Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of the branched-chain keto acid decarboxylase (KdcA) from *Lactococcus lactis* provides insights into the structural basis for the chemoselective and enantioselective carboligation reaction

Catrine L. Berthold,[a] Dörte Gocke,[b] Martin D. Wood,[c] Finian J. Leeper,[c] Martina Pohl[b] and Gunter Schneider[c]*

[a]Department of Medical Biochemistry and Biophysics, Karolinska Institutet, S-17177 Stockholm, Sweden, [b]Institute of Molecular Enzyme Technology, Heinrich Heine University Düsseldorf, Research Center Jülich, 52426 Jülich, Germany, and [c]Cambridge University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, England

Correspondence e-mail: gunter.schneider@ki.se

Received 31 August 2007
Accepted 15 October 2007

PDB References: KdcA–deazaThDP, 2vbg, r2vbgsf; holo-KdcA, 2vbf, r2vbfsf.

The thiamin diphosphate (ThDP) dependent branched-chain keto acid decarboxylase (KdcA) from *Lactococcus lactis* catalyzes the decarboxylation of 3-methyl-2-oxobutanoic acid to 3-methylpropanal (isobutyraldehyde) and CO_2. The enzyme is also able to catalyze carboligation reactions with an exceptionally broad substrate range, a feature that makes KdcA a potentially valuable biocatalyst for C—C bond formation, in particular for the enzymatic synthesis of diversely substituted 2-hydroxyketones with high enantioselectivity. The crystal structures of recombinant holo-KdcA and of a complex with an inhibitory ThDP analogue mimicking a reaction intermediate have been determined to resolutions of 1.6 and 1.8 Å, respectively. KdcA shows the fold and cofactor–protein interactions typical of thiamin-dependent enzymes. In contrast to the tetrameric assembly displayed by most other ThDP-dependent decarboxylases of known structure, KdcA is a homodimer. The crystal structures provide insights into the structural basis of substrate selectivity and stereoselectivity of the enzyme and thus are suitable as a framework for the redesign of the substrate profile in carboligation reactions.

1. Introduction

ThDP-dependent enzymes appear to be particularly promising as catalysts in bioorganic synthesis owing to their large diversity in substrates and diverse stereospecificity (Müller & Sprenger, 2004). The ability of many ThDP-dependent enzymes to catalyze the ligation of aldehydes into chiral 2-hydroxyketones as a side reaction, the so-called carboligation reaction, has been used in large-scale organic synthesis for decades (Sprenger & Pohl, 1999). The mode of action of the cofactor in ThDP-dependent decarboxylases is well understood (Kern *et al.*, 1997) and the carboligation reaction is most likely to proceed through the same mechanism (Fig. 1; Lobell & Crout, 1996; Siegert *et al.*, 2005).

A general feature of the family of ThDP-dependent enzymes is a high variation of the active-site residues, with hardly any strict conservation, despite the conserved overall reaction mechanism. This provides the basis for the considerable variation in the chemoselectivity of the carboligation reaction, which is ultimately dependent on the size and composition of the substrate-binding pocket. It is also noteworthy that most of the enzymes only catalyze the synthesis of (*R*)-enantiomers of 2-hydroxyketones and that the corresponding (*S*)-products are not readily accessible. Exceptions

have been reported concerning the ligation of benzaldehyde derivatives with acetaldehyde using benzoylformate decarboxylase (BFD) from *Pseudomonas putida* as catalyst, resulting in (*S*)-hydroxyketones (Iding *et al.*, 2000).

The stereochemical outcome of the carboligation reaction is dependent on whether the *Re* or *Si* face of the acceptor substrate approaches the α-carbanion/enamine intermediate; an approach by the *Si* face results in the (*R*)-enantiomer and the *Re* face yields the (*S*)-product (Iding *et al.*, 2000). Steric reasons have been put forward to explain the preference for (*R*)-enantioselectivity of the product (Knoll *et al.*, 2006). Analysis of the structures of several ThDP-dependent decarboxylases suggested that during the C—C bond-forming step the O atoms of the two substrates point roughly in the same direction to allow proton transfer facilitated by a histidine residue and that the presence or absence of a pocket, denoted the S-pocket, determines the enantioselectivity in the carboligation reaction (Knoll *et al.*, 2006; Fig. 2). Formation of the (*S*)-enantiomer in high enantiomeric excess only occurs when the acceptor aldehyde has an optimal fit in this S-pocket, otherwise the (*R*)-product is formed. Rational redesign of the S-pocket in BFD indeed resulted in an (*S*)-specific enzyme variant accepting longer aliphatic aldehydes, demonstrating the structure-guided approach for engineering enantioselectivity in ThDP-dependent enzymes (Gocke *et al.*, unpublished results).

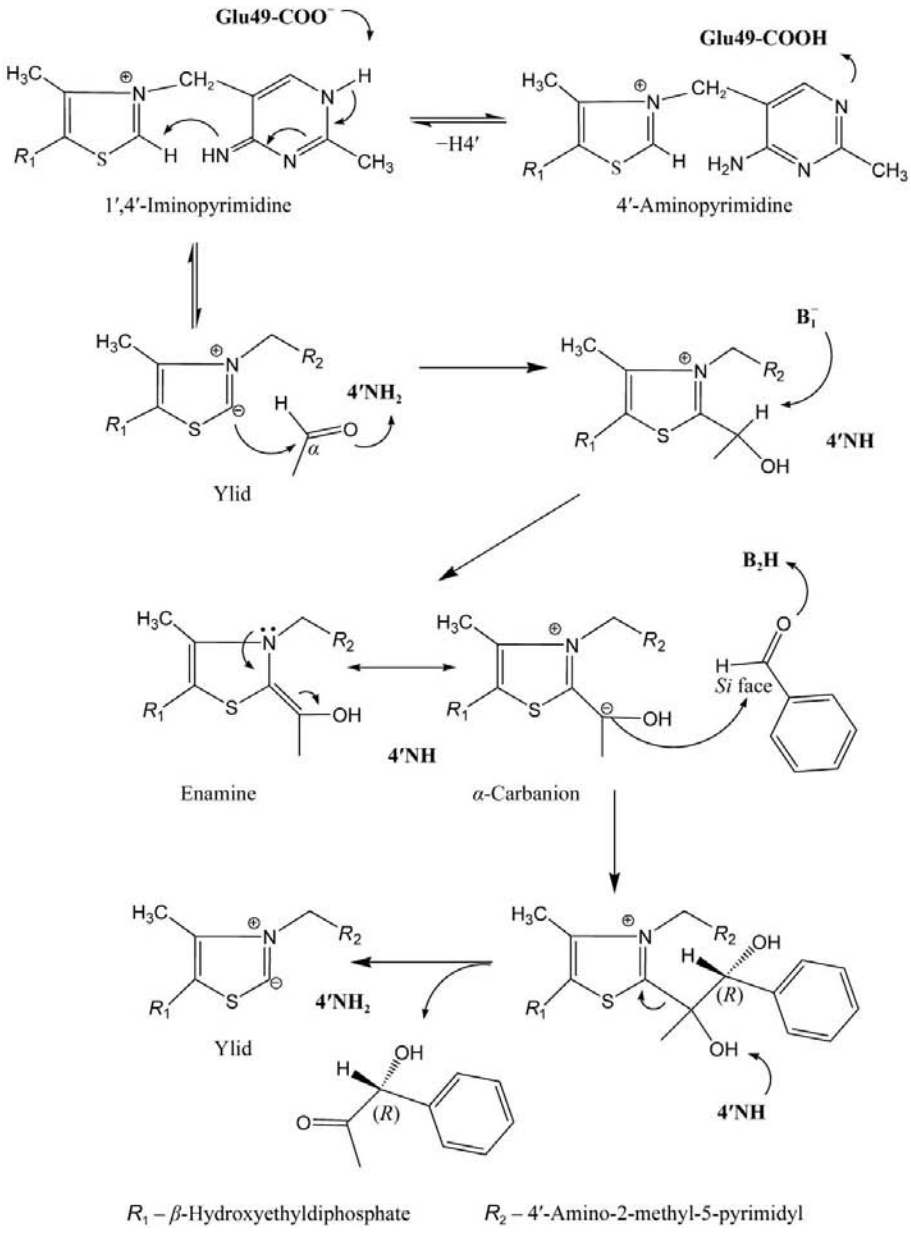

Figure 1
The mechanism of the carboligation reaction of ThDP-dependent enzymes with acetaldehyde as acyl donor and benzaldehyde as the acceptor substrate. KdcA catalyzes the formation of the (*R*)-enantiomer of phenylacetylcarbinol in 92% enantiomeric excess (Gocke *et al.*, 2007). Initially, the cofactor is activated by the enzyme by a conserved glutamic acid residue that stabilizes the imino form of the pyrimidine ring, in which 4′NH can abstract a proton from the C2 atom of the thiazolium ring and create the reactive ylid. Next follows activation of the acyl donor aldehyde by covalent linkage to the C2 atom of ThDP. During this step 4′NH_2 is thought to stabilize the negative charge generated on the substrate carbonyl group. After deprotonation of C^{α}, which forms the enamine/α-carbanion intermediate, ligation with the acceptor aldehyde follows and then product release. Note that the roles of donor and acceptors can be reversed (*i.e.* benzaldehyde is the acyl donor and acetaldehyde is the acceptor substrate), which results in the formation of (*R*)-2-hydroxypropiophenone.

The branched-chain keto acid decarboxylase (KdcA) from *Lactococcus lactis* is a recently discovered member of the ThDP-dependent enzyme family which was originally identified as taking part in flavour formation during cheese ripening (Smit *et al.*, 2005). A sequence comparison suggests that KdcA belongs to the ThDP-dependent decarboxylases (Yep *et al.*, 2006), also named the POX group (Duggleby, 2006) after pyruvate oxidase, the first member of this subgroup of ThDP-dependent enzymes of known structure (Muller & Schulz, 1993). KdcA catalyses the decarboxylation of 3-methyl-2-oxobutanoic acid into isobutyraldehyde, but other substrates are also accepted for decarboxylation (Smit *et al.*, 2005; Yep *et al.*, 2006; Gocke *et al.*, 2007). A recent study of the carboligation activity of KdcA revealed an exceptionally broad substrate range, including various aromatic aldehydes and CH-acidic aldehydes as well as alipathic aldehydes (Gocke *et al.*, 2007). KdcA is the only enzyme so far identified that accepts substrates as large as indole-3-acetaldehyde, produced *in situ* by decarboxylation of

indole-3-pyruvate, as an acyl donor for carboligation. These features make KdcA a valuable addition to the toolbox of ThDP-dependent enzymes for C—C bond formation (Müller & Sprenger, 2004), in particular the enzymatic synthesis of diversely substituted 2-hydroxyketones with high enantioselectivity. Here, we provide high-resolution crystal structures of recombinant holo-KdcA and of a complex with an inhibitory ThDP analogue that mimics a reaction intermediate. Based on these two structures, we compare and discuss the carboligation activity of KdcA and other homologous enzymes for which the reaction has been characterized.

2. Methods

2.1. Protein production

Full details of the cloning, expression and purification of KdcA from *L. lactis* have been described elsewhere (Gocke *et al.*, 2007). In brief, the gene coding for KdcA was cloned into the pET28a (Novagen) vector containing a hexahistidine tag and recombinant protein was produced in *Escherichia coli* strain BL21(DE3). The protein contains a 23-amino-acid N-terminal extension, MGSSHHHHHHSSGLVPRGSHMAS, and could be purified to homogeneity by a single step of Ni–NTA chromotography.

2.2. Preparation of apoenzyme and inhibition by a deazaThDP analogue

The apo form of KdcA was produced by overnight dialysis against 50 m*M* Tris–HCl pH 8.5, 1 m*M* EDTA and 1 m*M* DTT. The buffer was exchanged to 20 m*M* MES buffer pH 6.8 containing 1 m*M* DTT prior to protein concentration by centrifugation. No decarboxylase activity was observed for the apoenzyme using the assay described previously (Gocke *et al.*, 2007), while activity was restored upon addition of ThDP and Mg^{2+}. Incubation of the apoenzyme with 0.1 m*M* 2-[(*R*)-1-hydroxyethyl]deazaThDP prior to addition of ThDP and Mg^{2+} did not result in active holoenzyme, demonstrating the strong binding of the inactive deazaThDP analogue.

2.3. Crystallization

Crystallizations were performed using the vapour-diffusion method. A protein solution of 3–5 mg^{-1} KdcA in 20 m*M* MES buffer pH 6.8 containing 2.5 m*M* $MgSO_4$, 0.1 m*M* ThDP and 1 m*M* DTT was used for crystallization of the holoenzyme. Initial conditions were found using commercial crystallization screens; the small irregular crystals obtained from these could be optimized to single 0.1 × 0.1 × 0.2 mm crystals by additive screening. The optimized crystals were produced in 24-well hanging-drop plates with a reservoir solution containing 25–26%(*v*/*v*) PEG 200 in 0.1 *M* MES buffer pH 6.2. The drops were set up by mixing 2 µl protein solution with 2 µl well solution to which 0.5 µl of the additive spermine tetrahydrochloride (0.1 *M*) had been added. The crystals grew to full size within 1 d and were flash-frozen in liquid nitrogen without any additional cryoprotectant.

The complex of KdcA with 2-[(*R*)-1-hydroxyethyl]deaza-ThDP was crystallized using the same condition as for the holoenzyme. The deazaThDP derivative was synthesized as described by Leeper *et al.* (2005). A solution of the apoenzyme (5 mg ml^{-1} in 20 m*M* MES buffer pH 6.8) was incubated for several hours with 1 m*M* 2-[(*R*)-1-hydroxyethyl]deazaThDP and 2.5 m*M* $MgSO_4$ in order to form the complex. These crystals grew more slowly and heavy precipitant appeared among the few sporadically formed crystals.

2.4. Data collection and structure determination

X-ray data were collected in a nitrogen stream on beamlines ID14-EH1 and ID14-EH3 at the European Synchrotron Research Facility, Grenoble, France. A summary of the data statistics can be found in Table 1. The data were processed with *MOSFLM* (Leslie, 1992) and scaled with *SCALA* from the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994). Molecular replacement was carried out with *MOLREP* (Vagin & Teplyakov, 1997). The monomer of the closest homologue of known structure, indolepyruvate decarboxylase (IPDC) from *Enterobacter cloacae* (PDB code 1ovm; Schütz *et al.*, 2003), was used as a search model for the initial structure determination of the holoenzyme. Two monomers were found in the asymmetric unit, comprising the biological dimer. The biological sequence was fitted to the electron density with the program *ARP*/*wARP* (Perrakis *et al.*,

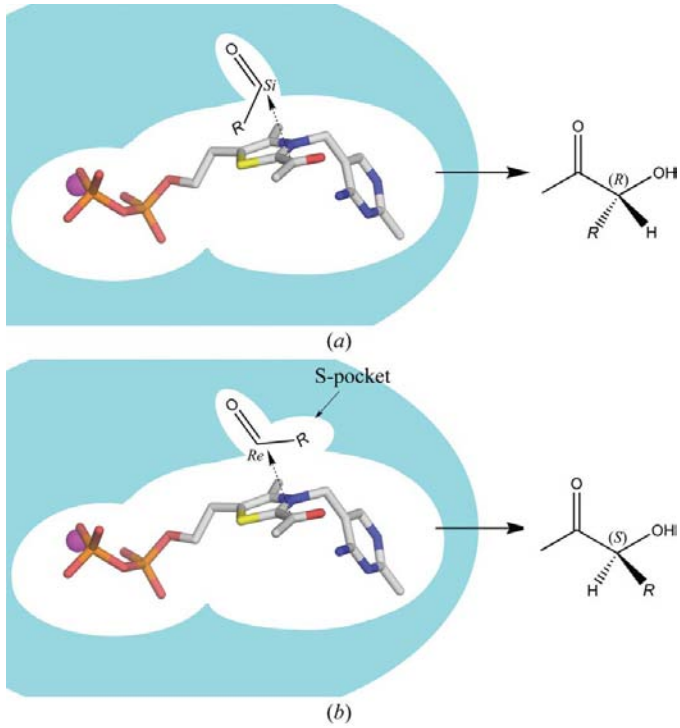

Figure 2
A schematic representation of the S-pocket. (*a*) When there is no S-pocket or an S-pocket that is too small to fit the acceptor substrate side chain *R*, it will bind with the *Si* face towards the enamine and the (*R*)-enantiomer will be formed. (*b*) Perfect fit of the acceptor substrate side chain in the S-pocket above the thiazolium ring will allow the substrate to align, resulting in the (*S*)-enantiomer.

1999). Refinement was carried out with *REFMAC*5 (Murshudov *et al.*, 1997); *Coot* (Emsley & Cowtan, 2004) was employed for model building as well as assignment of water molecules. The structure of KdcA with the deazaThDP inhibitor was determined using the dimer of the holoenzyme as a search model. Restraints for the inhibitor molecule, obtained using the Dundee *PRODRG*2 server (Schüttelkopf & van Aalten, 2004), were loosened towards the end of the refinement. The validation tools in *Coot* as well as *PROCHECK* (Laskowski *et al.*, 1993) were used to monitor the geometry of the structures. All protein-molecule figures were produced using *PyMOL* (W. L. DeLano; http://www.pymol.org).

2.5. Modelling of carboligation substrates

Indole-3-pyruvate, which is the largest known substrate that (after decarboxylation) can undergo ligation to an acceptor aldehyde in KdcA, was modelled into the holoenzyme. Alignment of the C^{α} atom and the carboxyl group was performed by superimposing the structure of KdcA with those of substrate complexes of a homologous decarboxylase determined previously (Berthold *et al.*, 2007). The planar enamine intermediates were modelled based on the structure of the 2-[(*R*)-1-hydroxyethyl]deazaThDP–KdcA complex.

3. Results and discussion

3.1. Three-dimensional structure of the KdcA subunit

The structure of holo-KdcA from *L. lactis* was determined by molecular replacement to a resolution of 1.6 Å. The crystals belonged to space group $P2_12_12_1$ and contained one dimer per asymmetric unit. All residues (1–547) could be modelled into the electron-density map, with the exception of the loop 182–187, for which no interpretable electron density could be detected. In addition, residues 538–547 of the C-terminal helix were not well defined in the electron-density map and were refined with occupancies of 0.5 in one of the monomers. The bound cofactors ThDP and Mg^{2+} are well defined in both monomers, which allowed a precise definition of their interactions with enzymic groups.

The KdcA monomer shares the conserved fold of three domains of α/β topology common to members of the ThDP-dependent enzyme family (Muller *et al.*, 1993; Fig. 3*a*). ThDP is bound in a cleft formed at the interface between the PYR (residues 1–181) and PP (residue 346–547) domains from two different subunits and displays the characteristic V-conformation, stabilized by a conserved large hydrophobic residue (Ile404) positioned below the thiazolium ring of the cofactor (Lindqvist & Schneider, 1993). The R domain (residues 182–345) in KdcA does not display the nucleotide-binding site found in other members of the family (Muller & Schulz, 1993; Pang *et al.*, 2002; Berthold *et al.*, 2005) and the corresponding binding pocket is instead filled by large aromatic residues. The closest homologue IPDC (Schütz *et al.*, 2003) shows 41% sequence identity to KdcA and superposition of the subunits results in a root-mean-square deviation of 1.5 Å over 516 C^{α} atoms. The crystal structure of KdcA was further compared with those of pyruvate decarboxylase from *Zymomonas mobilis* (*Zm*PDC; Dobritzsch *et al.*, 1998), BFD from *P. putida* (Hasson *et al.*, 1998) and benzaldehyde lyase from *P. fluorescens* (BAL; Mosbacher *et al.*, 2005), which are homologues with well characterized carboligation activity (Knoll *et al.*, 2006; Janzen *et al.*, 2006; Siegert *et al.*, 2005). Superposition of the KdcA monomer results in a root-mean-square deviation of 2.1 Å over 513 C^{α} atoms with *Zm*PDC (PDB code 1zpd), 2.5 Å over 442 C^{α} atoms with BFD (PDB code 1bfd) and 2.2 Å over 438 C^{α} atoms with BAL (PDB code 2ag0).

Table 1
Crystallographic data-collection and refinement summary.

Values in parentheses are for the highest resolution shell.

	Holo-KdcA	KdcA–deazaThDP
Data collection		
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters		
a (Å)	65.6	66.0
b (Å)	108.4	108.6
c (Å)	146.6	146.5
Molecules in ASU	2	2
Resolution (Å)	1.60 (1.69-1.60)	1.8 (1.9-1.8)
R_{sym}	0.067 (0.476)	0.087 (0.376)
Mean $I/\sigma(I)$	11.3 (2.2)	13.1 (2.1)
Completeness (%)	92.4 (63.8)	97.4 (85.2)
Wilson *B* factor (Å^2)	18	16
Refinement		
Reflections in working set	114611	85927
Reflections in test set	6436	4792
R factor (%)	16.3	16.3
R_{free} (%)	20.0	20.5
Atoms modelled	9806	9634
No. of amino acids in model	1092	1090
No. of ligands	2	2
No. of magnesium ions	2	2
No. of water molecules	1108	950
Average *B* factor (Å^2)		
Protein	19.5	15.3
Ligands	12.7	9.7
Magnesium ions	13.2	10.8
Waters	32.3	26.0
R.m.s. deviations from ideals		
Bonds (Å^2)	0.009	0.008
Angles (°)	1.207	1.131
Ramachandran distribution (%)		
Most favoured	91.2	91.4
Additionally allowed	8.6	8.6
Generously allowed	0.2	0
Disallowed	0	0

After completion of the model, two stretches of unassigned electron density remained in surface pockets located between the R and PYR domains from different subunits in the dimer (Fig. 4). These electron densities fitted well to the sequence stretch HSSGL (residues −14 to −10) of the N-terminal tag, but no electron density was observed between residue −10 and the N-terminus of the enzyme subunit. The distances between the N-termini of the two subunits in the KdcA dimer and these two peptide stretches are too large to be covered by the remaining nine residues of the tag, *i.e.* the bound peptide has to come from another molecule related by crystallographic symmetry. Indeed, the N-terminus of a symmetry-related adjacent molecule is located at a distance (∼22 Å) that could be bridged by the nine disordered residues of the tag. These

interactions between adjacent molecules in the lattice suggest that the peptide tags might contribute to crystal packing.

3.2. Quaternary structure

The smallest catalytic unit of ThDP-dependent enzymes is a dimer owing to the location of the active site at the interface of two subunits (Muller *et al.*, 1993). Most members of the POX subclass of ThDP-dependent enzymes, including IPDC, form tetramers in solution and in the crystal, with the packing being best described as a dimer of dimers (Duggleby, 2006). The asymmetric unit of the crystals of KdcA contains two subunits which are related by a molecular noncrystallographic twofold axis. This molecular twofold axis generates a dimeric molecule with the packing common to other ThDP-dependent enzymes (Fig. 3*b*). However, there are no packing interactions in the crystals that might suggest a higher oligomeric form for KdcA. The crystal structure is thus consistent with size-exclusion chromatography studies indicating that KdcA is a dimer in solution (Gocke *et al.*, 2007), in contrast to most other members of this enzyme family. A comparison of the areas involved in the tetramer interface in IPDC with KdcA shows very low sequence conservation; many of the smaller hydrophobic residues in IPDC are replaced by glutamate or lysine residues in KdcA, with flexible charged side chains commonly being found on protein surfaces.

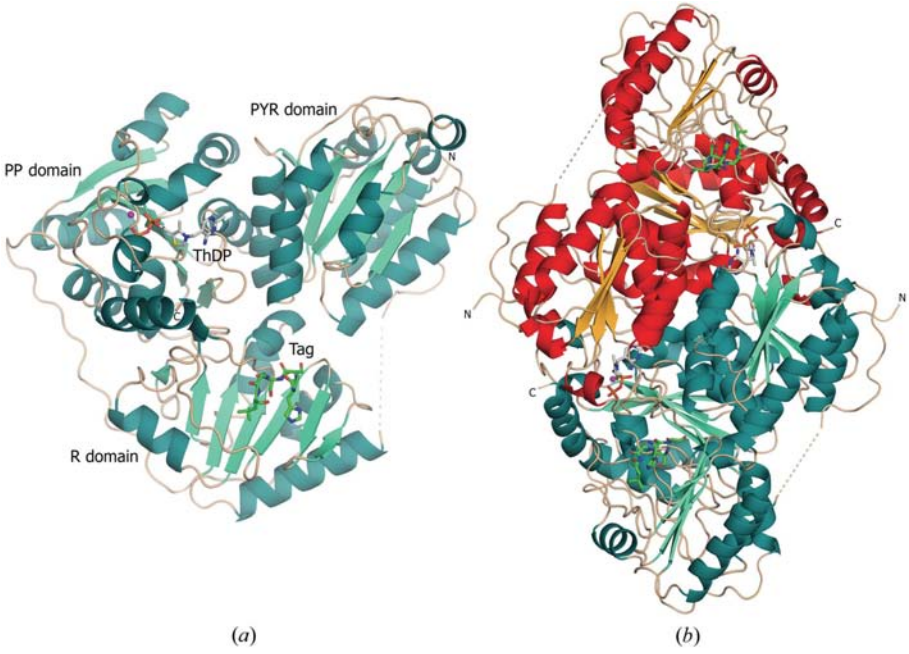

Figure 3
KdcA architecture: (*a*) monomer, (*b*) dimer. ThDP and the ordered N-terminal tag are shown as stick models with C atoms in grey and bright green, respectively. The Mg^{2+} ion is coloured magenta.

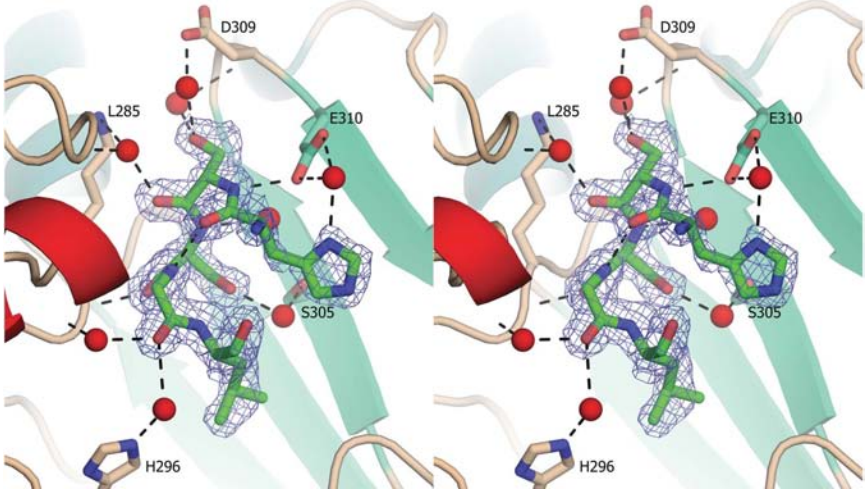

Figure 4
A stereoview of the interactions between the N-terminal tag and a crystallographically related molecule. Residues involved in hydrogen-bonding interactions with residues from the N-terminal tag are shown as stick models. All but two hydrogen bonds (indicated by dashed black lines) are mediated by ordered water molecules (red spheres). The $2F_o - F_c$ electron-density map is contoured at 1σ.

3.3. Active site

The active-site residues of KdcA are highly conserved with respect to IPDC, including the two catalytic histidines (His112 and His113), an aspartic acid (Asp26) and a glutamic acid (Glu462) positioned close to the thiazolium ring (KdcA numbering). The active-site pocket, which is lined by polar residues on one side and hydrophobic side chains on the other, only displays one distinct difference compared with the IPDC structure: a leucine residue in the latter is replaced by a phenylalanine (Phe542) in KdcA. Replacement by this bulkier amino acid results in closure of the entrance to the active site, which in the IPDC holo structure is open and connects the active site to the surface.

3.4. Inhibitor complex

2-[(*R*)-1-Hydroxyethyl]deazaThDP is an inhibitor of many ThDP-dependent enzymes and mimics the covalent reaction intermediate 2-[(*R*)-1-hydroxyethyl]ThDP that occurs in the decarboxylation of pyruvate (Leeper *et al.*, 2005). Binding of this deazaThDP derivative totally abolished the activity of KdcA as a consequence of the impaired ylid formation of this compound. We determined the structure of the complex of KdcA with this analogue of the reaction intermediate to 1.8 Å resolution in order to obtain a template that was more suitable than the holoenzyme

for modelling of substrate/intermediate complexes (Fig. 5). The overall structure of the enzyme in this complex is very similar to that of the holoenzyme with ThDP, with an r.m.s. deviation of 0.11 Å based on 541 C^{α} atoms. The only exceptions are the last residues of the C-terminal helix, which appear to be better ordered in the complex.

The position of the bound 2-[(*R*)-1-hydroxyethyl]deazaThDP ligand superimposes well with that of ThDP in the holoenzyme. The covalent hydroxyethyl moiety of the ligand extends into the active-site channel and the α-hydroxyl group forms a hydrogen bond to the 4′-amino group of the pyrimidine ring of ThDP. The geometry around the C^{α} atom of the hydroxyethyl moiety is slightly distorted from tetrahedral, with a smaller C2—C^{α}—OH angle (99–100°) and a larger C2—C^{α}—C^{β} angle (117–119°). Furthermore, the C^{α} atom of the ligand lies slightly out of the thiazolium plane (26°). The out-of-plane feature of the intermediate has previously been observed in freeze-trapping experiments in which substrates were trapped as postdecarboxylation intermediates in the decarboxylase subunit of the human α-ketoacid dehydrogenase complex (Machius *et al.*, 2006) and oxalyl-CoA decarboxylase (Berthold *et al.*, 2007).

3.5. Carboligation features and modelling of substrates

The carboligation reactions catalyzed to various extents by ThDP-dependent decarboxylases are of particular interest for biocatalytic purposes since they allow the chemoenzymatic formation of C—C bonds in a stereoselectively defined manner. The identification of the features of these enzymes that determine substrate selectivity and stereoselectivity is therefore essential if the substrate spectrum and enantioselectivity of these enzymes are to be modified by rational design. The structure of KdcA provides insights into the structural basis of several unique properties of the enzyme with respect to carboligation.

The position of the bound aldehyde substrate that also determines the stereochemistry of the product is determined by two factors. One is the catalytically relevant orientation of the aldehyde carbonyl group, which has to point towards catalytic residues implied in proton-transfer steps in ThDP-dependent decarboxylases such as *Zm*PDC (Schenk *et al.*, 1997; Dobritzsch *et al.*, 1998), BAL (Kneen *et al.*, 2005), IPDC (Schütz *et al.*, 2003) and BFD (Polovnikova *et al.*, 2003). In KdcA, His112 (corresponding to His115 in IPDC, His113 in *Zm*PDC and His70 in BFD) is structurally conserved and is most likely to have a similar

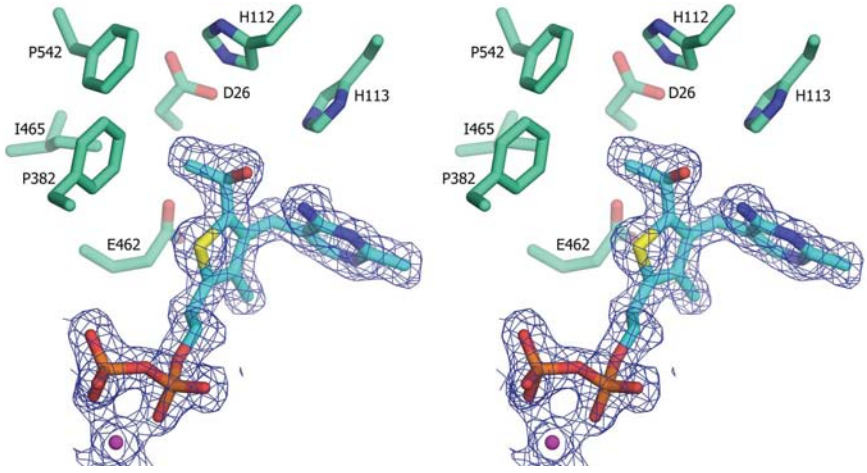

Figure 5
Part of the $2F_o - F_c$ electron-density map of the complex of KdcA with the bound analogue of the reaction intermediate 2-[(*R*)-1-hydroxyethyl]deazaThDP. The electron-density map at the positions of the bound ligand and Mg^{2+} ion is contoured at 1σ. Active-site residues in the proximity of the hydroxyethyl moiety are shown as stick models.

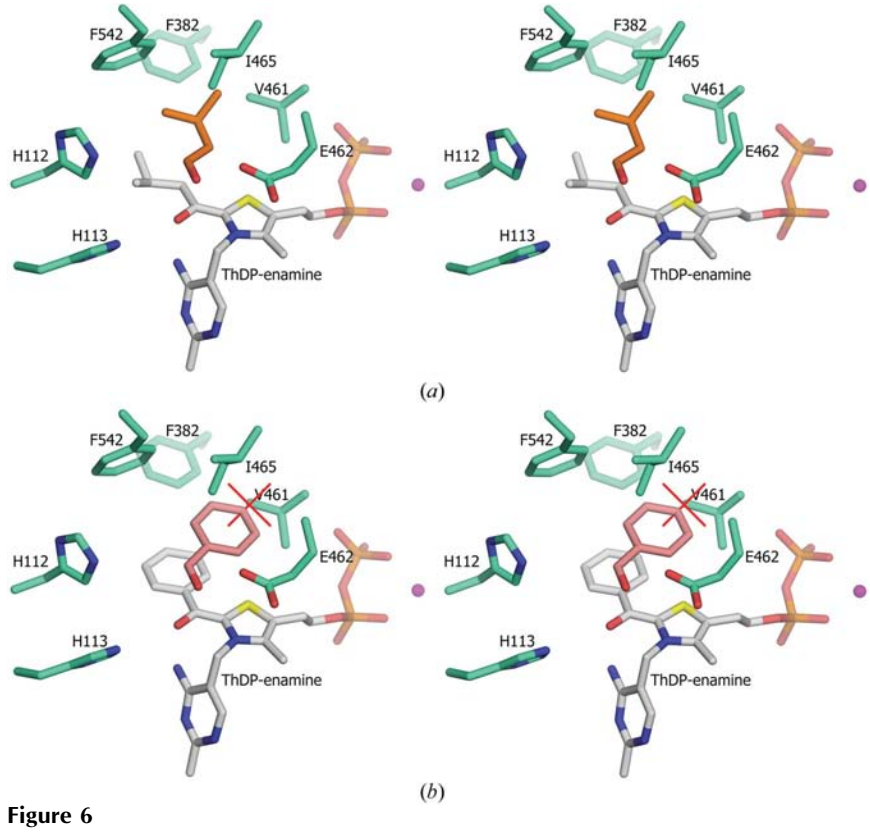

Figure 6
The S-pocket in KdcA controls enantioselectivity. (*a*) The self-ligation of isovaleraldehyde results in a slight excess of the (*S*) enantiomers (30–47%; Gocke *et al.*, 2007). Here, the acyl donor is modelled covalently bound to ThDP in the form of the enamine (grey) and the acceptor (orange) is fitted into the S-pocket with the *Re* face towards the enamine, resulting in the unusual (*S*)-enantiomer. (*b*) Self-ligation of benzaldehyde, on the other hand, results in pure (*R*)-benzoin. Modelling of the acceptor substrate (pink) into the S-pocket results in unavoidable clashes.

function in catalysis. Secondly, the aldehyde side chain can point (i) towards the substrate channel, resulting in an attack on the *Si* face of the carbonyl group (giving *R*-selectivity), or (ii) bind in the S-pocket, leading to attack on the *Re* face [*i.e.* (*S*)-enantiomer] (Fig. 2). Comparison of the S-pockets of various ThDP-dependent enzymes, defined in KdcA by residues Phe542, Ile465, Glu462 and Val461, and their accessibilities shows that KdcA has an average-sized S-pocket. BAL, on the other hand, has no S-pocket, BFD has a small pocket with a large entrance (substitution of Val461 in KdcA by alanine) and *Zm*PDC has an S-pocket about as large as KdcA, but the entrance is closed by a bulky isoleucine residue (Ile472). For *Zm*PDC it was shown that by replacing Ile472 at the entrance by site-directed mutagenesis the S-pocket could be made more accessible and the enantioselectivity was shifted from (*R*) to (*S*) (Siegert *et al.*, 2005). In a similar way, BFD was engineered to increase the size of the S-pocket, which is only accessible to acetaldehyde in the wild-type enzyme. The resulting BFD variants prefer propanal over acetaldehyde and produce (*S*)-2-hydroxy-1-phenylbutan-1-one (Gocke *et al.*, unpublished results). In KdcA, isovaleraldehyde is the only substrate for which the (*S*)-enantiomer of the corresponding 2-hydroxyketone can be produced to some extent (Gocke *et al.*, 2007). Modelling shows that the side chain of this substrate fits well into the S-pocket of the enzyme with a number of hydrophobic interactions (Fig. 6*a*), whereas the pocket is restricted in size owing to residues Ile465, Val461 and Phe542, which would clash with larger substrates such as benzaldehyde (Fig. 6*b*).

The substrate channel in KdcA, which is lined by Phe542 and Met538, is narrower than the wide and straight BFD channel but larger than that in *Zm*PDC, where a tryptophan residue (Trp392) limits the access to the active site. The small donor pocket of *Zm*PDC is known to optimally fit an acetaldehyde molecule, while KdcA on the other hand can adopt larger donor aldehydes (Gocke *et al.*, 2007). Indole-3-pyruvate, which after decarboxylation KdcA can ligate to acetaldehyde, was fitted into the structure of the holoenzyme as a model of the Michaelis complex. The substrate could be modelled straightforwardly such that the substrate C^{α} atom was aligned for nucleophilic attack by the cofactor, while at the same time favourable stacking interactions of the indole moiety were made to the side chains of Phe542 and Phe382 (Fig. 7). The shape of the substrate-binding pocket allows an optimal fit of the indole moiety, which makes KdcA the only biocatalyst identified so far that accepts substrates as large as indole-3-acetaldehyde in carboligation reactions.

KdcA displays the intriguing feature that aliphatic substrates larger than acetaldehyde only act as acyl donors when ligated with benzaldehyde as acceptor. In ligation reactions with acetaldehyde, however, benzaldehyde acts as an acyl donor in about 60% of the cases, whereas the larger substituted benzaldehyde 3,5-dichlorobenzaldehyde can only act as an acyl donor (Gocke *et al.*, 2007). These features of the carboligation reaction involving benzaldehyde and benzaldehyde derivatives can be explained by the structure of KdcA. Fig. 8 shows a model of the covalent ThDP–benzaldehyde adduct with a second benzaldehyde molecule bound in the acceptor position. The stacking interactions between the benzene rings favour binding of the second benzaldehyde molecule, *i.e.* the carboligation reaction between two benzaldehyde molecules yielding benzoin. Binding of a substituted benzene ring in the acceptor position, however, would result in several unfavorable interactions with surrounding hydrophobic residues, which prevents these compounds from acting as acceptors in the carboligation reaction. Indeed, the formation of benzoin derivatives from substituted benzaldehydes has not been observed with KdcA. In mixed carboligations with aliphatic and aromatic aldehydes, the nonplanarity of the aliphatic chains would result in clashes with either the donor benzene ring and/or the two bulky phenylalanine residues 382 and 542 and thus disfavour carboligation with aliphatic aldehydes as acceptors. Larger aliphatic aldehydes are therefore predominantly bound as acyl donors with benzaldehyde as acceptor.

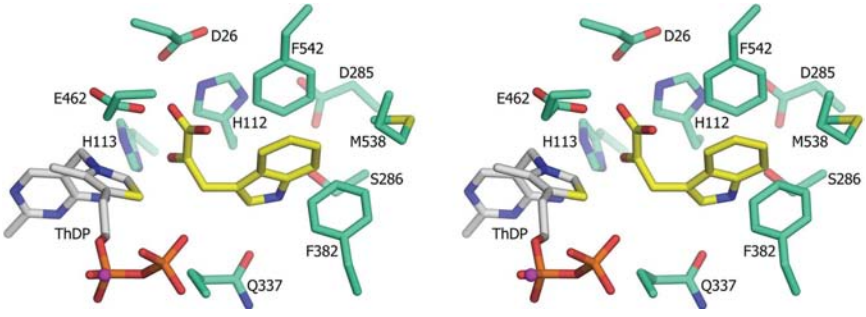

Figure 7
Structural basis of chemoselectivity. Stereoview of the Michaelis complex of KdcA with the bound donor substrate indole-3-pyruvate. The substrate-binding pocket allows an optimal fit of the indole moiety.

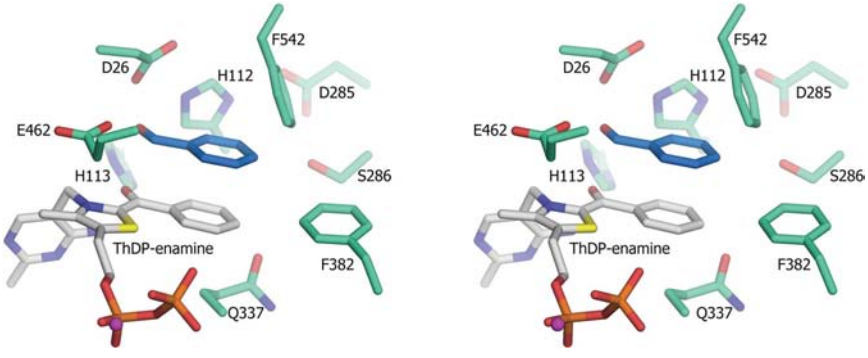

Figure 8
Benzaldehyde modelled as acyl donor aldehyde covalently bound to ThDP in the form of the enamine (grey) and as acceptor substrate (blue) with the *Si* side facing the enamine.

We gratefully acknowledge access to synchrotron radiation at the European Synchrotron Research Facility, France. This work was supported by the Swedish Science Council. MDW was supported by a studentship from the BBRSC and Syngenta.

References

Berthold, C. L., Moussatche, P., Richards, N. G. & Lindqvist, Y. (2005). *J. Biol. Chem.* **280**, 41645–41654.

Berthold, C. L., Toyota, C. G., Moussatche, P., Wood, M. D., Leeper, F., Richards, N. G. & Lindqvist, Y. (2007). *Structure*, **15**, 853–861.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Dobritzsch, D., König, S., Schneider, G. & Lu, G. (1998). *J. Biol. Chem.* **273**, 20196–20204.

Duggleby, R. G. (2006). *Acc. Chem. Res.* **39**, 550–557.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Gocke, D., Nguyen, C., Pohl, M., Stillger, T., Walter, L. & Müller, M. (2007). *Adv. Synth. Catal.* **349**, 1425–1435.

Hasson, M. S., Muscate, A., McLeish, M. J., Polovnikova, L. S., Gerlt, J. A., Kenyon, G. L., Petsko, G. A. & Ringe, D. (1998). *Biochemistry*, **37**, 9918–9930.

Iding, H., Dünnwald, T., Greiner, L., Liese, A., Müller, M., Siegert, P., Grötzinger, J., Demir, A. S. & Pohl, M. (2000). *Chemistry*, **6**, 1483–1495.

Janzen, E., Müller, M., Kolter-Jung, D., Kneen, M. M., McLeish, M. J. & Pohl, M. (2006). *Bioorg. Chem.* **34**, 345–361.

Kern, D., Kern, G., Neef, H., Tittmann, K., Killenberg-Jabs, M., Wikner, C., Schneider, G. & Hübner, G. (1997). *Science*, **275**, 67–70.

Kneen, M. M., Pogozheva, I. D., Kenyon, G. L. & McLeish, M. J. (2005). *Biochim. Biophys. Acta*, **1753**, 263–271.

Knoll, M., Müller, M., Pleiss, J. & Pohl, M. (2006). *ChemBioChem*, **7**, 1928–1934.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Leeper, F. J., Hawksley, D., Mann, S., Perez Melero, C. & Wood, M. D. (2005). *Biochem. Soc. Trans.* **33**, 772–775.

Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.

Lindqvist, Y. & Schneider, G. (1993). *Curr. Opin. Struct. Biol.* **3**, 896–901.

Lobell, M. & Crout, D. H. G. (1996). *J. Am. Chem. Soc.* **118**, 1867–1873.

Machius, M., Wynn, R. M., Chuang, J. L., Li, J., Kluger, R., Yu, D., Tomchick, D. R., Brautigam, C. A. & Chuang, D. T. (2006). *Structure*, **14**, 287–298.

Mosbacher, T. G., Müller, M. & Schulz, G. E. (2005). *FEBS J.* **272**, 6067–6076.

Müller, M. & Sprenger, G. A. (2004). *Thiamine: Catalytic Mechanisms in Normal and Disease States*, edited by F. Jordan & M. S. Patel, pp. 77–91. New York: Marcel Dekker.

Muller, Y. A., Lindqvist, Y., Furey, W., Schulz, G. E., Jordan, F. & Schneider, G. (1993). *Structure*, **1**, 95–103.

Muller, Y. A. & Schulz, G. E. (1993). *Science*, **259**, 965–967.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

Pang, S. S., Duggleby, R. G. & Guddat, L. W. (2002). *J. Mol. Biol.* **317**, 249–262.

Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.

Polovnikova, E. S., McLeish, M. J., Sergienko, E. A., Burgner, J. T., Anderson, N. L., Bera, A. K., Jordan, F., Kenyon, G. L. & Hasson, M. S. (2003). *Biochemistry*, **42**, 1820–1830.

Schenk, G., Leeper, F. J., England, R., Nixon, P. F. & Duggleby, R. G. (1997). *Eur. J. Biochem.* **248**, 63–71.

Schüttelkopf, A. W. & van Aalten, D. M. F. (2004). *Acta Cryst.* D**60**, 1355–1363.

Schütz, A., Sandalova, T., Ricagno, S., Hübner, G., König, S. & Schneider, G. (2003). *Eur. J. Biochem.* **270**, 2312–2321.

Siegert, P., McLeish, M. J., Baumann, M., Iding, H., Kneen, M. M., Kenyon, G. L. & Pohl, M. (2005). *Protein Eng. Des. Sel.* **18**, 345–357.

Smit, B. A., van Hylckama Vlieg, J. E., Engels, W. J., Meijer, L., Wouters, J. T. & Smit, G. (2005). *Appl. Environ. Microbiol.* **71**, 303–311.

Sprenger, G. A. & Pohl, M. (1999). *J. Mol. Catal. B*, **6**, 145–159.

Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.

Yep, A., Kenyon, G. L. & McLeish, M. J. (2006). *Bioorg. Chem.* **34**, 325–336.