nature chemical biology

ARTICLE

PUBLISHED ONLINE: 26 SEPTEMBER 2010 | DOI: 10.1038/NCHEMBIO.447

Rational assignment of key motifs for function guides *in silico* enzyme identification

Matthias Höhne[1,3], Sebastian Schätzle[1,3], Helge Jochens[1], Karen Robins[2] & Uwe T Bornscheuer[1]*

Biocatalysis has emerged as a powerful alternative to traditional chemistry, especially for asymmetric synthesis. One key requirement during process development is the discovery of a biocatalyst with an appropriate enantiopreference and enantioselectivity, which can be achieved, for instance, by protein engineering or screening of metagenome libraries. We have developed an *in silico* strategy for a sequence-based prediction of substrate specificity and enantiopreference. First, we used rational protein design to predict key amino acid substitutions that indicate the desired activity. Then, we searched protein databases for proteins already carrying these mutations instead of constructing the corresponding mutants in the laboratory. This methodology exploits the fact that naturally evolved proteins have undergone selection over millions of years, which has resulted in highly optimized catalysts. Using this *in silico* approach, we have discovered 17 (*R*)-selective amine transaminases, which catalyzed the synthesis of several (*R*)-amines with excellent optical purity up to >99% enantiomeric excess.

Enzyme catalysis represents one cornerstone in the area of white (industrial) biotechnology[1–5]. The usually excellent chemo-, regio- and enantioselectivity of biocatalysts[6,7] facilitates and simplifies many chemical processes for the production of a broad range of products. The production of optically pure building blocks for the pharmaceutical and fine chemicals industries are the most valuable areas[8–12] of use for enzyme catalysis. These processes are environmentally friendlier than other options because additional reaction steps can be avoided and the processes are usually performed under milder reaction conditions with fewer organic solvents and/or in the absence of hazardous materials[13]. The biocatalytic synthesis of optically pure compounds can either be performed as a kinetic resolution or as an asymmetric synthesis[14], as shown in **Scheme 1** for transaminase-catalyzed reactions.

The kinetic resolution of the racemic amine with an (*S*)-selective transaminase produces the (*R*) enantiomer. The disadvantage of this approach is that a maximum yield of only 50% can be achieved (**Scheme 1**, left). The atom efficiency of such a process is low, as the ketone generated has little value and the recycling of this compound requires chemical reductive amination to produce the racemic amine for subsequent resolutions. Synthetically, an asymmetric synthesis is much more economical because yields of up to 100% are possible. However, in this case the (*S*)-amine would be produced (**Scheme 1**, right). An asymmetric synthesis strategy is thus clearly favored with respect to the atom efficiency of the process but has the major disadvantage that only one specific enantiomer can be accessed with an enzyme having a distinct enantiopreference. As both enantiomers of a building block are often needed for targets with different absolute configuration, an enzyme platform providing either (*R*)- or (*S*)-specific enzymes is highly desired. Several strategies can be used to provide access to enzymes with complementary enantiopreference. This includes classical screening of strain collections for the identification of complementary enzymes[15]. Indeed, (*R*)- and (*S*)-selective hydroxy nitrile lyases[16], hydantoinases[17] and keto reductases[18] have been described. A much more promising option is the use of protein engineering tools[19–22], with which changes in substrate specificity[21], stability improvements[23] and also switches in enantiopreference[24–29] can be achieved. A recent example of the combination of rational protein design and directed evolution concepts is the inversion of the enantiopreference of a *Bacillus subtilis* esterase by a double mutation identified in a library of only 2,800 variants[30].

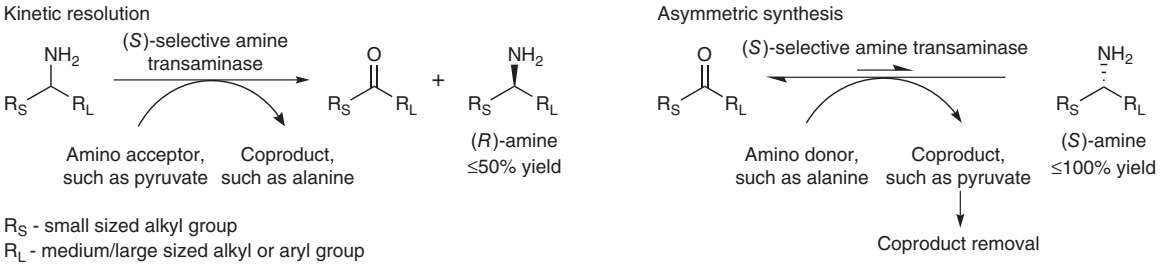
In this contribution, we have explored a third approach, in which we take advantage of the numerous protein sequences already deposited in databases. One practical advantage when compared with classical screening is that the complete nucleotide—and hence, protein—sequence is already known, and after identification of promising candidates, the gene can easily be cloned from the original source organism or obtained as a synthetic gene. The major challenge is, however, the retrieval of desired sequences if no enzyme with desired enantiopreference has been described in the literature and the easy and ideal starting point for this *in silico* approach is thus missing. This is indeed the case for amine-pyruvate transaminases (referred to here as amine transaminases, EC 2.6.1.18): in contrast to various (*S*)-selective amine transaminases, only two enzymes with (*R*)-selectivity have been reported[31–33], and at the submission of this study no information about nucleotide or amino acid sequences was available. In the meantime, the amino acid sequence of a previously commercially available (*R*)-selective amine transaminase has been published[34].

Transaminases belong to fold classes I and IV of pyridoxal-5′-phosphate (PLP)–dependent enzymes[35,36]. In contrast to α-amino acid aminotransferases (referred to throughout as α-transaminases, EC 2.6.1), which are ubiquitous enzymes found in all organisms, the small group of amine transaminases also converts substrates lacking an α-carboxylic acid moiety (**Scheme 1**). This makes them very attractive for organic synthesis of optically active amines[37], especially as the product range is not limited to α-amino acids. Hence, there is an urgent need for the discovery of (*R*)-selective amine transaminases to access a broad range of (*R*)-amines by efficient asymmetric synthesis (**Scheme 1**, right).

We first considered protein engineering to create an (*R*)-selective amine transaminase by inverting the enantiopreference of an (*S*)-selective amine transaminase (**Fig. 1a**), but there were several limitations, as no crystal structure of any amine transaminase is available and thus rational protein design is very difficult to perform.

One option would be to start from the crystal structure of an (*S*)-selective α-transaminase and reengineer the substrate-recognition site to create a variant that accepts substrates lacking the carboxylic

[1]Department of Biotechnology and Enzyme Catalysis, Institute of Biochemistry, Greifswald University, Greifswald, Germany. [2]Lonza AG, Valais Works, Visp, Switzerland. [3]These authors contributed equally to this work. *e-mail: uwe.bornscheuer@uni-greifswald.de



Scheme 1 | Strategies for the synthesis of optically active amines using amine transaminase. In a kinetic resolution (left), the amine transaminase converts in the ideal case only one of the amine enantiomers to the corresponding ketone. The remaining enantiomer can be isolated in high optical purity and at a (theoretical) maximum yield of 50%. In an asymmetric synthesis (right), a prostereogenic ketone is aminated enantioselectively, yielding directly the optically active amine. The most common cosubstrates for amine transaminases are pyruvate and alanine. As the equilibrium favors ketone formation, high yields in asymmetric synthesis can only be achieved by shifting the equilibrium, for example by enzymatic removal of the formed coproduct pyruvate[49].

function (**Fig. 1a**), resulting in an (*R*)-selective amine transaminase (owing to the switch in priority according the Cahn-Ingold-Prelog (CIP) rule[38]). Unfortunately, the substrate's α-carboxyl group plays an important role for the domain closure of the α-transaminase during substrate binding[39,40]. Thus, to solve this challenge, many additional mutations might be required, which could not be predicted because of the complexity of the problem.

Instead of performing directed evolution to create and identify an enzyme with the desired selectivity from a random mutant library, we developed a strategy to find enzymes with complementary enantiopreference by searching *in silico* in protein databases (**Fig. 1b**). Essentially, the method is based on two steps: (i) identification and prediction of important amino acid residues on the basis of structural information from related enzymes and (ii) data mining

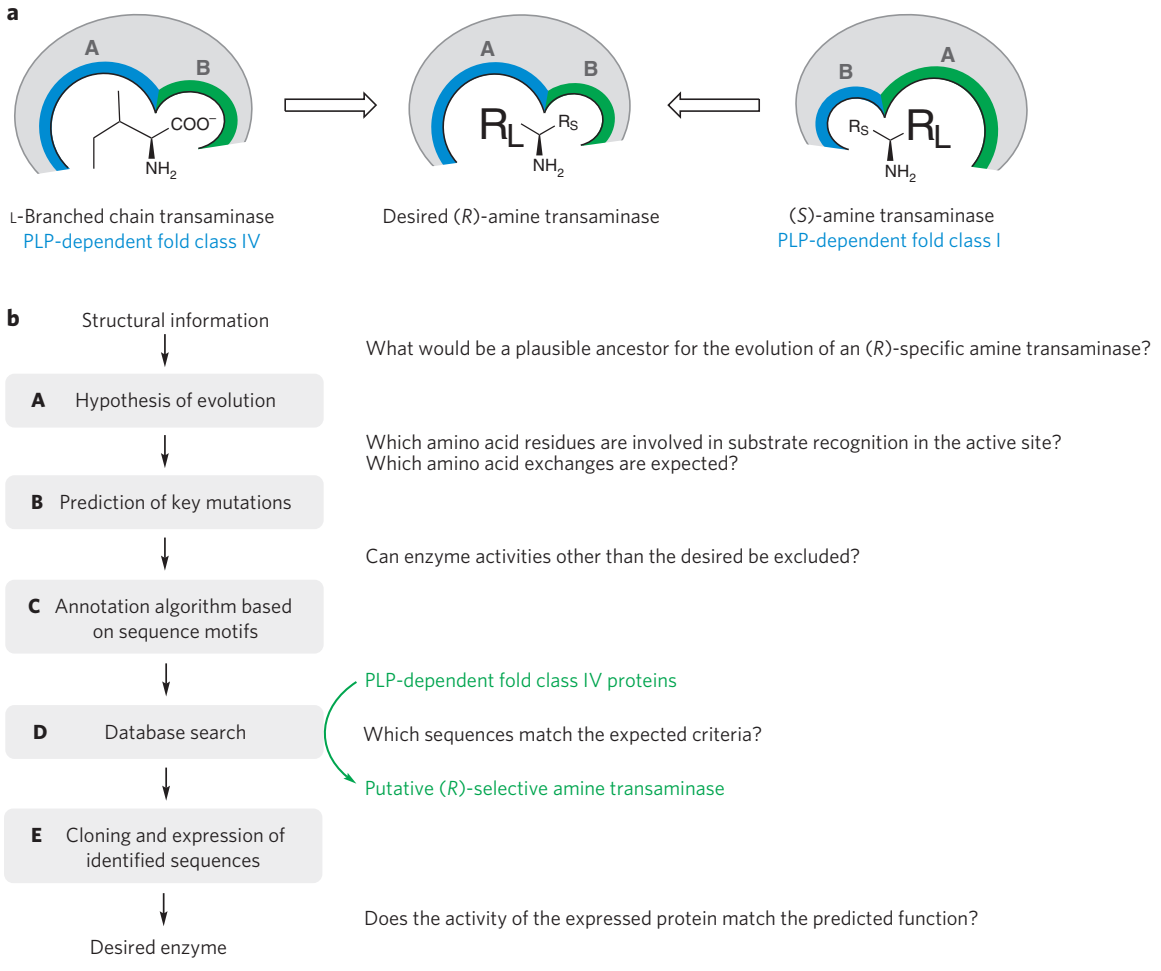


Figure 1 | Strategies for protein engineering. (**a**) Possible ancestors of amine transaminase with (*R*)-enantiopreference can be used to engineer an (*R*)-selective amine transaminase (center). This can be achieved by modification of the amino acids in the carboxyl group–binding pocket of an α-transaminase (left), such as an L-branched chain transaminase of the PLP-dependent fold class IV, or by engineering of the binding pockets of an (*S*)-selective amine transaminase (right) from PLP-dependent fold class I. It was assumed that according to the CIP rule[38], the large substituent (R_L) has a higher priority than the small substituent (R_S). (**b**) Flow scheme of the *in silico* approach for the identification of transaminases with inverted enantiopreference, with steps A–E.

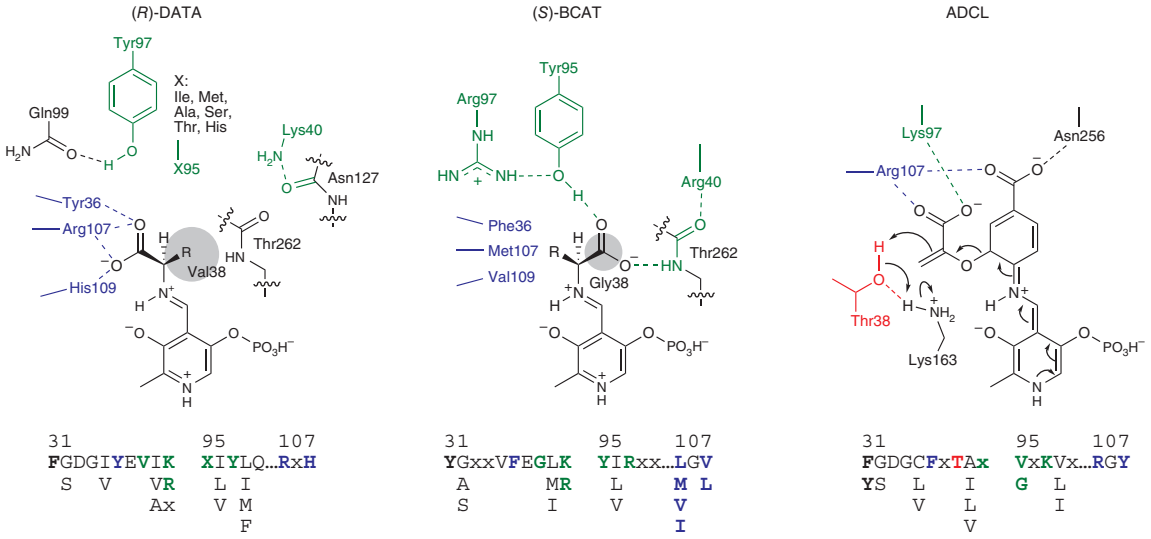


Figure 2 | Identification of key amino acid motifs that allow prediction of function of PLP-dependent fold class IV proteins. Top: schematic drawing of amino acids contributing to substrate binding in DATAs, BCATs and ADCLs. The external aldimines, which are formed after binding and transimination of the respective substrates with the PLP cofactor of the enzymes, are shown. Blue amino acids are part of binding pocket A (**Fig. 1a**), and amino acids of binding pocket B (**Fig. 1a**) are shown in green. The gray shaded circle represents the glycine or valine, depending on the transaminase, that is located behind the carboxyl group. The red threonine residue is important in the catalytic mechanism for shuttling of a proton during the reaction transition state. Bottom: sequence motifs derived from multiple-sequence alignments are given using the same color code, showing that amino acids important for substrate binding in the active site are rather conserved and can be used for a prediction of the substrate specificity of PLP-dependent fold class IV enzymes as well as for the enantiopreference of the transaminases within this fold class.

in protein sequences. This concept takes advantage of the currently available knowledge about the function of proteins combined with the experimentally uncharacterized huge diversity stored in protein sequence databases.

RESULTS

In the first step, we carefully analyzed structural information of related enzymes to evaluate a possible hypothesis for the evolution of an (*R*)-selective amine transaminase (step A; see **Fig. 1b** for a schematic representation of the steps). Then, we predicted important positions and suggestions for amino acid exchanges (step B). Next, we performed identification of putative—and so far unknown—sequence motifs to exclude unwanted enzyme activities, and, as the key step, we developed an annotation algorithm on the basis of these sequence motifs (step C). Protein-coding sequences thus identified in a database search (step D) as matching the predicted criteria were then cloned from synthetic genes. The resulting enzymes were investigated biochemically for transaminase activity and desired (*R*)-selectivity (step E).

Analysis of structural information

Two fold classes have been described for transaminases[35]. All α-transaminases belonging to fold class I of PLP-dependent enzymes known so far are L-selective, and no D-selective amino acid transaminases have been described within this fold class. However, D-amino acid aminotransferases[41] (DATAs) are members of fold class IV, and notably, L-branched chain amino acid aminotransferases (BCATs)[42] belong to the same fold class[43]. This indicates that within this protein fold class there is a certain flexibility in the architecture of the active site with regard to substrate recognition, and therefore we focused our search on fold class IV aminotransferases. Apart from DATAs and BCATs, only 4-amino-4-deoxychorismate lyase (ADCL) is currently known as a further member of fold class IV.

The opposite enantiopreference of DATAs and BCATs can be explained—via their crystal structures[41,42]—by the difference in substrate coordination in the active site, which consists of two binding pockets. If the α-carboxyl group is positioned in binding pocket B (**Fig. 1a**), the enzyme converts L-amino acids and thus shows (*S*)-preference. If, however, the α-carboxyl group of the substrate is accommodated in binding pocket A, this transaminase is (*R*)-selective and thus converts D-amino acids.

This indicates that the desired diversity is perhaps already available in nature and that it only needs to be identified and exploited. Following this line of reasoning, an (*R*)-selective amine transaminase could already have been evolved from an (*S*)-selective α-transaminase if modifications of the α-carboxyl group–binding pocket had occurred, allowing small hydrophobic side chains to bind instead of the α-carboxyl functionality. The plausible ancestor of an (*R*)-specific amine transaminase must therefore be an L-(*S*)-selective BCAT, as, according to the CIP rule[38], the substitution of the carboxyl group of an L-amino acid by a methyl group yields an (*R*)-amine.

Prediction of key features of the desired enzyme

Working from the known crystal structures, we studied the coordination of the α-carboxyl group in BCATs and DATAs. This allowed us to predict the necessary differences in protein sequences within PLP-dependent fold class IV proteins that determine the desired switch in substrate specificity from α-amino acids to amines, in line with the formal switch in enantiopreference. In the case of BCATs, the substrate binding in pocket B is realized in a more subtle manner than it is in DATAs and other α-transaminases (for more details see **Supplementary Results** and **Supplementary Fig. 1**). There is no direct contact of the carboxyl group oxygen atoms to any basic amino acid side chain such as that of an arginine (**Fig. 2**). Instead, one carboxyl oxygen atom is coordinated by the Tyr95 hydroxyl group, which is polarized by the coordination of an adjacent Arg97, and the other oxygen is bonded by two backbone amide nitrogen atoms of Thr262 and Ala263, which are activated by the coordination of their adjacent carbonyl groups by Arg40 (**Fig. 2** and **Supplementary Fig. 1**; note that in related sequences a lysine is at position 40).

Thus, an amine transaminase should differ from BCATs in the following residues. First, Tyr95 should be exchanged with

Table 1 | Section of a multiple-sequence alignment of putative (*R*)-selective amine transaminases.

Entry, GeneID	Organism	Sequence motif 1 (31 36 40)		Sequence motif 2 (95 99 107)	
1 ecDATA	*Bacillus* sp.	26	FGDGVYEVVKVYN	77	HIYFQVTRGTSPRAHQFP
2 ecBCAT	*Eschericha coli*	26	YGTSVFEGIRCYD	86	YIRPLIFVGDVGMGVNPP
3 ecADCL	*Eschericha coli*	21	FGDGCFTTARVID	78	VLKVVISRGSGGRGYSTL
4 115385557	*Aspergillus terreus*	55	HSDLTYDVPSVWD	106	FVELIVTRGLKGVRGTRP
5 211591081	*Penicillium chrysogenum*	53	HSDLTYDVPSVWD	104	FVEIIVTRGLKGVRGSRP
6 145258936	*Aspergillus niger*	53	RSDLTYDVISVWD	104	YVALIVTRGLQSVRGAKP
7 169768191	*Aspergillus oryzae*	53	HSDLTYDVPSVWD	104	FVELIVTRGLKGVRGNKP
8 70986662	*Aspergillus fumigatus*	53	HSDLTYDVISVWD	104	FVEVIVTRGLTGVRGSKP
9 119483224	*Neosartorya fischeri*	53	HGDLTYDVTTVWD	104	FVEVIVTRGLTGVRGSKP
10 46109768	*Gibberella zeae*	53	HGDLTYDVPAVWD	104	FVELIVTRGLKPVREAKP
11 114797240	*Hyphomonas neptunium*	53	HSDLTYDVPAVWN	104	YVEIIVTRGLKFLRGAQA
12 120405468	*Mycobacterium vanbaalenii*	69	HSDLTYTVAHVWH	120	FVNLTITRGYGKRKGEKD
13 13471580	*Mesorhizobium loti*	54	HSDATYDTVHVWN	105	YVEMLCTRGASPTFSRDP
14 20804076	*Mesorhizobium loti*	53	HSDATYDTVHVWE	104	YVEMICTRGGSPTFSRDP
15 86137542	*Roseobacter* sp.	45	HSDATYDVAHVWK	96	YVEFICTRGTSPTFSRDP
16 87122653	*Marinomonas* sp.	44	HSDATYDVVHVWQ	95	YVEMICTRGNSPDFSRDP
17 190895112	*Rhizobium etli*	38	RSDACQDTVSVWD	90	YVQIIMTRGRPPIGSRDL
18 89899273	*Rhodoferax ferrireducens*	34	RSDATYDVVTVWD	85	YVEMICTRGQPPWGSRDP
19 89053613	*Jannaschia* sp.	32	HSDIAYDVVPVWR	83	YVAMVAARGRNPVPGSRD
20 EEE43073	*Labrenzia alexandrii*	42	HSDITYDVVPVLD	93	YVAMVTSRGVNQVPGSRD
21 78059900	*Burkholderia* sp.	53	HADAAYDVVTVSR	104	YVWWCVTRGPLSVDRRDR
22 ABK12047	*Burkholderia cenocepacia*	48	HSDVTYDTVHVWN	99	YVEMLCTRGVSPTFSRDP
23 ZP_01448442	Alpha proteobacterium	22	HSDATYDVAHVWG	73	YVEFICTRGTSPNFSRDP
24 219677744	Gamma proteobacterium	30	LGDGVFDVVSAWK	81	SIRFIVTRGEPKGVVADP

The first three entries are the DATA, BCAT and ADCL sequences used for the identification of sequence motifs. Entries 4–24 refer to the newly discovered enzymes.

a hydrophobic residue that is incapable of forming a hydrogen bond with the carboxyl group. Second, Arg40 should be changed to a residue that cannot activate the amide backbone nitrogens of Thr262 and Ala263 by coordination of the neighboring backbone carbonyl oxygens. However, the situation regarding position 40 in PLP-dependent fold class IV proteins is complex. DATAs also have a basic amino acid, Lys40, but in contrast to the Arg40 in BCATs, Lys40 in DATAs adopts an altered conformation so that its ε-amino group forms a hydrogen bond to the backbone of an adjacent loop (**Fig. 2** and **Supplementary Fig. 1**). Thus, in spite of the presence of Lys40 in DATAs, the activation of the amide backbone nitrogens, critical for carboxyl-group binding in binding pocket B, is prevented. From this observation, we concluded that if an Arg40 or Lys40 is found in one of the sequences in a fold class IV protein, whether this residue facilitates the binding of a substrate carrying an α-carboxyl group or not cannot be predicted. In summary, the presence of a hydrophobic amino acid in position 95 and the absence of an arginine or lysine residue at position 40 would be a clear hint of altered substrate specificity toward amines.

Design and application of a sequence-based algorithm

Next, we developed a sequence-based prediction of the substrate specificity—conversion of amines versus α-D- or α-L-amino acids—to identify putative (*R*)-amine transaminases within PLP-dependent fold class IV proteins. Aside from the key residues considered important for amine transaminase activity, we compared residues involved in substrate coordination in the different enzymes. To simplify the structural description, we introduced a general numbering scheme for the amino acid residues of the different PLP-dependent fold class IV proteins. This was based on a multiple-sequence alignment (**Supplementary Fig. 2**). Fortunately, the amino acids that are in direct contact with the substrate in the active site are arranged in two relatively short sequence blocks. The first block is located at positions 36–40, and the second block comprises six amino acids at positions 95–97 and 107–109. Most of these amino acids fold into a β-sheet; only residues 107–109 are part of a loop. This information is important because during evolution, insertions or deletions take place preferentially in loop sequences without destroying enzyme activity but do not usually take place in α-helices or β-sheets[44]. Thus, it was observed that residues contributing to substrate recognition of a PLP-dependent fold class IV enzyme aligned well within the different enzymes, except in the motif at positions 107–109, where we sometimes observed insertions or deletions of one to two amino acids. Alignments, which include all known BCAT, DATA and ADCL proteins with experimentally verified enzyme activity, showed that the amino acids involved in substrate recognition in the active site seem to be quite conserved in these three groups. This allowed us to formulate different sequence motifs characteristic of DATA, BCAT and ADCL activity (**Fig. 2**; for a summary of the structure-function relationship of individual amino acids of the sequence motifs see **Supplementary Tables 1** and **2**; for the multiple-sequence alignments see **Supplementary Figs. 3–5**). On the basis of these comparative considerations, an annotation algorithm was developed (**Supplementary Fig. 6**) using the amino acid sequence motifs identified. This enabled an easy exclusion of all enzyme candidates that could clearly be designated as BCATs, ADCLs or DATAs. The analysis of the remaining sequences aimed at identifying transaminase sequences that fulfilled the requirements for the desired (*R*)-selective amine transaminase activity.

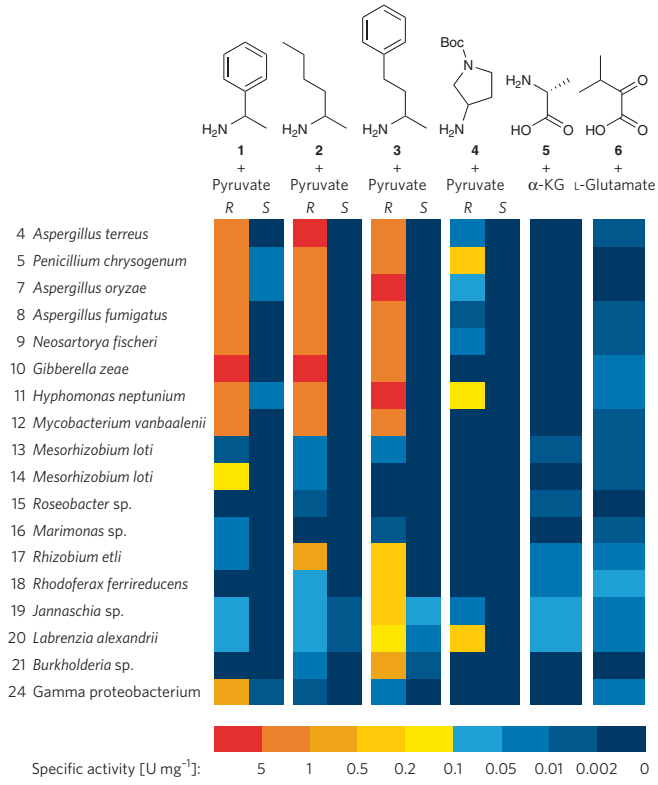


Figure 3 | Characterization of the discovered (*R*)-selective amine transaminases. Specific activities toward (*R*) and (*S*) enantiomers of amines **1**–**4** and α-amino acids **5** (D-alanine) and **6** (L-glutamate) are indicated by a color gradient. The numbering of the proteins corresponds to that given in **Table 1** (details on specific activity and expression levels are given in **Supplementary Tables 2** and **3**). α-KG, α-ketoglutarate.

Using this algorithm, we analyzed about 5,700 sequences annotated as BCATs and 280 protein sequences annotated as PLP-dependent fold class IV proteins from the US National Center for Biotechnology Information (NCBI) Protein Database. This search identified 21 sequences, 7 from eukaryotes and 14 from prokaryotes (**Table 1**). These sequences matched our criteria for the selection of putative (*R*)-selective amine transaminases with the predicted enantiopreference: the absence of arginine or lysine at position 40 in all sequences and a Phe95 instead of a Tyr95 in 8 out of the 21 sequences indicated that substrates having a small alkyl group instead of a carboxyl group in the α-position might be bound to the active site. In the other 13 sequences, a putative BCAT activity as suggested by a Tyr95 residue is unlikely because of the presence of alanine, glutamine, aspartic acid or glutamic acid at residue 97 instead of the required polarizing arginine. Thus, the coordination of a small alkyl group in binding pocket B seemed much more likely than binding of an α-carboxyl group, which we considered strong evidence that the respective proteins are amine transaminases rather than BCATs. To convey (*R*)-selectivity, binding pocket B should only allow the binding of a small alkyl group. The fact that in position 95 a tyrosine or phenylalanine is found suggests that the size of the binding pocket is not large compared to BCATs, as this could be obtained by mutation of Tyr95 to a small hydrophobic amino acid, such as alanine or valine.

Thus, we considered the criteria for the desired switch in substrate specificity toward (*R*)-amines to be fulfilled. It is noteworthy that sometimes the annotated function of a protein included in a database did not match the prediction made by our sequence motif approach. For example, out of 26 proteins annotated in the curated NCBI database as DATAs, 7 could clearly be identified as BCATs and one as an (*R*)-amine transaminase (**Supplementary Fig. 7**).

Confirmation of predicted activity and enantiopreference

In the next step we ordered all the putative (*R*)-selective amine transaminase genes as codon-optimized sequences for expression in *Escherichia coli*. They were subcloned and expressed in *E. coli* BL21 and underwent His-tag purification. Then we investigated the purified enzymes with respect to activity, enantioselectivity and enantiopreference toward a range of amines (**1**–**4**) (**Fig. 3**). Additionally, we also examined whether they possessed BCAT or DATA activity (conversion of D-alanine **5** or L-glutamate **6** with concomitant formation of valine). Seventeen of the 21 putative (*R*)-selective amine transaminase genes that had been identified were found to be (*R*)-selective amine transaminases (**Fig. 3** and **Supplementary Tables 3** and **4**). Three proteins could not be expressed in *E. coli* in sufficient amounts, and one protein was found to have very low activity on the substrates studied. Specific activity and substrate range differed greatly among all of the enzymes investigated, which we find unsurprising considering that the protein sequences originate from various microorganisms, that recombinant expression was never reported before and that specific activities within their natural functions as well as their natural substrates are unknown. Nevertheless, 10 out of the 21 proteins showed a specific activity >0.5 U mg^{-1} toward at least one of the investigated amines (**Fig. 3**), which is in the same activity range of known (*S*)-selective amine transaminases[45–47]. (One unit (U) of activity was defined as the amount of enzyme that produced 1 μmol ketone product per minute.)

Consequently, these newly identified (*R*)-selective transaminases can be used in asymmetric synthesis to yield optically pure (*R*)-amines. This was confirmed in preliminary experiments for the asymmetric synthesis of 2-aminohexane (**2**), 2-amino-4-phenylbutane (**3**), 1-*N*-Boc-3-aminopyrrolidine (**4a**) and 1-*N*-Boc-3-aminopiperidine (**4b**) from the corresponding ketones with three of the amine transaminases and resulted in low to moderate yields with excellent enantiomeric excess (ee) up to 99.6% (see **Supplementary Table 5**).

DISCUSSION

The strength of this new *in silico* approach is that new enzymes can be discovered very quickly. In contrast, directed evolution requires several rounds of random mutagenesis or iterative saturation mutagenesis to alter the enantioselectivity or substrate specificity, even if structural information is available to identify hot spots for mutagenesis. Usually more than 10^3–10^4 variants have to be screened for the identification of a mutant with desired properties. In contrast to this very low 'hit rate', at least 50% of the putative proteins identified in our study turned out to be useful biocatalysts.

Hence, before creating mutants or libraries stemming from rational predictions in the laboratory, it is worthwhile to investigate whether nature has already designed variants and possibly optimized these catalysts over millions of years of natural evolution.

The newly identified (*R*)-selective amine transaminases are an ideal starting point for further fine-tuning and optimization by protein engineering, as the requirements for industrial processes are often different from the enzyme's original function in nature. This was demonstrated in a recent study in which the substrate specificity of an (*R*)-amine transaminase was broadened so that bulky substrates could also be converted[34]. Also, other important properties, such as stability at higher temperatures and tolerance of high cosolvent, substrate and product concentration, can be improved significantly and can facilitate the application of amine transaminases for industrial-scale biotransformations.

In the example shown here, we took advantage of the enormous diversity of structures, elucidated mechanisms and substrate specificities already reported for PLP-dependent proteins. Although

such diversity is not yet explored in depth for all enzyme classes, the possibilities for the discovery of new catalysts using this concept will steadily increase on a daily basis as the number of protein sequences and solved structures available in the databases grows.

A second important requirement for the successful application of *in silico* enzyme discovery is a reliable prediction of the necessary key amino acid differences that determine the desired substrate specificity. In this study, the molecular mechanism for the change in substrate specificity from binding a carboxyl group to a small alkyl group—and the related predicted, desired switch in enantiopreference—could be easily rationalized. Method development for understanding the molecular basis and prediction of substrate specificity involving homology modeling and molecular docking are important research fields. Additionally, the application of more complex search algorithms and bioinformatics tools will help to refine such *in silico* enzyme discovery and facilitate its application in cases in which more complex solutions might be required.

METHODS

Alignments. All pairwise and multiple amino acid sequence alignments were done with the computer program STRAP using the ClustalW3D algorithm with standard parameters. The protein sequences annotated as BCATs or PLP-dependent fold class IV enzymes from the NCBI protein database were used for the database search and aligned to *E. coli* BCAT.

Cloning and expression of amine transaminase. The codon-optimized open reading frames (ORFs) encoding proteins 4, 5, 6, 7, 11, 12, 14, 16, 17, 18, 20, 21 and 24 (**Table 1**) were inserted into pGASTON between the NdeI and BamHI restriction sites. The codon-optimized ORFs encoding all other proteins were ordered subcloned in pET-22b. Transformed *E. coli* BL21 (DE3) strains were grown in 400 ml LB medium supplemented with ampicillin (100 μg ml^{-1}). Cells were incubated initially at 37 °C on a gyratory shaker until the OD_{600} reached 0.7. The cells were then induced by addition of 0.2% (w/v) rhamnose (pGASTON) or 0.1 mM IPTG (pET-22b), respectively. At the same time the incubation temperature was decreased to 20 °C, and cultivation continued for 20 h.

Purification of proteins. The cell pellet (~3 g wet weight) was washed twice with phosphate buffer (pH 7.5, 50 mM) containing 0.1 mM PLP at 4 °C. After disruption (French press), the cell suspension was centrifuged (10,000*g*, 30 min), and the resulting supernatant was passed through a 0.5-μm filter before chromatography. Chromatography was performed using an ÄKTA Purifier (GE Healthcare). The filtered cellular extract was applied to a 5-ml column of IMAC Sepharose 6 Fast Flow (GE Healthcare). The column was washed at a flow rate of 5 ml min^{-1} with 10 column-volumes of phosphate buffer (pH 7.5, 50 mM, containing 300 mM NaCl, 0.1 mM PLP and 30 mM imidazole to avoid nonspecific binding), and the active protein was eluted with 10 column-volumes of phosphate buffer (pH 7.5, 50 mM, containing 300 mM NaCl, 0.1 mM PLP and 300 mM imidazole at a flow rate of 5 ml min^{-1}). The fractions with the desired protein were pooled and desalted via gel chromatography with a 20 mM tricine buffer, pH 7.5, containing 0.01 mM PLP. The purified enzymes were stored at 4 °C. Protein concentrations were determined using the BCA assay kit (Uptima) after gel chromatography.

Characterization of substrate specificity. For an initial confirmation of activity, α-methylbenzyl amine (**1**, α-MBA) served as substrate in an acetophenone-based microplate assay[47]: a solution of 2.5 mM (*R*)- or (*S*)-α-MBA and pyruvate was reacted in the presence of the purified enzyme, and the increase in absorbance at 245 nm was correlated to the formation of acetophenone. The conversions of the other amines (**2–4**) were monitored using a conductivity assay[48]: a solution containing 10 mM amine and pyruvate was reacted in the presence of the purified amine transaminase, and the decrease in conductivity was related to the conversion of substrate. For all measurements, either an appropriate amount of purified enzyme was applied, dependent on the specific activity, or the highest concentration possible was applied (0.3–2.2 mg ml^{-1} purified enzyme, dependent on expression and purification) while measuring low activities.

To verify DATA or BCAT activity, the decrease of NADH was measured spectrophotometrically at 340 nm using dehydrogenase-coupled microplate assays: a solution of 5 mM α-ketoglutaric acid and D-alanine **5** was reacted in the presence of the purified transaminase, and 1 U ml^{-1} lactate dehydrogenase and 0.5 mM NADH were used for measuring DATA activity. For measuring BCAT activity, a solution containing 5 mM 3-methyl-2-oxobutyric acid and L-glutamate **6**, 10 mM ammonium chloride, 1 U ml^{-1} glutamate dehydrogenase and 0.5 mM NADH was used. All reactions took place in tricine buffer (pH 7.5, 20 mM) containing 0.01 mM PLP. The pH of the buffer was adjusted with 1,8-diazabicyclo[5.4.0]undec-7-ene. The specific activity was expressed as units per milligram protein, and one unit of activity was defined as the amount of enzyme that produced 1 μmol ketone product per minute.

Asymmetric synthesis of amines 1–4. Preliminary asymmetric syntheses were performed at 30 °C for 24 h in sodium phosphate buffer (100 mM, pH 7) containing PLP (1 mM) and NAD^+ (1 mM) in 1.5-ml Eppendorf tubes. The reaction mixture contained 50 mM ketone, L-alanine (5 equiv., 250 mM), lactate dehydrogenase from bovine heart (90 U), glucose (150 mM) and glucose dehydrogenase (15 U). Amine transaminases from *Aspergillus terreus*, *Mycobacterium vanbaalenii* and *Mesorhizobium loti* (entries 4, 12 and 14 in **Table 1**) were expressed in *E. coli* BL21 as described above, frozen in aliquots and applied directly as whole-cell biocatalysts (~0.05 g wet cell weight per ml) without further purification. The conversion was measured by detection of the formed amines (**1**, gas chromatography; **2–4**, capillary electrophoresis). Chiral analysis of **2–4** was performed using capillary electrophoresis as reported previously[49]. The percent enantiomeric excess values for **1** were analyzed by gas chromatography. After extraction of the amine with ethyl acetate, derivatization to the trifluoroacetamide was performed by adding a 20-fold excess of trifluoroacetic acid anhydride. After purging with nitrogen to remove excess anhydride and residual trifluoroacetic acid, the derivatized compound was dissolved in ethyl acetate (50 μl) and baseline separated using a Shimadzu GC14A that was equipped with a heptakis-(2,3-di-*O*-acetyl-6-*O*-*tert*-butyldimethylsilyl)-β-cyclodextrin column (25 m by 0.25 mm). The retention times were 16.0 min ((*S*)-**1**) and 16.2 min ((*S*)-**2**) using the following oven temperature program: 80 °C for 10 min, heating with 20 °C per min to 180 °C, maintained at 180 °C for a further 10 min.

Received 23 November 2009; accepted 23 August 2010; published online 26 September 2010

References

1. Bornscheuer, U.T. & Kazlauskas, R.J. *Hydrolases in Organic Synthesis* (Wiley-VCH, Weinheim, Germany, 2005).
2. Breuer, M. *et al.* Industrial methods for the production of optically active intermediates. *Angew. Chem. Int. Edn Engl.* **43**, 788–824 (2004).
3. Buchholz, K., Kasche, V. & Bornscheuer, U.T. *Biocatalysis and Enzyme Technology* (Wiley-VCH, Weinheim, Germany, 2005).
4. Schmid, A. *et al.* Industrial biocatalysis today and tomorrow. *Nature* **409**, 258–268 (2001).
5. Schoemaker, H.E., Mink, D. & Wubbolts, M.G. Dispelling the myths—Biocatalysis in industrial synthesis. *Science* **299**, 1694–1697 (2003).
6. Bommarius, A.S. & Riebel, B.R. *Biocatalysis: Fundamentals and Applications* (Wiley-VCH, Weinheim, Germany, 2004).
7. Grunwald, P. *Biocatalysis: Biochemical Fundamentals and Applications* (Imperial College Press, 2009).
8. Liese, A., Seelbach, K. & Wandrey, C. (eds.) *Industrial Biotransformations* (Wiley-VCH, Weinheim, London, UK, 2006).
9. Patel, R.N. (ed.) *Biocatalysis in the Pharmaceutical and Biotechnology Industries* (CRC Press, Boca Raton, Florida, USA, 2006).
10. Pollard, D.J. & Woodley, J.M. Biocatalysis for pharmaceutical intermediates: the future is now. *Trends Biotechnol.* **25**, 66–73 (2007).
11. Straathof, A.J.J., Panke, S. & Schmid, A. The production of fine chemicals by biotransformations. *Curr. Opin. Biotechnol.* **13**, 548–556 (2002).
12. Tao, J., Lin, G.-Q. & Liese, A. (eds.) *Biocatalysis for the Pharmaceutical Industry: Discovery, Development, and Manufacturing* (Wiley-VCH, Weinheim, Germany, 2009).
13. Hou, C.T. (ed.). *Handbook of Industrial Biocatalysis* (CRC Press, Boca Raton, Florida, USA, 2005).
14. Faber, K. *Biotransformations in Organic Chemistry* (Springer-Verlag, Berlin, 2004).
15. Mugford, P.F., Wagner, U.G., Jiang, Y., Faber, K. & Kazlauskas, R.J. Enantiocomplementary enzymes: classification, molecular basis for their enantiopreference, and prospects for mirror-image biotransformations. *Angew. Chem. Int. Edn Engl.* **47**, 8782–8793 (2008).
16. Griengl, H., Schwab, H. & Fechter, M. The synthesis of chiral cyanohydrins by oxynitrilases. *Trends Biotechnol.* **18**, 252–256 (2000).
17. Cheon, Y.-H. *et al.* Crystal structure of D-hydantoinase from *Bacillus stearothermophilus*: insight into the stereochemistry of enantioselectivity. *Biochemistry* **41**, 9410–9417 (2002).
18. Gröger, H. *et al.* Enantioselective reduction of ketones with designer cells at high substrate concentrations: highly efficient access to functionalized optically active alcohols. *Angew. Chem. Int. Edn Engl.* **45**, 5677–5681 (2006).
19. Fox, R.J. *et al.* Improving catalytic function by ProSAR-driven enzyme evolution. *Nat. Biotechnol.* **25**, 338–344 (2007).
20. Kazlauskas, R.J. & Bornscheuer, U.T. Finding better protein engineering strategies. *Nat. Chem. Biol.* **5**, 526–529 (2009).
21. Lutz, S. & Bornscheuer, U.T. (eds.) *Protein Engineering Handbook* (Wiley VCH, Weinheim, Germany, 2009).
22. Turner, N.J. Directed evolution drives the next generation of biocatalysts. *Nat. Chem. Biol.* **5**, 567–573 (2009).
23. Reetz, M.T., Carballeira, J.D. & Vogel, A. Iterative saturation mutagenesis on the basis of B factors as a strategy for increasing protein thermostability. *Angew. Chem. Int. Edn Engl.* **45**, 7745–7751 (2006).

24. Ivancic, M., Valinger, G., Gruber, K. & Schwab, H. Inverting enantioselectivity of *Burkholderia gladioli* esterase EstB by directed and designed evolution. *J. Biotechnol.* **129**, 109–122 (2007).
25. Koga, Y., Kato, K., Nakano, H. & Yamane, T. Inverting enantioselectivity of *Burkholderia cepacia* KWI-56 lipase by combinatorial mutation and high-throughput screening using single-molecule PCR and in vitro expression. *J. Mol. Biol.* **331**, 585–592 (2003).
26. Magnusson, A.O., Takwa, M., Hamberg, A. & Hult, K. An S-selective lipase was created by rational redesign and the enantioselectivity increased with temperature. *Angew. Chem. Int. Edn Engl.* **44**, 4582–4585 (2005).
27. May, O., Nguyen, P.T. & Arnold, F.H. Inverting enantioselectivity by directed evolution of hydantoinase for improved production of l-methionine. *Nat. Biotechnol.* **18**, 317–320 (2000).
28. Williams, G.J., Woodhall, T., Farnsworth, L.M., Nelson, A. & Berry, A. Creation of a pair of stereochemically complementary biocatalysts. *J. Am. Chem. Soc.* **128**, 16238–16247 (2006).
29. Zha, D.X., Wilensek, S., Hermes, M., Jaeger, K.E. & Reetz, M.T. Complete reversal of enantioselectivity of an enzyme-catalyzed reaction by directed evolution. *Chem. Commun. (Camb.)* 2664–2665 (2001).
30. Bartsch, S., Kourist, R. & Bornscheuer, U.T. Complete inversion of enantioselectivity towards acetylated tertiary alcohols by a double mutant of a *Bacillus subtilis* esterase. *Angew. Chem. Int. Edn Engl.* **47**, 1508–1511 (2008).
31. Iwasaki, A., Yamada, Y., Ikenaka, Y. & Hasegawa, J. Microbial synthesis of (*R*)- and (*S*)-3,4-dimethoxyamphetamines through stereoselective transamination. *Biotechnol. Lett.* **25**, 1843–1846 (2003).
32. Matcham, G.W. & Bowen, A.R.S. Biocatalysis for chiral intermediates: Meeting commercial and technical challenges. *Chim. Oggi* **14**, 20–24 (1996).
33. Koszelewski, D., Lavandera, I., Clay, D., Rozzell, D. & Kroutil, W. Asymmetric synthesis of optically pure pharmacologically relevant amines employing ω-transaminases. *Adv. Synth. Catal.* **350**, 2761–2766 (2008).
34. Savile, C.K. *et al.* Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **329**, 305–309 (2010).
35. Jansonius, J.N. Structure, evolution and action of vitamin B6-dependent enzymes. *Curr. Opin. Struct. Biol.* **8**, 759–769 (1998).
36. Percudani, R. & Peracchi, A. The B6 database: a tool for the description and classification of vitamin B6-dependent enzymatic activities and of the corresponding protein families. *BMC Bioinformatics* **10**, 273 (2009).
37. Höhne, M. & Bornscheuer, U.T. Biocatalytic routes to optically active amines. *ChemCatChem* **1**, 42–51 (2009).
38. Cahn, R.S., Ingold, C. & Prelog, V. Specification of molecular chirality. *Angew. Chem. Int. Edn Engl.* **5**, 385–415 (1966).
39. Mizuguchi, H. *et al.* Strain is more important than electrostatic interaction in controlling the pK(a) of the catalytic group in aspartate aminotransferase. *Biochemistry* **40**, 353–360 (2001).
40. Okamoto, A., Nakai, Y., Hayashi, H., Hirotsu, K. & Kagamiyama, H. Crystal structures of *Paracoccus denitrificans* aromatic amino acid aminotransferase: A substrate recognition site constructed by rearrangement of hydrogen bond network. *J. Mol. Biol.* **280**, 443–461 (1998).
41. Sugio, S., Petsko, G.A., Manning, J.M., Soda, K. & Ringe, D. Crystal structure of a D-amino acid aminotransferase: how the protein controls stereoselectivity. *Biochemistry* **34**, 9661–9669 (1995).
42. Goto, M., Miyahara, I., Hayashi, H., Kagamiyama, H. & Hirotsu, K. Crystal structures of branched-chain amino acid aminotransferase complexed with glutamate and glutarate: true reaction intermediate and double substrate recognition of the enzyme. *Biochemistry* **42**, 3725–3733 (2003).
43. Mehta, P.K., Hale, T.I. & Christen, P. Aminotransferases: demonstration of homology and division into evolutionary subgroups. *Eur. J. Biochem.* **214**, 549–561 (1993).
44. Wolf, Y., Madej, T., Babenko, V., Shoemaker, B. & Panchenko, A.R. Long-term trends in evolution of indels in protein sequences. *BMC Evol. Biol.* **7**, 19 (2007).
45. Hanson, R.L. *et al.* Preparation of (*R*)-amines from racemic amines with an (*S*)-amine transaminase from *Bacillus megaterium*. *Adv. Synth. Catal.* **350**, 1367–1375 (2008).
46. Shin, J.-S., Yun, H., Jang, J.-W., Park, I. & Kim, B.-G. Purification, characterization, and molecular cloning of a novel amine:pyruvate transaminase from *Vibrio fluvialis* JS17. *Appl. Microbiol. Biotechnol.* **61**, 463–471 (2003).
47. Schätzle, S., Höhne, M., Redestad, E., Robins, K. & Bornscheuer, U.T. Rapid and sensitive kinetic assay for characterization of ω-transaminases. *Anal. Chem.* **81**, 8244–8248 (2009).
48. Schätzle, S., Höhne, M., Robins, K. & Bornscheuer, U.T. A conductometric method for the rapid characterization of the substrate specificity of amine-transaminases. *Anal. Chem.* **82**, 2082–2086 (2010).
49. Höhne, M., Kühl, S., Robins, K. & Bornscheuer, U.T. Efficient asymmetric synthesis of chiral amines by combining transaminase and pyruvate decarboxylase. *ChemBioChem* **9**, 363–365 (2008).

Author contributions

K.R. and U.T.B. initiated the project. M.H. designed the *in silico* strategy, devised the annotation algorithm and performed the database search and identification of the putative amine transaminases. M.H. expressed and confirmed amine transaminase activity and (*R*)-selectivity for the first three proteins. S.S. coordinated the comparative characterization of all proteins and performed cloning, expression, purification, data collection and data analysis. H.J. contributed to gene cloning, protein expression and activity measurements. U.T.B. and M.H. cowrote the paper, and all authors read and edited the manuscript.

Competing financial interests

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturechemicalbiology/.

Additional information

Supplementary information is available online at http://www.nature.com/naturechemicalbiology/. Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/. Correspondence and requests for materials should be addressed to U.T.B.