi, YJ, Kang, TH, Lee, KH, Cha, SS, Kim, GH, Lee, HS, Kim, KT and Kim, KJ. (2005). Highly efficient protein expression and purification using bacterial hemoglobin fusion vector. *Plasmid* **53**:274-282.

Laber, B, Gomis-Ruth, FX, Romao, MJ and Huber, R. (1992). *Escherichia coli* dihydrodipicolinate synthase. Identification of the active site and crystallization. *Biochem J* **288 (Pt 2)**:691-695.

References

Lamble, HJ, Danson, MJ, Hough, DW and Bull, SD. (2005a). Engineering stereocontrol into an aldolase-catalysed reaction. *Chem Commun*:124-126.

Lamble, HJ, Heyer, NI, Bull, SD, Hough, DW and Danson, MJ. (2003). Metabolic pathway promiscuity in the archaeon *Sulfolobus solfataricus* revealed by studies on glucose dehydrogenase and 2-keto-3-deoxygluconate aldolase. *J Biol Chem* **278**:34066-34072.

Lamble, HJ, Royer, SF, Hough, DW, Danson, MJ, Taylor, GL and Bull, SD. (2007). A thermostable aldolase for the synthesis of 3-deoxy-2-ulosonic acids. *Advanced Synthesis & Catalysis* **349**:817-821.

Lamble, HJ, Theodossis, A, Milburn, CC, Taylor, GL, Bull, SD, Hough, DW and Danson, MJ. (2005b). Promiscuity in the part-phosphorylative Entner-Doudoroff pathway of the archaeon *Sulfolobus solfataricus*. *FEBS Lett* **579**:6865-6869.

Lassila, JK, Keeffe, JR, Oelschlaeger, P and Mayo, SL. (2005). Computationally designed variants of Escherichia coli chorismate mutase show altered catalytic activity. *Protein Eng Des Sel* **18**:161-163.

Lawrence, MC, Barbosa, JA, Smith, BJ, Hall, NE, Pilling, PA, Ooi, HC and Marcuccio, SM. (1997). Structure and mechanism of a sub-family of enzymes related to N-acetylneuraminate lyase. *J Mol Biol* **266**:381-399.

Lebbink, JH, Kaper, T, Bron, P, van der Oost, J and de Vos, WM. (2000). Improving low-temperature catalysis in the hyperthermostable *Pyrococcus furiosus* beta-glucosidase CelB by directed evolution. *Biochemistry (Mosc)* **39**:3656-3665.

Leung, DW, Chen, EW and Goeddel, DV. (1989). A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique* **1**:11-15.

Li, JW, Zheng, JL, Wang, XW, Jin, M and Chao, FH. (2007). Directed molecular evolution of nitrite oxido-reductase by DNA-shuffling. *Biomed Environ Sci* **20**:113-118.

Lin, W, Zhang, X, He, Z, Jin, Y, Gong, L and Mi, A. (2002). Reduction of azides to amines or amides with zinc and ammonium chloride as reducing agent. *Synthetic Communications* **32**:3279-3284.

Lipovsek, D, Antipov, E, Armstrong, KA, Olsen, MJ, Klibanov, AM, Tidor, B and Wittrup, KD. (2007). Selection of horseradish peroxidase variants with enhanced enantioselectivity by yeast surface display. *Chem Biol* **14**:1176-1185.

Lippow, SM and Tidor, B. (2007). Progress in computational protein design. *Curr Opin Biotechnol* **18**:305-311.

Littlechild, JA, Guy, J, Connelly, S, Mallett, L, Waddell, S, Rye, CA, Line, K and Isupov, M. (2007). Natural methods of protein stabilization: thermostable biocatalysts. *Biochem Soc Trans* **35**:1558-1563.

Liu, JJ, Hsu, CC and Wong, CH. (2004). Sequential aldol condensation catalyzed by DERA mutant Ser238Asp and a formal total synthesis of atorvastatin. *Tetrahedron Lett* **45**:2439-2441.

Looger, LL, Dwyer, MA, Smith, JJ and Hellinga, HW. (2003). Computational design of receptor and sensor proteins with novel functions. *Nature* **423**:185-190.

Looger, LL and Hellinga, HW. (2001). Generalized dead-end elimination algorithms make large-scale protein side-chain structure prediction tractable: implications for protein design and structural genomics. *J Mol Biol* **307**:429-445.

Lutz, S, Ostermeier, M, Moore, GL, Maranas, CD and Benkovic, SJ. (2001). Creating multiple-crossover DNA libraries independent of sequence identity. *Proc Natl Acad Sci U S A* **98**:11248-11253.

Machajewski, TD and Wong, CH. (2000). The Catalytic Asymmetric Aldol Reaction. *Angew Chem Int Ed Engl* **39**:1352-1375.

Machielsen, R, Leferink, NG, Hendriks, A, Brouns, SJ, Hennemann, HG, Daubetamann, T and van der Oost, J. (2008). Laboratory evolution of *Pyrococcus furiosus* alcohol dehydrogenase to improve the production of (2S,5S)-hexanediol at moderate temperatures. *Extremophiles* **12**:587-594.

MacKerell, AD, Bashford, D, Bellott, M and other authors. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B* **102**:3586-3616.

Massant, J, Verstreken, P, Durbecq, V, Kholti, A, Legrain, C, Beeckmans, S, Cornelis, P and Glansdorff, N. (2002). Metabolic channeling of carbamoyl phosphate, a thermolabile intermediate: evidence for physical interaction between carbamate kinase-like carbamoyl-phosphate synthetase and ornithine carbamoyltransferase from the hyperthermophile *Pyrococcus furiosus*. *J Biol Chem* **277**:18517-18522.

Mazelis, M, Whatley, FR and Whatley, J. (1977). The enzymology of lysine biosynthesis in higher plants. The occurrence, characterization and some regulatory properties of dihydrodipicolinate synthase. *FEBS Lett* **84**:236-240.

McRee, DE. (1999). XtalView/Xfit--A versatile program for manipulating atomic coordinates and electron density. *J Struct Biol* **125**:156-165.

Merz, A, Yee, MC, Szadkowski, H, Pappenberger, G, Crameri, A, Stemmer, WP, Yanofsky, C and Kirschner, K. (2000). Improving the catalytic activity of a thermophilic enzyme at low temperatures. *Biochemistry (Mosc)* **39**:880-889.

Midelfort, CF, Gupta, RK and Meloche, HP. (1977). Specificity of 2-keto-3-deoxygluconate-6-P aldolase for open chain form of 2-keto-3-deoxygluconate-6-P. *J Biol Chem* **252**:3486-3492.

Mirwaldt, C, Korndorfer, I and Huber, R. (1995). The crystal structure of dihydrodipicolinate synthase from *Escherichia coli* at 2.5 A resolution. *J Mol Biol* **246**:227-239.

Mitsakos, V, Dobson, RC, Pearce, FG, Devenish, SR, Evans, GL, Burgess, BR, Perugini, MA, Gerrard, JA and Hutton, CA. (2008). Inhibiting dihydrodipicolinate synthase across species: towards specificity for pathogens? *Bioorg Med Chem Lett* **18**:842-844.

Miura, T and Kajimoto, T. (2001). Application of L-threonine aldolase-catalyzed reaction to the preparation of protected 3R,5R-dihydroxy-L-homoproline as a mimetic of idulonic acid. *Chirality* **13**:577-580.

Murshudov, GN, Vagin, AA and Dodson, EJ. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**:240-255.

Murshudov, GN, Vagin, AA, Lebedev, A, Wilson, KS and Dodson, EJ. (1999). Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr D Biol Crystallogr* **55**:247-255.

Nagano, N, Orengo, CA and Thornton, JM. (2002). One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. *J Mol Biol* **321**:741-765.

Nakahara, K, Yamamoto, H, Miyake, C and Yokota, A. (2003). Purification and characterization of class-I and class-II fructose-1,6-bisphosphate aldolases from the cyanobacterium Synechocystis sp. PCC 6803. *Plant Cell Physiol* **44**:326-333.

Ness, JE, Kim, S, Gottman, A, Pak, R, Krebber, A, Borchert, TV, Govindarajan, S, Mundorff, EC and Minshull, J. (2002). Synthetic shuffling expands functional protein diversity by allowing amino acids to recombine independently. *Nat Biotechnol* **20**:1251-1255.

References

O'Maille, PE, Bakhtina, M and Tsai, MD. (2002). Structure-based combinatorial protein engineering (SCOPE). *J Mol Biol* **321**:677-691.

Ostermeier, M, Shim, JH and Benkovic, SJ. (1999). A combinatorial approach to hybrid enzymes independent of DNA homology. *Nat Biotechnol* **17**:1205-1209.

Otten, LG and Quax, WJ. (2005). Directed evolution: selecting today's biocatalysts. *Biomol Eng* **22**:1-9.

Papaleo, E, Riccardi, L, Villa, C, Fantucci, P and De Gioia, L. (2006). Flexibility and enzymatic cold-adaptation: a comparative molecular dynamics investigation of the elastase family. *Biochim Biophys Acta* **1764**:1397-1406.

Park, KW, Webster, DA, Stark, BC, Howard, AJ and Kim, KJ. (2003). Fusion protein system designed to provide color to aid in the expression and purification of proteins in Escherichia coli. *Plasmid* **50**:169-175.

Patel, RN. (2001). Enzymatic synthesis of chiral intermediates for Omapatrilat, an antihypertensive drug. *Biomol Eng* **17**:167-182.

Pauluhn, A, Ahmed, H, Lorentzen, E, Buchinger, S, Schomburg, D, Siebers, B and Pohl, E. (2008). Crystal structure and stereochemical studies of KD(P)G aldolase from Thermoproteus tenax. *Proteins* **72**:35-43.

Pearce, FG, Perugini, MA, McKerchar, HJ and Gerrard, JA. (2006). Dihydrodipicolinate synthase from *Thermotoga maritima*. *Biochem J* **400**:359-366.

Petersen, M and Kiener, A. (1999). Biocatalysis Preparation and functionalization of N-heterocycles. *Green Chemistry* **april**:99-106.

Pollard, DJ and Woodley, JM. (2007). Biocatalysis for pharmaceutical intermediates: the future is now. *Trends Biotechnol* **25**:66-73.

Ramsaywak, PC, Labbe, G, Siemann, S, Dmitrienko, GI and Guillemette, JG. (2004). Molecular cloning, expression, purification, and characterization of fructose 1,6-bisphosphate aldolase from Mycobacterium tuberculosis--a novel Class II A tetramer. *Protein Expr Purif* **37**:220-228.

Ran, N, Draths, KM and Frost, JW. (2004). Creation of a shikimate pathway variant. *J Am Chem Soc* **126**:6856-6857.

Ran, N and Frost, JW. (2007). Directed evolution of 2-keto-3-deoxy-6-phosphogalactonate aldolase to replace 3-deoxy-D-arabino-heptulosonic acid 7-phosphate synthase. *J Am Chem Soc* **129**:6130-6139.

Reetz, MT. (2001). Combinatorial and Evolution-Based Methods in the Creation of Enantioselective Catalysts. *Angew Chem Int Ed Engl* **40**:284-310.

Reetz, MT, Bocola, M, Carballeira, JD, Zha, D and Vogel, A. (2005). Expanding the range of substrate acceptance of enzymes: combinatorial active-site saturation test. *Angew Chem Int Ed Engl* **44**:4192-4196.

Reetz, MT, Wilensek, S, Zha, D and Jaeger, K-E. (2001). Directed evolution of an enantioselective enzyme through combinatorial multiple-cassette mutagenesis. *Angewandte Chemie International Edition* **40**:3589-3591.

Reher, M and Schonheit, P. (2006). Glyceraldehyde dehydrogenases from the thermoacidophilic euryarchaeota *Picrophilus torridus* and *Thermoplasma acidophilum*, key enzymes of the non-phosphorylative Entner-Doudoroff pathway, constitute a novel enzyme family within the aldehyde dehydrogenase superfamily. *FEBS Lett* **580**:1198-1204.

Reiter, WD, Palm, P and Zillig, W. (1988). Analysis of transcription in the archaebacterium *Sulfolobus* indicates that archaebacterial promoters are homologous to eukaryotic pol II promoters. *Nucleic Acids Res* **16**:1-19.

Renslo, AR. (2007). The echinocandins: Total and semi-synthetic approaches in antifungal drug discovery. *Anti-Infective Agents in Medicinal Chemistry* **6**:201-212.

Roberts, SJ, Morris, JC, Dobson, RC and Gerrard, JA. (2003). The preparation of (S)-aspartate semi-aldehyde appropriate for use in biochemical studies. *Bioorg Med Chem Lett* **13**:265-267.

Rossi, M, Rella, R, Pensa, M, Bartolucci, S, De Rosa, M, Gambacorta, A, Raia, CA and Dell'Aversano Orabona, N. (1986). Structure and Properties of a Thermophilic and Thermostable DNA Polymerase Isolated from *Sulfolobus solfataricus*. *Syst Appl Microbiol* **7**:337-341.

Samland, AK and Sprenger, GA. (2006). Microbial aldolases as C-C bonding enzymes-unknown treasures and new developments. *Appl Microbiol Biotechnol* **71**:253-264.

Sanchez-Moreno, I, Francisco Garcia-Garcia, J, Bastida, A and Garcia-Junceda, E. (2004). Multienzyme system for dihydroxyacetone phosphate-dependent aldolase catalyzed C-C bond formation from dihydroxyacetone. *Chem Commun*:1634-1635.

Saraf, MC, Moore, GL, Goodey, NM, Cao, VY, Benkovic, SJ and Maranas, CD. (2006). IPRO: an iterative computational protein library redesign and optimization procedure. *Biophys J* **90**:4167-4180.

Scapin, G, Reddy, SG, Zheng, R and Blanchard, JS. (1997). Three-dimensional structure of *Escherichia coli* dihydrodipicolinate reductase in complex with NADH and the inhibitor 2,6-pyridinedicarboxylate. *Biochemistry (Mosc)* **36**:15081-15088.

Scheib, H, Pleiss, J, Stadler, P, Kovac, A, Potthoff, AP, Haalck, L, Spener, F, Paltauf, F and Schmid, RD. (1998). Rational design of Rhizopus oryzae lipase with modified stereoselectivity toward triradylglycerols. *Protein Eng* **11**:675-682.

Schmidt-Dannert, C and Arnold, FH. (1999). Directed evolution of industrial enzymes. *Trends Biotechnol* **17**:135-136.

Seebeck, FP and Hilvert, D. (2003). Conversion of a PLP-dependent racemase into an aldolase by a single active site mutation. *J Am Chem Soc* **125**:10158-10159.

Selig, M, Xavier, KB, Santos, H and Schonheit, P. (1997). Comparative analysis of Embden-Meyerhof and Entner-Doudoroff glycolytic pathways in hyperthermophilic archaea and the bacterium *Thermotoga*. *Arch Microbiol* **167**:217-232.

Sen, S, Dasu, VV and Mandal, B. (2007). Developments in directed evolution for improving enzyme functions. *Appl Biochem Biotechnol* **143**:212-223.

Serianni, AS, Pierce, J, Huang, S and Barker, R. (1982). Anomerization of furanose sugars: Kinetics of Ring-opening Reactions by ^{1}H and ^{13}C Saturation-Transfer NMR Spectroscopy. *J Am Chem Soc* **104**:4037-4044.

Shao, Z and Arnold, FH. (1996). Engineering new functions and altering existing functions. *Curr Opin Struct Biol* **6**:513-518.

Shao, Z, Zhao, H, Giver, L and Arnold, FH. (1998). Random-priming in vitro recombination: an effective tool for directed evolution. *Nucleic Acids Res* **26**:681-683.

She, Q, Singh, RK, Confalonieri, F and other authors. (2001). The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc Natl Acad Sci U S A* **98**:7835-7840.

Shedlarski, JG and Gilvarg, C. (1970). The pyruvate-aspartic semialdehyde condensing enzyme of *Escherichia coli*. *J Biol Chem* **245**:1362-1373.

References

Shelton, MC, Cotterill, IC, Novak, STA, Poonawale, RH, Sudarshan, S and Toone, EJ. (1996). 2-keto-3-deoxy-6-phosphogluconate aldolases as catalysts for stereocontrolled carbon-carbon formation. *J Am Chem Soc* **118**:2117-2125.

Shi, C, Lu, X, Ma, C, Ma, Y, Fu, X and Yu, W. (2008). Enhancing the thermostability of a novel beta-agarase AgaB through directed evolution. *Appl Biochem Biotechnol* **151**:51-59.

Sieber, V, Martinez, CA and Arnold, FH. (2001). Libraries of hybrid proteins from distantly related sequences. *Nat Biotechnol* **19**:456-460.

Siebers, B, Brinkmann, H, Dorr, C, Tjaden, B, Lilie, H, van der Oost, J and Verhees, CH. (2001). Archaeal fructose-1,6-bisphosphate aldolases constitute a new family of archaeal type class I aldolase. *J Biol Chem* **276**:28710-28718.

Siebum, A, van Wijk, A, Schoevaart, R and Kieboom, T. (2006). Galactose oxidase and alcohol oxidase: Scope and limitations for the enzymatic synthesis of aldehydes. *Journal of Molecular Catalysis B-Enzymatic* **41**:141-145.

Sorensen, HP, Sperling-Petersen, HU and Mortensen, KK. (2003). Production of recombinant thermostable proteins expressed in Escherichia coli: completion of protein synthesis is the bottleneck. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **786**:207-214.

Sriprapundh, D, Vieille, C and Zeikus, JG. (2003). Directed evolution of *Thermotoga neapolitana* xylose isomerase: high activity on glucose at low temperature and low pH. *Protein Eng* **16**:683-690.

Stahly, DP. (1969). Dihydrodipicolinic acid synthase of *Bacillus licheniformis*. *Biochim Biophys Acta* **191**:439-451.

Stedman, KM, Schleper, C, Rumpf, E and Zillig, W. (1999). Genetic requirements for the function of the archaeal virus SSV1 in *Sulfolobus solfataricus*: construction and testing of viral shuttle vectors. *Genetics* **152**:1397-1405.

Stemmer, WP. (1994a). DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc Natl Acad Sci U S A* **91**:10747-10751.

Stemmer, WP. (1994b). Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370**:389-391.

Straathof, AJ, Panke, S and Schmid, A. (2002). The production of fine chemicals by biotransformations. *Curr Opin Biotechnol* **13**:548-556.

Sugiyama, M, Hong, Z, Liang, PH, Dean, SM, Whalen, LJ, Greenberg, WA and Wong, CH. (2007). D-Fructose-6-phosphate aldolase-catalyzed one-pot synthesis of iminocyclitols. *J Am Chem Soc* **129**:14811-14817.

Sukumaran, J and Hanefeld, U. (2005). Enantioselective C-C bond synthesis catalysed by enzymes. *Chem Soc Rev* **34**:530-542.

Suzuki, T, Iwasaki, T, Uzawa, T, Hara, K, Nemoto, N, Kon, T, Ueki, T, Yamagishi, A and Oshima, T. (2002). *Sulfolobus tokodaii* sp. nov. (f. *Sulfolobus* sp. strain 7), a new member of the genus *Sulfolobus* isolated from Beppu Hot Springs, Japan. *Extremophiles* **6**:39-44.

Takayama, S, McGarvey, GJ and Wong, CH. (1997). Microbial aldolases and transketolases: new biocatalytic approaches to simple and complex sugars. *Annu Rev Microbiol* **51**:285-310.

Tam, PH, Phenix, CP and Palmer, DR. (2004). MosA, a protein implicated in rhizopine biosynthesis in *Sinorhizobium meliloti* L5-30, is a dihydrodipicolinate synthase. *J Mol Biol* **335**:393-397.

Terwilliger, TC. (2004). Using prime-and-switch phasing to reduce model bias in molecular replacement. *Acta Cryst D* **60**:2144-2149.

Theodossis, A, Walden, H, Westwick, EJ, Connaris, H, Lamble, HJ, Hough, DW, Danson, MJ and Taylor, GL. (2004). The structural basis for substrate promiscuity in 2-keto-3-deoxygluconate

aldolase from the Entner-Doudoroff pathway in *Sulfolobus solfataricus*. *J Biol Chem* **279**:43886-43892.

Tsuchida, T, Kubota, K and Yoshinaga, F. (1986). Studies on the Fermentative Production of L-Proline .6. Improvement of L-Proline Production by Sulfaguanidine Resistant Mutants Derived from L-Glutamic Acid-Producing Bacteria. *Agric Biol Chem* **50**:2201-2207.

Tsujimoto, N, Gunji, Y, Ogawa-Miyata, Y, Shimaoka, M and Yasueda, H. (2006). l-Lysine biosynthetic pathway of *Methylophilus methylotrophus* and construction of an l-lysine producer. *J Biotechnol* **124**:327-337.

Tudor, DW, Lewis, T and Robins, DJ. (1993). Synthesis of the Trifluoroacetate Salt of Aspartic-Acid Beta-Semialdehyde, an Intermediate in the Biosynthesis of L-Lysine, L-Threonine, and L-Methionine. *Synthesis-Stuttgart*:1061-1062.

Valetti, F and Gilardi, G. (2004). Directed evolution of enzymes for product chemistry. *Nat Prod Rep* **21**:490-511.

van den Heuvel, RHH, van den Berg, WAM, Rovida, S and van Berkel, WJH. (2004). Laboratory-evolved vanillyl-alcohol oxidase produces natural vanillin. *J Biol Chem* **279**:33492-33500.

van Loo, B, Spelberg, JH, Kingma, J, Sonke, T, Wubbolts, MG and Janssen, DB. (2004). Directed evolution of epoxide hydrolase from *A. radiobacter* toward higher enantioselectivity by error-prone PCR and DNA shuffling. *Chem Biol* **11**:981-990.

van Sint Fiet, S, van Beilen, JB and Witholt, B. (2006). Selection of biocatalysts for chemical synthesis. *PNAS* **103**:1693-1698.

Vauterin, M and Jacobs, M. (1994). Isolation of a poplar and an *Arabidopsis thaliana* dihydrodipicolinate synthase cDNA clone. *Plant Mol Biol* **25**:545-550.

Vieille, C, Hess, JM, Kelly, RM and Zeikus, JG. (1995). xylA cloning and sequencing and biochemical characterization of xylose isomerase from *Thermotoga neapolitana*. *Appl Environ Microbiol* **61**:1867-1875.

Wahler, D and Reymond, JL. (2001). High-throughput screening for biocatalysts. *Curr Opin Biotechnol* **12**:535-544.

Wernic, D, Dimaio, J and Adams, J. (1989). Enantiospecific Synthesis of L-Alpha-Aminosuberic Acid - Synthetic Applications in Preparation of Atrial Natriuretic Factor Analogs. *J Org Chem* **54**:4224-4228.

Williams, GJ, Woodhall, T, Farnsworth, LM, Nelson, A and Berry, A. (2006). Creation of a pair of stereochemically complementary biocatalysts. *J Am Chem Soc* **128**:16238-16247.

Williams, GJ, Woodhall, T, Nelson, A and Berry, A. (2005). Structure-guided saturation mutagenesis of N-acetylneuraminic acid lyase for the synthesis of sialic acid mimetics. *Protein Eng Des Sel* **18**:239-246.

Wolterink-van Loo, S, van Eerde, A, Siemerink, MA, Akerboom, J, Dijkstra, BW and van der Oost, J. (2007). Biochemical and structural exploration of the catalytic capacity of *Sulfolobus* KDG aldolases. *Biochem J* **403**:421-430.

Xu, HF, Zhang, XE, Zhang, ZP, Zhang, YM and Cass, AEG. (2003). Directed evolution of *E-coli* alkaline phosphatase towards higher catalytic activity. *Biocatalysis and Biotransformation* **21**:41-47.

Xue, Y, Xu, Y, Liu, Y, Ma, Y and Zhou, P. (2001). *Thermoanaerobacter tengcongensis* sp. nov., a novel anaerobic, saccharolytic, thermophilic bacterium isolated from a hot spring in Tengcong, China. *Int J Syst Evol Microbiol* **51**:1335-1341.

References

Yu, H, Chokhawala, HA, Huang, S and Chen, X. (2006). One-pot three-enzyme chemoenzymatic approach to the synthesis of sialosides containing natural and non-natural functionalities. *Nat Protoc* **1**:2485-2492.

Yugari, Y and Gilvarg, C. (1965). The condensation step in diaminopimelate synthesis. *J Biol Chem* **240**:4710-4716.

Zha, D, Eipper, A and Reetz, MT. (2003). Assembly of designed oligonucleotides as an efficient method for gene recombination: a new tool in directed evolution. *Chembiochem* **4**:34-39.

Zhao, H and Arnold, FH. (1997). Optimization of DNA shuffling for high fidelity recombination. *Nucleic Acids Res* **25**:1307-1308.

Zhao, H, Giver, L, Shao, Z, Affholter, JA and Arnold, FH. (1998). Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat Biotechnol* **16**:258-261.

Ziegenhorn, J, Senn, M and Bucher, T. (1976). Molar absorptivities of beta-NADH and beta-NADPH. *Clin Chem* **22**:151-160.

Zillig, W, Stetter, KO and Janekovic, D. (1979). DNA-dependent RNA polymerase from the archaebacterium *Sulfolobus acidocaldarius*. *Eur J Biochem* **96**:597-604.

Curriculum vitae

Suzanne van Loo werd geboren op 11 april 1979 te Helmond. In 1997 haalde zij haar atheneum-diploma aan de SG Jan van Brabant. In september datzelfde jaar begon ze haar studie biologie aan de Landbouwuniversiteit in Wageningen. Zij heeft daarin de cellulaire/moleculaire specialisatie richting gevolgd. Afstudeervakken werden uitgevoerd aan de laboratoria voor virologie en genetica. Voor een stage in de evolutionaire genetica ging zij naar de Emory University van Atlanta, USA. In september 2002 kreeg zij haar diploma.

Vervolgens begon zij met haar promotieonderzoek bij de werkgroep Bacteriële genetica aan het laboratorium voor Microbiologie aan de Wageningen Universiteit (juli 2003-december 2008). De resultaten van dit onderzoek staan beschreven in dit proefschrift. Vanaf 1 september is zij als Post-Doc werkzaam bij de onderzoeksgroep molecular & developmental genetics aan de Universiteit Leiden.

List of publications

Van Eerde, A, Wolterink-van loo, S, van der Oost, J, and Dijkstra BW. (2006). Fortuitous structure determination of 'as isolated' *Escherichia coli* bacterioferritin in a novel crystal form. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **62**:1061-1066

Wolterink-van Loo, S, van Eerde, A, Siemerink, MA, Akerboom, J, Dijkstra, BW and van der Oost, J. (2007). Biochemical and structural exploration of the catalytic capacity of *Sulfolobus* KDG aldolases. *Biochem J* **403**:421-430

Wolterink-van Loo, S, Levisson, M, Cabrières, MC, Franssen, MC and van der Oost, J. (2008). Characterization of thermostable dihydrodipicolinate synthase from *Thermoanaerobacter tengcongensis*. *Extremophiles* **12**:461-469

Wolterink-van Loo, S, Siemerink M.A.J., Perrakis, G, Kaper, T, Kengen, SWM and van der Oost, J. Optimizing low-temperature activity of Sulfolobus acidocaldarius 2-keto-3-deoxygluconate aldolase. *Submitted for publication.*

Schurink, M, Wolterink-van Loo, S, Kaper, T, Ruigrok, VJB, Looger, LL, van der Oost, J and Franssen, MCR. Stereoselectivity of *Sulfolobus* KDG aldolases and engineering of substrate specificity towards azido-substituted aldehydes. *Submitted for publication.*

Overview of completed training activities

Discipline specific activities	**Year**
Courses	
Bioinformatics I, Wageningen University	2003
Protein engineering, VLAG	2004
Advanced course on biocatalysis, Delft	2004
Bionanotechnology, VLAG	2005
Organic chemistry, Wageningen University	2005
Meetings and congresses	
Annual Molecular Genetics meeting, Lunteren	2003
Dutch biotechnology congress (NBC), Ede (poster)	2004
Annual Protein meetings, Lunteren (poster)	2003-2007
NCCC VII (presentation)	2006
International Congress on Biocatalysis, Hamburg (presentation)	2006

General courses	
VLAG introductionary course	2004
Organizing and supervising thesis work	2003

Optional activities	
Preparing PhD research proposal	2003
Interdisciplinary IBOS meetings (poster-presentation)	2003-2007
Vlag PhD trip, California	2007
Research & PhD meetings, microbiology	2003-2007

Dankwoord

Waar zouden wij zijn zonder de trein? Eigenlijk geen idee, maar in ieder geval heb ik er genoeg tijd om na te denken wie ik op deze plek wil bedanken. Ten eerste natuurlijk mijn promotor en begeleider John, die mij begeleid heeft naar dit (eind)station. Je gaf me ondersteuning met ideeën waar nodig, maar ook vooral bedankt voor het in mij gestelde vertrouwen. Willem, we hadden niet zoveel contact tijdens mijn aio-tijd. Dus voornamelijk bedankt voor de puntjes op i's van dit proefschrift.

Tijdens het project stapten er ook studenten op de aldolase-trein. Marco, Mark, Georgios en Vincent bedankt voor jullie bijdrage aan dit proefschrift! Marco, na een afstudeervak op het zij-traject van aminotransferases (waar verder weinig mee gedaan is), had je een goede suggestie voor een stageproject, nu wel met aldolases. Je ging hiervoor naar Groningen voor een screen van random-mutanten en nu sta je ook nog eens als paranimf naast me, bedankt! Thijs en Harm, bedankt voor de praktische hulp tijdens deze screens. Als we dan toch in Groningen zijn, André en Bauke geweldig bedankt voor de inwijding in de kristallografie.

Verder wil ik alle mensen binnen het IBOS project bedanken die op hun manier een bijdrage aan dit proefschrift hebben geleverd door discussies tijdens de projectmeetings en ook voor de bewandelde zijsporen die niet in dit boekje terecht gekomen zijn. Extra bedankt natuurlijk Maurice en Marloes voor de grote bijdrage aan met name hoofdstuk 2 en 5. Theo Sonke, jij ook enorm bedankt voor de suggesties tijdens meetings en het altijd weer even kritisch bekijken van de manuscripten die door DSM 'goedgekeurd' moesten worden.

Thijs Kaper, je schreef je eigen veni-project, gebaseerd op aldolases en het 'computational design' werk. Ik vond het fijn met je samen te werken. Helaas maakte je het project niet af (omdat je een goede baan kon krijgen). Bij deze wil ik je geweldig bedanken voor jou bijdrage aan hoofdstuk 4 en 5.

Door de goede sfeer binnen microbiologie en met name BacGen heb ik met veel plezier aan dit proefschrift kunnen werken. Natuurlijk wil ik alle BacGenners bedankten voor de kritische opmerkingen maar ook voor de gezellige tijd (borrels, 'kerstetentjes' etc.). Op het lab lange tijd samen met Ans en Marke. Ans nog extra bedankt voor de praktische 'inleiding' in de microbiologische laboratoriumtechnieken en de vele bestellingen. Marke, bedankt voor gezellige tijd en extra bedankt omdat je hier mijn paranimf wil zijn. Bedankt ook voor de goede werksfeer op de kamer, in eerste instantie: Krisztina, Ronnie, Johan, Odette, Hao en Servé. Servé, nog eens extra bedankt voor de hulp bij enzymkinetiek. Later waren er Mark, Marco, Colin, Bart, Marcel en John R, die voor een goede sfeer zorgde op de werkkamer. Ook aan de overige BacGenners Thijs E, Harmen Jasper A, Jasper W, Stan, Matthijs, Pierpaolo, Magnus, Nicolas, Katrin en Teunie wil ik zeggen: Bedankt voor de fijne tijd bij

BacGen. Verder wil ik alle microbiologen nog bedanken voor gezelligheid en kritische bijdragen tijdens de woensdagochtend meetings. Dan zijn er nog enkele mensen over op het lab die niet mag missen om te bedanken. Wim voor het oplossen van de computer-problemen (zelfs toen ik er officieel niet meer werkte). Jannie, die altijd maar weer zorgde dat alles wat wij smerig maken weer schoon en steriel is. Nees, bedankt voor de bestellingen. Renée voor het terugvinden van hopeloos oude, maar wel nuttige boeken in de bibliotheek. Dan wil ik natuurlijk nog Francis en Anja te bedanken, voor de assistentie en het altijd klaar staan voor het afhandelen van administratieve zaken.

Dan rest mij nog mijn familie te bedanken, mijn familie, waar ik altijd op kan rekenen. Linda en Mark bedankt voor jullie interesse. Bert, ook jij promoveerde op enzymen. Je bent dan ook al die tijd naast mijn grote broer, mijn grote voorbeeld geweest. Papa en mama, natuurlijk bedankt voor de support die ik van jullie kreeg. Mama, je offerde heel wat vrije uurtjes op om op Denise te passen, zodat ik verder kon schrijven – bedankt hiervoor, anders had het vast nog veel langer geduurd. Papa nog extra bedankt, omdat je mama door mij moest missen en natuurlijk voor de inhoudelijke bijdrage aan de Nederlandse samenvatting. Verder natuurlijk ook de schoonfamilie geweldig bedankt voor de getoonde interesse en afleiding tijdens de gezellige familie-feestjes.

Dan nog na-tuur-lijk Arthur. Je hebt me op deze rit geweldig geholpen. Aan de ene kant gewoon omdat je er was, maar ook praktisch door allerlei andere taken thuis over te nemen. Ik was ook vast niet erg gezellig de laatste tijd. Hopelijk hebben we nu weer weekenden voor ons zelf! Als laatste dan Denise, je kwam er nog net voordat mijn promotie-tijd voorbij zou zijn. Daardoor duurde het nog een beetje langer... Hopelijk heb ik nu ook meer tijd voor jou!

134