ChemPubSoc Europe

DOI: 10.1002/cctc.201500014

CHEMCATCHEM Full Papers

Genome Mining for Innovative Biocatalysts: New Dihydroxyacetone Aldolases for the Chemist's Toolbox

Christine Guérard-Hélaine,[a, b] Véronique de Berardinis,[c, d, e] Marielle Besnard-Gonnet,[c, d, e] Ekaterina Darii,[c, d, e] Marine Debacker,[a, b] Adrien Debard,[c, d, e] Carlos Fernandes,[a, b] Virgil Hélaine,[a, b] Aline Mariage,[c, d, e] Virginie Pellouin,[c, d, e] Alain Perret,[c, d, e] Jean-Louis Petit,[c, d, e] Martine Sancelme,[a, b] Marielle Lemaire,*[a, b] and Marcel Salanoubat*[c, d, e]

Stereoselective carboligating enzymes were discovered by a genome mining approach to extend the biocatalysis toolbox. Seven hundred enzymes were selected by sequence comparison from diverse prokaryotic species as representatives of the aldolase (FSA) family diversity. The aldol reaction tested involved dihydroxyacetone (DHA) and glyceraldehyde-3-phosphate. The hexose-6-phosphate formation was monitored by mass spectrometry. Eighteen enzymes annotated either as transaldolases or aldolases were found to exhibit a DHA aldolase activity. Remarkably, six of them proven as aldolases, and not transaldolases, shared very limited similarities with those currently described. Multiple sequence alignment performed on all enzymes revealed a Tyr in the new DHA aldolases as found in FSA_{coli} instead of a Phe usually found in transaldolases. Four of these DHA aldolases were biochemically characterised in comparison with FSA_{coli}. In particular, an aldolase from *Listeria monocytogenes* exhibited interesting catalytic properties.

Introduction

The need for innovative biocatalysts with improved or unnatural properties is still crucial thanks to the great potential of enzymes with synthetically useful properties. The bottleneck of the biocatalytic processes is often limited by the availability of suitable enzymes.[1] To meet this requirement, three main strategies are accessible: i) evolution of known enzymes,[2] ii) de novo design of new biocatalysts by computational design,[3] iii) search for new enzymes in the biodiversity.[2a, 3a, 4] Owing to the tremendous microbial biodiversity, it is highly probable that many valuable activities have still to be found among wild-type enzymes that have evolved over billions of years.[5] Aldolases belong to a well-established class of enzymes catalysing one of the most important reactions in organic synthesis, stereoselective carbon–carbon ligation. They proved to be attractive and powerful for the synthesis of natural products and complex bioactive small molecules.[6] Despite their broad tolerance to the aldehyde acceptor (electrophile), their use for synthesis is limited by the narrow ketone donor (nucleophile) specificity. Thus there is a real need to find new aldolases able to display broader donor specificities, in particular, for nonphosphorylated substrates to increase the chemist's toolbox of versatile biocatalysts.[6c, g] The discovery of fructose-6-phosphate aldolases ($FSAA_{coli}$ and $FSAB_{coli}$) in *Escherichia coli* has opened up exciting new perspectives.[7] Indeed, $FSAs_{coli}$ possesses the unique capacity to react efficiently with at least four different non-phosphorylated donor substrates (dihydroxyacetone, DHA; hydroxyacetone, HA; hydroxybutanone and glycolaldehyde) and various acceptors.[7a, 8] Given the extraordinary potential of FSAs, using wild-type or variants, many syntheses were described to prepare especially antidiabetic (glycosidase inhibitors), anti-infective compounds or phosphorylated sugars as metabolites.[6c, 9] Since their discovery in 2001, no new natural efficient DHA aldolases have been identified[6g] from other microorganisms unlike most other aldolases. However, *E. coli* transaldolase B ($TalB_{coli}$) and DHAP-dependent rhamnulose aldolase display a slight aldol promiscuous activity towards DHA that has been improved by mutagenesis.[10]

Here we used a universal large-scale LC–MS screening method to prospect among different enzyme families for DHA aldolases using D,L-glyceraldehyde 3-phosphate (D,L-G3P) as

[a] Dr. C. Guérard-Hélaine,[+] M. Debacker, Dr. C. Fernandes, Dr. V. Hélaine, M. Sancelme, Prof. M. Lemaire[+]
Clermont Université, Université Blaise Pascal
ICCF, BP 10448, 63000 Clermont-Ferrand (France)
E-mail: marielle.lemaire@univ-bpclermont.fr

[b] Dr. C. Guérard-Hélaine,[+] M. Debacker, Dr. C. Fernandes, Dr. V. Hélaine, M. Sancelme, Prof. M. Lemaire[+]
CNRS, UMR 6296,BP 8002663177 Aubière (France)

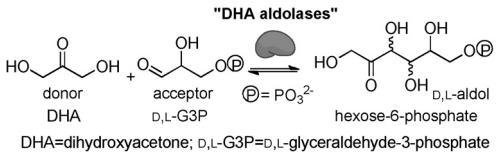
[c] Dr. V. de Berardinis,[+] M. Besnard-Gonnet, Dr. E. Darii, A. Debard, A. Mariage, V. Pellouin, Dr. A. Perret, J.-L. Petit, Dr. M. Salanoubat[+]
CEA, DSV, IG, Genoscope
2 rue Gaston Crémieux, 91057 Evry (France)
E-mail: salanou@genoscope.cns.fr

[d] Dr. V. de Berardinis,[+] M. Besnard-Gonnet, Dr. E. Darii, A. Debard, A. Mariage, V. Pellouin, Dr. A. Perret, J.-L. Petit, Dr. M. Salanoubat[+]
CNRS-UMR 8030
2 rue Gaston Crémieux, 91057 Evry (France)

[e] Dr. V. de Berardinis,[+] M. Besnard-Gonnet, Dr. E. Darii, A. Debard, A. Mariage, V. Pellouin, Dr. A. Perret, J.-L. Petit, Dr. M. Salanoubat[+]
Université d'Evry Val d'Essonne
boulevard François Mitterrand, 91057 Evry (France)

[+] These authors contributed equally to this work.

Supporting Information for this article is available on the WWW under http://dx.doi.org/10.1002/cctc.201500014.



Scheme 1. Aldolisation catalysed by DHA aldolases.

the acceptor substrate (Scheme 1). In this study, to explore the DHA aldolase families, we applied a sequence-driven approach that successfully found new nitrilases from the natural diversity.[11] To our knowledge no such large genome mining approach for new aldolases discovery have been described in literature.

Results and Discussion

Enzyme selection and preliminary characterisation

We reasoned that natural promiscuous catalytic properties are widespread in several enzyme families. Thus sequence comparisons against UniprotKB proteins with already known aldolases (EC 4.1.2 with the exception PLP dependent glycine aldolases), including FSA from *E. coli* and transaldolases (EC 2.2.1), were conducted and allowed selection of 1148 candidate proteins from 313 different prokaryote species representative of the aldolase/transaldolase diversity. Among the candidate proteins, 731 genes (64%) from approximately 230 species were successfully cloned in an expression vector using automated platform facilities. Out of these, overexpression of 571 proteins (79%) were visible on sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE). The sequences of nine new DHA aldolases and five new transaldolases have been submitted to European Nucleotide Archive and UniprotKB Id are reported in the Supporting Information Table 2b.

Proteins of interest were identified by LC–MS by monitoring hexulose-6-phosphate formation produced by the aldol addition of dihydroxyacetone (DHA) as the donor and D,L-G3P as the acceptor (Figure 1 (1)).

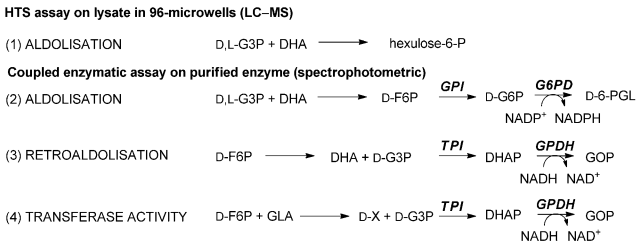


Figure 1. High-throughput screening (HTS) methodology and initial catalytic activity characterisation. G3P = glyceraldehyde-3-phosphate, F6P = fructose-6-phosphate, DHA = dihydroxyacetone, DHAP = dihydroxyacetone phosphate, GOP = glycerolphosphate, G6P = glucose-6-phosphate, 6-PGL = 6-phosphonoglucono-1,5-lactone, X = xylulose, GLA = glycolaldehyde, TPI = triose phosphate isomerase, GPDH = glycerol-3-phosphate dehydrogenase, GPI = glucose-6-phosphate isomerase, G6PD = glucose-6-phosphate dehydrogenase

On each microtiterplate, *E. coli* cell lysate was used as a negative control with $FSAA_{coli}$ as a positive control. Hits were considered as positive when the signal was at least three-fold higher than the negative control. This methodology led to the selection of 19 positive hits (Supporting Information, Table 5). To confirm these results, the corresponding genes were cloned with a His-tag and 17 enzymes were successfully purified by nickel affinity chromatography (Supporting Information, Table 5). Among the two unsuccessful purifications, A0A0E4G3U3 ($ALD_{listeria}$), one of the strongest hits, was purified in its native form (untagged) by fast protein liquid chromatography (FPLC). The formation of a hexulose-6-phosphate was confirmed by LC–MS analysis for these 18 purified enzymes. Among them, only two were annotated as aldolases (FSA), whereas the other 16 were annotated as transaldolases in UniprotKB database (Supporting InformationTable 5). To analyse these two activities further, known spectrophotometric assays were applied as described in Figure 1 (2)–(4). For aldolase activity, fructose-6-phosphate (F6P) formation was monitored by following NADPH appearance (Figure 1 (2)). Retroaldolisation of D-F6P was also assayed through D-G3P formation by monitoring NADH disappearance (Figure 1, entry 3).[7a] $FSAA_{coli}$ was used as a positive control. Transaldolase activity was assayed by using F6P and glycolaldehyde as an acceptor. D-G3P formation was monitored as for retroaldolase activity assay (Figure 1 (4)). $TalB_{coli}$ was used as a positive control.

DHA aldolase activity was clearly detected for the two annotated as FSA and for six enzymes annotated as transaldolases. The transaldolase activity was confirmed for 10 out of the 16 putative transaldolases. The six enzymes annotated as "transaldolases" active for both aldolisation and retroaldolisation reactions did not display any transaldolase activity (Table 1). This is corroborated by amino acid sequence comparisons that show that these enzymes revealed low sequence identity (<30% identity) with various transaldolases but display a higher sequence similarity with $FSAA_{coli}$ (<41% identity, Supporting Information, Table 6). Five among these six aldolases are from various *Streptococcus* species and shared sequence identities between 60% and 79%. Interestingly, $ALD_{listeria}$ from *Listeria monocytogenes* that seemed to be the most efficient enzyme during the preliminary studies shares no sequence identity with $FSAA_{coli}$. The family of transaldolases and related aldolases are divided into five subfamilies. Subfamily 1 contains the "classical" transaldolases, whereas subfamily 5 contains FSAs from *E. coli*.[12] FSAs are specified by the presence of the three signatures from the protein sequence analysis and classification Interpro database: IPR001585 corresponding to transaldolases and related aldolases families, IPR04731 and finally IPR023001 (see Supporting Information Figure 1). Only B7NAG9 corresponded to this context, the seven other DHA aldolases simply shared the IPR001585 (See Supporting Information, Table 7).

A multiple sequence alignment including these enzymes and $FSAA_{coli}$ and $TalB_{coli}$ as references demonstrate that essential catalytic and substrate binding residues[10a,13] are highly conserved in these eight aldolases despite a low level of sequence conservation with $FSAA_{coli}$ for seven of them (Figure 2).

Table 1. Experimental validation of the 18 new aldolases/transaldolases performed on purified enzymes. Enzymes from *E. coli* strain K12 are used as references.

UniprotKB id	Annotation	Genome	Aldolase activity	Retroaldolisation activity	Transferase activity
P78055[a]	FSA A[b]	*E. coli (strain K12)*	+	+	–
B7NAG9	FSA 1[b]	*E. coli O17:K52:H18*	+	+	–
A0A0D6J3Z8	FSA[b]	*Streptococcus pneumoniae*	+	+	–
A0A0E4G3U3 ($ALD_{listeria}$)	transaldolase	*Listeria monocytogenes serotype 4b str. H7858*	+	+	–
A0A0E4C393	transaldolase	*Streptococcus suis*	+	+	–
ALD_{Smut}	transaldolase	*Streptococcus mutans serotype c*	+	+	–
A0A0D6H018 (ALD_{Spyog})	transaldolase	*Streptococcus pyogenes serotype M1*	+	+	–
A0A0E4C363 (ALD_{Sgord})	transaldolase	*Streptococcus gordonii*	+	+	–
Q8E738	transaldolase family protein	*Streptococcus agalactiae*	+	+	–
P0A870[a]	transaldolase B	*Escherichia coli (strain K12)*	–	–	+
Q7M7U6	transaldolase	*Wolinella succinogenes*	–	–	+
Q4JVK0	transaldolase	*Corynebacterium jeikeium*	–	–	+
A0A0E3VZZ2	transaldolase	*Shewanella baltica*	–	–	+
A0A0E4C370	transaldolase	*Pseudomonas mendocina*	–	–	+
Q3IG49	transaldolase	*Pseudoalteromonas haloplanktis*	–	–	+
Q8NQ64	transaldolase	*Corynebacterium glutamicum*	–	–	+
A9E0W9	probable transaldolase	*Oceanibulbus indolifex*	–	–	+
A1A1M8	transaldolase	*Bifidobacterium adolescentis*	–	–	+
A0A0E4C394	transaldolase	*Neisseria meningitidis*	–	–	+
Q1H0R4	transaldolase	*Methylobacillus flagellatus*	–	–	+

[a] Known enzymes served as references; [b] FSA = fructose-6-phosphate aldolase.

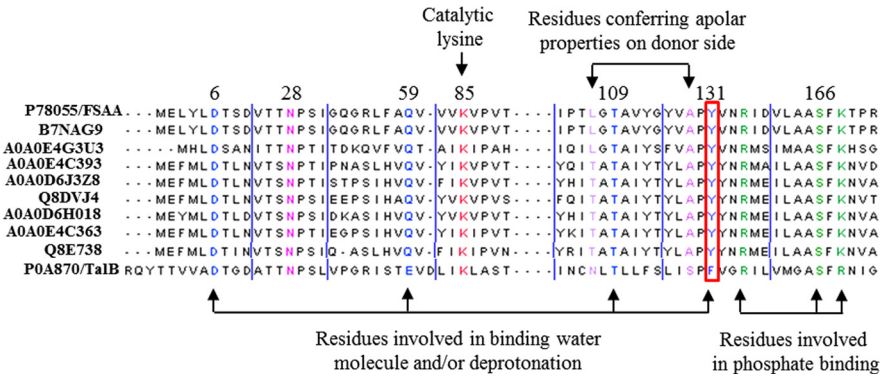


Figure 2. Specific residues identified in the active site. Numbers correspond to $FSAA_{coli}$ sequence. Residues involved in binding water molecules and/or deprotonation are shown in blue. Residues conferring apolar properties on the donor side are shown in purple. Residues involved in phosphate binding are shown in green. The catalytic lysine is highlighted in red. A 3D structure of $FSAA_{coli}$ with these highlighted conserved residues was proposed (Supporting Information, Figure 3).

Interestingly, the eight DHA aldolases have a tyrosine as in $FSAA_{coli}$ (Tyr 131) instead of a phenylalanine usually found in transaldolases.[14] In $FSAA_{coli}$, this tyrosine contributes to the aldolase activity by binding a water molecule together with Gln 59 and Thr 109 in the active site.[13] To reinforce the correlation between the presence of the tyrosine residue and the aldolase activity, we purified the 11 annotated transaldolases from the collection for which LC–MS screening results did not allow their initial selection. Multiple sequence alignment revealed that four of them have a tyrosine and one from *Microscilla marina* (A1ZYC6) a leucine instead of a phenylalanine (Supporting Information, Figure 2). These four enzymes (C2DGL7, A0A0E4G3V8, A0A0E4C372, Q8Y990) with a tyrosine were originated from phylogenetically diverse organisms (*Aeromonas*, *Enterococcus*, *Clostridium* etc.) extending the diversity of organism far beyond the *Listeria* and the *Streptococcus* genera (Supporting Information, Table 8).

Biochemical characterization of these 11 enzymes confirmed that the presence of the tyrosine is related to DHA aldolase activity and that of the phenylalanine to transaldolase activity (Supporting Information, Table 9). For C2DGL7 and A0A0E4G3V8, no expression could be detected on SDS PAGE. Notably, no significant activity could be detected for A1ZYC6 harbouring a leucine in the same position despite a good overexpression.

As illustrated by the phylogenetic tree (Figure 3), new transaldolases characterised in this study were found in all the different subfamilies of already described transaldolases except in subfamily 2 which is composed of transaldolases of higher plants.[14] We suggest that the new DHA aldolases could be either grouped in subfamily 5 or comprise a new subfamily since they did not share any of the two FSA Interpro signature (Figure 3, in squared dash points).

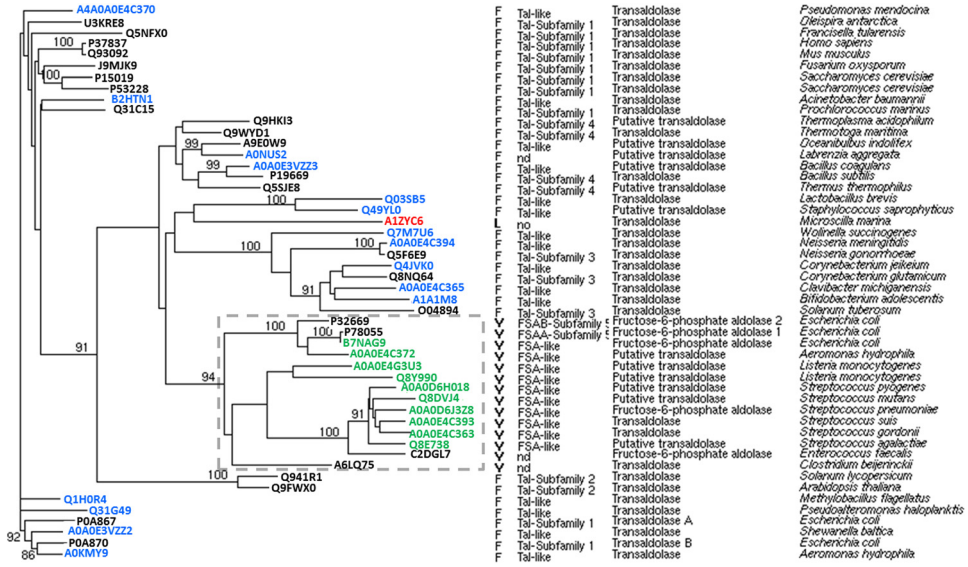


Figure 3. Phylogenetic tree including reference enzymes of each transaldolase subfamily (in black)[14] with biochemical data or 3D structures with the enzymes from our study shown in colour (blue: transaldolase activity; green: aldolase activity; red: no activity). Column 1 indicates the sequence variation corresponding to Tyr131 in FSA; column 2 indicates the respective family, if applicable, or the activity determined in this study (FSA-like or transaldolase-like); column 3 and 4 indicate the annotation in databases according to sequence similarities and the source organism. nd: not done.

Genes encoding transaldolase-like proteins have been found associated with catabolic operons in some organisms, but their functions remain unknown.[15] An analysis of the genomic context of the new DHA aldolases (Supporting Information, Table 10) revealed that a large majority (nine enzymes) are included in a gene cluster with predicted genes involved in sugar transport systems such as $FSAB_{coli}$. The bacterial sugar phosphotransferase system (PTS) catalyses the transfer of the phosphoryl group from phosphoenolpyruvate to its sugar substrates concomitant with the translocation of these sugars across the bacterial membrane. The DHA aldolases are generally annotated as Tal-like proteins in these clusters and many of them are flanked by a pyruvate formate lyase and a glycerol dehydrogenase (Supporting Information, Table 10).

Characterization of the most promising new DHA aldolases

A biochemical characterisation of four aldolases was then conducted to evaluate their biocatalytic capacity (A0A0D6H018-from *S. pyogenes* (ALD_{Spyog}), A0A0E4C363 from *S. gordonii* (ALD_{Sgord}), Q8DVJ4 from *S. mutans* (ALD_{Smut})and $ALD_{listeria}$). Kinetic parameters were determined and compared to those of $FSAA_{coli}$. Specificity spectra towards the acceptor substrates were studied by varying their hydrophobicity, polarity and chain length. These assays were conducted on purified enzymes (see Experimental Section).

First we determined the stereoselectivity of these enzymes. The spectrophotometric assays based on the retroaldolisation of F6P (Figure 1) indicated that the (3*S*, 4*R*) configuration could be processed but did not exclude the processing of the three others. To analyse this further, small-scale reactions were performed in the same conditions as for the four enzymes by using hydroxyacetone (HA) and D-G3P, which are the best substrates of $FSAA_{coli}$.[7b,9d] By analogy with $FSAA_{coli}$, which is highly stereoselective, the expected product should be exclusively 1-deoxy-D-F6P (d-D-F6P). The aldols were isolated as their barium salts and studied by ^{13}C NMR spectroscopy. In all cases, only one compound was detected and identified as d-D-F6P. Thus, the new aldolases displayed a high stereoselectivity leading to the (3*S*, 4*R*) adduct. With this information in hand, deeper biochemical investigations were conducted.

The specific activities for the retroaldol reaction were determined on purified enzymes (Supporting Information, Table 12). $ALD_{listeria}$ displayed the highest specific activity, 5-fold higher than the one found for $FSAA_{coli}$. Besides this result, ALD_{Sgord} and ALD_{Smut} were found 3-fold more active than $FSAA_{coli}$.

$FSAA_{coli}$ was found to be particularly thermostable even if not originating from a thermophilic microorganism. To continue the comparison, thermostability properties of these four enzymes were studied. The activity of all enzymes but one (ALD_{Smut}) remained unchanged after a 40 min incubation at 70 °C, which is similar to that of $FSAA_{coli}$ (Supporting Information, Figure 3). Thus this property could be exploited to simplify the purification process of these three enzymes. Moreover, despite its presence in the *E. coli* genome, $FSAA_{coli}$, is not expressed under usual growth conditions and will not contaminate the heat-stable overexpressed enzymes.[7b]

The pH profiles of these enzymes were revealed to be relative similar in between and with $FSAs_{coli}$, showing an optimum at 8.5. Remarkably, ALD_{Sgord} , and ALD_{Spyog} retained 97 and 90% of their respective activity when lowering the pH to 7. This result could be exploited to conduct aldol reactions in the presence of acceptors sensitive to alkaline pH. Five buffers (phosphate; *N*-(2-hydroxyethyl)piperazine-*N*′-propanesulfonic acid, HEPES; tetraethylammonium, TEA; tris(hydroxymethyl)-aminomethane, Tris and glycylglycine, glygly) were then tested, and no significant difference was revealed except for ALD_{Smut} for which 50% of activity was lost in the presence of HEPES.

Steady-state kinetic parameters were determined for the four aldolases by using $FSAA_{coli}$ as a reference. The results are presented in Table 2. The retroaldol reaction was tested on three substrates. Among them, d-D-F6P was revealed as the best substrate for all the aldolases, in particular for $FSAA_{coli}$. $ALD_{listeria}$ was the more efficient on D-F6P and D-arabinose-5-phosphate (D-A5P), followed by ALD_{Sgord}. Interestingly, the result that D-A5P was found as a substrate suggests that glycolalde-

Table 2. Kinetic parameters determination for various substrates at pH 8.0 in 50 mM glycylglycine buffer (k_{cat} calculated for one active site)

Substrate	Kinetic parameters	$FSAA_{coli}$	ALD_{Sgord}	ALD_{Smut}	ALD_{Spyog}	$ALD_{listeria}$ [b]
D-F6P[a]	K_m [mM]	19±1	28±1	39±6	18±2	8.8±0.4
	k_{cat} [s^{-1}]	0.60±0.01	4.1±0.2	3.1±0.5	1.2±0.1	2.3±0.1
	k_{cat}/K_m [s^{-1} mM^{-1}]	0.32±0.002	0.14±0.01	0.08±0.02	0.06±0.01	0.26±0.02
d-D-F6P[a]	K_m [mM]	0.28±0.02	0.97±0.06	0.67±0.07	0.39±0.02	0.56±0.04
	k_{cat} [s^{-1}]	25±1	32±1	12±1	10±1	17±1
	k_{cat}/K_m [s^{-1} mM^{-1}]	89±6	33±2	18±2	26±2	30±2
D-A5P[a]	K_m [mM]	0.40±0.05	0.29±0.01	0.41±0.03	0.12±0.01	0.05±0.001
	k_{cat} [s^{-1}]	0.43±0.05	0.55±0.02	0.40±0.03	0.25±0.01	0.46±0.01
	k_{cat}/K_m [s^{-1} mM^{-1}]	1.1±0.2	1.9±0.1	1.0±0.1	2.1±0.1	9.2±0.2
D,L-G3P[a]	K_m [mM]	0.5±0.1	0.23±0.03	0.54±0.09	0.49±0.07	0.12±0.01
	k_{cat} [s^{-1}]	6±2	19±2	11±2	13±2	13±1
	k_{cat}/K_m [s^{-1} mM^{-1}]	12±5	80±10	20±5	27±5	108±10
DHA[a]	K_m [mM]	16±1	50±5	27±2	24±2	64±3
	k_{cat} [s^{-1}]	6.9±0.1	18±2	6.6±0.4	8.2±0.5	22±1
	k_{cat}/K_m [s^{-1} mM^{-1}]	0.43±0.01	0.36±0.05	0.24±0.02	0.34±0.03	0.34±0.02

[a] F6P = fructose-6-phosphate, G3P = glyceraldehyde-3-phosphate, d-D-F6P = 1-deoxy-D-fructose-6-phosphate, A5P = arabinose-5-phosphate, DHA = dihydroxyacetone. [b] Native protein purified by FPLC.

hyde is an acceptor of the aldol reaction (retroaldolisation reaction lead to glycolaldehyde). This activity is exceptional among aldolase families and until present was only demonstrated for $FSAs_{coli}$.

In addition, all the enzymes exhibited similar specificity for DHA. As the sole acceptor assayed, D,L-G3P was revealed as an excellent acceptor substrate, particularly for $ALD_{listeria}$ and ALD_{Sgord}. $ALD_{listeria}$ showed the highest catalytic efficiency for the aldol reaction with G3P (9-fold higher than $FSAA_{coli}$, the lowest one).

These results highlight that the broad tolerance towards various substrates described for $FSAA_{coli}$ is also a property of these new aldolases. $ALD_{listeria}$ and ALD_{Sgord} are the most promising enzymes for biocatalysis applications.

To evaluate each aldolase's synthetic potential, various acceptor substrates were reacted with HA, which is the best donor for FSA_{coli}. The reactions were monitored by thin layer chromatography (TLC) over a period of 24 h.

Reaction conditions were standardised[16] for all the enzymes to compare the results (see Experimental Section). In the case of phosphorylated acceptors, two TLCs were performed either in a very polar eluent adapted to phosphorylated compounds or in a classical one to follow HA disappearance. Depending on the enzyme assayed, this methodology enabled classification of the substrates into four categories: high, medium, low, and no substrate. The aldehydes were chosen according to their hydrophobicity (benzaldehyde, pentanal) or hydrophilicity (phosphorylated or polyhydroxylated compounds) and carbon-chain length. Results are summarised in Table 3. The five enzymes accept alkyl aldehydes as substrates except benzaldehyde. The smallest one, formaldehyde, was fully converted in 24 h with all tested enzymes. $FSAA_{coli}$ and $ALD_{listeria}$ gave the highest conversion rates. Pentanal was partially converted after 24 h incubation by all enzymes except $FSAA_{coli}$ for which a total reaction was observed. Concerning the aldose substrates, the increase of carbon-chain length reduced the conversion rates. Thus, glyceraldehyde was found to be the best acceptor followed by erythrose, which was a moderate substrate for ALD_{Smut} and ALD_{Spyog}. Aldehydes with more than four carbons were no acceptor. However, the C5 phosphorylated acceptor (ribose-5-phosphate) was accepted by all enzymes with the same conversion rate. The presence of this phosphate group enhanced the catalytic efficiency.

As previously illustrated,[9f] A5P gave no direct aldolisation with HA. The TLC analysis indicated after a few minutes that A5P was cleaved into glycolaldehyde and G3P. These two compounds quickly reacted with HA to give 1-deoxy-D-xylulose and d-D-F6P, respectively. The cleavage could not be avoided even by adding 10 equivalents of HA in order to saturate the active site with the donor.

To demonstrate the biocatalytic potential of these aldolases, small-scale synthesis of three compounds was performed. The best new aldolase $ALD_{listeria}$ was chosen for the proof of concept. Thus propanal, D-glyceraldehyde and 4-nitrobutanal were used as acceptor substrates with DHA as the donor (see Supporting Information, Scheme S1). These acceptors were selected because they lead to aldols previously prepared with $FSAA_{coli}$.[9b,16] To note, 4-nitrobutanal can lead to a nitrocyclitol

Table 3. TLC analysis for various enzymes.[a]

Acceptor	$FSAA_{coli}$	ALD_{Sgord}	ALD_{Smut}	ALD_{Spyog}	$ALD_{listeria}$ [c]
formaldehyde	■	■	■	■	■
benzaldehyde	□	□	□	□	□
pentanal	■	▪	▪	▪	▪
D-Gly[b]	■	■	■	■	■
D-erythrose	■	■	▪	▪	■
2-d-D-ribose[b]	□	□	□	□	□
D-glucose	□	□	□	□	□
D-R5P[b]	■	■	■	■	■
D-A5P[b]	□	□	□	□	□
D-G6P[b]	□	□	□	□	□

[a] Conditions: 50 mM acceptor, 50 mM donor (HA); enzyme (0.07 mg), H_2O pH 8; ■: good, total conversion after 3 h; ■: medium, total conversion after 12 h, ▪: low, incomplete conversion after 12 h, □: no conversion after 24 h. [b] D-Gly = D-glyceraldehyde; 2-d-D-ribose = 2-deoxy-D-ribose; D-R5P = D-ribose-5-phosphate, D-G6P = D-glucose-6-phosphate. [c] Native protein purified by FPLC.

as described in our previous work through an intramolecular Henry reaction[9b] and propanal was described as a poor substrate by Fessner et al.[16] The three acceptors were substrates and produced the corresponding expected aldol in 85%, 25% and 70% yields, respectively. The analytical data were consistent to those described in literature. The most interesting result relies on the successful conversion of propanal requiring eight time less enzyme, leading to a two times higher yield in a reaction three times faster than the same reaction performed with wild-type $FSAA_{coli}$.[16]

$ALD_{listeria}$ in its N-terminal His-tag version leads to the formation of highly aggregated proteins and cannot be purified properly. Considering its high catalytic properties, a study was performed to tackle this problem. Alternatives were tested such as the use of an *E. coli* BL21 strain optimised for expressing rare codons, a synthetic gene with optimised codons for *E. coli*, purification with and without an N-terminal His-tag and, finally, a His-tag introduced at the C-terminus. This last construction allowed us to purify $ALD_{listeria}$ efficiently by nickel affinity chromatography (Supporting Information, Figure 5 and Table 11). To ensure that the His-tag introduced in terminal position did not affect the catalytic properties, the K_m on F6P was determined. The presence of the His-tag did not change the biochemical properties as K_m was found at 10.3 mM (for native enzyme $K_m = 8.8$ mM, see Table 2).

Conclusions

A universal high-throughput screening strategy based on mining genomes and selection of enzymes representative of the biodiversity was applied to search for new aldolase activities. High-performance liquid chromatography/mass spectrometry revealed as a powerful and appropriate technique for the aldol detection, thus bypassing fastidious probe synthesis used in UV/Vis or fluorescence spectrometry. A total of 26 new biocatalysts, including transaldolases, were found and their activities were validated on purified enzymes. Among the 10 new DHA aldolases, eight were wrongly annotated as "transaldolases" in the UniprotKB database.

These enzymes such as $FSAs_{coli}$ possess a tyrosine as a third coordinator of the catalytic water molecule, whereas the transaldolases have a phenylalanine.[10a] The sequences of these new DHA aldolases do not harbour Interpro motifs specific to FSAs, and this feature could not allow FSA activity prediction. Nevertheless, their catalytic properties were found to be very close to those of the known $FSAA_{coli}$ and phylogenetic studies suggest proximity with transaldolase subfamily 5.

Four new DHA aldolases were more deeply characterised. They were stereoselective for the (3*S*, 4*R*) configuration, and exhibited similar acceptor substrate specificity. Their tolerance towards different donors, their various pH profiles, and better catalytic properties than those of $FSAA_{coli}$ make them promising biocatalysts for further application in organic synthesis. Indeed, the most active $ALD_{listeria}$ was confirmed to be a good catalyst for use in organic synthesis with dihydroxyacetone as a donor.

This approach has extended the chemist's enzymatic toolbox with new efficient stereoselective carboligating biocatalysts. Further investigations are underway on the search for other unusual biocatalysed aldol reactions.

Experimental Section

General

D-fructose-6-phosphate disodium salt hydrate, glycolaldehyde dimer, 1,3-dihydroxyacetone dimer, D,L-glucose 3-phosphate diethyl acetal barium salt, D-(+)-glyceraldehyde, D-ribose-5-phosphate disodium salt hydrate, D-glucose-6-phosphate dipotassium salt hydrate, β-nicotinamide adenine dinucleotide reduced disodium salt hydrate, β-nicotinamide adenine dinucleotide phosphate sodium salt, glycylglycine, HEPES, triethanolamine, sodium phosphate monobasic dihydrate and Trizma hydrochloride, α-glycerophosphate dehydrogenase/triosephosphate isomerase from rabbit muscle, glucose-6-phosphate dehydrogenase and phosphoglucose isomerase were purchased from Sigma–Aldrich. Hydroxyacetone was purchased from Fluka and purified by silica gel chromatography. Formaldehyde, benzaldehyde and D-glucose were purchased from Avocado. Valeraldehyde was purchased from Acros. D-erythrose was purchased from Alfa Aesar. 2-deoxy- D-ribose was purchased from Lancaster Synthesis. D-arabinose-5-phosphate and D-deoxyfructose-6-phosphate were synthesised in the laboratory as previously described.[9d] Synthetic gene was done by Genewiz. Oligonucleotides were from Sigma-Genosys.

Selection of enzymes

The collection of new aldolases was developed in a three-step process. First, a set of proteins was created by collecting information on known representative of the classes of aldolase and transaldolase. The list of EC numbers was selected on the basis of substrate structure: DHAP/DHA, pyruvate aldolases and transaldolases (as they display promiscuous aldolase activity) were chosen (Supporting Information, Table 1). Second, this set was compared to UniprotKB database and the metagenome from Genoscope using low stringency parameters (>30% of identity, on 80% of the length) resulting in the selection of 11 717 candidate enzymes. Third, a clustering based on 80% identity was applied to create putative isofunctional groups. Representatives of each cluster were selected for cloning and screening processes. The choice of the primers was based on the Primer3 program.[17] Specific extensions are added to the primers for cloning into our plasmid pET22b(+) (Novagen) modified for ligation independent cloning.[18]

Cloning and high-throughput screening

All steps, from primers purchase to cell lysate preparation, were performed in 96-microwell plates. The cloning was done in the modified plasmid pET22b(+), the protein production in *E. coli* BL21Star (DE3) plysE strains (Invitrogen) and BL21-CodonPlus (DE3)-RIPL (Agilent technologies), cell lysate preparations and protein concentrations were performed as previously described.[11] Primers used for His-tagged variants are described in Supporting Information, Table 2. When strains of the selected proteins were not available, PCR amplifications were performed on another DSM strain DNA and the clones were sequenced (Supporting Information Table 2 and Table 2 b).

Biochemical assays were performed in 96-well microtiterplates. Enzyme assays were performed in a final volume of 100 μL containing 1 mM dithyothreitol, 3 μL cell lysate (0.05 to 0.1 mg mL^{-1} of total proteins), 10 mM of each substrate (50 mM glycylglycine buffer, pH 7.5). Assays were performed overnight at RT and then reactions were stopped by adding 0.01 volume equivalents of trifluoroacetic acid. After centrifugation, a 1/20 dilution was done before LC–MS injection by transferring 10 μL volumes of each well of acidified reaction media in a 96-well daughter microtiterplate (Figure 1). The controls were prepared as described above replacing enzyme cell lysate by *E. coli* BL21 empty vector cell lysate.

Purification of aldolases

Selected enzymes were overexpressed and purified by loading the clear crude cell extract onto a Ni–NTA column (QIAGEN), according to the manufacturer's instructions. The wild-type $ALD_{listeria}$ was produced as an enzyme without a His-tag. Purification was conducted from a 400 mL culture as previously reported[19] except that the Ni–NTA purification step was replaced by a 25 mL Q-Sepharose Fast Flow (GE Lifesciences) ion-exchange chromatography. The enzyme purity was analysed by SDS PAGE. Protein concentrations were determined by the Bradford method with bovine serum albumin as the standard (Bio-Rad) and the samples were stored at −80 °C. $ALD_{listeria}$ was later produced with a C-terminal His-tag (see Supporting Information, Sections 7 and 8.)

Partially purified samples of non His-tagged $ALD_{listeria}$ were obtained by heating the cell free extract at 70 °C for 45 min before centrifugation. Then the protein was dialysed against 25 mM glycylglycine buffer, pH 8.0.

Enzymatic assay by LC–MS analysis

LC–MS–MS analyses were performed on a Hybrid triple quadrupole-linear ion trap mass spectrometer (QTRAP 5500 from ABSciex, Toronto, Canada) equipped with an ESI source and coupled to a Dionex UltiMate 3000 RS LC system (Thermo Scientific Dionex Corporation, Sunnyvale, CA, USA). The QTRAP mass spectrometer was operated in the ESI negative ion mode with the following parameters: ion source (IS) 4500 V, curtain gas (CUR) 20 a.u., temperature (TEM) 500 °C, gas 1 (GS1) 45 a.u., gas 2 (GS2) 60 a.u., CID medium. MS–MS experiments were performed by using MRM scan type.

For the first part of experiments, the ESIMS–MS method for the detection of an hexose-6-phosphate and G3P including two most intense transitions for each compound (monitoring of phosphate moieties ($[H_2PO_4]^-$ and $[PO_3]^-$) was employed (Supporting Information, Table 3). For final experiments, the MRM method including a set of transitions related to the monitoring of phosphate moieties and sugar cleavage product ions of hexose phosphates was used (Supporting Information, Table 4).

The reverse-phase chromatography method was used on an ACQUITY UPLC BEN C18 column (1.7 μm, 2.1×150 mm) from Waters (Milford, USA). The flow rate was 300 μL min^{-1}, the sample injection volume was 10 μL. The column was thermostated at 50 °C. The mobile phase A was 10 mM ammonium carbonate in water, and the mobile phase B was acetonitrile. Linear gradient from 0 to 100% of B was applied over a period of 7 min. The data processing was performed by using Analyst software (ABSciex).

Spectrophotometric enzymatic assay

One unit (U) of DHA aldolase is defined as the amount of enzyme that cleaves one micromole of D-F6P to afford D-G3P and DHA per minute at pH 8 (glycylglycine 50 mM buffer) and 25 °C. For transaldolase activity, glycolaldehyde is added as an acceptor substrate with F6P as the donor to afford D-G3P and xylulose.

Enzymatic assays were performed with purified enzymes. In each enzymatic assay, the formation of one product was measured using different coupled enzymatic systems for which NADH or NADPH consumption have been monitored at 340 nm (Figure 1). Aldolase activity in the direction of F6P formation from D,L-G3P and DHA was measured by monitoring F6P production using the coupled enzymatic system glucose-6-phosphate isomerase and glucose-6 phosphate dehydrogenase.[7a] Aldolase activity in the retroaldolisation direction from F6P and transferase activity from F6P and glycolaldehyde were analysed by monitoring G3P appearance using the coupled enzymatic system triose phosphate isomerase and glycerol-3-phosphate dehydrogenase.

Steady-state kinetic parameters for F6P, dF6P, A5P, D,L-G3P and DHA

Each kinetic assay (1 mL) was performed in glycylglycine buffer (50 mM pH 8.0). In the case of F6P, A5P, deoxyfructose-6-phosphate (dF6P), the substrate cleavage rate was measured by monitoring D-G3P appearance with glycerol-3-phosphate dehydrogenase triose phosphate isomerase auxiliary enzymes (10 U) and NADH (0.36 mM). In the case of DHA and D,L-G3P, the substrate formation rate was measured by monitoring F6P appearance with glucose-6 phosphate dehydrogenase (8 U)/glucose-6-phosphate isomerase (13.5 U) auxiliary enzymes and $NADP^+$ (1 mM)[7a] (for more information see Supporting Information).

Stereochemistry

Reaction: DHAP (0.5 mmol) was added to triose phosphate isomerase (150 U) and HA (3 eq, 1.5 mmol) in water (pH 8.0, 20 mL). Once the solution homogenised, the reaction was initiated by aldolase addition (20 U). The reaction proceeded at RT under gentle agitation (100–200 rpm). Every 20 min, 3 more equivalents of HA were added (total 9 equiv. over 80 min).

Purification protocol

The reaction was stopped by decreasing the pH to 3.0 resulting in partial enzyme precipitation. The pH was then adjusted to 6.0 and $BaCl_2 \cdot 2H_2O$ (2 equiv.) was added. The solution was centrifuged 15 min at 4000 rpm at 4 °C and the pellet was discarded. After partial concentration under vacuum, 5 volumes of ethanol were added. The solution was left to stand overnight at 4 °C and then centrifuged. The sugar barium salt was obtained as white powder after one ethanol rinsing step followed by two others with acetone. The stereochemistry was determined by NMR. All spectra samples were identical to the literature one (1-deoxy-D-fructose-6-phosphate)

Acceptor screening

Aldolase (0.07 mg) in glycylglycine buffer solution (50 mM pH 8.0) was added to a solution (200 μL total reaction volume) containing HA as donor (50 mM) and an acceptor compound (50 mM) in

water (pH 8.0). The reaction proceeded at RT under gentle agitation (100–200 rpm). The reaction was monitored at regular intervals by TLC (chloroform/methanol 9:1 or 8:2 and ammonium/ethanol 5:6) and worked up after 24–48 h depending on the substrate consumption. The relative conversion rates were determined on the basis of densitometric TLC monitoring.

Sequence comparison and phylogenetic analysis

All protein-versus-protein alignments were calculated with gapped BLASTp and the BLOSUM62 scoring matrix. Phylogenetic analysis was performed using MAFFT[20] for multiple sequence alignment, QUICKTREE[21] for tree building and TreeDyn[22] for tree rendering. All the enzyme UniprotKB identifiers are reported in the tree. The aldolases used as a reference were taken from Samland et al.[14] and others (J9MJK9 from *Fusarium oxysporum*).[23]

General procedure for preparative synthesis

A sample (1.5 mL, 25 mg of protein) of $ALD_{listeria}$ aldolase (synthetic gene, partially purified by heat treatment for 45 min at 70 °C in phosphate buffer pH 7.5, 100 mM) was added to a solution (18.5 mL) containing 130 mg of DHA (70 mM) and a final acceptor concentration of 50 mM (propanal, D-glyceraldehyde or nitrobutanal) at pH 7.5. After 8 h for propanal and nitrobutanal or 24 h for D-glyceraldehyde, the reaction was lyophilised. The solid obtained was suspended in EtOH and then centrifugated. After evaporation of the supernatant the crude oil was purified by column chromatography on silica gel to provide the expected pure compound. In the case of propanal the experimental procedure and the work-up were slightly different and are explained in detail in the corresponding paragraph below.

5,6-Dideoxy-D-*threo*-hexulose: Yield (85%, 182 mg), propanal (60 mg) was added followed by two other additions of a 60 and 150 mg portion after 2 and 4 h, respectively, until total DHA consumption. The volatile status of this acceptor was exploited and the column chromatography was not necessary in that case to obtain a pure compound. Simple concentration under high vacuum of the above mentioned EtOH solution afforded the desired aldol. The spectral data are identical to those described in Ref. [16]. ^{1}H NMR (400 MHz, D_2O): $\delta = 4.60$ (d, $J = 19.4$ Hz, 1H, 1-H_A), 4.48 (d, $J = 19.3$ Hz, 1H, 1-H_B), 4.34 (d, $J = 2.3$ Hz, 1H, 3-H), 3.89 (ddd, $J = 2.3$, 6.4, 7.7 Hz, 1H, 4-H), 1.67–1.48 (m, 2H, 5-H), 0.92 ppm (t, $J = 7.5$ Hz, 3H, 6-H). ^{13}C NMR (100 MHz, D_2O): $\delta = 213.2$ (C2), 77.1 (C3), 73.5 (C4), 65.9 (C1), 25.4 (C5), 9.4 ppm (C6).

D-Fructose: Yield (25%, 45 mg, low yield owing to purification process). Column chromatography on silica gel: 90:10 CH_2Cl_2/MeOH then 80:20 CH_2Cl_2/MeOH when DHA was removed. The ^{1}H NMR spectra is provided in Supporting Information.

(1*S*,2*S*,3*R*,6*R*)-1-(Hydroxymethyl)-6-nitrocyclohexane-1,2,3-triol: For a better control of the intramolecular Henry reaction, the enzymatic reaction was performed at pH 8.5. yield (70%, 145 mg). Column chromatography on silica gel: 90:10 CH_2Cl_2/MeOH. The spectral data are identical to those described in Ref. [9b]. ^{1}H NMR (400 MHz, MeOD): $\delta = 4.81$ (dd, $J = 3.9$, 12.9 Hz, 1H, 6-H), 3.83 (d, $J = 11.0$ Hz, 1H, 7-H), 3.75 (ddd, $J = 4.6$, 9.4, 11.6 Hz, 1H, 3-H), 3.39 (d, $J = 9.4$ Hz, 1H, 2-H), 3.35 (d, $J = 11.0$ Hz, 1H, 7-H), 2.49–2.36 (m, 1H, 5-H_{eq}), 2.10–1.86 (m, 2H, 5-H_{eq}, 4-H_{eq}), 1.38–1.26 ppm (m, 1H, 4-H_{ax}). ^{13}C NMR (100 MHz, MeOD): $\delta = 86.3$ (C_6), 77.0 (C_1), 75.0 (C_2), 70.9 (C_3), 61.8 (C_7), 29.5 (C_4), 24.6 ppm (C_5).

Acknowledgements

This research was supported by the French Agence Nationale pour la Recherche (ANR Genozyme, ANR-08-GENM-026). We thank Christine Pelle for $ALD_{listeria}$ *purification and Andrew Tolonen for English corrections.*

Keywords: biocatalysis · C–C coupling · enzymes · gene expression · mass spectrometry

[1] a) R. Wohlgemuth, *Curr. Opin. Biotechnol.* **2010**, *21*, 713–724; b) S. Wenda, S. Illner, A. Mell, U. Kragl, *Green Chem.* **2011**, *13*, 3007–3047; c) K. Drauz, H. Gröger, O. May, *Enzyme Catalysis in Organic Synthesis, 3rd ed.*, Wiley-VCH, **2012**.

[2] a) T. Davids, M. Schmidt, D. Bottcher, U. Bornscheuer, *Curr. Opin. Chem. Biol.* **2013**, *17*, 215–220; b) U. Bornscheuer, G. Huisman, R. Kazlauskas, S. Lutz, J. Moore, K. Robins, *Nature* **2012**, *485*, 185–194.

[3] a) H. Kries, R. Blomberg, D. Hilvert, *Curr. Opin. Chem. Biol.* **2013**, *17*, 221–228; b) L. Giger, S. Caner, R. Obexer, P. Kast, D. Baker, N. Ban, D. Hilvert, *Nat. Chem. Biol.* **2013**, *9*, 494–498; c) E. Althoff, L. Wang, L. Jiang, L. Giger, J. Lassila, Z. Wang, M. Smith, S. Hari, P. Kast, D. Herschlag, D. Hilvert, D. Baker, *Protein Sci.* **2012**, *21*, 717–726.

[4] a) G. Behrens, A. Hummel, S. Padhi, S. Schatzle, U. Bornscheuer, *Adv. Synth. Catal.* **2011**, *353*, 2191–2215; b) M. Guazzaroni, A. Beloqui, P. Golyshin, M. Ferrer, *World J. Microbiol. Biotechnol.* **2009**, *25*, 945–954.

[5] a) M. Ferrer, F. Martinez-Abarca, P. Golyshin, *Curr. Opin. Biotechnol.* **2005**, *16*, 588–593; b) P. Golyshin, *Curr. Opin. Biotechnol.* **2005**, *16*, 585–587.

[6] a) P. Clapés, W. D. Fessner, *Sci. Synth. Stereosel. Synth.* **2011**, *2*, 677–734; b) P. Clapés, W. D. Fessner, G. A. Sprenger, A. K. Samland, *Curr. Opin. Chem. Biol.* **2010**, *14*, 154–167; c) P. Clapés, X. Garrabou, *Adv. Synth. Catal.* **2011**, *353*, 2263–2283; d) M. Brovetto, D. Gamenara, P. Mendez, G. Seoane, *Chem. Rev.* **2011**, *111*, 4346–4403; e) M. Müller, *Adv. Synth. Catal.* **2012**, *354*, 3161–3174; f) K. Fesko, M. Gruber-Khadjawi, *ChemCatChem* **2013**, *5*, 1248–1272; g) P. Falcicchio, S. Wolterink-Van Loo, M. C. R. Franssen, J. van der Oost, *Extremophiles* **2014**, *18*, 1–13.

[7] a) M. Schürmann, G. A. Sprenger, *J. Biol. Chem.* **2001**, *276*, 11055–11061; b) I. Sánchez-Moreno, L. Nauton, V. Théry, A. Pinet, J.-L. Petit, V. de Berardinis, A. K. Samland, C. Guérard-Hélaine, M. Lemaire, *J. Mol. Catal. B* **2012**, *84*, 9–14.

[8] a) X. Garrabou, J. A. Castillo, C. Guérard-Hélaine, T. Parella, J. Joglar, M. Lemaire, P. Clapés, *Angew. Chem. Int. Ed.* **2009**, *48*, 5521–5525; *Angew. Chem.* **2009**, *121*, 5629–5633; b) M. Sugiyama, Z. Hong, P.-H. Liang, S. M. Dean, L. J. Whalen, W. A. Greenberg, C.-H. Wong, *J. Am. Chem. Soc.* **2007**, *129*, 14811–14817.

[9] a) J. A. Castillo, C. Guérard-Hélaine, M. Gutierrez, X. Garrabou, M. Sancelme, M. Schürmann, T. Inoue, V. Hélaine, F. Charmantray, T. Gefflaut, L. Hecquet, J. Joglar, P. Clapés, G. A. Sprenger, M. Lemaire, *Adv. Synth. Catal.* **2010**, *352*, 1039–1046; b) F. Camps Bres, C. Guérard-Hélaine, C. Fernandes, J. Castillo, M. Lemaire, *Tetrahedron: Asymmetry* **2013**, *24*, 1075–1081; c) W.-D. Fessner, D. Heyl, M. Rale, *Catal. Sci. Technol.* **2012**, *2*, 1596–1601; d) I. Sánchez-Moreno, V. Hélaine, N. Poupard, F. Charmantray, B. Légeret, L. Hecquet, E. Garcia-Junceda, R. Wohlgemuth, C. Guérard-Hélaine, M. Lemaire, *Adv. Synth. Catal.* **2012**, *354*, 1725–1730; e) M. Mifsud, A. Szekrenyi, J. Joglar, P. Clapés, *J. Mol. Catal. B* **2012**, *84*, 102–107; f) C. Guérard-Hélaine, M. Debacker, P. Clapés, A. Szekrenyi, V. Hélaine, M. Lemaire, *Green Chem.* **2014**, *16*, 1109–1113.

[10] a) S. Schneider, T. Sandalova, G. Schneider, G. A. Sprenger, A. K. Samland, *J. Biol. Chem.* **2008**, *283*, 30064–30072; b) S. Schneider, M. Gutierrez, T. Sandalova, G. Schneider, P. Clapés, G. A. Sprenger, A. K. Samland, *ChemBioChem* **2010**, *11*, 681–690; c) X. Garrabou, J. Joglar, T. Parella, R. Crehuet, J. Bujons, P. Clapés, *Adv. Synth. Catal.* **2011**, *353*, 89–99.

[11] C. Vergne-Vaxelaire, F. Bordier, A. Fossey, M. Besnard-Gonnet, A. Debard, A. Mariage, V. Pellouin, A. Perret, J. Petit, M. Stam, M. Salanoubat, J. Weissenbach, V. de Berardinis, A. Zaparucha, *Adv. Synth. Catal.* **2013**, *355*, 1763–1779.

[12] A. K. Samland, G. A. Sprenger, *Int. J. Biochem. Cell Biol.* **2009**, *41*, 1482–1494.

[13] S. Thorell, M. Schürmann, G. A. Sprenger, G. Schneider, *J. Mol. Biol.* **2002**, *319*, 161 – 171.
[14] A. K. Samland, S. Baier, M. Schürmann, T. Inoue, S. Huf, G. Schneider, G. A. Sprenger, T. Sandalova, *FEBS J.* **2012**, *279*, 766 – 778.
[15] a) C. Alcantara, L. A. Sarmiento-Rubiano, V. Monedero, J. Deutscher, G. Perez-Martinez, M. J. Yebra, *Appl. Environ. Microbiol.* **2008**, *74*, 5731 – 5740; b) M. Tangney, J. K. Brehm, N. P. Minton, W. J. Mitchell, *Appl. Environ. Microbiol.* **1998**, *64*, 1612 – 1619; c) M. J. Yebra, A. Veyrat, M. A. Santos, G. Perez-Martinez, *J. Bacteriol.* **2000**, *182*, 155 – 163; d) J. Reizer, A. Reizer, M. H. Saier, Jr., *Microbiology* **1995**, *141*, 961 – 971.
[16] M. Rale, S. Schneider, G. Sprenger, A. Samland, W. Fessner, *Chem. Eur. J.* **2011**, *17*, 2623 – 2632.
[17] S. Rozen, H. Skaletsky, *Methods Mol. Biol.* **2000**, *132*, 365 – 386.
[18] C. Aslanidis, P. J. de Jong, *Nucleic Acids Res.* **1990**, *18*, 6069 – 6074.
[19] A. Kreimeyer, A. Perret, C. Lechaplais, D. Vallenet, C. Medigue, M. Salanoubat, J. Weissenbach, *J. Biol. Chem.* **2007**, *282*, 7191 – 7197.
[20] K. Katoh, K. Misawa, K. Kuma, T. Miyata, *Nucleic Acids Res.* **2002**, *30*, 3059 – 3066.
[21] K. Howe, A. Bateman, R. Durbin, *Bioinformatics* **2002**, *18*, 1546 – 1547.
[22] F. Chevenet, C. Brun, A. L. Banuls, B. Jacq, R. Christen, *BMC Bioinf.* **2006**, *7*, 439.
[23] E. Kourtoglou, D. Mamma, E. Topakas, P. Christakopoulos, *Process Biochem.* **2008**, *43*, 1094 – 1101.

Received: January 7, 2015
Published online on June 11, 2015