A process for the identification and preparation of a (R)-specific omega-transaminase

Description

The present invention relates to processes for screening, characterization and preparation of (R)-specific ω-transaminases, to the ω-transaminases obtained thereby and their use in various transamination processes.

Chiral amines play an important role in the pharmaceutical, agrochemical and chemical industry. They are frequently used as intermediates or synthones for the preparation of various physiologically, for instance pharmaceutically active substances, such as cephalosporine or pyrrolidine derivatives. In a great number of the various applications of chiral amines, only one particular optically active form, either the (R) or the (S) enantiomer is physiologically active. There is accordingly a need to provide processes for the preparation of chiral amines in an optically active form.

These needs are partially met by preparing chiral amines by crystallisation of diastereomeric salts through adding of chiral carboxylic acids (Breuer et al., Angewandte Chemie (2004) 116, 806-843). Other chemical methods use enantioselective synthesis by reducing prochiral precursors with C=N-double bonds.

Among several enzymatic methods that have been employed for the synthesis of optically active amino acids and amines, ω-transaminases (ω-TAs) have recently received increased attention, because of their potential for the asymmetric synthesis of opti-

cally active amines, which are frequently used as building blocks for the preparation of numerous pharmaceuticals.

ω-transaminases are PLP (pyridoxal phosphate) dependent enzymes that catalyze amino group transfer reactions. When employing ω-transaminases, so-called enantioenriched and/or pure optically active amines can principally be produced via two reaction strategies (i) the asymmetric synthesis starting from ketones, and (ii) the kinetic resolution starting from racemic amines. Although ω-transaminases exhibit good enantioselectivity in general, they have not been used frequently in asymmetric synthesis, although in this case a 100% yield is theoretically possible. One specific requirement in an asymmetric synthesis employing ω-transaminases is to shift the equilibrium to the product side, especially when using an amino acid like alanine as amino donor, since in this case the equilibrium is on the side of the substrates (ketone, alanine) and not on the side of the products (amine, pyruvate); another requirement is that the stereoselectivity of the enzyme has to be perfect, which is not always the case for ω-transaminases. Therefore, ω-transaminases are mainly used for the kinetic resolution of racemic amines, where enantioselectivity does not necessarily need to be perfect. Thus, although kinetic resolutions of racemic amines have been investigated, the limitation to a maximum yield of 50 % considerably hampers their application. On the other hand, asymmetric synthesis requires methods to shift the unfavourable equilibrium towards synthesis of single enantiomers of optically pure amines for which several methods were developed, which is one key prerequisite for efficient processes to enable the use of transaminases in industrial scale. Such methods are disclosed in WO 2007/093372.

EP 0 987 332 A1 discloses a process for producing optically active amino compounds, namely (R)-amino compounds, by means of a microbial enzyme, in particular a (R)-selective ω-transaminase derived from Arthrobacter sp. EP 1 038 953 A1 discloses a further (R)-selective ω-transaminase, which, however, is derived from Mycobacterium aurum.

Iwasaki et al. (Biotechnol. Let. 2003 (25), 1843-1846), Koszelewski et al. (Adv. Synth. Catal. 2008 (350), 2761-2766) and Iwasaki et al. (Appl. Microbiol. Biotechnol. 2006 (69), 499-505) disclose a R-specific transaminase from Arthrobacter sp. The identification of microorganism providing useful (S)- or (R)-selective transaminases is usually done by selecting microorganism obtained from for instance soil samples and enriching them in culture (Jun et al., App. Microbiol. Biotechnol. 2004 (70), 2529-2534 and Shin et al. (Biosc. Biotechnol. Biochem. 2001 (65), 1782-1788)). Especially all described R-selective ω-transaminases were obtained by enrichment culture. Such methods are time-consuming, since for an efficient process it is often preferred to overexpress the enzyme in a heterologeous organism like Escherichia coli. This requires the isolation of the gene sequence from the wild type organisms but the cloning of the DNA sequence is not always successful and a very time-consuming process.

In particular, during process development or for the identification of novel ω-TA, purification of the enzyme and characterization of their enzymatic properties is of great interest. Since, however, the ω-TA activity is usually determined with low throughput methods like HPLC or capillary electrophoresis (CE), the determination of the enzyme properties is rather often the limiting step.

In general in comparison to (S)-selective ω-transaminases the number of available (R)-selective ω-transaminase is much more limited. This stands in sharp contrast to the high demand for (R)-selective ω-transaminases which are highly desirable for the asymmetric synthesis of (R)-enantiomers of various chiral-amines. Thus, there is still the need to provide further (R)-selective transaminases and means and methods to obtain them in order to produce R-enantiomers, in particular optically active amines, for instance for pharmaceutical or agrochemical applications, which are up to now either not available at all or not in an economically feasible process, preferably in industrial scale.

The present invention is therefore based on the technical problem to provide simple, fast and reliable means and methods to identify, characterize and obtain (R)-selective ω-transaminases, in particular for the production of desired (R)-enantiomers, preferably in optically pure form, which are preferably suitable for identification, characterization and preparation in an industrial scale.

The present invention solves its technical problem by the provision of the teaching of the independent claims, in particular the claimed process, the products and uses obtained thereby.

Thus, the present teaching provides in a particular embodiment a process for the preparation of a (R)-selective ω-transaminase (also called ω-TA), comprising the following steps:

a) providing at least one query biomolecule sequence of at least one transaminase or lyase, preferably a transaminase or lyase which belongs to fold-type IV of PLP-dependent enzymes, and at least one biomolecule bank,

b) searching the biomolecule bank with the query biomolecule sequence to identify a group of first target biomolecule sequences, wherein the first target biomolecule sequences have a degree of sequence identity of at least 20%, preferably 32%, to the query biomolecule sequence, calculated on amino acid level,

c) selecting in the group of first target biomolecule sequences a group of second biomolecule target sequences, which do not comprise, on amino acid level, at least one of the following amino acid sequence motives c1) to c3) with

c1) at position 95 to 97 an amino acid sequence Tyr Xa1 Xa2, with Xa1 being an amino acid Ile, Val, Leu, Met, Phe, and Xa2 being an amino acid Arg or Lys or

c2) at position 97 to 99 an amino acid sequence Tyr Xaa Gln, with Xaa being an amino acid, preferably a conventional amino acid and in the region from position 105 to 111, preferably at position 107 to 109, an amino acid sequence Arg Xaa Xa3, Xa3 being an amino acid, preferably being His or

c3) at position 38 Thr, at position 97 Lys and at position 107 to 109 an amino acid sequence Arg Xa4 Xa5, Xa4 being an amino acid, preferably being Gly, and Xa5 being an amino acid, preferably being Tyr,

and which comprise

c4) at position 95 another amino acid than Tyr, Arg, Lys, or at position 95 Tyr, but at position 97 no Arg or Lys and

c5) at position 40 no Lys or Arg and

c6) in the region from position 161 to 165, preferably at position 163, Lys

to identify a group of second target biomolecule sequences and

d) providing, preferably preparing, a biomolecule having the second biomolecule target sequence identified in step c) and being or coding at least partially a protein with the activity of a (R)-selective ω-transaminase.

The term "a biomolecule being or coding at least partially a protein" means preferably that the biomolecule may either be a protein with the activity of a (R)-selective ω-transaminase or, in case the biomolecule is a nucleotide sequence, at least a part of said nucleotide sequence codes said protein, preferably a full-length protein. Accordingly, in case the provided biomolecule is a nucleotide sequence molecule, at least a part of said nucleotide sequence molecule codes a protein with the activity of a (R)-selective ω-transaminase, whereby possibly a further part of said nucleotide molecule has some other function, for instance regulatory or replicative function. Thus, the term "providing a biomolecule having the second biomolecule target sequence identified in step c) and being or coding at least partially a protein with the activity of a (R)-selective ω-transaminase" is preferably equivalent to the term "providing a biomolecule having the second biomolecule target sequence identified in step c) which may

be a protein with the activity of a (R)-selective ω-transaminase or which may be a nucleotide sequence molecule which includes a sequence coding said protein as long as the protein has the activity of a (R)-selective ω-transaminase".

In one embodiment of the present invention in step c2) the position of the Arg Xaa Xa3-motive may vary by 1-2 amino acids from positions 107 to 109, i. e. may be at positions 105 to 111.

In one embodiment of the present invention in step c6) the position of Lys may vary by 1-2 amino acids at position 163, i. e. may be at positions 161 to 165.

In the context of the present invention each amino acid sequence, in particular amino acid sequence motive, or DNA sequence or DNA sequence motive given is given in the amino terminal to carboxy terminal or 5′ to 3′ direction, whichever applies, and unless otherwise specified. Preferably, the sequences given in the given direction are given as a continuous stretch without any intervening nucleotides or amino acids, whichever applies.

In a preferred embodiment of the present invention the at least one query biomolecule sequence of at least one transaminase as used in step a) is preferably a (R)-selective ω-transaminase. In a further preferred embodiment the at least one query biomolecule sequence of the at least one transaminase as used in step a) is selected from the group consisting of a (R)-selective ω-transaminase, a branched chain amino acid amino transferase (BCAT) and a D-amino acid transaminase (DATA). Preferably, the branched chain amino acid amino transferase is from E. coli. In a preferred embodiment of the present invention the at least one query biomolecule sequence of the

at least one lyase as used in step a) is an amino deoxychorismat-lyase (ADCL).

The present invention provides a process for the preparation of a (R)-selective ω-transaminase in a fast, efficient, simple, reliable and easy manner. The present invention therefore allows to identify and prepare (R)-selective ω-transaminase, which have not been known or accessible before, opening up numerous ways to provide chiral amines, preferably (R)-enantiomers, in particular in very efficient asymmetric synthesis routes. It could be successfully shown that even proteins with a low degree of sequence identity, for instance only 35%, to known (R)-selective ω-transaminase in contrast to the expectations are in fact (R)-selective ω-transaminases. In a particularly advantageous and unexpected manner the present invention allows preferably to provide (R)-selective ω-transaminases from Aspergillus terreus and Mesorhizobium loti, whereby the transaminase from Aspergillus is the first eucaryotic ω-transaminase and it converts substrates with (R)-selectivity.

Basically, the present invention is based on the technical teaching that proteins or its encoding nucleotide sequences, even with a low degree of sequence identity to known transaminases or lyases, in particular (R)-selective ω-transaminases, can be used as a potential source for identifying and preparing (R)-selective ω-transaminases, whereby the putative (R)-selective ω-transaminases are screened and prepared by discriminating undesired and thereby positively selecting for proteins displaying particular sequence motives in the amino acid sequence, which attribute to the desired enzymatic activity. Thus, the present invention specifically uses particular structural features, in particular the absence and presence of particularly

amino acids in putative (R)-selective ω-TA, to identify and prepare (R)-selective ω-transaminases.

Accordingly, in a first step a) of the present process, a query biomolecule sequence of at least one transaminase or lyase, for instance a known (R)-selective ω-transaminase, or at least a characteristic sequence part for such a lyase or transaminase, preferably of the (R)-specific ω-transaminase, which preferably is in particular able to identify a PLP-dependent enzyme of fold type IV, is provided together with at least one biomolecule bank. Preferably, the query biomolecule sequence is an ORF (open reading frame) of a transaminase or lyase, in particular of the (R)-selective ω-TA, coding nucleotide sequence, or a characteristic part thereof. In another embodiment the query biomolecule sequence is the ORF-amino acid sequence itself or a characteristic part thereof.

In the context of the present invention a characteristic part of the query biomolecule sequence of at least one transaminase or lyase, in particular of the known (R)-selective ω-transaminase, is a biomolecule sequence which in form of its DNA sequence molecule hybridises under the following conditions to the full length query biomolecule sequence, in particular to the DNA sequence of the ORF.

Methods for the hybridization of nucleic acids such as DNA are well known and are described for example in Molecular Cloning, Third Edition (2001); Methods for General and Molecular Bacteriology, ASM Press (1994); Immunology methods manual, Academic Press (Molecular), and many other standard textbooks.

An example for a hybridization under stringent conditions is as follows. A filter with a nucleic acid immobilized thereon and the nucleic

acid used as probe are incubated in a solution comprising 50% formamide, 5×SSC (750 mM sodium chloride and 75 mM sodium citrate), 50 mM sodium phosphate (pH 7.6), 5×Denhardt's solution, 10% dextran sulfate and 20 µg/l denatured salmon sperm DNA at 42° C overnight. After incubation, the filter is washed in 0.2×SSC solution (ca. 65° C). These stringent hybridizations conditions can be modified by adjusting the concentration of formamide (the conditions become less stringent as the concentration of formamide is lowered) and by changing the salt concentrations and the temperature conditions.

Hybridization under less stringent conditions is carried out, for example, by incubating a filter with a nucleic acid immobilized thereon and a nucleic acid used as probe in a solution comprising 6×SSCE (20×SSCE: 3 mol/l sodium chloride, 0.2 mol/l sodium dihydrogenphosphate and 0.02 mol/l EDTA, pH 7.4), 0.5% SDS, 30% formamide and 100 µg/l denatured salmon sperm DNA at 37° C overnight, and washing the filter with 1×SSC solution containing 0.1% SDS (50° C).

In a particularly preferred embodiment the present invention understands under the term "at least one query biomolecule sequence of at least one transaminase or lyase" a biomolecule sequence, which is suitable to select for biomolecule sequences in accordance with the presence or absence of sequence motives c1) to c6) as identified herein. Accordingly, such a query biomolecule sequence is a sequence screening for and identifying those biomolecule sequences which do not comprise on amino acid level at least one of the amino acid sequence motives c1) to c3) and which comprise on amino acid level sequence motives c4), c5) and c6). Such a query biomolecule

sequence may be embodied as an amino acid sequence information or as a DNA sequence molecule or DNA sequence information.

In a preferred embodiment the characteristic part of the query biomolecule sequence used in step a) encompasses, preferably consists of, the region from positions 30 to 170, most preferably 30 to 120 of the ORF of a transaminase, preferably a (R)-selective ω-transaminase, or a lyase.

The biomolecule bank is in a subsequent step b) searched with the query biomolecule sequence to identify a group of first target biomolecule sequences, which show at least a minimum degree of sequence identity of at least 20%, preferably 25%, preferably 32%, preferably at least 33%, most preferably 34%, at least 35%, at least 36%, at least 40%, at least 50%, at least 60%, at least 70% and at least 80%, at least 90% or at least 95% to the query biomolecule sequence, based on amino acid level, and wherein said group of first target biomolecules represents a first selection from the biomolecule bank used in step a). Said degree of sequence identity is preferably a sequence identity between at least the characteristic part of the, preferably essentially the complete, in particular the complete, ORF of the query biomolecule sequence and at least the characteristic part of the, preferably essentially the complete, in particular the complete ORF of the biomolecule sequences screened. Subsequent to said step b) in this group of first target biomolecule sequences, those sequences are selected in a step c), which do not comprise as a sequence motive c1) at position 95 to 97 an amino acid sequence Tyr Xa1 Xa2, with Xa1 being an amino acid Ile, Val, Leu, Met, Phe, and Xa2 being an amino acid Arg or Lys or which do not comprise as sequence motive c2) at position 97 to 99 an amino acid sequence Tyr

Xaa Gln, with Xaa being an amino acid, preferably a usual amino acid, and in the region from position 105 to 111, preferably at position 107 to 109, an amino acid sequence Arg Xaa Xa3, Xa3 being an amino acid, preferably an usual amino acid, preferably being His or which do not comprise sequence motive c3) at position 38 Thr, at position 97 Lys and at position 107 to 109 an amino acid sequence Arg Xa4 Xa5, Xa4 being an amino acid, preferably an usual amino acid, preferably being Gly, and Xa5 being an amino acid, preferably an usual amino acid, preferably being Tyr, thereby discriminating and discharging those sequences, which do display at least one of the above-identified sequence motives c1), c2) or c3). In a further subsequent or simultaneous selection those biomolecule sequences are selected, which according to sequence motive c4) do show at position 95 another amino acid than Tyr, Arg, Lys, or at position 95 Tyr, but at position 97 no Arg and no Lys and which according to sequence motive c5) have at position 40 no Lys and no Arg and which according to sequence motive c6) have in the region from position 161 to 165 at least one Lys, preferably one Lys, preferably at position 163 Lys, so as thereby to select and identify a group of second biomolecule target sequences. The group of second target biomolecule sequences identified and screened for the above-identified sequence motives represent biomolecule sequences being or encoding a protein with the activity of an (R)-selective ω-transaminase, which is provided thereby.

In the context of the present invention, a transaminase is a pyridoxalphosphate-dependent enzyme catalysing the transfer of amino groups, being preferably classified in folding type IV. Transaminases are classified in E.C. 2.6.1.X. In a particularly preferred embodiment of the present invention, the transaminase is a (R)-selective trans-

aminase, particular is in a preferred embodiment an ω-transaminase. In the context of the present invention a protein with the activity of an (R)-selective ω-transaminase is a protein which is able under appropriate reaction conditions to catalyse a transfer of nitrogenous groups such as amino groups from the donor to an acceptor such as a (R)-selective ω(omega)-transaminase (beta-alanine-pyruvate transaminase) is able to do so. In context of the present invention a (R)-selective ω-transaminase is preferably an enzyme with the classification code E.C.2. 6.1.18.

In the context of the present invention the term optically active chiral amine relates to the same subject-matter as the term enantiomerically active chiral amine. These terms in particular refer to a preparation which is essentially free, in an even more preferred embodiment free of the undesired enantiomer. Accordingly, an optically active chiral amine essentially comprises an excess of one enantiomer or even consists of only one enantiomer.

In particular, in the context of the present invention, an optically active chiral amine has an optical purity of at least 70, 80, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.5, 99.6, 99.7, 99.8 and in particular at least 99.9%.

In the present invention the optical purity is given in % excess of one enantiomer over the other enantiomer. Thus, the optical purity in % is the quotient of the difference between the (R) and the (S) enantiomer concentrations and the sum of the concentrations of both enantiomers (optical purity of A in % = ([A]-[B]): ([A]+[B]) x 100, wherein A and B represent the concentrations of the (R) and (S) enantiomers or vice versa).

In the context of the present invention a biomolecule bank is a source of biomolecules itself or a collection of sequence data of biomolecules, in particular polynucleotide or polypeptide sequence data.

In the context of the present invention a biomolecule is preferably a polynucleotide molecule carrying genetic information, in particular a DNA molecule. In a further preferred embodiment of the present invention a biomolecule is a polypeptide, in particular protein, comprising a number of amino acids. Thus, a biomolecule bank may either be a physical source of biomolecules, in particular may be a gene bank, in particular a cDNA- or genomic library or is a collection of information about said biomolecules, in particular a collection of sequence data, in particular amino acid sequences or polynucleotide sequence, in particular DNA sequences.

The present invention refers also to biomolecule sequences, amino acid sequences, polynucleotide sequences or nucleotide sequences, in particular DNA sequences, whereby said wording designated on one hand the physical substance per se, that means a protein, a polypeptide or a DNA molecule, and on the other hand the information content of said chemical substance, which, in case a polynucleotide is referred to is the type, order and number of its nucleotides and, in case a polypeptide is referred to, is the type, order and number of the single amino acids forming the polypeptide.

In the context of the present invention the terms “biomolecule”, “DNA molecule”, “polynucleotide molecule”, “polypeptide” or “protein” refers to chemical substances per se. In case, the present invention specifically refers to an amino acid sequence information, a DNA sequence information, a polynucleotide sequence information or a

biomolecule sequence information, it is referred not to the physical form of a biomolecule, but rather to the information contained therein, i. e. the type, order and number of its constituents, namely the amino acids or nucleotides.

The present invention is, in one preferred embodiment A), in its steps a) to c) applicable to a biomolecule bank being a collection of sequence data of, in one embodiment, polynucleotide sequences, in particular DNA sequences or, in another embodiment, amino acid sequences, both of which are searched in a step b) with sequence alignment tools, preferably such as BLAST, which is a basic local alignment search tool, to identify a group of first target biomolecule sequences (Altschul, S.F. et al. 1990. J. Mol. Biol. 215:403; Altschul, S.F. et al. 1997. Nucleic Acid Res. 25:3389-3402). Other suitable programs include GAP, BESTFIT and FASTA in the Wisconsin Genetics Software Package (Genetics Computer Group (GCG), Madison, WI, USA). In a further step c) the group of identified first biomolecule sequences is subjected to further selection steps in the course of which the identified sequence motives are used to identify and select negatively with sequence motives c1) to c3) and select positively with sequence motives c4) to c6) the desired biomolecule sequence. Once the second group of biomolecule sequences is identified, the present invention uses said sequence information to prepare corresponding oligo- or polynucleotide molecules, for instance hybridisation primers, to screen and select in physical forms of biomolecule banks DNA sequence molecules encoding the enzymes identified in the second group. Thus, in a preferred embodiment the sequence information obtained in step c) can be used for cloning of a corresponding gene from a genomic DNA or cDNA library and expression of the protein in, for instance E. coli, Sac-

charomyces cerevisiae, Pichia pastoris or others. In another preferred embodiment of the present invention the sequence information obtained in step c) is used to de novo synthesize the desired (R)-selective ω-transaminase. In such a preferred embodiment it is possible to use the sequence information obtained in step c) for the synthesis of a gene, with for instance an optimized codon usage and mRNA stability, and cloning and expression of such a gene. According to such a preferred embodiment it is possible to use as the at least one biomolecule bank in step a) either a biomolecule bank containing DNA sequence information or amino acid sequence information. In case, the biomolecule bank contains DNA sequence information said information either has to be translated into amino acid sequence information in order to be processed according to steps b) to c), in which amino acid sequences are used as a query biomolecule sequence and as amino acid sequence motives or steps b) to c), which are carried out in the DNA sequence bank using query biomolecule sequences and sequence motives c1) to c6) translated back into DNA sequence information.

In another preferred embodiment of the present invention the present teaching is applied in an embodiment B) in step a) to a biomolecule bank which is present in physical form of a gene bank, in particular gene library such as a cDNA-, metagenome or genomic library, using DNA molecules encoding all or at least the characteristic part of at least one (R)-selective ω-transaminase, to search for, identify and select a group of first biomolecule sequences, in particular DNA sequence molecules, which, calculated on amino acid level, have a degree of sequence identity of at least 20%, preferably 32% to the query DNA sequence molecule. The subsequent step c) is preferably carried out in said group of first DNA molecules using nu-

cleotide molecule primers identifying and positively and negatively selecting for the desired sequence motives c1) to c6).

In the context of the present invention the term "sequence motive" refers to the specific selective characteristics of the amino acid sequence of the putative (R)-selective ω-transaminase as identified specifically in step c) of the claimed process. In particular, the first sequence motive c1) is the sequence motive identified at position 95 to 97 being in this order an amino acid sequence Tyr Xa1 Xa2, with Xa1 being an amino acid Ile, Val, Leu, Met, Phe, and Xa2 being an amino acid Arg or Lys and which should according to the present invention not be present in the finally selected biomolecule sequence. The second sequence motive c2) is the sequence motive identified at position 97 to 99 being in this order an amino acid sequence Tyr Xaa Gln, with Xaa being an amino acid and in the region from position 105 to 111, preferably at position 107 to 109, an amino acid sequence being in this order Arg Xaa Xa3, with Xaa being an amino acid and Xa3 being an amino acid, preferably being His , and which also should not be present in the finally selected biomolecule sequence. The third sequence motive c3) is the sequence motive characterized by the presence of Thr at position 38, at position 97 Lys and at position 107 to 109 an amino acid sequence Arg Xa4 Xa5, Xa4 being an amino acid, preferably being Gly and Xa5 being an amino acid, preferably being Tyr , which motive c3) should not be present in the finally selected biomolecule sequence.

The fourth sequence motive c4) requires that at position 95 another amino acid than Tyr, Arg, Lys, or at position 95 Tyr, but at position 97 no Arg or Lys is present, which qualifies for a putative (R)-selective ω-transaminase prepared according to the present invention. The

fifth sequence motive c5) requires that at position 40 no Lys and no Arg is present in the finally prepared and identified putative (R)-selective ω-transaminase. The sixth sequence motive c6) requires that in the region from position 161 to 165 at least one Lys, preferably one Lys, preferably at position 163 Lys, is present in the finally prepared and identified (R)-selective ω-TA.

In the context of the present invention, the identification and location of the amino acid positions is determined as follows.The first target biomolecule sequences are aligned, preferably multiply aligned to each other and the query biomolecule sequence. Alignment can be done with conventional alignment software, such as STRAP, in particular ClustalW, preferably ClustalW3D. In another embodiment it is also possible to align the sequences pairwise, i. e. each first target biomolecule sequence to the query biomolecule sequence. In a preferred embodiment the first target biomolecule sequences are aligned to the BCAT of E. coli. In another preferred embodiment the first target biomolecule sequences are aligned to another query biomolecule sequence used in the present invention, for instance the (R)-selective ω-transaminase.

The annotation of the amino acid positions as given in the present invention is determined by the position of the corresponding sequence motive in the query biomolecule sequence used as the positional standard. Alignment as described above aligns the corresponding amino acid positions of the first target biomolecule sequences to the sequence motives present in the query biomolecule sequence.

As an example, using the BCAT of E. coli as standard for the annotation of the position the known E. coli BCAT amino acid sequence from position 92 to 100, namely TSAYIRPLI (SEQ ID no. 9) is used to identify positions 95 to 99 in the putative ω-transaminase, which is analysed for the absence of the sequence motives c1), c2), c3) and in the presence of c4). The known amino acid sequence from position 35 to 42, namely VFEGIRCY (SEQ ID no. 10) of the E. coli BCAT marks the position of the G at position 38 for sequence motive c3) and c5). The amino acid sequence DVGMGVNP in the E. coli BCAT amino acid sequence (SEQ ID no. 11) marks position 104 to 111 for sequence motive c3) at positions 105 to 111. The amino acid sequence PTAAKAGGN from positions 159 to 167 of the known E. coli BCAT (SEQ ID no. 12) marks position 163 as being a K for sequence motive c6).

Thus, in a preferred embodiment of the present invention the biomolecule is a protein and the biomolecule sequence is an amino acid sequence. Accordingly, in one preferred embodiment the biomolecule bank is a bank, in particular database, which bank is a bank with collected information on various proteins, in particular amino acid sequence information. Preferred data base banks are the NCBI protein data bank, the UniProtKB/SwissProt and the UniProtKB/TrEMBL data bank.

In a preferred embodiment it is also possible that the biomolecule is a DNA molecule and the biomolecule sequence is a DNA sequence. Accordingly, in one preferred embodiment, the biomolecule bank is a bank, in particular database, with collected information on various polynucleotide sequences, in particular DNA sequences.

In a preferred embodiment the present invention uses a process according to the above, wherein the biomolecule bank is a biomolecule database and the biomolecule database is searched in step b) with a biomolecule sequence alignment program, in particular BLAST. In a preferred embodiment it is foreseen that if the biomolecule bank is a biomolecule database, the biomolecule database is searched in step b) with either an amino acid sequence, if the biomolecule database is an amino acid sequence database, or with a DNA sequence, if the biomolecule database is a DNA sequence database.

The invention foresees in step d) finally to provide a biomolecule, preferably a DNA molecule or an amino acid sequence molecule, i.e. protein, either by de novo synthesis of the transaminase or by isolating from a physical gene bank polynucleotides encoding the desired transaminase with the help of primers, being defined on the base of the second biomolecule sequences. The obtained polynucleotides, in particular DNA sequence molecules, are used to be expressed under appropriate conditions in an appropriate culture medium to express the desired transaminase.

The present invention also relates to a process for the screening and identification of a (R)-selective ω-transaminase, comprising the above-identified steps a) to c), in particular providing at least one query biomolecule sequence of at least one (R)-selective ω-transaminase and at least one biomolecule bank, searching the biomolecule bank with the query biomolecule sequence to identify a group of first target biomolecule sequences, wherein the first biomolecule sequences have a degree of sequence identity of at least 20%, preferably 25%, preferably 32% to the query biomolecule sequence, calculated on amino acid level, selecting in the group of first

target biomolecule sequences biomolecule sequences, which do not comprise, on amino acid level, anyone of the sequence motives c1) to c3) with c1) being at position 95 to 97 an amino acid sequence Tyr Xa1 Xa2, with Xa1 being an amino acid Ile, Val, Leu, Met, Phe, and Xa2 being an amino acid Arg or Lys or c2) being at position 97 to 99 an amino acid sequence Tyr Xaa Gln, with Xaa being an amino acid and in the region from position 105 to 111, preferably 107 to 109, an amino acid sequence Arg Xaa Xa3, Xa3 preferably being an amino acid, preferably His or c3) being at position 38 Thr, at position 97 Lys and at position 107 to 109 an amino acid sequence Arg Xa4 Xa5, Xa4 being an amino acid, preferably being Gly, and Xa5 being an amino acid, preferably being Tyr, and which do comprise sequence motives c4), c5) and c6) with c4) being on amino acid level at position 95 another amino acid than Tyr, Arg, Lys, or at position 95 Tyr, but at position 97 no Arg or Lys and c5) being at position 40 no Lys or Arg and c6) being in the region from position 161 to 165, preferably at position 163, Lys to identify a group of second biomolecule sequences, which are biomolecule sequences of a (R)-selective ω-TA.

In a furthermore preferred embodiment of the present invention the present teaching provides (R)-selective ω-transaminases being obtainable according to the preparation processes of the present invention. In particular, the present invention provides proteins and DNA sequences, in particular DNA molecules, encoding said protein from Mesorhizobium loti and Aspergillus terreus as identified in SEQ ID no. 1 and 2 for Mesorhizobium loti and 3 and 4 for Aspergillus terreus. The present invention provides in particular and in a most preferred embodiment the teaching that the (R)-selective ω-transaminases obtainable or prepared according to the present in-

vention, for instance those identified in SEQ ID no. 1 and 3, can be used in a transamination reaction, in particular can be used in a process for the preparation of an optically active chiral amine, preferable an (R)-enantiomer of a chiral amine, comprising reacting at least one amino acceptor and at least one amino donor with the (R)-selective ω-transaminase according to the present invention, in particular of SEQ ID no. 1 or 3, and obtaining the optically active chiral amine. In a preferred embodiment the process for the preparation of an optically active chiral amine is an asymmetric synthesis using as a keto-group containing compound preferably a ketone and an amino donor. In another preferred embodiment the preparation method is using a kinetic resolution reaction. With an amino donor being preferably a mixture of racemic amines and an amino acceptor, preferably ketones, as educts.

In the preferred embodiment of the present invention according to which an optically active chiral amine is synthesized by using a (R)-selective ω-transaminase of the present invention in an asymmetric synthesis starting from ketones the preferred degree of conversion into the desired optical active chiral amine, i.e. the (R)-enantiomer is at least 70%, 80%, 90%, 95%, 98%, 99%, 99.5% and most preferably 100%.

In another preferred embodiment according to which the (R)-selective ω-transaminase of the present invention is used in a kinetic resolution reaction starting from racemic amines the preferred degree of conversion into the optically active chiral amine, preferably the (S)-enantiomer, is at least 30, 40, 45, 46, 47, 48, 49, in particular 50%.

The concentrations for analysing the optical purity and the conversion can be determined for instance using HPLC, capillary electrophoresis (CE), gas chromatography (GC) or photo- or fluorimetric methods.

Thus, in a preferred embodiment, the present invention relates to a process for the preparation of an optical active chiral amine, said process comprising reacting an amino acceptor compound comprising a keto group and a racemic mixture of an amine in the presence of a transaminase, in particular a (R)-selective ω-transaminase, preferably according to the present invention, to obtain an (S)-enantiomer of the chiral amine.

In another preferred embodiment of the present invention there is provided a process for the preparation of an optical active chiral amine, in particular an (R)-enantiomer of said amine that process comprising reacting an amino acceptor compound comprising a ketogroup and an amino donor in the presence of a (R)-selective ω-transaminase, in particular obtainable according to the present invention, to obtain an (R)-enantiomer of the amine.

In the context of the present invention an amino acceptor is a molecule capable of accepting an amino group transferred from an amino donor by a transaminase, in particular an ω-transaminase. In a particularly preferred embodiment of the present invention the amino acceptor contains a ketone functionality.

In a particularly preferred embodiment of the present invention the amino acceptor is selected from the group consisting of phenylpyruvic acid, a salt thereof, pyruvic acid, a salt thereof, acetophenone, 2-ketoglutarate, 3-oxobutyrate, 2-butanone, 3-oxopyrrolidine (3-OP),

3-pyridylmethylketone (3-PMK), 3-oxobutyric acid ethyl ester (3-OBEE), 3-oxopentanoic acid methyl ester (3-OPME), N-1-boc-oxopiperidinone, N-1-boc-3-oxopyrrolidine (B3OP), 3-oxo-piperidine, N-1-boc-3-oxopiperidine (B3OPi), 1-Cbz-3-oxopiperidine (C3OPi), 1-Cbz-3-oxopyrrolidine (C3OP), alkyl-3-oxo-butonoates, methoxyacetone and 1-oxotetralone.

In the context of the present invention an amino donor is a molecule capable of providing an amino group to an amino acceptor using a transaminase, in particular an ω-transaminase. In a particular preferred embodiment the amino donor is an amine or amino acid.

In a particularly preferred embodiment the amino donor is selected from the group consisting of β-alanine, alanine, α-methylbenzylamine (α-MBA), glutamate, phenylalanine, glycin, 3-aminobutyrate, isopropylamine, 2-aminobutane and γ-aminobutyrate or a salt, for instance a chloride, of any one thereof. In a particularly preferred embodiment the obtained ketone product may be phenylpyruvic acid, a salt thereof, pyruvic acid, a salt thereof, glyoxylic acid, a salt thereof, acetophenone, 2-ketoglutarate, acetone, 3-oxobutyrate, 2-butanone, 3-oxopyrrolidine (3-OP), 3-pyridylmethylketone (3-PMK), 3-oxobutyric acid ethyl ester (3-OBEE), 3-oxopentanoic acid methyl ester (3-OPME), N-1-boc-oxopiperidinone and N-1-boc-3-oxopyrrolidine (B3OP) or a salt, for instance a chloride, of any one thereof.

In a further preferred embodiment the present invention relates to a process for the preparation of an optically active chiral amine which is selected from the group of amines having an optically active amino group, in particular amines with alkylgroups, branched alkylgroups or arylalkylgroups. In particular, these amines, in particular mono- or

bicyclic amines, are in particular amines of 5 to 6-membered cyclic or S-, O-, or N-substituted heterocyclic hydrocarbons or aromatic amines, in particular alkyl- or alkoxy-substituted aromatic amines. In a preferred embodiment, the obtained chiral amines are selected from the group consisting of phenylalanine, alanine, 3-aminopiperidine, alkyl-3-amino-butanoates, 3-aminopyrrolidine (3-AP), 3-pyridyl-1-ethylamine (3-PEA), N-1-boc-3-aminopyrrolidine (B3AP), N-1-boc-3-aminopiperidine (B3APi), 1-Cbz-3-amino-piperidine (C3APi), 1-Cbz-3-aminopyrrolidine (C3AP), 3-aminobutyric acid ethyl ester (3-ABEE), 3-aminopentanoic acid methyl ester (3-APME), α-methylbenzylamine (α-MBA), 1-aminotetraline, 3,4-dimethoxy phenyl acetone, α-methyl-4-(3-pyridyl)-butanamine, γ-aminobutyrate, glutamate, isopropylamine, β-aminobutyrate, sec-butylamine, methoxyisopropylamine, derivatives of 3-aminopyrrolidine, 1-N-Boc-3-aminopiperidine, cephalosporine and derivatives of cephalosporine.

In a particularly preferred embodiment the present invention therefore foresees reacting 3OP with an (R)-selective ω-transaminase and an amino donor to obtain optically active (R)-3AP.

In a further preferred embodiment, the present invention foresees reacting 3-PMK with an (R)-selective ω-transaminase and an amino donor to obtain optically active (R) 3-PEA.

In a further preferred embodiment of the present invention, the invention foresees reacting 3-OBEE with an (R)-selective ω-transaminase and an amino donor to obtain optically active (R) 3-ABEE.

In a further preferred embodiment the invention foresees reacting 3-OPME with an (R)-selective ω-transaminase and an amino donor to obtain optically active (R) 3-APME.

In a further preferred embodiment the invention foresees reacting B3OP with an (R)-selective ω-transaminase and an amino donor to obtain optically active (R)-B3AP.

In a further preferred embodiment, the present invention foresees reacting B3OPi with an (R)-selective ω-transaminase and an amino donor to obtain optically active (R)-B3APi.

In a further preferred embodiment, the invention foresees reacting C3OPi with an (R)-selective ω-transaminase and an amino donor to obtain optically active (R)-C3APi.

In a further preferred embodiment, the invention foresees reacting C3OP with an (R)-selective ω-transaminase and an amino donor to obtain optically active (R)-C3AP.

In a further preferred embodiment of the present invention the invention foresees reacting acetophenone with an (R)-selective ω-transaminase and an amino donor to obtain optically active (R) α-MBA.

In a further preferred embodiment the present invention foresees reacting as an amino acceptor, in particular mono- or bicyclic, oxogroup-containing 5 to 6 membered cyclic or S-, O-, or N-substituted heterocyclic hydrocarbons or aromatics, in particular alkyl- or alkoxy-substituted aromatics with an amino donor and an (R)-selective ω-transaminase to obtain amines, in particular mono-

or bicyclic amines, in particular amines of 5 to 6 membered cyclic or S-, O-, or N-substituted heterocyclic hydrocarbons or aromatic amines, in particular alkyl- or alkoxy-substituted aromatic amines, in particular in (R) form.

In a particularly preferred embodiment of the present invention, the amino acceptor and the amino donor are reacted with the transaminase in aqueous medium, for example physiological buffer. In a particularly preferred embodiment the transamination reaction is carried out at a pH in the range from 5 to 9, in particular from 7 to 8,5. In a particular preferred embodiment, the reaction is carried out in a temperature range from 10 to 65°C, preferably 20 to 50°C, in particular 18 to 25°C, preferably room temperature or 34°C to 39°C, in particular 37°C. In a further preferred embodiment of the present invention the amino acceptor and the amino donor are provided in a molar ratio from 1:1 to 1:5, in particular from 1:1 to 1:2. In a preferred embodiment of the present invention the enzymatic activity may be from 1 to 20.000 µmol/min.

The present invention also relates to a process for the analysis of a transaminase, in particular for the characterization of properties of a transaminase, comprising the following steps:

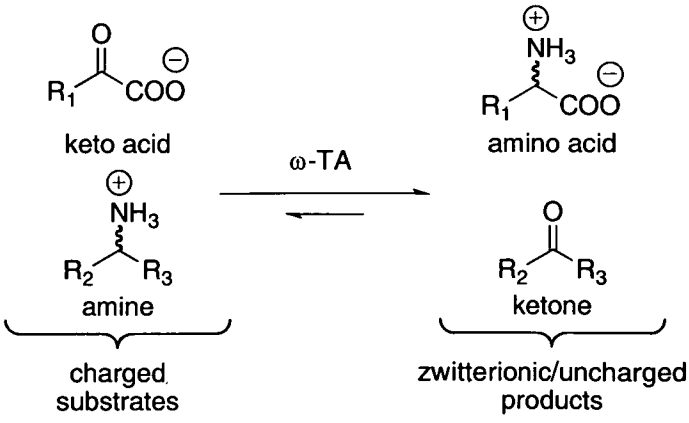
i. providing a charged amino acceptor, a charged amino donor and a transaminase, preferably a ω-transaminase, most preferably a (R)-selective ω-transaminase obtainable according to the present invention

ii. reacting the amino acceptor and the amino donor with the transaminase in a reaction medium, and thereby

iii. determining the conductivity of the reaction medium under a first set of reaction conditions and

iv. subsequently to step iii) determining the conductivity of the reaction medium under a second set of reaction conditions, so as to obtain at least two conductivity values reflecting the properties of the transaminase.

In the course of a transaminase, preferably ω-TA-catalyzed, reaction of the charged substrates amino donor, preferably an amine, and the amino acceptor, preferably a keto acid, for instance pyruvate, the conductivity decreases since a non-charged ketone and the zwitterionic amino component, preferably amino acid, for instance alanine are formed.



The present process for analysis allows a simple measurement of the reaction progress. Preferably, a low conductivity buffer, particularly the low conducting CHES (N-Cyclohexyl-2-aminoethanesulfonic acid) buffer is most suitable in order to avoid a too high initial conductivity. Preferably, a calibration of the conductivity process can be done by simulation of different conversions. As an example, for the standard substrate pair α-methylbenzylamine and pyruvate, a 1 mM

conversion corresponds to a change of 44 µS. A validation of the present process by comparing measured reaction rates with capillary electrophoresis yielded an excellent conformity. Cell extracts do not significantly interfere with the present process. Since pyruvate is the common amino acceptor of virtually all ω-TA, the present process can be used for investigations of the transaminase activity towards different amino donors. Moreover, also information about enantioselectivity of the enzyme can be obtained.

In this embodiment the invention provides a process for analysing a transaminase, in particular for the characterisation of properties of a transaminase, which allows to analyse the activity of the transaminase in dependency upon for instance the pH-value or the temperature of the reaction medium or allows to analyse the stability of the reaction, the effect of additives or buffer compositions.

According to the present process for analysis a first measurement of the conductivity of the reaction medium is carried out under a first set of reaction conditions and thereafter at least one second measurement of the conductivity of the reaction medium is carried out in order to be able to compare both of the obtained conductivity values and draw conclusions on the activities and properties of the transaminase tested. A decrease in the conductivity shows in a transaminase reaction according to the present invention the activity of said transaminase. Recognising a reduced decrease, an accelerated decrease or no decrease of the conductivity allows drawing conclusions on the properties of the transaminase.

In the context of the present invention a set of reaction conditions is preferably a set of reaction conditions the conditions of which are

preferably selected from the group consisting of temperature, pH-value and composition of the reaction medium, preferably reaction conditions as identified above, except the concentration of educts and products. In one embodiment of the present invention the set of reaction conditions are kept constant over the reaction time. In another embodiment of the invention the set of reaction conditions may be different over the reaction time.

In a preferred embodiment, the invention therefore also relates to such a process for the analysis of a transaminase, wherein the reaction medium is a low conductivity buffer. In a particularly preferred embodiment of the process the charged amino acceptor is pyruvate.

In a preferred embodiment of the present invention the present process for the analysis of a transaminase is used subsequently to the process for the preparation of an (R)-selective ω-transaminase of the present invention and extends the teaching of the present invention to not only providing new and advantageous (R)-selective ω-transaminases but also allow to determine their characteristics.

Further preferred embodiments of the present invention are the subject-matter of subclaims.

The present invention is illustrated in more detail in the following examples and the accompanying sequence listing.

SEQ ID no. 1 shows the full amino acid sequence (ORF) of the (R)-selective ω-TA from Mesorhizobium loti,

SEQ ID no. 2 shows the DNA-sequence encoding SEQ ID no. 1,

SEQ ID no. 3 shows the full amino acid sequence (ORF) of the (R)-selective ω-TA from Aspergillus terreus,

SEQ ID no. 4 shows the DNA sequence encoding SEQ ID no. 3,

SEQ ID no. 5 shows the full amino acid sequence (ORF) of the (R)-selective ω-transaminase from Mycobacterium aurum,

SEQ ID no. 6 shows the DNA sequence encoding SEQ ID no. 5,

SEQ ID no. 7 shows the full amino acid sequence (ORF) of the (R)-selective ω-transaminase from Arthrobacter sp.,

SEQ ID no. 8 shows the DNA sequence encoding SEQ ID no. 7,

SEQ ID no. 9 shows the sequence motive of E. coli BCAT used for determination of positions 95 to 99,

SEQ ID no. 10 shows the sequence motive of E. coli BCAT used for determination of position 38,

SEQ ID no. 11 shows the sequence motive of E. coli BCAT used for determination of position 107,

SEQ ID no. 12 shows the sequence motive of E. coli BCAT used for determination of position 163,

SEQ ID no. 13 shows the DNA-sequence encoding the (R)-selective ω-TA from Penicillium chrysogenum,

SEQ ID no. 14 shows the full amino acid sequence (ORF) of SEQ ID no. 13,

SEQ ID no. 15 shows the DNA-sequence encoding the (R)-selective ω-TA from Aspergillus niger,

SEQ ID no. 16 shows the full amino acid sequence (ORF) of SEQ ID no. 15,

SEQ ID no. 17 shows the DNA-sequence encoding the (R)-selective ω-TA from Aspergillus oryzae,

SEQ ID no. 18 shows the full amino acid sequence (ORF) of SEQ ID no. 17,

SEQ ID no. 19 shows the DNA-sequence encoding the (R)-selective ω-TA from Aspergillus fumigatus,

SEQ ID no. 20 shows the full amino acid sequence (ORF) of SEQ ID no. 19,

SEQ ID no. 21 shows the DNA-sequence encoding the (R)-selective ω-TA from Neosartorya fischeri,

SEQ ID no. 22 shows the full amino acid sequence (ORF) of SEQ ID no. 21,

SEQ ID no. 23 shows the DNA-sequence encoding the (R)-selective ω-TA from Gibberella zeae,

SEQ ID no. 24 shows the full amino acid sequence (ORF) of SEQ ID no. 23,

SEQ ID no. 25 shows the DNA-sequence encoding the (R)-selective ω-TA from Hyphomonas neptunium,

SEQ ID no. 26 shows the full amino acid sequence (ORF) of SEQ ID no. 25,

SEQ ID no. 27 shows the DNA-sequence encoding the (R)-selective ω-TA from Mesorhizobium loti MAFF303099,

SEQ ID no. 28 shows the full amino acid sequence (ORF) of SEQ ID no. 27,

SEQ ID no. 29 shows the DNA-sequence encoding the (R)-selective ω-TA from Roseobacter sp.,

SEQ ID no. 30 shows the full amino acid sequence (ORF) of SEQ ID no. 29,

SEQ ID no. 31 shows the DNA-sequence encoding the (R)-selective ω-TA from Marinomonas sp.,

SEQ ID no. 32 shows the full amino acid sequence (ORF) of SEQ ID no. 31,

SEQ ID no. 33 shows the DNA-sequence encoding the (R)-selective ω-TA from Rhizobium etli,

SEQ ID no. 34 shows the full amino acid sequence (ORF) of SEQ ID no. 33,

SEQ ID no. 35 shows the DNA-sequence encoding the (R)-selective ω-TA from Rhodoferax ferrireducens,

SEQ ID no. 36 shows the full amino acid sequence (ORF) of SEQ ID no. 35,

SEQ ID no. 37 shows the DNA-sequence encoding the (R)-selective ω-TA from Jannaschia sp.,

SEQ ID no. 38 shows the full amino acid sequence (ORF) of SEQ ID no. 37,

SEQ ID no. 39 shows the DNA-sequence encoding the (R)-selective ω-TA from Labrenzia alexandrii,

SEQ ID no. 40 shows the full amino acid sequence (ORF) of SEQ ID no. 39,

SEQ ID no. 41 shows the DNA-sequence encoding the (R)-selective ω-TA from Burkholderia sp.,

SEQ ID no. 42 shows the full amino acid sequence (ORF) of SEQ ID no. 41,

SEQ ID no. 43 shows the DNA-sequence encoding the (R)-selective ω-TA from Burkholderia cenocepacia,

SEQ ID no. 44 shows the full amino acid sequence (ORF) of SEQ ID no. 43,

SEQ ID no. 45 shows the DNA-sequence encoding the (R)-selective ω-TA from alpha proteobacterium,

SEQ ID no. 46 shows the full amino acid sequence (ORF) of SEQ ID no. 45,

SEQ ID no. 47 shows the DNA-sequence encoding the (R)-selective ω-TA from gamma proteobacterium,

SEQ ID no. 48 shows the full amino acid sequence (ORF) of SEQ ID no. 47,

SEQ ID no. 49 shows the DNA-sequence encoding the (R)-selective ω-TA from Mycobacterium vanbaalenii and

SEQ ID no. 50 shows the full amino acid sequence (ORF) of SEQ ID no. 49.

Figure 1 shows a chromatogram of B3APi obtained by a standard synthesis.

Figure 2 shows a chromatogram of B3APi obtained by the asymmetric synthesis according to the invention.

Figure 3 shows a chromatogram of C3AP obtained by the asymmetric synthesis according to the invention.

Figure 4 shows a chromatogram of MPPA obtained by the asymmetric synthesis according to the invention.

Figure 5 shows a chromatogram of B3AP obtained by the asymmetric synthesis according to the invention.

Examples:

Example 1 – Identification of (R)-selective ω-transaminases

The amino acid sequence of the (R)-selective ω-TA of Mycobacterium aurum as given in EP 1 038 953 A1 (SEQ ID no. 5) and the amino acid sequence of the (R)-selective ω-TA Arthrobacter sp. as given in EP 0 987 332 A1 (SEQ ID No. 7) are used as query bio-

molecule sequence. The biomolecule bank used in this example is the pubmed protein data bank of the NCBI (http://www.ncbi.nlm.nih.gov/pubmed (13 July 2009).

Using a BLAST search with amino acid sequence from M. aurum or Arthobacter sp. ω-TA (SEQ ID no. 5 or 7) as query using standard parameters (BLOSUM62 scoring matrix, word size: 3, gap costs: existence – 11, extension – 1) a first group of 100 various amino acid sequences from different organisms have been identified, which all have a minimum degree of 30 % sequence identity to the query sequence.

For the BLAST the "non-redundant protein sequences (nr)" (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&BLAST_PROGRAMS=blastp&PAGE_TYPE=BlastSearch&SHOW_DEFAULTS=on&LINK_LOC=blasthome) have been used on 13 July 2009. In this first group of 100 various amino acids representing the first target biomolecule sequences, those sequences have been searched for and identified by a sequence motive search which do not show sequence motives c1), c2) or c3) and which do show sequence motives c4), c5) and c6). 21 ORF could be identified and are listed in Table 1 below.

Table 1

No.	Source Organisms	% Identity with AS-ωTA SEQ ID7	% Identity with MA-ω-TA SEQ ID5	Gene-ident Nr./ Protein-ident. Nr. (NCBI database)
1	*Aspergillus terreus NIH2624*	44	40	115385557
2	*Penicillium chrysogenum Wisconsin 54-1255*	44	42	211591081
3	*Aspergillus niger*	40	36	145258936
4	*Aspergillus oryzae RIB40*	41	40	169768191
5	*Aspergillus fumigatus Af293*	41	38	70986662

6	*Neosartorya fischeri NRRL 181*	41	38	119483224
7	*Gibberella zeae PH-1*	40	39	46109768
8	*Hyphomonas neptunium ATCC 15444*	44	40	114797240
9	*Mycobacterium vanbaalenii PYR-1*	50	91	120405468
10	*Mesorhizobium loti MAFF303099*	38	37	13471580
11	*Mesorhizobium loti*	35	37	20804076
12	*Roseobacter sp. MED193*	37	37	86137542
13	*Marinomonas sp. MED121*	36	34	87122653
14	*Rhizobium etli CIAT 652*	33	32	190895112
15	*Rhodoferax ferrireducens T118*	38	36	89899273
16	*Jannaschia sp. CCS1*	37	32	89053613
17	*Labrenzia alexandrii DFL-11*	42	36	*EEE43073*
18	*Burkholderia sp. 383*	32	32	*78059900*
19	*Burkholderia cenocepacia HI2424*	36	33	*ABK12047*
20	*Alpha proteobacterium HTCC2255*	36	34	*ZP_01448442*
21	*Gamma proteobacterium*	27	26	
	Mycobacterium aurum	49	100	SEQ ID 5
	Arthrobacter sp.	100	49	SEQ ID 7

AS: Aspergillus sp., MA: Mycobacterium aurum

Example 2 – Preparation and analysis of the transaminases of Aspergillus terreus, Mycobacterium vanbaalenii and Mesorhizobium loti

2.1 The ω-TA from Mycobacterium vanbaalenii, entry 9 in Table 1 (SEQ ID no. 49 and 50), from Aspergillus terreus, entry 1 in table 1 (SEQ ID no. 3 and 4), and Mesorhizobium loti, entry 11 in table 1 (SEQ ID no. 1 and 2), have been obtained and used in codon usage adapted form (to E. coli) to express the enzymes in Escherichia coli. The Mycobacterium vanbaalenii transaminase is called in the following Mva-TA, the Aspergillus terreus transaminase Ate-TA and the Mesorhizobium loti Mlo-TA.

2.2 Acetophenonassay

The Mva-TA and Ate-TA converted (R)-α-MBA at least 100 times faster than the (S)-enantiomer in an acetophenonassay with 2,5 mM amine and 2,5 mM pyruvate at pH 7,5 and 30°C.

Assay: The increase in absorbtion of the acetophenone formed during the reaction was monitored at 245 nm.

The Mlo-TA did not convert either (R)- nor (S)-α-MBA.

In the following, further amines have been tested in the presence of pyruvate or α-ketoglutarate as amino acceptors (10 mM amine, 10 mM pyruvate, 0,1 mM PLP, phosphate buffer pH 7,5, incubation 24 h at 30°C, analysis by thin layer chromatography).

A conversion could be seen for 2-aminoheptane, 2-aminopentane, 1,3-dimethybutylamine and 4-phenyl-2-aminobutane. Also, a minimal conversion of isopropylamine could be detected.

No conversion with other amino donors such as D-alanine, L-valine, γ-aminobutyrate, ethylamine, benzyl amine, putrescine, 2-amino-3-methylbutane and 3,3-dimethyl-2-aminobutane could be detected.

Thus, all three proteins were proven to be ω-TA.

In particular, using (R)- and (S)-2-aminohexane as substrate, only the (R)-enantiomer was significantly converted. Thus, also Mlo-TA is a (R)-selective ω-TA. No DATA (D-amino acid transaminase) or

BCAT (branched chain aminotransferase) activity was seen for all three proteins.

2.3 Conductivity assay

Also, 1-N-boc-3-aminopyrrolidine (B3AP), 1-N-boc-3-aminopiperidine (B3APi) and 1-Cbz-3-aminopiperidine (C3APi) were used as substrates to determine the relative activities of these substances against the model substrate α-MBA.

During the reaction of the amine and pyruvate (at pH 7,5 both substrates are charged) to alanine and ketone (the ketone has no charge, alanine is a zwitterionic compound and does not contribute to conductivity) monitoring of the kinetics of the conductivity allows to conclude on the conversion rates.

Before starting the reaction, a calibration was carried out by determining different conversions in dependence of various concentrations of alanine, pyruvate, ketone, and amine.

The reduction in conductivity was per mM conversion of α-MBA, B3AP, B3APi and C3APi 44 µS, 50 µS, 48,5 µS and 49,3 µS.

In addition to the three recombinantly expressed transaminases, a commercially available (R)-selective ATA-117 from Codexis was tested using the following reaction conditions: 50 mM CHES buffer, pH 7,5, pH adjustment with BIS-TRIS (=Bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methane), 0,1 mM PLP, 5-6 mM amine and pyruvate, reaction at 25°C.

A conversion of the substrates could be shown for Ate-TA. The relative activities were 2 % for (R)-B3AP and (R,S)-C3APi and 1 % for (R)-B3APi compared to (R)-α-MBA.

2.4 Determination of the enantioselectivity by asymmetric synthesis of amines of both (R)-selective ω-TA from Aspergillus terreus and Mesorhizobium loti

By an independent method it is shown that Mlo and Ate-TA are (R)-selective and convert the substrates to the desired products with excellent enantioselectivity.

For the definitive proof of (R)-enantioselectivity of both transaminases an asymmetric synthesis of different amines were carried out and the optical purity thereof was determined.

As amino donor a hundred-fold excess of alanine was used. The conversion was not exactly determined, but roughly estimated. In further experiments different methods for increasing the conversion (PDC) could be developed. High to excellent enantioselectivities were obtained with both transaminases, except with C3APi, wherein only a very low enantioselectivity was obtained.

	Ate-TA		Mlo-TA	
Amine	%ee	%c	%ee	%c
B3AP	99,8	40 ± 20	-	-
C3AP	99,6	40 ± 20	-	-
B3APi	>99,9	30 ± 20	-	-
C3APi	49	30 ± 20		

MPPA	-	-	98	15 ± 10

Table 2

Methods:

The transaminases were expressed in E. coli BL21.

The culture medium (50 ml LD-Amp medium) was induced with 0.2% of rhamnose, when an optical density (OD) of 0.5 was obtained, and was cultivated for 12 hours at 25 °C. Subsequently, the cells were washed with a sodium phosphate buffer, centrifuged, suspended in 1 ml sodium phosphate buffer (pH 7.5) and frozen in 200 µl aliquots.

For the biocatalysis 5 µl of a 500 mM ketone solution in DMSO, 22 mg D-alanine and 10 µl of a 10 mM PLP were added to these aliquots. The reaction mixture was filled up to a total volume of 500 µl with a 50 mM sodium phosphate buffer (pH 7.5). The reaction mixtures were incubated over night at 25 °C and 500 rpm. For determination of the enantiomeric excess by means of CE 200 µl of a 1 N NaOH was added to 400 µl of the sample, extracted with 400 µl of dichloromethane and the organic phase was separated. The organic phase was then extracted with 100 µl of a 5 mM triethylammonium phosphate buffer (pH 3). Subsequently, the obtained aqueous phase was injected in the CE.

The program of separation:

CE-Capillare 30 cm, 50 µm inner diameter, temperature 15 °C

- rinsing with triethylammoniumphosphat buffer pH 3, 1 min, 30 psi
- rinsing with 5 % highly sulfated cyclo dextrin (HSαCD or HSγCD), 1 min, 10 psi
- injection: 5-10 s, 1 psi
- water dip
- separation: voltage 10 or 15 kV
- detection: MPPA, C3AP and C3APi at 200 nm; B3AP, B3APi at 190 nm

Conditions of separation:

	Separating segment [cm]	Chiral selector	Migration time [min] S-Amin	R-Amin	Voltage while separating
MPPA	10	HSγCD	2.4	2.8	15 kV
B3AP	10	HSγCD	4.55	6.1	15 kV
B3APi	20	HSγCD	12.0	12.4	15 kV
C3AP	20	HSαCD	10.2	11.3	10kV
C3APi	20	HSαCD	6.8	7.2	15 kV

Table 3

Example 4 – Preparation and analysis of all transaminases identified in Table 1

4.1 Expression and purification of the transaminases

The codon optimized open reading frames encoding proteins with entry numbers 1, 2, 3, 4, 8, 9, 11, 13, 14, 15, 17, 18 and 21 in Table 1 were inserted into pGASTON between the NdeI and BamHI restriction sites using a ligation independent cloning strategy. The codon optimized ORF encoding all other proteins were ordered already subcloned into pET-22b. Transformed E. coli BL21 (DE3) strains

were grown in 400 ml LB medium supplemented with ampicillin (100 µg/ml). Cells were incubated initially at 37°C on a gyratory shaker until the OD_{600} reached 0.7. The cells were then induced by the addition of 0.2% rhamnose (pGASTON) or 0.1 mM IPTG (pET-22b), respectively, and at the same time the incubation temperature was decreased to 20°C. After induction the incubation was continued for a further 20 h. Aliquots were withdrawn at several points of time after induction to follow the expression.

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 (4000 × g, 30 min) and the resulting supernatant was passed through a 0.5 µm filter prior to chromatography. Chromatography was performed using an ÄKTA Purifier (GE Healthcare). The filtered cellular extract was applied to a 5 ml column of IMAC SepharoseTM 6 Fast Flow (GE Healthcare). The column was washed at a flow rate of 5 ml min-1 with a 10 column-volume of 50 mM phosphate buffer, pH 7.5, containing 300 mM NaCl, 0.1 mM PLP and 30 mM imidazol (to avoid unspecific binding) and the ATA activity 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 imidazol (flow rate of 5 ml min-1). The activity containing fractions 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.

The amount of each protein purified from approximately 3 g cells (wet weight) is given in Table 4 below.

Table 4

Entry	Enzyme source	Protein yield after purification [mg]
1	Aspergillus terreus	8.6
2	Penicillium chrysogenum	26.2
3	Aspergillus niger	-
4	Aspergillus oryzae	20.6
5	Aspergillus fumigatus	14.8
6	Neosartorya fischeri	23.3
7	Gibberella zeae	4.8
8	Hyphomonas neptunium	6.5
9	Mycobacterium vanbaalenii	8.9
10	Mesorhizobium loti MAFF303099	6.9
11	Mesorhizobium loti	5.3
12	Roseobacter sp.	27.5
13	Marimonas sp.	23.7
14	Rhizobium etli	6.5
15	Rhodoferax ferrireducens	7.5
16	Jannaschia sp.	24.8
17	Labrenzia alexandrii	12.5
18	Burkholderia sp.	41.6
19	Burkholderia cenocepacia	-
20	alpha proteobacterium	-
21	gamma proteobacterium	2.6

4.2 Characterization of substrate specificity of (R)-selective ω-TA

For determining activity towards α-methylbenzyl amine (α-MBA) in the initial screen of the expressed proteins, an acetophenone-based assay was used: 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 amines 2-aminohexane, 4-phenyl-2-aminobutane and 1-N-Boc-3-aminopyrrolidine were monitored using a conductivity assay: 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 investigating DATA- and BCAT-activity the decrease of NADH was measured spectrophotometrically at 340 nm using dehydrogenase coupled assays: a solution of 5 mM α-ketoglutaric acid and D-alanine was reacted in the presence of the purified transaminase, 1 U/ml lactate dehydrogenase and 0.5 mM NADH for measuring DATA-activity. A solution containing 5 mM 3-methyl-2-oxobutyric acid and L-glutamate, 10 mM ammonium chloride, 1 U/ml glutamate dehydrogenase and 0.5 mM NADH was used for measuring BCAT-activity.

All reactions took place in 20 mM Tricine buffer pH 7.5 containing 0.01 mM PLP. The pH of the buffer was adjusted with 1,8-Diazabicyclo[5.4.0]undec-7-en.

Results are given in Table 5 below.

Table 5: Specific activities for various substrates.

Substrates	pyruvate **1**		pyruvate **2**		pyruvate **3**		pyruvate **4**		2KG D-Ala	MOB L-Glu
Entry	R	S	R	S	R	S	R	S		
1	15.2	<0.001	2.91	<0.001	9.7	<0.001	0.031	<0.001	<0.001	0.003
2	1.3	<0.001	1.1	0.044	5.6	<0.001	0.264	<0.001	<0.001	<0.001
3	-[a)]	-	-	-	-	-	-	-	-	-
4	3.7	0.001	1.4	0.023	5.2	0.002	0.051	0.002	<0.001	<0.001
5	4.1	<0.001	2.4	<0.001	4.5	<0.001	0.009	<0.001	<0.001	0.005
6	4.5	<0.001	7.4	<0.001	6.0	<0.001	0.013	<0.001	<0.001	0.005
7	18.6	<0.001	19.6	<0.001	8.2	<0.001	<0.001	<0.001	<0.001	0.016

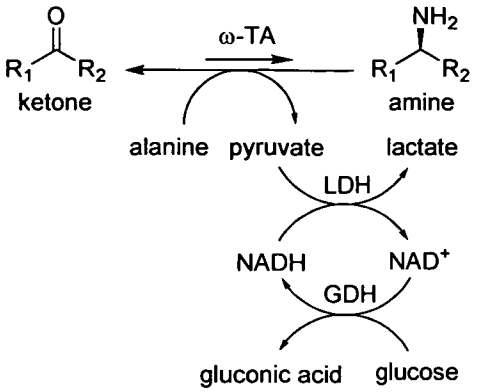
8	3.6	<0.001	3.2	0.225	20.7	<0.001	0.163	<0.001	<0.001	0.012
9	4.7	<0.001	5.6	<0.001	2.6	<0.001	<0.001	<0.001	<0.001	0.003
10	0.011	<0.001	0.003	<0.001	0.010	<0.001	0.001	<0.001	0.004	0.004
11	0.013	<0.001	0.124	<0.001	0.002	<0.001	<0.001	<0.001	<0.001	0.005
12	0.003	<0.001	0.001	<0.001	0.001	<0.001	0.001	<0.001	0.003	0.002
13	0.002	<0.001	0.020	<0.001	0.003	<0.001	<0.001	<0.001	<0.001	0.003
14	0.867	<0.001	0.012	<0.001	0.260	<0.001	<0.001	<0.001	0.020	0.016
15	0.056	<0.001	0.001	<0.001	0.307	<0.001	<0.001	<0.001	0.010	0.098
16	0.059	0.007	0.071	0.002	0.370	0.068	0.022	<0.001	0.062	0.020
17	0.060	0.003	0.073	0.001	0.120	0.027	0.205	0.002	0.063	0.023
18	0.017	<0.001	0.002	<0.001	1.1	0.007	<0.001	<0.001	<0.001	0.001
19	-[a)]	-	-	-	-	-	-	-	-	-
20	-	-	-	-	-	-	-	-	-	-
21	0.028	<0.001	0.610	0.004	0.034	<0.001	<0.001	<0.001	<0.001	0.031

1 – aminohexane, **2** – α-MBA, **3** – 4-phenyl-2-aminobutane, **4** – 1-N-Boc-3-aminopyrrolidine, **2KG** – 2-ketoglutarate, **D-Ala** – D-alanine, **L-Glu** – L-glutamate, **MOB** – 3-methyl-2-oxobutyric acid. Entry number corresponds to Table 1. All measurements were done at least in duplicates. The deviation of single measurements from the mean value was < 10 %.

a) Measurement was not possible since protein yield during expression was very low/protein was unstable during purification.

4.3 Asymmetric Synthesis of (R)-amines **1-4** (see legend of Table 5 above) with ω-TAs from Aspergillus terreus, Mesorhizobium loti and Mycobacterium vanbaalenii

Asymmetric syntheses were performed at 30°C for 24 hours in sodium phosphate buffer (100 mM, pH 7) containing pyridoxal-5'-phosphate PLP monohydrate (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). ω-TA from Aspergillus terreus, Mesorhizobium loti and Mycobacterium vanbaalenii (entry 1, 11 and 9 in Table 1) were expressed in E. coli BL21 as described above, frozen in aliquots and applied directly as whole cell biocatalysts without further purification. The conversion was measured by detection of the formed amines (1, gas chromatography (GC); 2-4, capillary electrophoresis (CE)). Chiral analysis of 2-4 was performed using CE as described above. The enantiomeric excess (%ee) value for 1 was analysed by GC. 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 by using a Shimadzu GC14A that was equipped with a Heptakis-(2,3-di-O-acetyl-6-O-tert-butyldimethylsilyl)-β-cyclodextrin column (25 m x 0.25 mm). The retention times were 16.0 min ((S)-1) and 16.2 min ((S)-2) at an oven temperature gradient of 80 °C/10 min//20 °C//180 °C/10min.

Results are given in Table 6 below.

Table 6

formed amines	ω -TA	Conversion [%][b]	Enantiomeric excess [%eeP][c]
1	Ate	32	>99
1	Mlo	41	>99
1	Mva	35	>99
2	Ate	15	>99
2	Mlo	1	95.0
2	Mva	2	>99
3	Ate	14	>99
4	Ate	11	>99

[a]Reaction conditions: 50 mM ketone, 250 mM D-alanine, 100 mM sodium phosphate buffer pH 7.0, 1 mM PLP, 1 mM NADH. The co-product pyruvate of the reaction was removed with lactate dehydrogenase (LDH). For cofactor recycling, glucose dehydrogenase (GDH) was used. [b]Conversions were not optimized. The deviation of a single measurement from the mean value did not exceed 10 %. Compound 4 was only converted by Ate-TA. [c](R)-enantiomers.

Claims

1. A process for the preparation of a (R)-selective ω-transaminase, comprising the following steps:

a) providing at least one query biomolecule sequence of at least one transaminase or lyase and at least one biomolecule bank,

b) searching the biomolecule bank with the query biomolecule sequence to identify a group of first target biomolecule sequences, wherein the first target biomolecule sequences have a degree of sequence identity of at least 20% to the query biomolecule sequence, calculated on amino acid level,

c) selecting in the group of first target biomolecule sequences a group of second biomolecule target sequences, which do not comprise, on amino acid level, at least one of the following amino acid sequence motives c1) to c3) with

c1) at position 95 to 97 an amino acid sequence Tyr Xa1 Xa2, with Xa1 being an amino acid Ile, Val, Leu, Met, Phe, and Xa2 being an amino acid Arg or Lys or

c2) at position 97 to 99 an amino acid sequence Tyr Xaa Gln, with Xaa being an amino acid and in the region from position 105 to 111 an amino acid sequence Arg Xaa Xa3, Xa3 being an amino acid, preferably being His or

c3) at position 38 Thr, at position 97 Lys and at position 107 to 109 an amino acid sequence Arg Xa4 Xa5, Xa4 being an amino acid, preferably being Gly and Xa5 being an amino acid, preferably being Tyr

and which comprise

c4) at position 95 another amino acid than Tyr, Arg, Lys, or at position 95 Tyr, but at position 97 no Arg or Lys and

c5) at position 40 no Lys or Arg and

c6) in the region from position 161 to 165 Lys

to identify a group of second target biomolecule sequences and

d) providing a biomolecule having the second target biomolecule sequence identified in step c) and being or coding at least partially a protein with the activity of a (R)-selective ω-transaminase.

2. A process for screening for a (R)-selective ω-transaminase comprising steps a) to c) of claim 1.

3. The process according to claim 1 or 2, wherein the biomolecule is a protein and the biomolecule sequence is an amino acid sequence.

4. The process according to claim 1 or 2, wherein the biomolecule is a DNA molecule and the biomolecule sequence is a DNA sequence.

5. The process according to any one of the preceding claims, wherein the biomolecule bank is a biomolecule database and the biomolecule database is searched in step b) with a biomolecule sequence alignment tool, in particular BLAST.

6. The process according to claims 1, 3, 4 and 5, wherein the protein or DNA molecule provided in step d) is provided by de novo synthesis.

7. The process according to any one of claims 1, 2 and 4, wherein the biomolecule bank is a gene bank and the gene bank is searched in step b) with a query DNA sequence molecule.

8. The process according to claim 7, wherein in step c) DNA-sequence primers are used to select for the group of second biomolecule sequences.

9. The process according to any one of the preceding claims, wherein the query biomolecule sequence is a sequence representing all or a part of a functional (R)-selective ω-transaminase.

10. A (R)-selective ω-transaminase being obtainable according to any one of the methods of claims 1 and 3 to 9, preferably selected from the group of biomolecule sequences consisting of SEQ ID no. 1, 2, 3 and 4.

11. A process for analysing a transaminase, in particular for the characterization of properties of a transaminase, comprising the following steps:

i) providing a charged amino acceptor, a charged amino donor and a transaminase,

ii) reacting the amino acceptor and the amino donor with the transaminase in a reaction medium, and thereby

iii) determining the conductivity of the reaction medium under a first set of reaction conditions and

iv) subsequently to step iii) determining the conductivity of the reaction medium under a second set of reaction conditions, so as to obtain at least two conductivity values reflecting the properties of the transaminase.

12. The process according to claim 11, wherein the conductivity of the reaction medium is determined at least three times under at least a third set of reaction conditions.

13. The process according to claims 11 or 12, wherein the sets of reaction conditions are the same.

14. The process according to any one of claims 11 to 13, wherein the reaction medium is a low conductivity buffer.

15. The process according to any one of claims 11 or 14, wherein the charged amino acceptor is pyruvate.

16. A process for the screening or preparation of a (R)-selective ω-transaminase according to any one of claims 1 to 9, wherein subsequent to the screening or preparation of the (R)-selective ω-

transaminase a process for the analysis of a transaminase according to any one of claims 11 to 15 is performed.

17. A process for the preparation of an optically active chiral amine, comprising reacting at least one amino acceptor and at least one amino donor with a (R)-selective ω-transaminase according to claim 10 and obtaining the optically active chiral amine.